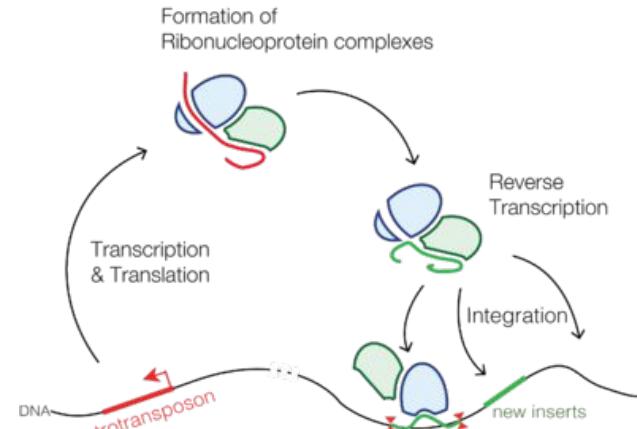
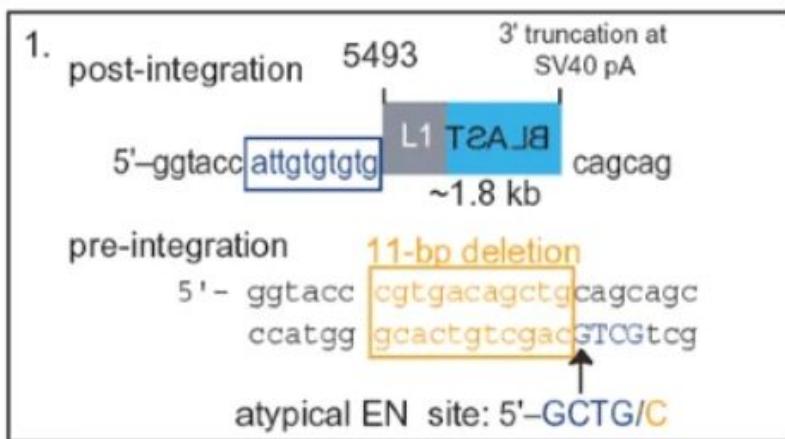


# BIOLOGIA MOLECULAR COMPUTACIONAL

Análise de domínios de transcriptase reversa em  
retrotranspósons non-LTR

# RETROTRANSPOSONS

- Elementos genéticos que se copiam.
- LTR - long terminal repeats - algumas centenas de bases
- Reparo de DNA independente de endonucleases



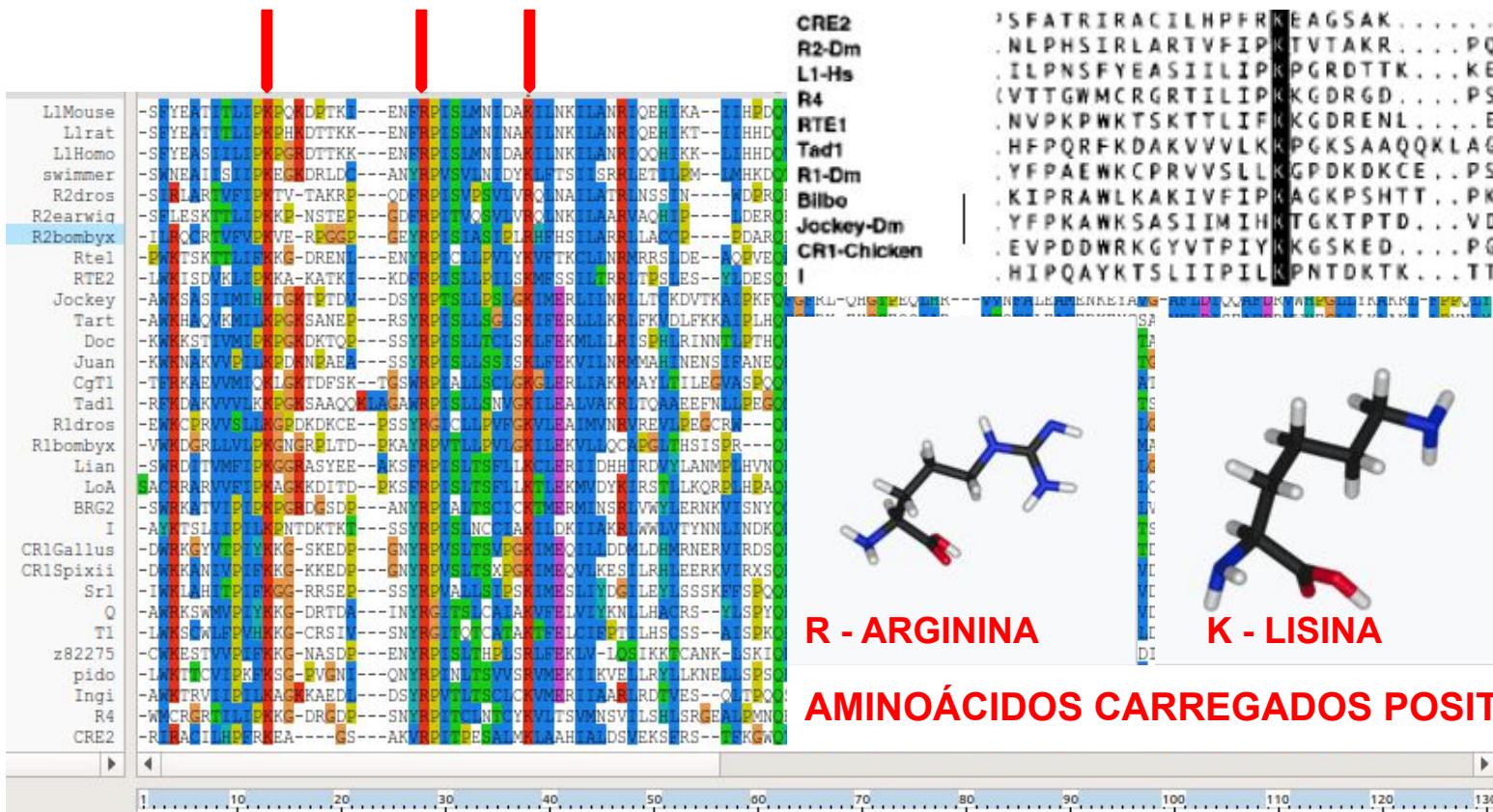
# **PARTE 1**

- Arquivo com organismos e suas respectivas sequências de proteínas do domínio transcriptase reversa de retrotransposons do tipo non-LTR
- Alinhamento Múltiplo no Clustal
- Análise de regiões de interesse
- Construção de árvore filogenética a partir do alinhamento múltiplo

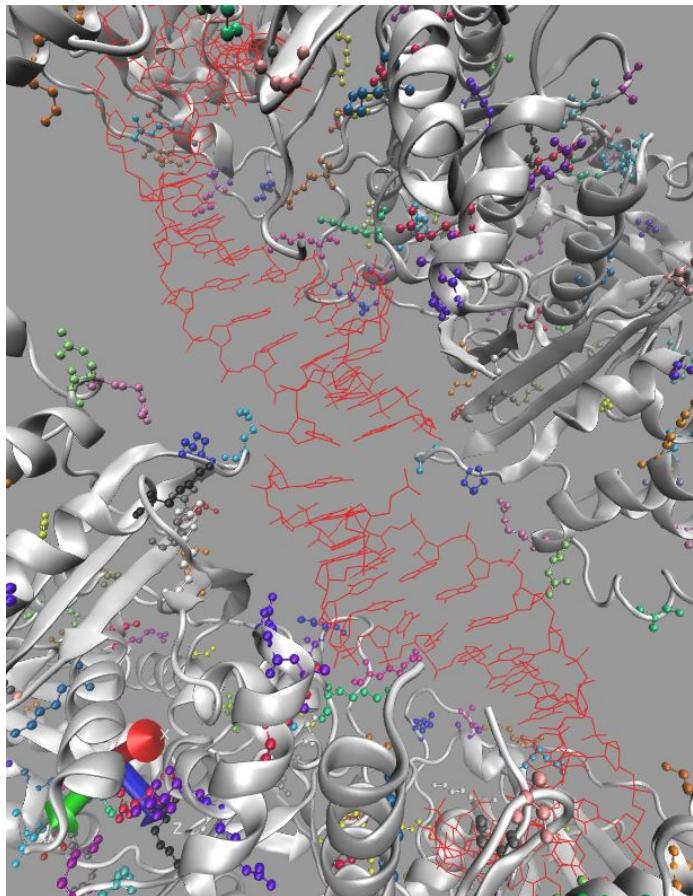
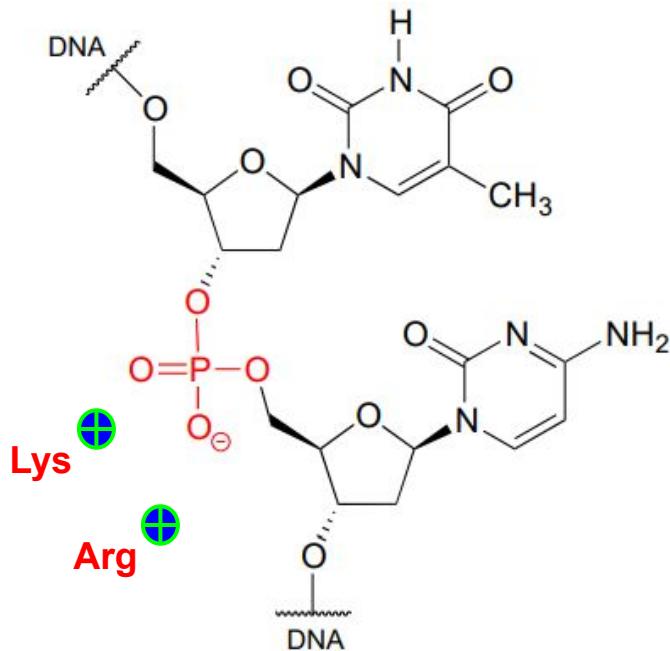
# Alinhamento Múltiplo no Clustal



# REGIÃO I

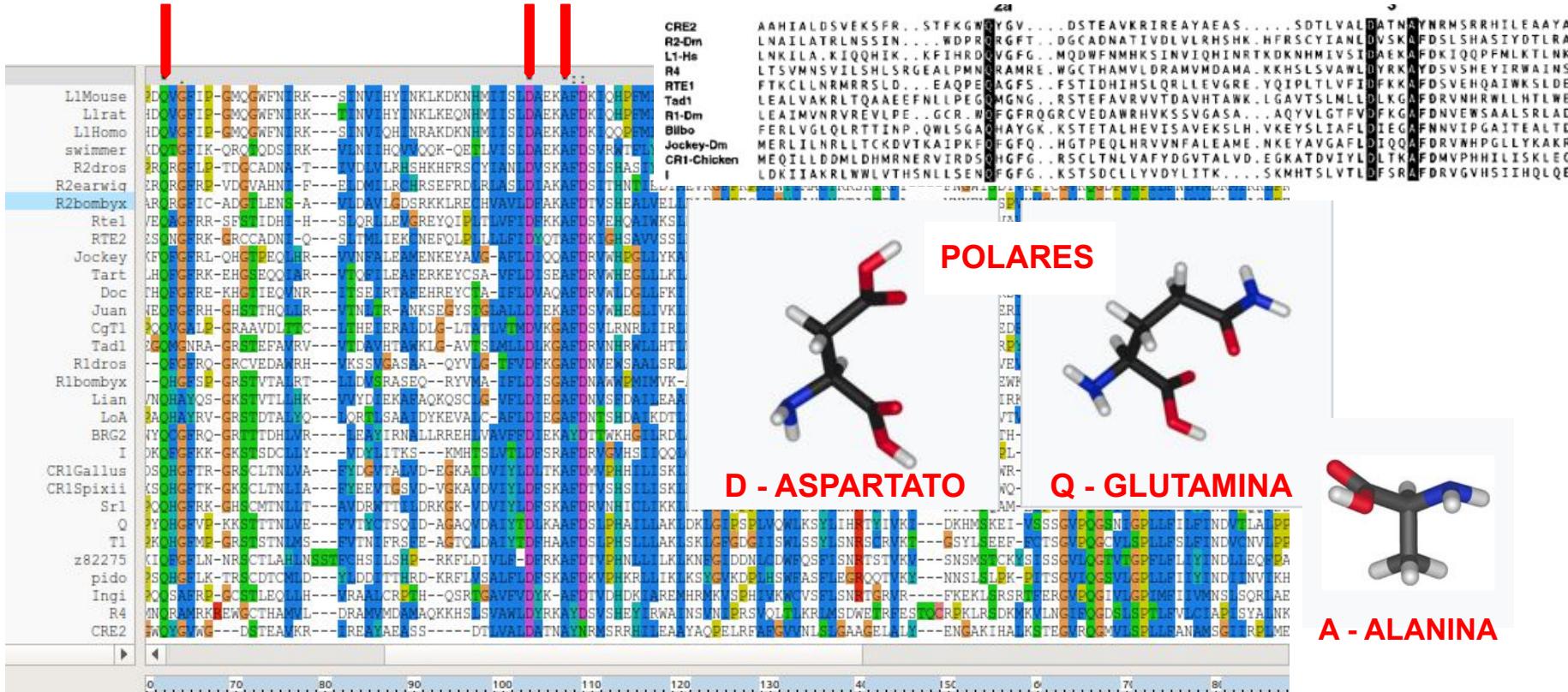


# REGIÃO I

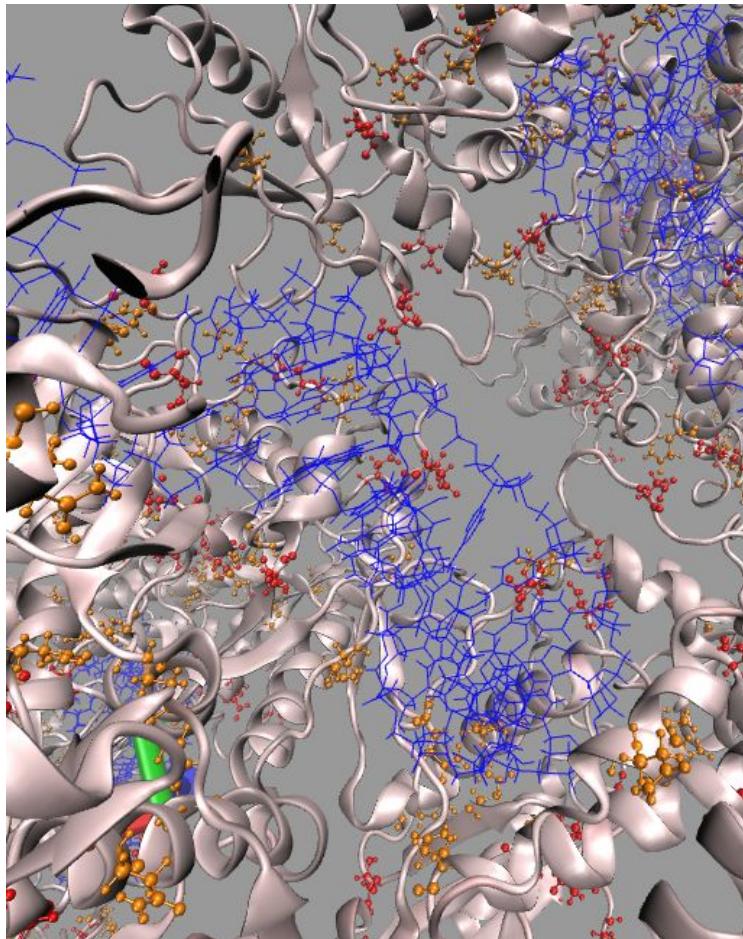
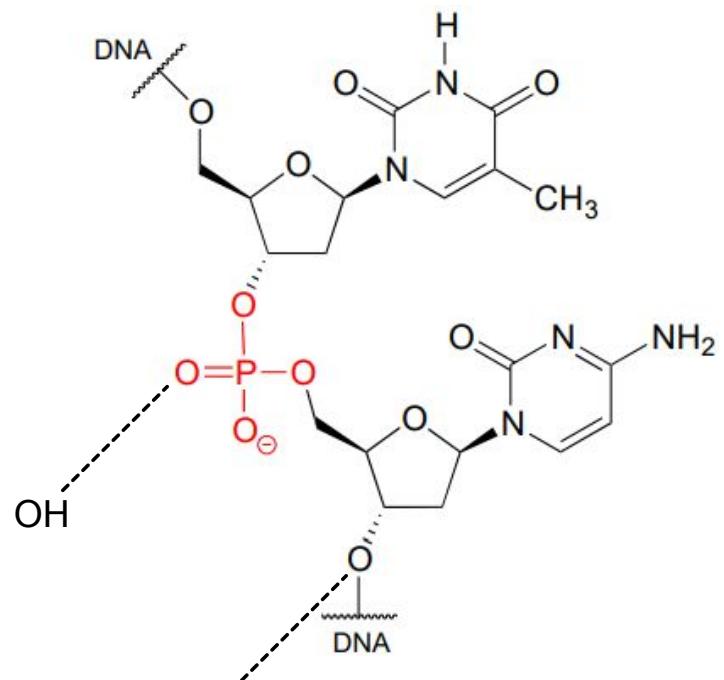


PDB CODE: 4OL8

# REGIÃO II



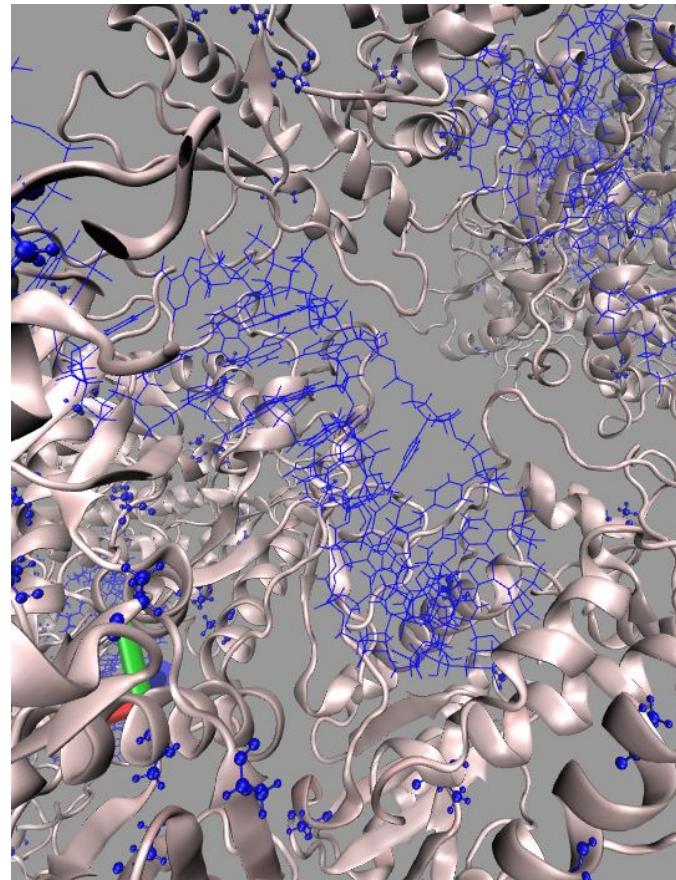
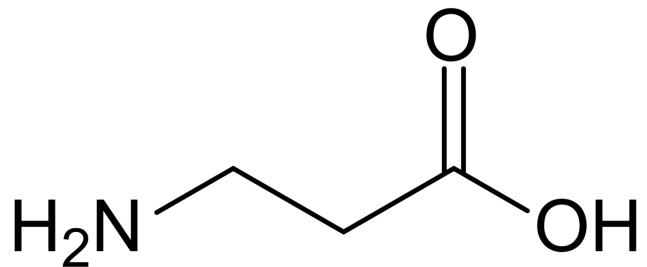
# REGIÃO II



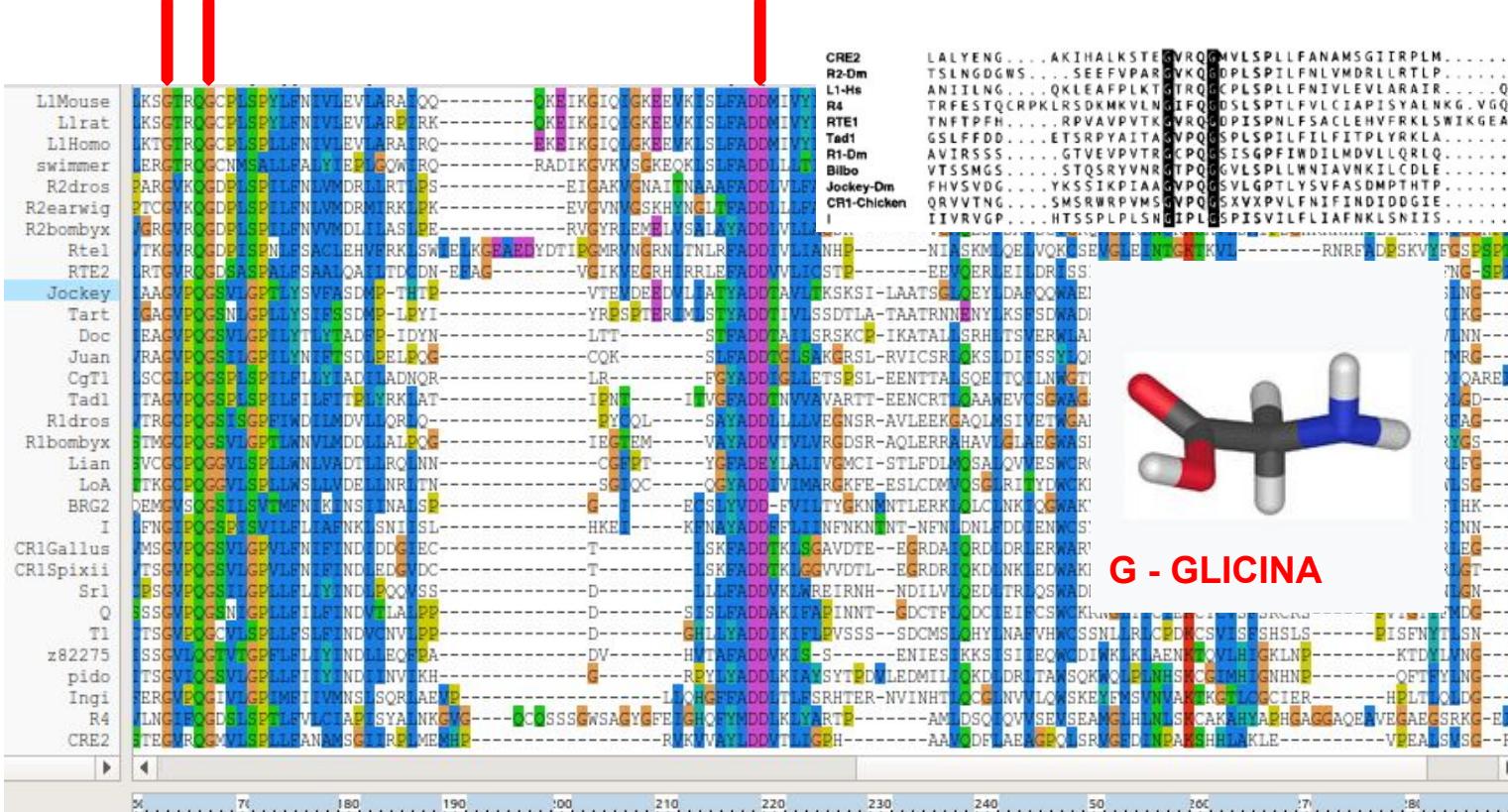
PDB CODE: 1HVU

# REGIÃO II

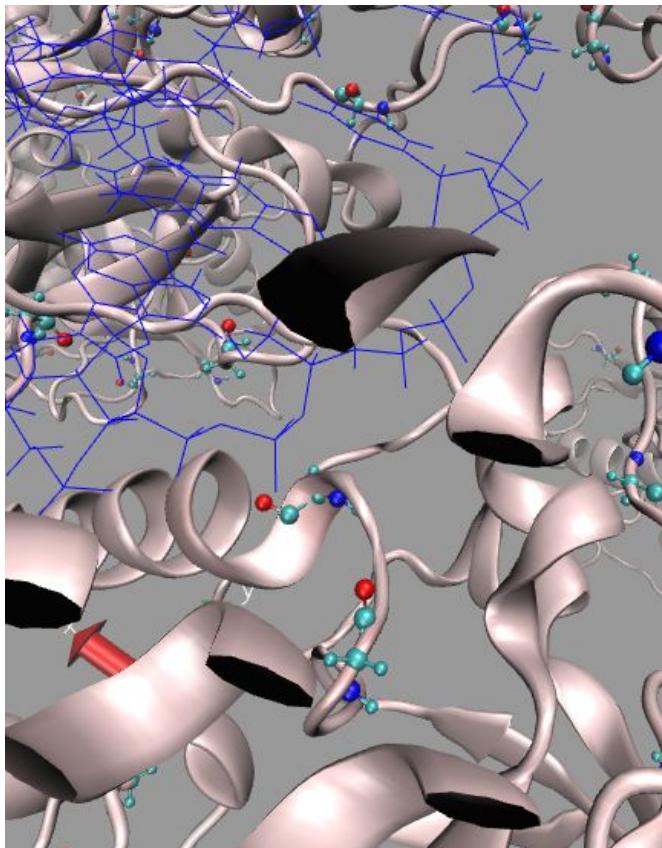
ALANINAS APOLARES



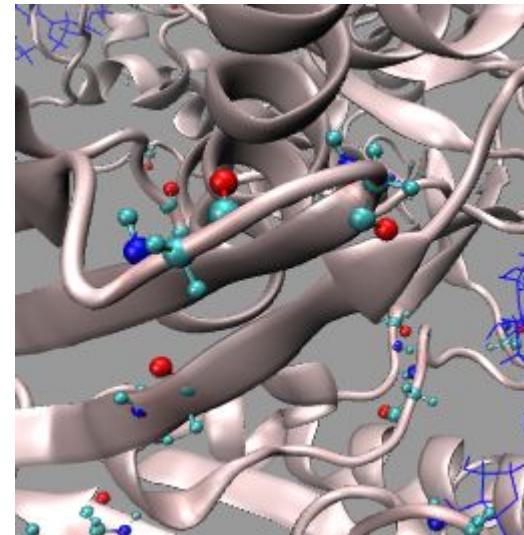
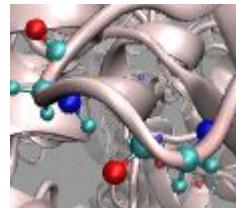
# REGIÃO III



# REGIÃO III



LOOPS e TERMINAÇÕES DE ESTRUTURA SECUNDÁRIA



MOTILIDADE DA PROTEÍNA PRÓXIMO A SUPERFÍCIE  
DE INTERAÇÃO COM RNA

# REGIÃO III

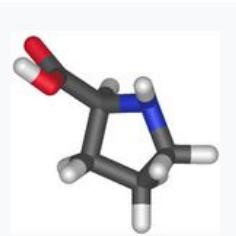


	KSGT PROG
L1Mouse	KSGT PROG
L1rat	KSGT PROG
L1Homo	KTG PROG
swimmer	ERGV PROG
R2dros	PARGV KOG
R2earwig	PTCCV KOG
R2bombyx	PTGV PROG
Rtel	TKGV PROG
RTE2	PTGV PROG
Jockey	IAAGV POG
Tart	IAGV POG
Doc	IEAGV POG
Juan	IRAGV POG
CgTl	ISCCV POG
Tad1	ITAGV POG
Rldros	ITRGCV POG
Ribombyx	ITMGCV POG
Lian	IVGCCV POG
LoA	ITKGV POG
BRG2	DEMGV S G
I	FNGV POG
CR1Gallus	MSGV POG
CR1Spixii	ITSGV POG
Srl	MSGV POG
Q	SSSGV POG
T1	ITSGV POG
z82275	SSSGV ILOG
pid0	ITSGV ILOG
Ingi	ERGV POG
R4	ILNGV POG
CRE2	STEGV PROG

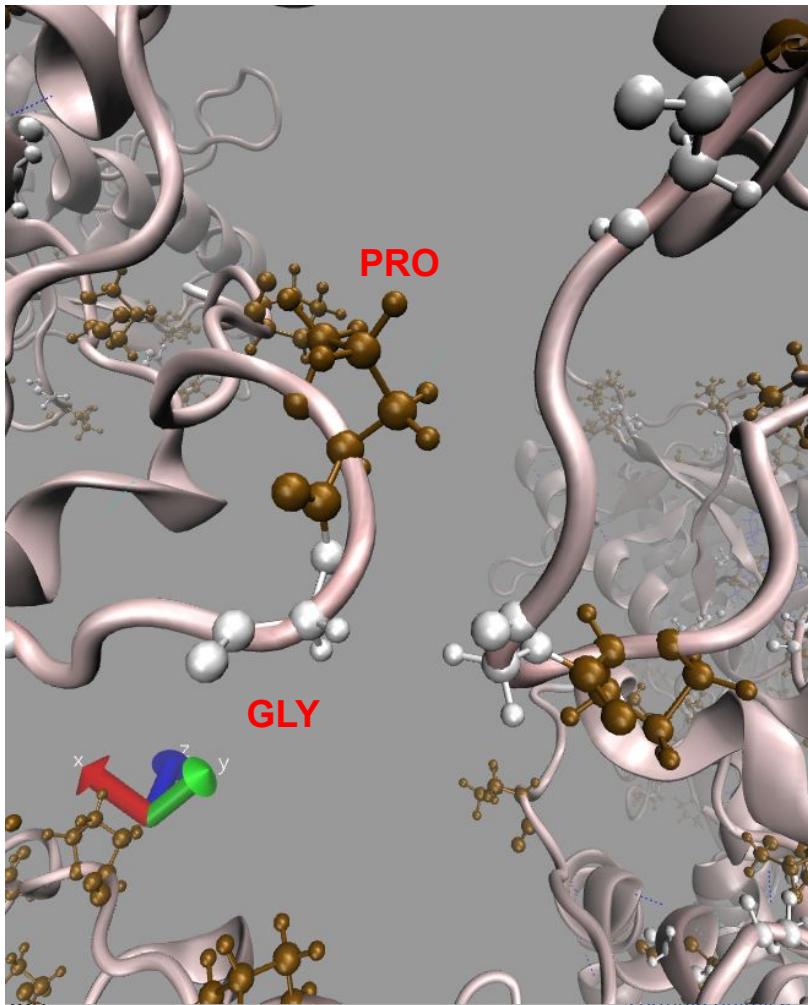
**PONTES PRO-GLY  
ALTERAM A ENERGIA  
DE ATIVAÇÃO DA  
FORMAÇÃO DE LOOPS  
CURTOS.**

<https://doi.org/10.1021/ja042798>

i

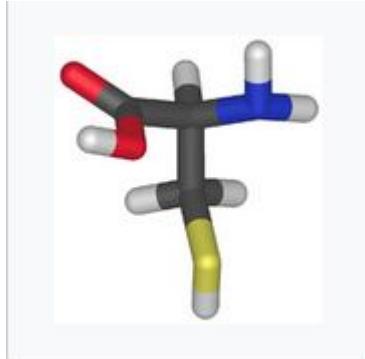


**P - PROLINA**



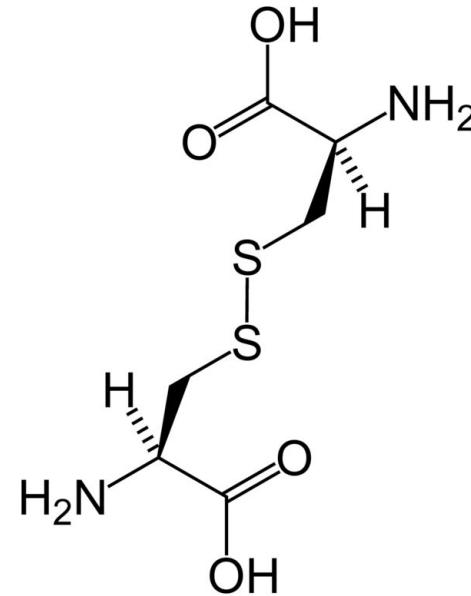
# REGIÃO III

	KSGVPROG
LlMouse	KSGVPROG
Llrat	KSGVPROG
LlHomo	KTGIVPROG
swimmer	ERGVPROG
R2dros	PARGVKQG
R2earwig	PTCCVKQG
R2bombyx	PRGVPROG
Rtel	TKGVPROG
RTE2	RTGVPROG
Jockey	IAAGVPOG
Tart	SAGVPOG
Doc	EAQVPOG
Juan	RAQVPOG
CgT1	SCCLPOG
Tad1	TAGVPOG
Rldros	ITRGVPOG
Rlbombyx	IMGCQPOG
Lian	WCQVPOG
LoA	TKGVPOG
BRG2	DEMCGPOG
I	FNGVPOG
CR1Gallus	MSGVPOG
CR1Spixii	TSGVPOG
Srl	MSGVPOG
Q	SSSGVPOG
T1	TSQCVPOG
z82275	SSSGVPOG
pid0	TSQCVPOG
Ingi	ERGVPOG
R4	LMGIVPOG
CRE2	TEGVPROG

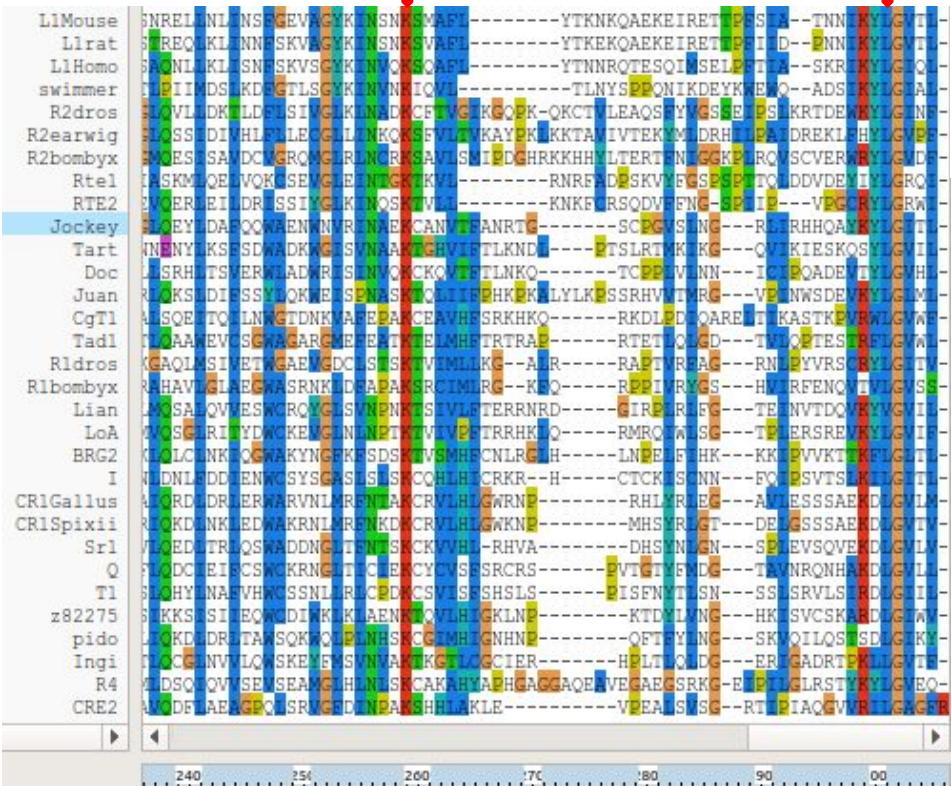


C - CISTEÍNA

POSSÍVEIS FUTURAS PONTES DISSULFETO

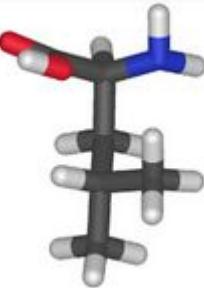


# REGIÃO IV



CRE2  
R2-Dm  
L1-Hs  
R4  
RTE1  
Tad1  
R1-Dm  
Bilbo  
Jockey-Dm  
CR1-Chicken  
I

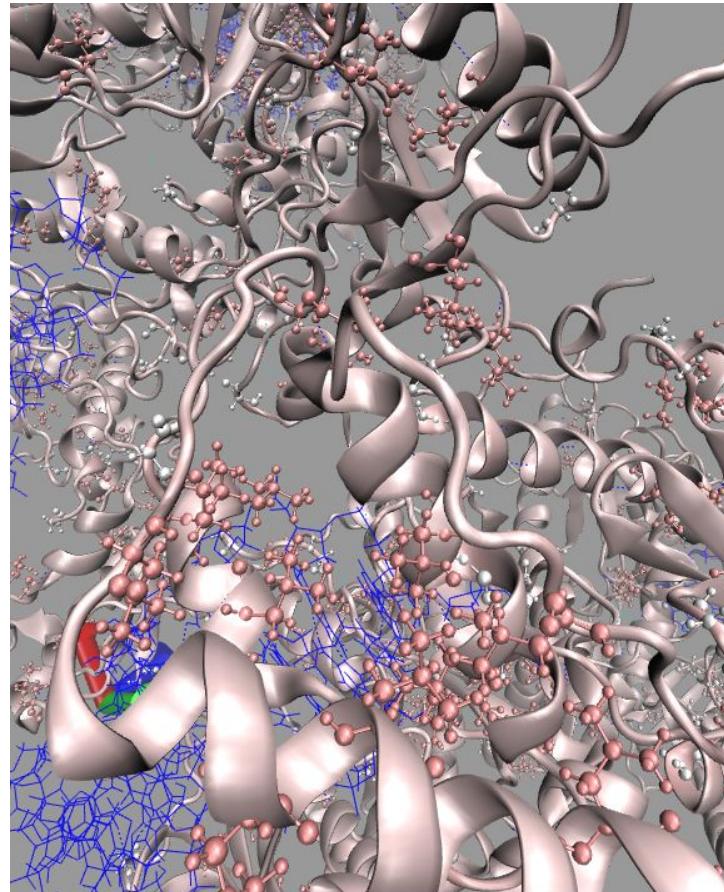
A G P Q L S R V G F D I N P A I S H H L A K L E . . . . . V P E A L S V S G R T I P I A Q G V V R I I G A . . . .  
T L D F L S I V G L K L N A D I C F T V G I K G Q P K Q K C T V L E A Q S F Y V G S S E I P S L K R T D E W K Y L G I . . . .  
I S N F S K V S G Y K I N V Q I S Q A F L Y T N . . . . . N R Q T E S Q I M N E L P F T I A S K R K I Y L G I . . . .  
V S E V S E A M G L H L N L S I C A K A H Y A P H G A G G A Q E A V E G A E G S R K G E I P I L G I R S T Y K Y L G V . . . .  
L V Q K C S E V G L E I N T G I T K V L R N R F . . . . . A D P S E V Y F G S P S P T Q L D D V D E Y I Y I G R . . . .  
C S G W A G A R G M E F E A T E T E L M H F T R . . . . . T R A P T E T L Q L G D T V L Q P T E S T R F L G V I W L D . . . .  
V E I W G A E V G D C L S T S I T V I M L L K G . . . . . A L R R A P T V R F A G R N L P Y V R S C R Y L G I T V S . . . .  
L S E W I K S R G L G I N P S I T E L V L F T N . . . . . K Y K I P P L N P P I L G R L S F S D S A S Y I G L I V I D . . . .  
F Q Q W A E N W V R I N A E I C A N V T F A N . . . . . R T G S C P G V S L N G R L I R H Q A Y K Y L G I T L D . . . .  
L E R W A R V N L M R F N T A C I C R V L H L G . . . . . W R N P R H L Y R L E G A V L E S S A E K D L G V I L M D . . . .  
I G N W G S Y S G A S L S L S I C Q H L H I C K . . . . . K H H C T S K I C S N N I Q I P T V T S L K I L G I T



L-LEUCINA

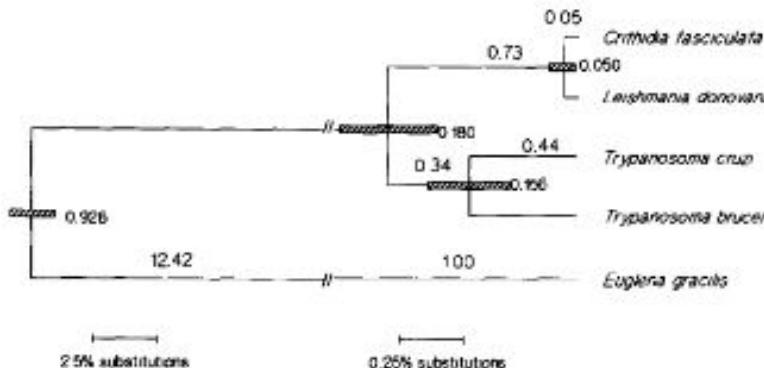
# REGIÃO IV

FOLHAS BETA E FINS DE LOOP



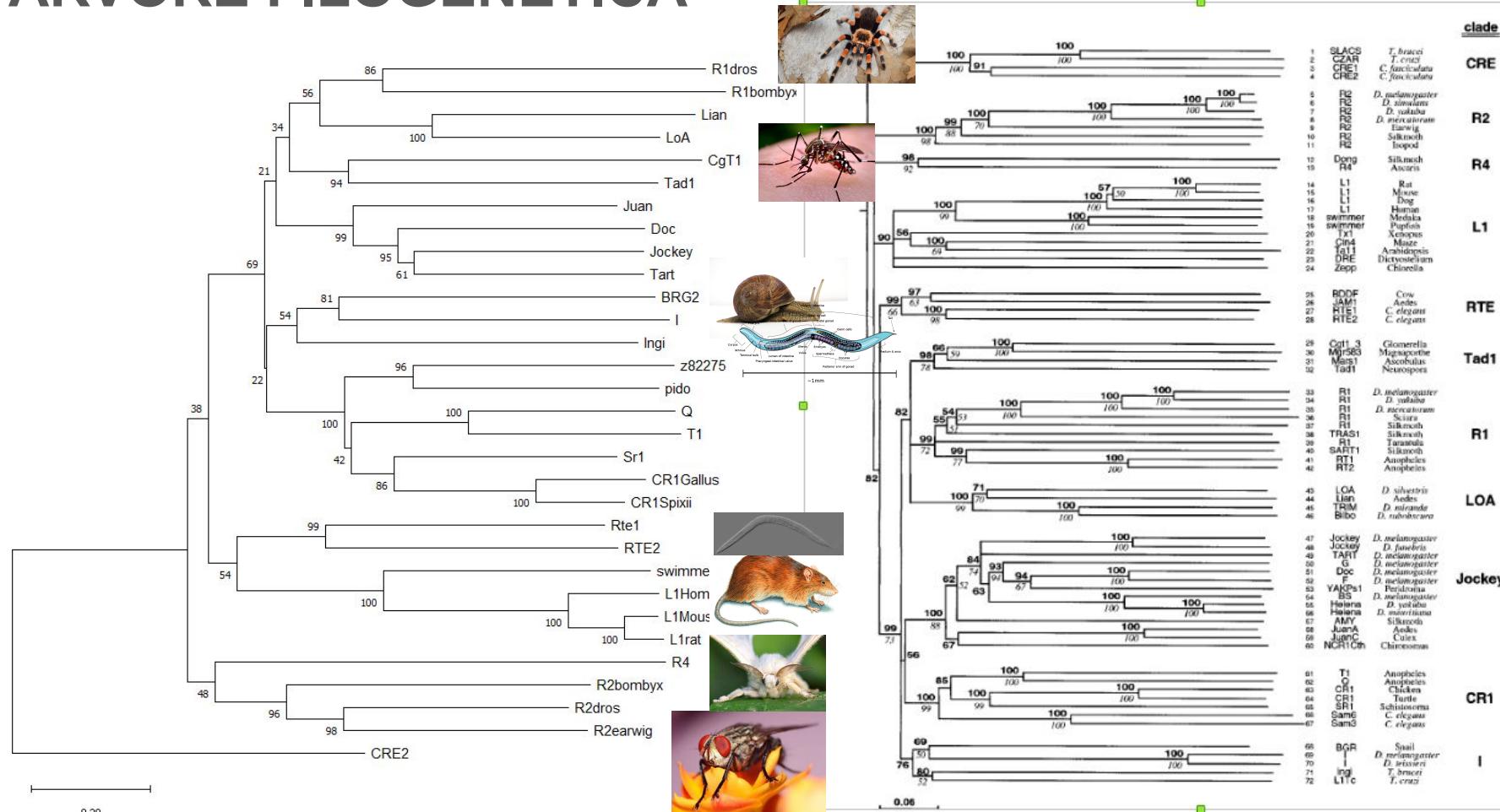
# ÁRVORE FILOGENÉTICA

- Métodos, parâmetros e escolha de raiz para construção da árvore filogenética
1. Neighboor joining method
  2. Valor de Bootstrap 300
  3. OUTGROUP: **CRE2 - *Critidia fasciculata***



**RNA RIBOSSOMAL**

# ÁRVORE FILOGENÉTICA



# **PARTE 2**

- Sequências de interesses / Sequencias Problemas
- Busca do domínio transcriptase reversa das sequências de interesse através de sites de predição de domínio
- Novo arquivo de filogenia: Sequências da Parte 1 + Sequências de Interesse
- Alinhamento Múltiplo
- Análise de regiões de interesse
- Construção de Árvore filogenética

# BUSCA PELOS DOMÍNIOS DE INTERESSE

#### **Resultados da pesquisa de sequência**

**Nostra** a descrição detalhada desta página de resultados

Encontramos 2 correspondências Rfam-A para sua sequência de pesquisa (1 significativa e 1 insignificante).

BVT 1

**Mostre** as opções de pesquisa e a sequência que você em

[Retorne](#) ao formulário de pesquisa para p

## Resultados da pesquisa de sequência

#### **Resultados da pesquisa de sequência**

Encontraram-se 4 novos candidatos PFCs. A soma que contribui de pessoas (2 significativas e 2 indiscutíveis)

**Mostre** as opções de pesquisa e a sequênci  
Retorne ao formulário de pesquisa para p

TPA: poliproteína de *Schistosoma mansoni*

GenBank: CA100246.1

## Gráficos de proteínas idênticas de GenPept

> CAJ00246.1: 498-749 TPA: poliproteína [*Schistosoma mansoni*]  
 PKVSSPSVPEFRPISVAPVATRCLLKLAKMLPFLQRFQARLNRDGFCAEVNLHSVRHVHTR  
 HTGASFALLDLSRDFATVDSHSDSIIIRAKRAYGAPELLCPRCLYNNYRTRSTCVNTELTHPTCGVKGDPLSP  
 LLFVIMVLDEVEGLDPMTHLTVDGSENLVIAYADALVUWNLNLARLQRKLDRDISILHLEAGHWSVNPESKR  
 TIUTSGHSKTTA1 SOTETTFAAGMTPRLP1 SAADTENYI GTK

## Resultados da pesquisa de sequência

#### **Resultados da pesquisa de sequência**

Encontraram-se 4 novos candidatos PFCs. A soma que contribui de pessoas (2 significativas e 2 indiscutíveis)

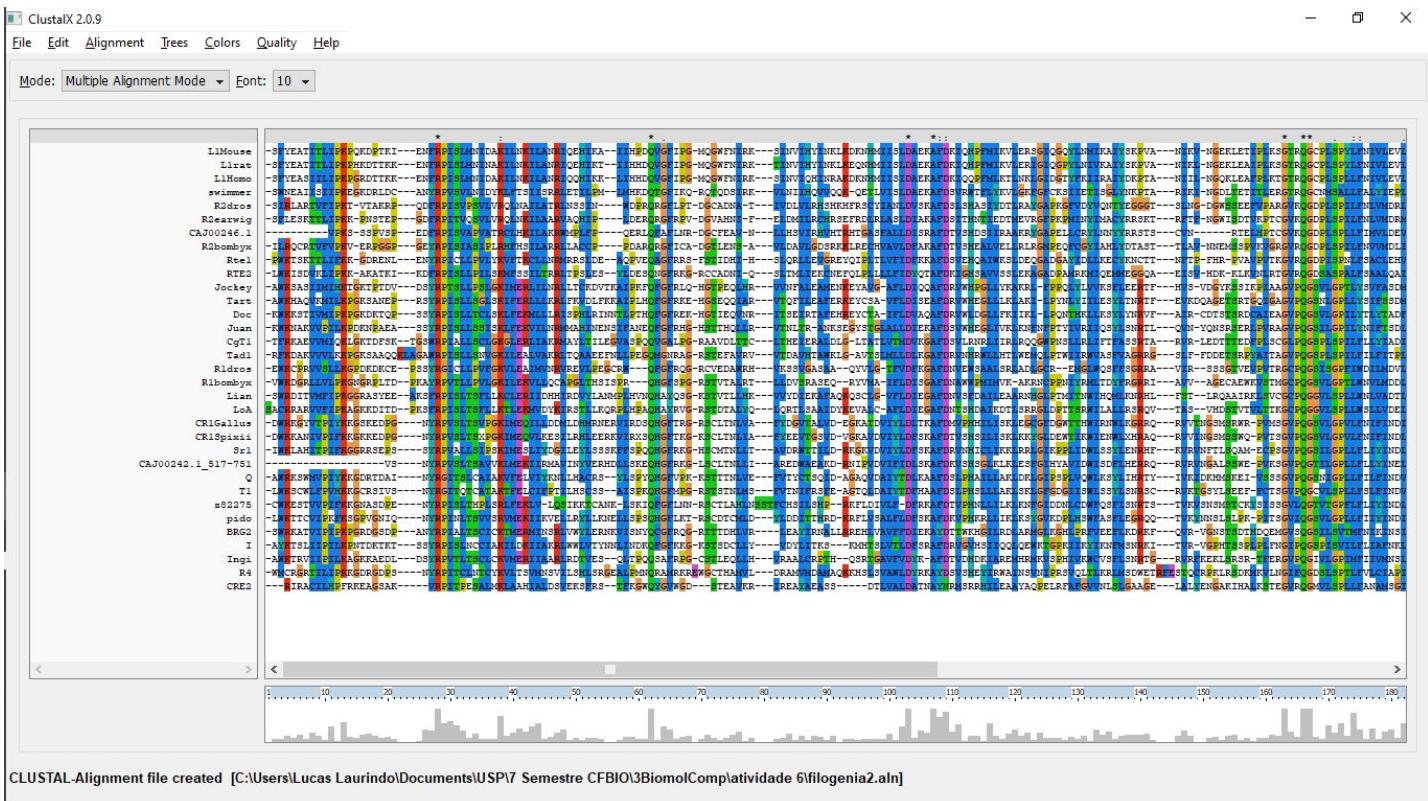
**Mostre** as opções de pesquisa e a sequênci  
Retorne ao formulário de pesquisa para p

TPA: endonuclease-transcriptase reversa [Schistosoma mansoni]

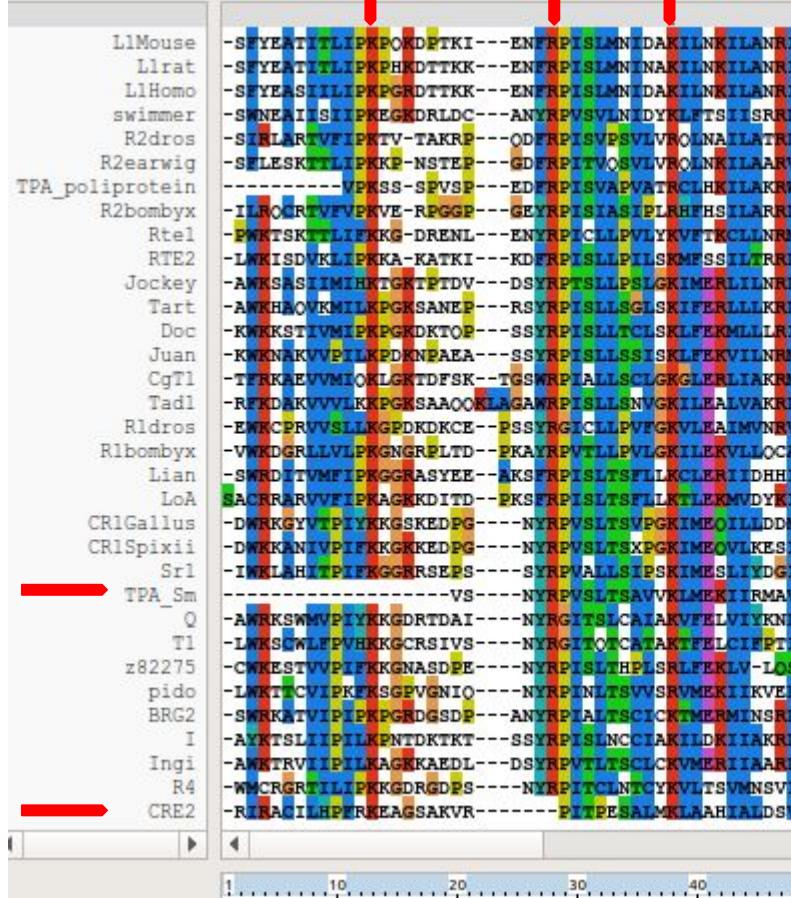
GasBank CA 100242

Gráficos de proteínas idênticas do GenBank.

# ALINHAMENTO MÚLTIPO COM SEQUÊNCIAS DE INTERESSE



# REGIÃO I



CRE2  
R2-Dm  
L1-Hs  
R4  
RTE1  
Tad1  
R1-Dm  
Bilbo  
Jockey-Dm  
CR1-Chicken

?SFATRIRACILHPFREAGSAK.....VRPITPESALM  
.NLPHSIRLARTVFIPKTVTAKR....PQDFRPISVPSVLV  
.ILPNSFYEASIIILIPKPGDRTTK...KENFRPISLMNIDAK  
(VTTGWMCRGRFTILIPKKGDRGD....PSNYRPITCLNT(YX  
.NVPKPWKTSKTTLIFIKGDRENL...ENYRPICLLPVLYX  
.HFPQRFKDAKVVVLKKPGKSAAQQKLAGAWRPISLLSNVG  
.YFFPAEWKCPRVVSVLLKGPDKDKE..PSYSYRGICLLPVFG  
.KIPRAWLKAKIVFIPKAGKPSHTT..PKDFRPISLSSFLX  
.YFPKAWSASIIMIHKTGKPTD...VDSYRPTSLLPGLX  
.EVPPDDWRKGYVTPYIKGSKED...PGNYRPVSLTSVPGX  
.HIPQAYKT5LIIPILKPNTDKTK...TTSYRPISLNCCIAX

## SEQUÊNCIAS PROBLEMA

- CAJ00242
- CAJ00246

Há perda de aminoácidos carregados

# REGIÃO II

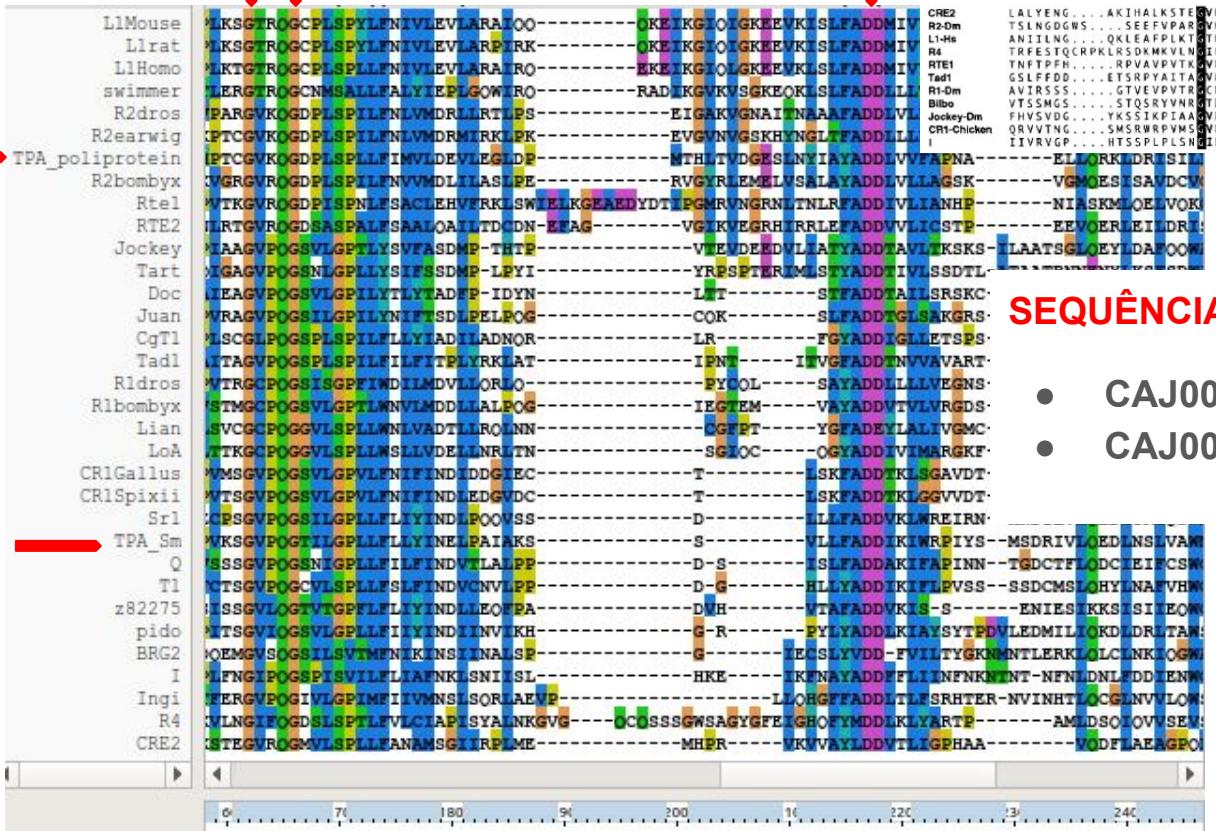


## SEQUÊNCIAS PROBLEMA

- CAJ00242
- CAJ00246

Conservado

# REGIÃO III

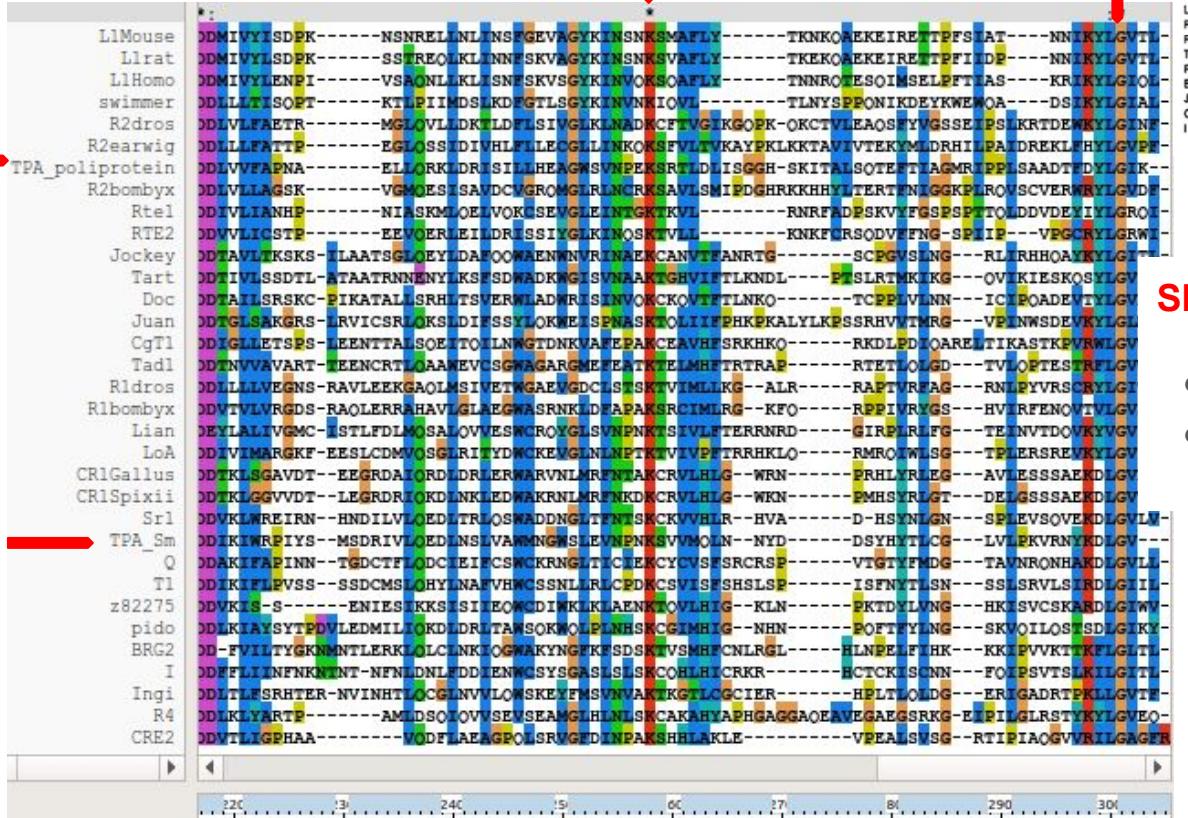


## SEQUÊNCIAS PROBLEMA

- CAJ00242
- CAJ00246

Conservado

# REGIÃO IV



```

CRE2          AGPQLSRVGFIDINPAKSHHLAKL.....VPEALSMSGRTIPIAQGVVIRLG...GFRGDTAS
R2-0m         TLDPLSIYGLKLNAQCFVTVGKIGKOPKOKCTVLEAOSFYVGSSIEPSLKRTEWKYLG...NFTATG
R1-Ms         ISNFVKSVSGYKINVQISQAFLYTN.....NROTESQIMNELPFTIASKRIKYLG...QLTRDWK
R4           VSEVEAMGLHNLNSQCAKHYAPHGAGGAQEAVEGAEGSRKG-----EORLPMEV
RTE1          LVQKCSEVGLGLENITGCTVLRNRF.....ADPSEVYFGSPSPPTQLDDVDEYTYLG...RQTNQANQ
Tad1          CSGWAGARGMEEFATVTELHMHTR.....TRAPRTETLQLGDTVLQPTESTRFLG...WLDRLKLNTVRAH
R1-0m         VEIWGAEVGDCLSTSIVIMLLKG.....ALKRPTVRFAGRNLPTVRSCRYLG...TVSEGKMFLL
Bill          LSEWTKSRLG1NPSHIELVLFN.....KYKIPPLNPPLINGCRLSFSDUSASYLG...LVIDKKLSSWN
Jockey-Dm   FQOWAENNNVRINAECANVTFAN.....RTGSCPVGVLNGRLIRHHQAYXYLG...LTLDRKLTFSKR
CRT-Chicken  IERWAVRNLMRFNTAICRVLHIG.....WRNPRHLYRLEGAVLESSSAEKG...LMDLKDLNNSQ
I             IGNWGSYSGASLSLSKQHLLICK.....KHHCTSKISCNNIQTPTVTSKL...LTLNNKYKWNTH

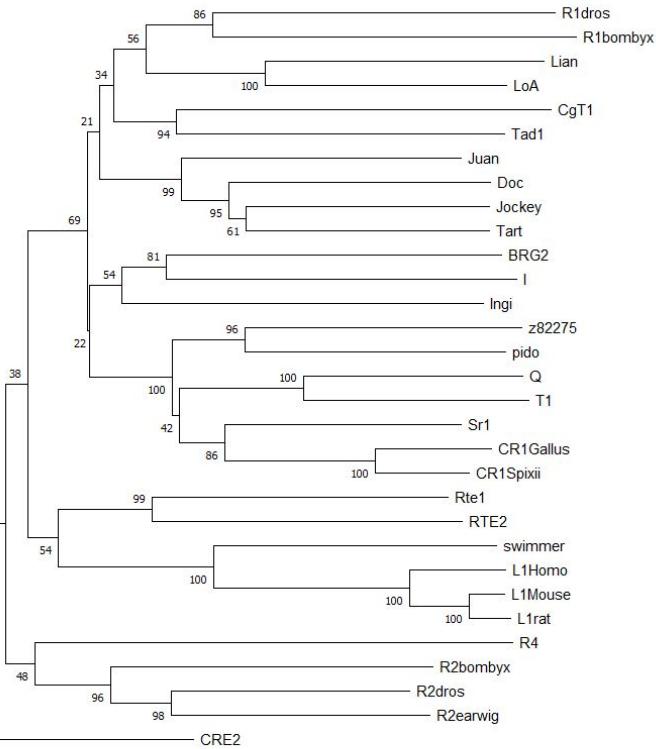
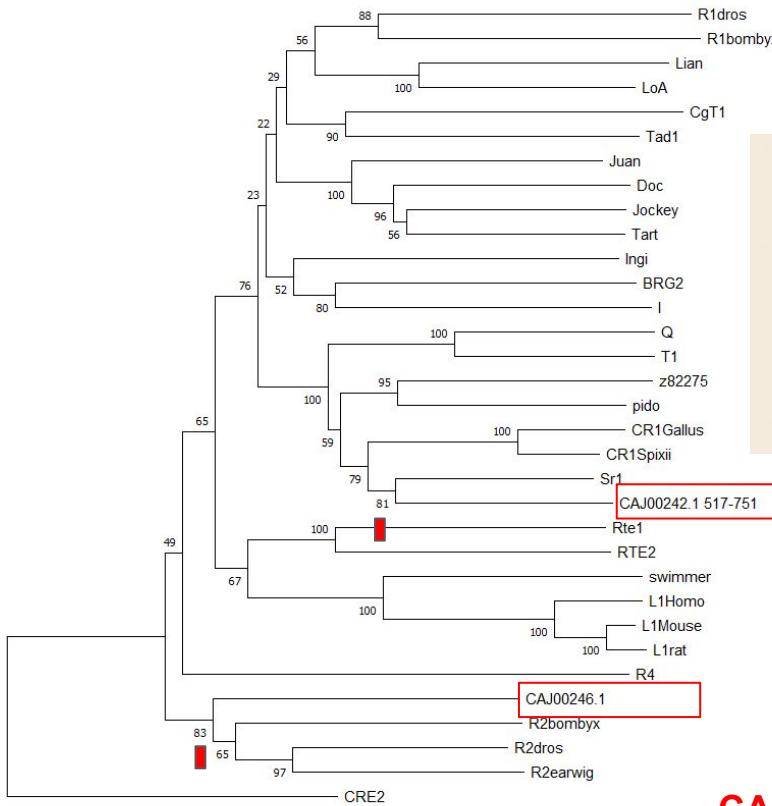
```

## SEQUÊNCIAS PROBLEMA

- CAJ00242
- CAJ00246

Conservado

# ÁRVORE FILOGENÉTICA



**CAJ00242-1 517-751 - CR1**  
**CAJ00246.1 - R2**

# CONCLUSÃO

As sequências problema apresentam perda de aminoácidos do possível sítio ativo do domínio. Isso pode acontecer pela perda de informação na inserção dos retrotransposons no DNA. Isso faz com que elas apareçam em grupos diferentes na filogenia, pois estes aminoácidos são altamente conservados e portanto há um grande peso em seu alinhamento.