Protomata-Learner

Learning Automata on Protein Sequences

François Coste, Goulven Kerbellec

Rencontres Bioinformatique de la Plateforme Genouest,
le 3 décembre 2007

Outline

Introduction

Bioinformatic problem

- Biological question :
- Computer science answer :

Bioinformatic problem

- Biological question :
  How to define signatures of known protein families ?
- Computer science answer :

Bioinformatic problem

- Biological question :
  How to define signatures of known protein families ?
- Computer science answer :
  Using machine learning algorithms !

Protein families

- Amino acid alphabet
  $\rightarrow$ {A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V}

Protein families

- Amino acid alphabet
  $\rightarrow$ {A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V}
- Protein sequence

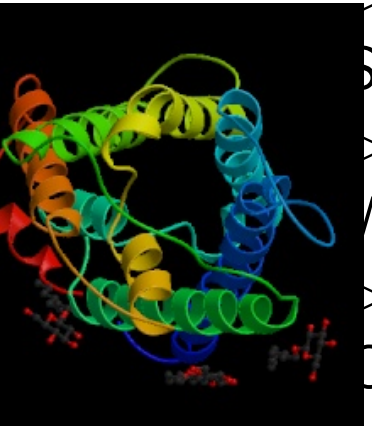
  >AQP1_bovin MASEFKKKLFWRAVVAEFLAMILFIFISIG-
  SALGFHYPIKSNQTTGAVQDNVKVSLAFGLSI...

Protein families

- Amino acid alphabet
  $\rightarrow$ {A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V}
- Protein sequence

  >AQP1_bovin MASEFKKKLFWRAVVAEFLAMILFIFISIG-
  SALGFHYPIKSNQTTGAVQDNVKVSLAFGLSI...
- Protein data set

  >AQP1_bovin MASEFKKKLFWRAVVAEFLAMILFIFISIG-
  SALGFHYPIKSNQTTGAVQDNVKVSLAFGLSI...
  >AQP2_rat MWELRSIAFSRAVLAEFLATLLFVFFGLGSALQ-
  WASSPPSVLQIAVAFGLGIGILVQALGH...
  >AQP3_mouse MGRQKELMNRCGEMLHIRYRLLRQALAE-
  CLGTLILVMFGCGSVAQVVLSRGTHGGFLT...



Protein family & Natural selection properties

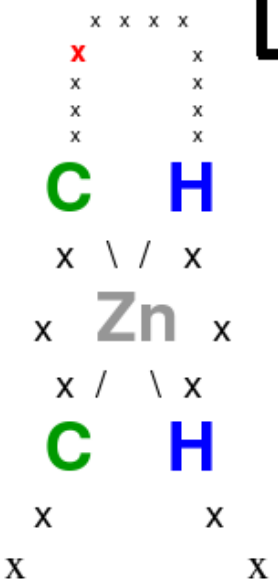
- Common function
- Common topology (3D structure)
- Common signature

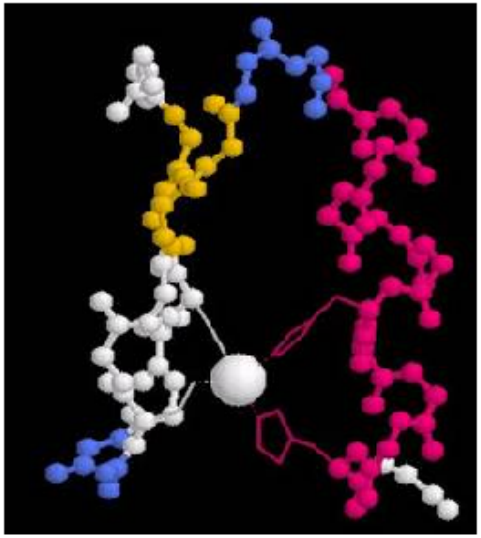
Pattern of the zinc finger protein family

```
ZBT11 ...Csi..CgrtLpklyslriHmlk..H...
ZBT10 ...Cdi..CgklFtrrehvkrHslv..H...
ZBT34 ...Ckf..CgkkYtrkdqleyHirg..H...
```

Zinc Finger Pattern

C-x(2,4)-**C**-x(3)-[LIVMFYWC]-x(8)-**H**-x(3,5)-**H**





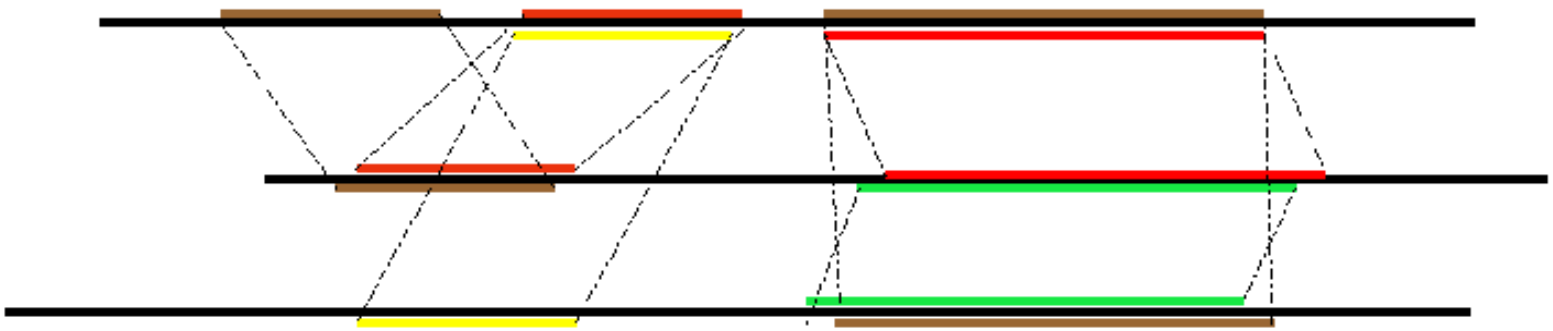
Characterization

Significantly similar areas

- The first phase of the approach is based on the similarity in protein sequences.
- The problem is how to describe and modelize the significantly similar areas.
- And how to optimize the process of identification.

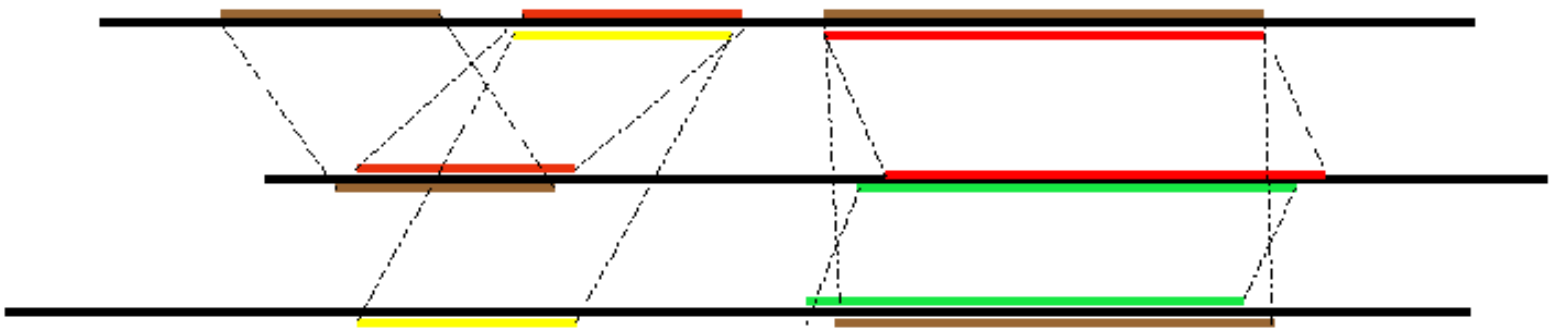
Similar Fragment Pairs Score

Significantly Similar Fragment Pair (SFP) :



Similar Fragment Pairs Score

Significantly Similar Fragment Pair (SFP) :

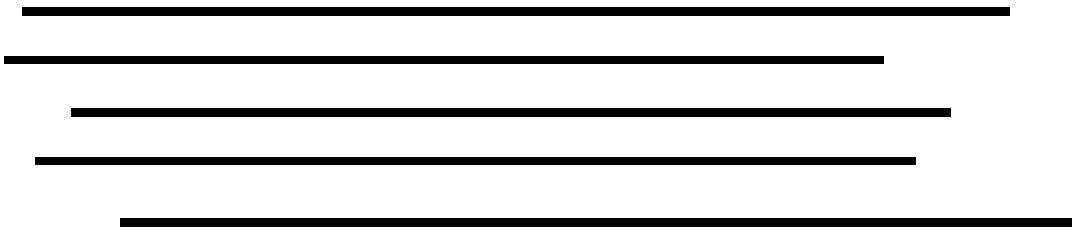


SFP : (f_1, f_2) s.t. $w(f_1, f_2) > 0$ DIALIGN [Morgenstern et al...]

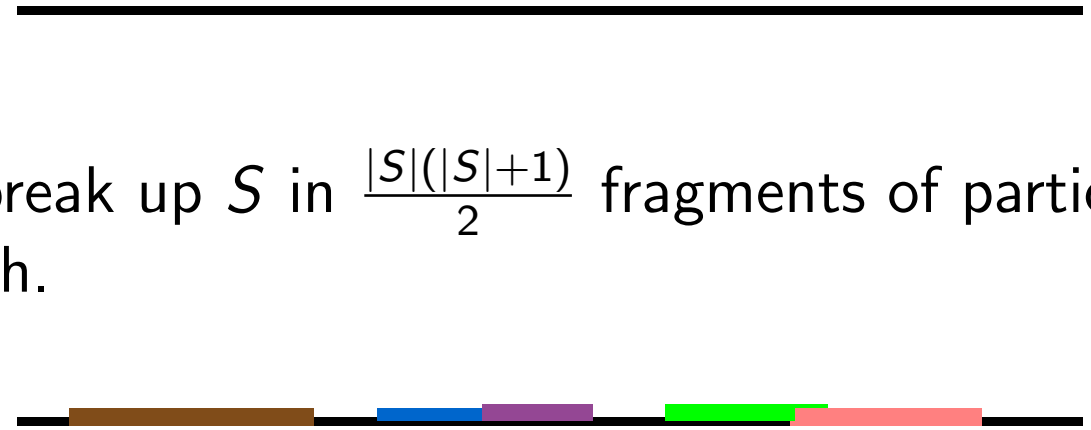
$w(f_1, f_2)$: weight of a fragment pair $= -logP(s, l)$
$P(s, l)$: probability for a random fragment pair of length l to have a similarity greater than s

s : similarity of (f_1, f_2), l : length of f_1 and f_2

From a set of protein sequences P ... to a graph of fragments G

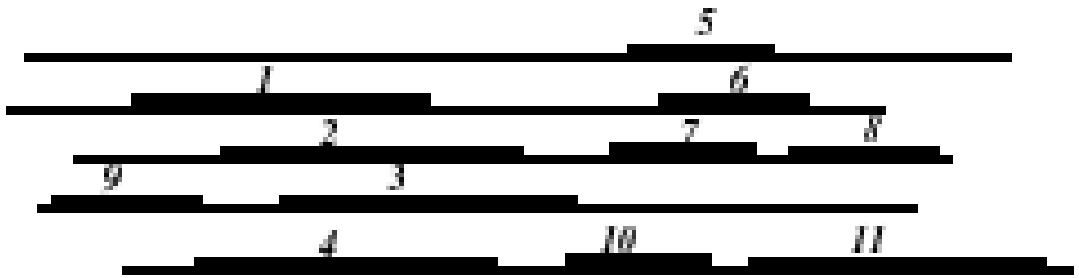


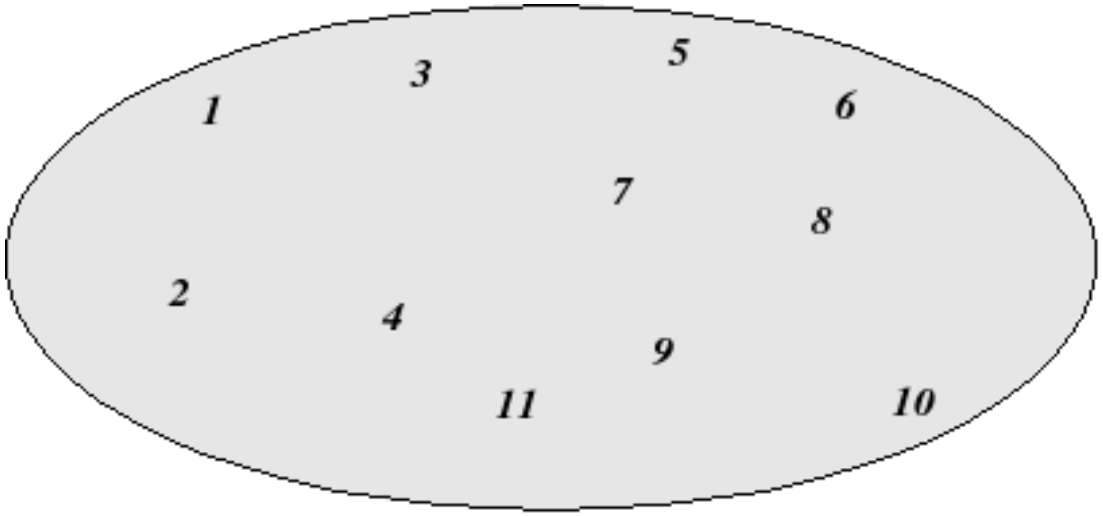
For any sequence S from P



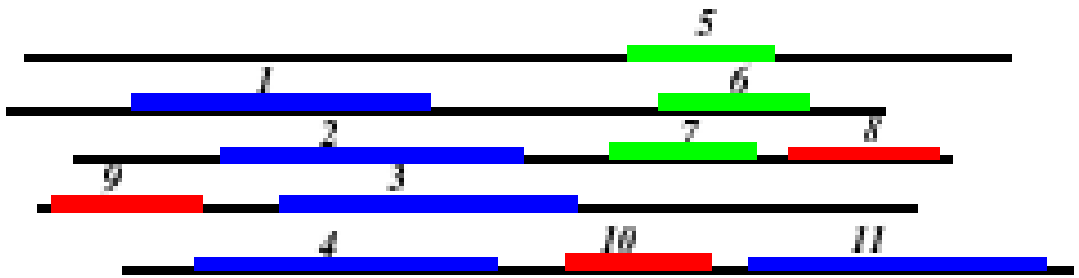
- We can break up S in $\frac{|S|(|S|+1)}{2}$ fragments of particular positions and length.

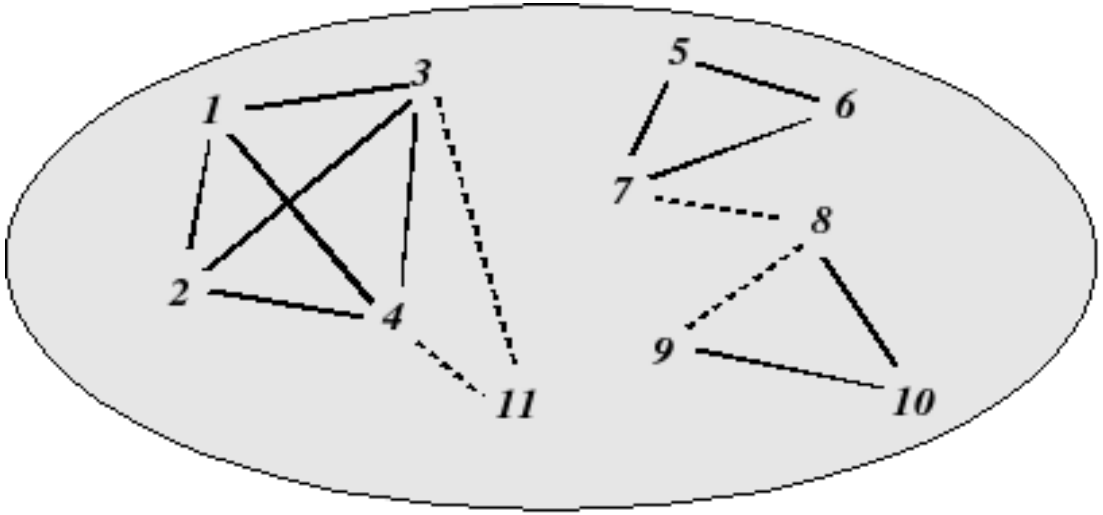
The graph G where nodes are the fragments of P





The graph G where edges shows the similarity level between fragments





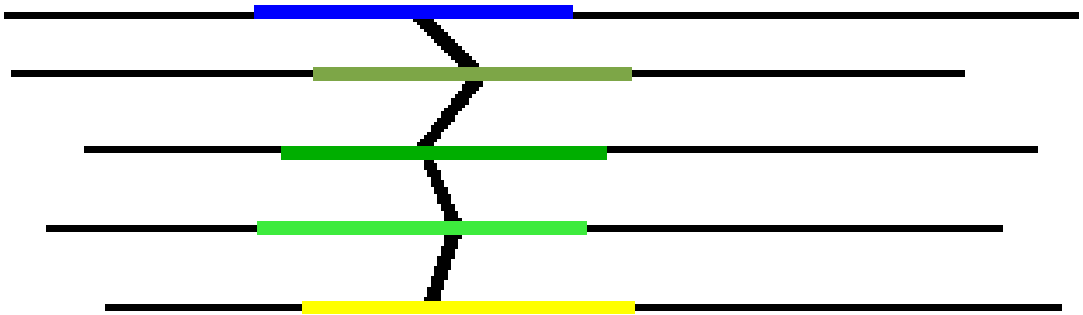
PLMA block

Definition

A PLMA (Partial Local Multiple Alignment) block is a significantly similar set of fragments that testifies a characteristic area in a set of biological sequences. It's a connected subgraph of the graph of fragments.

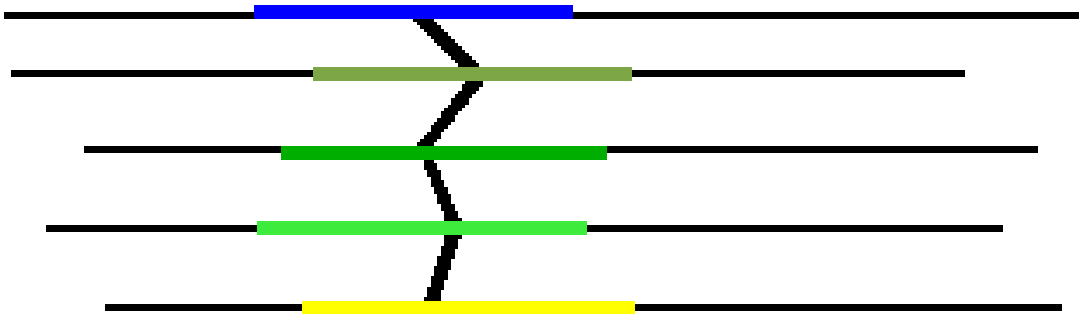
Weak or Strong consensus ?

- A pathwise connected involves a weak consensus by transitivity

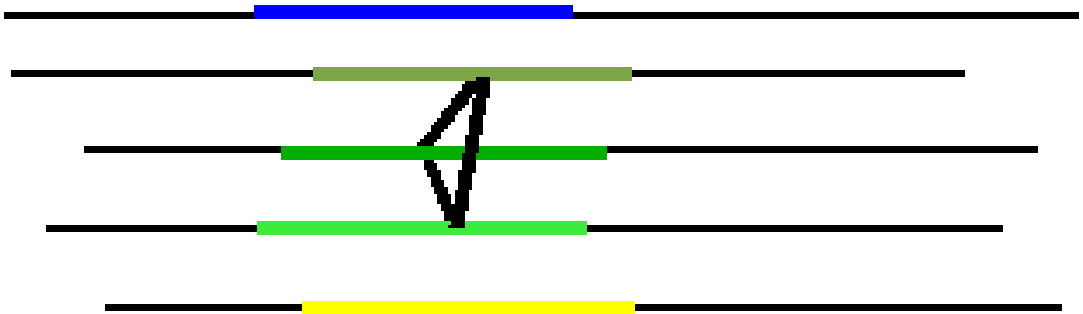


Weak or Strong consensus ?

- A pathwise connected involves a weak consensus by transitivity

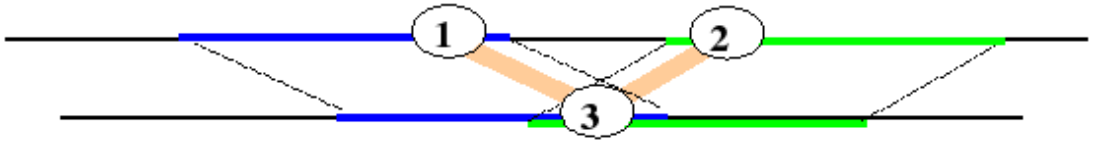


- A clique involves a strong consensus



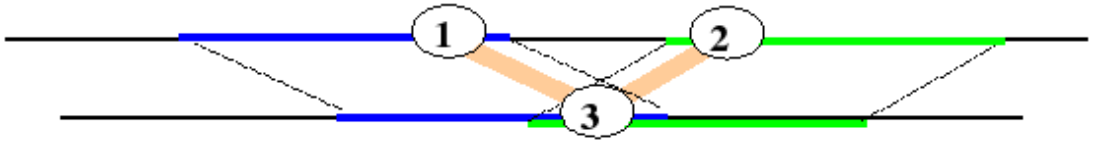
Incompatibilities between SFPs

- Preservation

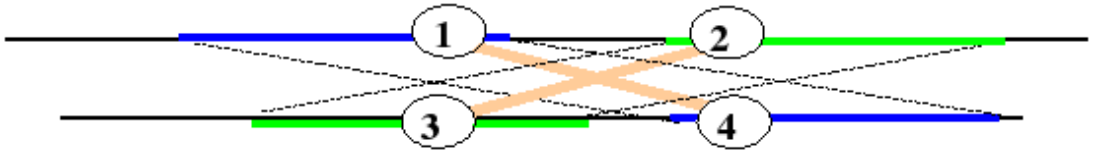


Incompatibilities between SFPs

- Preservation

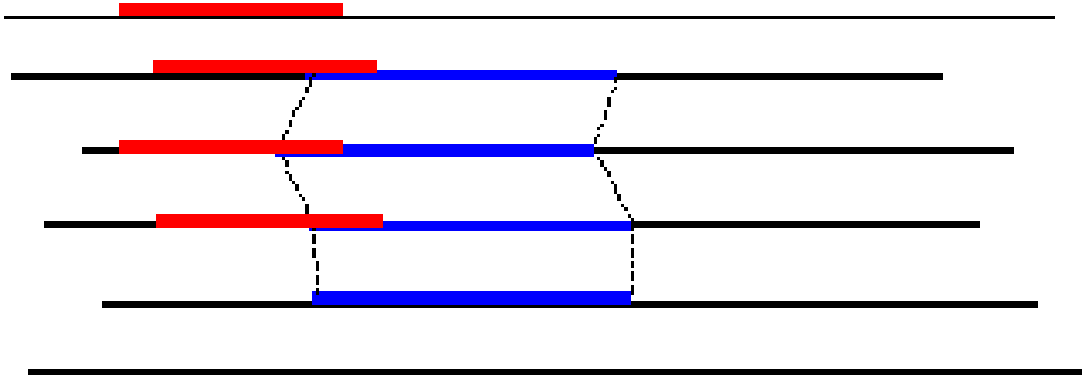


- Inconsistency



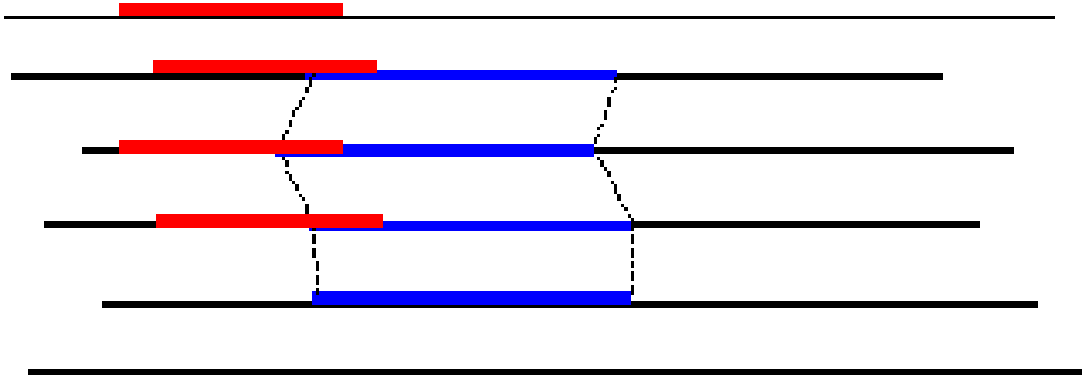
Incompatibilities between PLMA blocks

- Interference

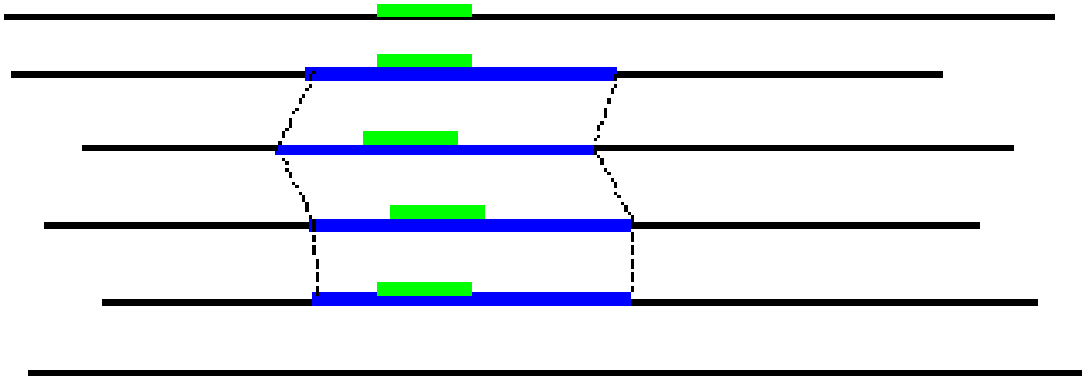


Incompatibilities between PLMA blocks

- Interference

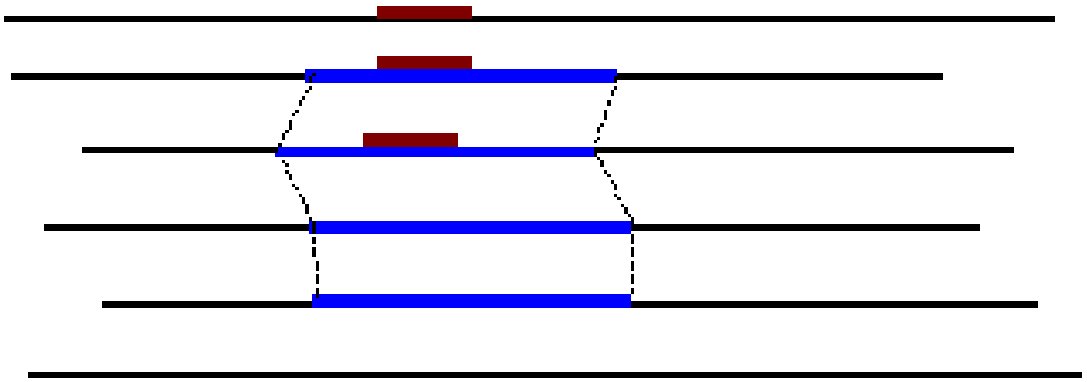


- And yet



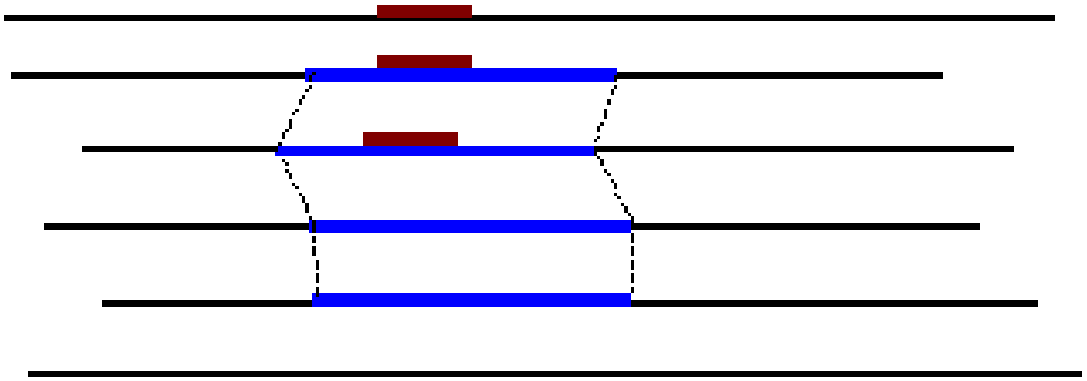
Incompatibilities between PLMA blocks

- Intersection

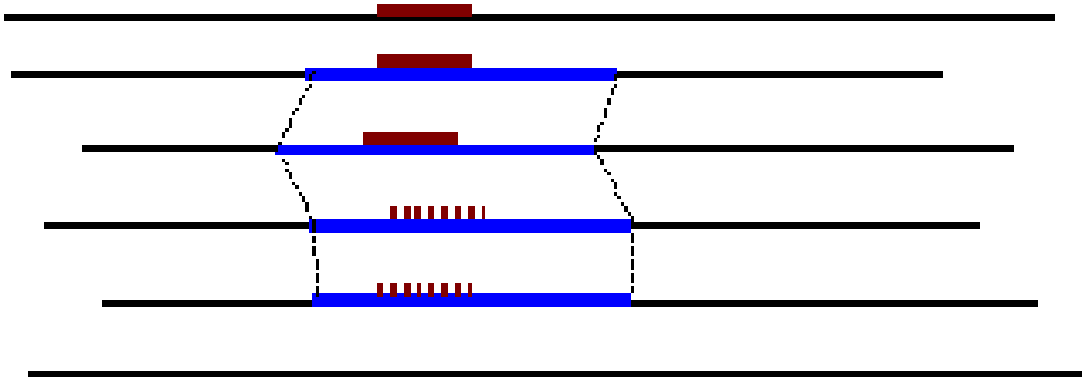


Incompatibilities between PLMA blocks

- Intersection



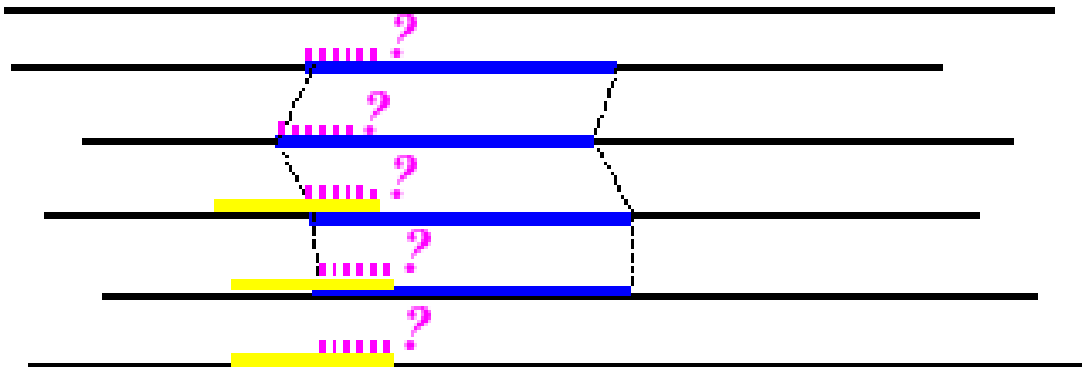
- Allowed



Preservation property of a PLMA block

Invariant

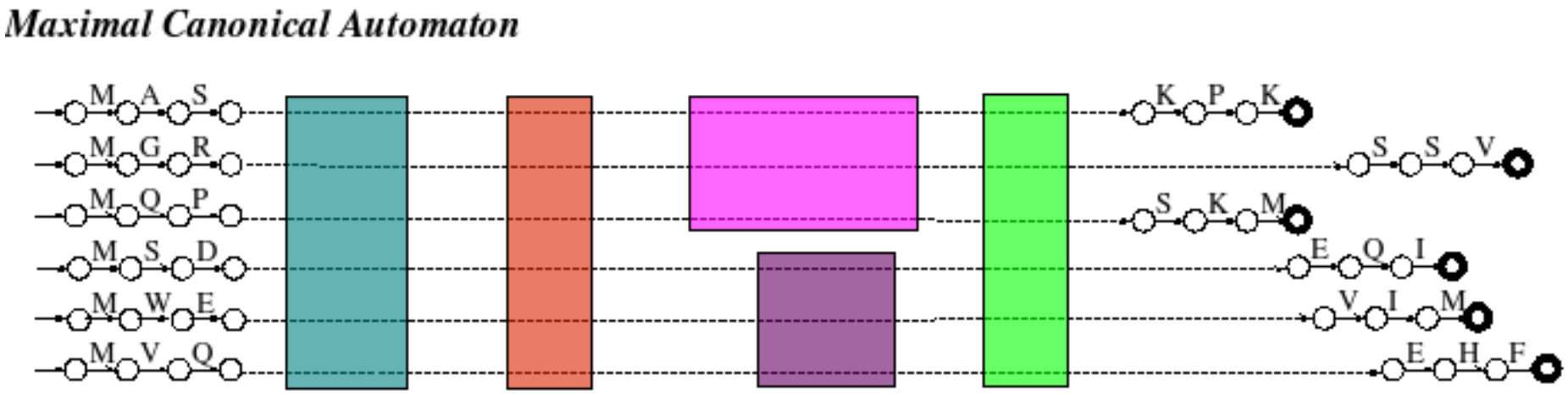
For each pair of aligned positions $P1$ and $P2$,
There exists an SFP that aligns $P1$ and $P2$.



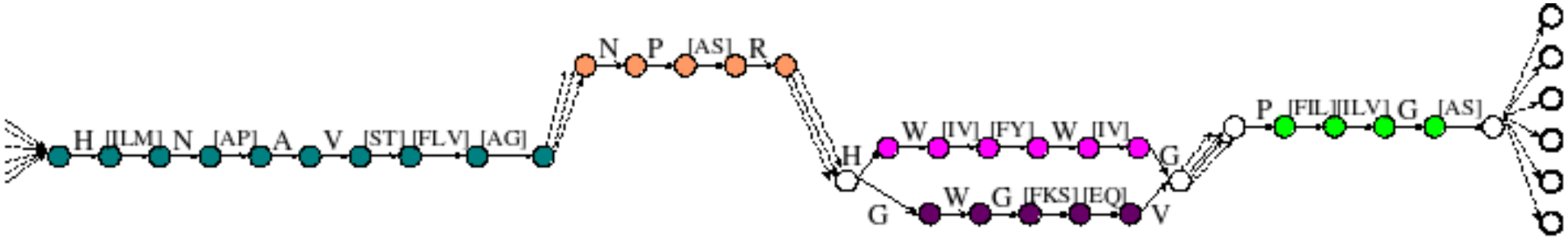
Generalization

PLMA blocks merging

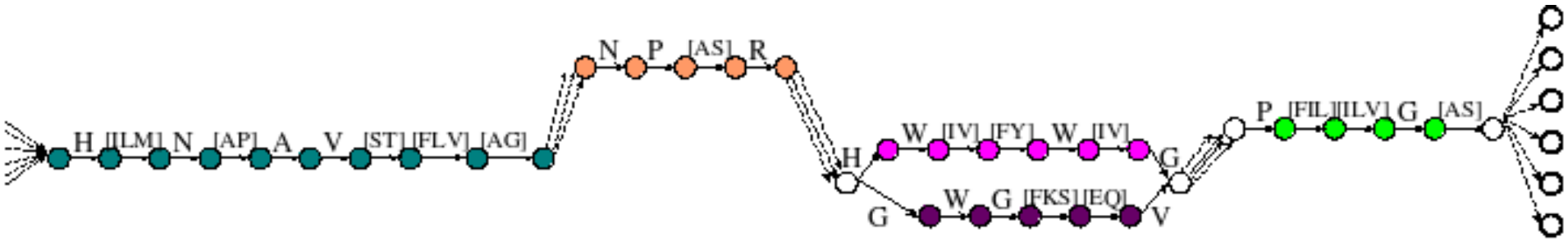
Characterization stage → Partial Local Multiple Alignments (PLMA) :



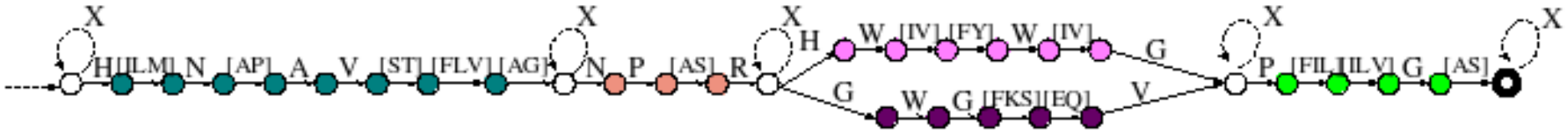
Merging PLMA blocks :



Gap generalization



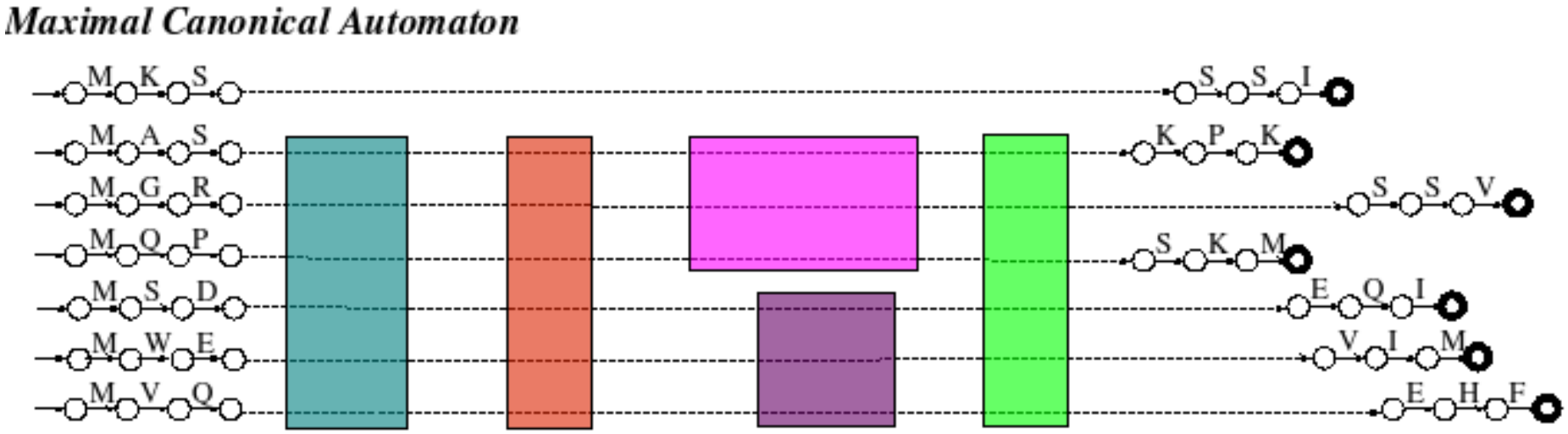
Transitions used less than a given quorum are treated as gaps :
(if not in an exception path)



Representative segments : sub-paths used more than quorum

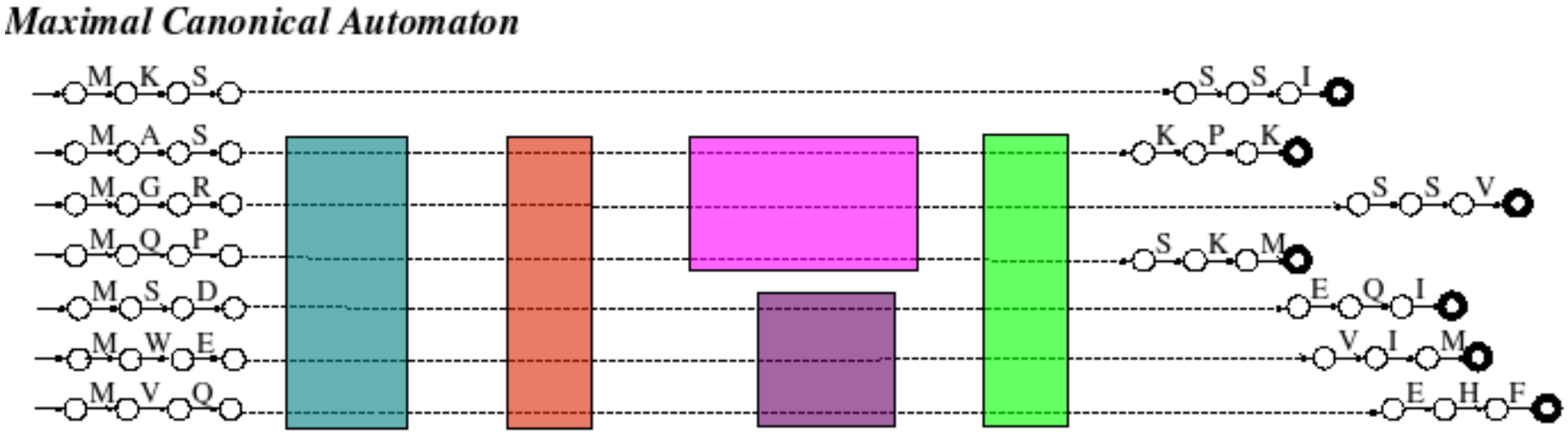
What if ?

Characterization stage → Partial Local Multiple Alignment (PLMA) :

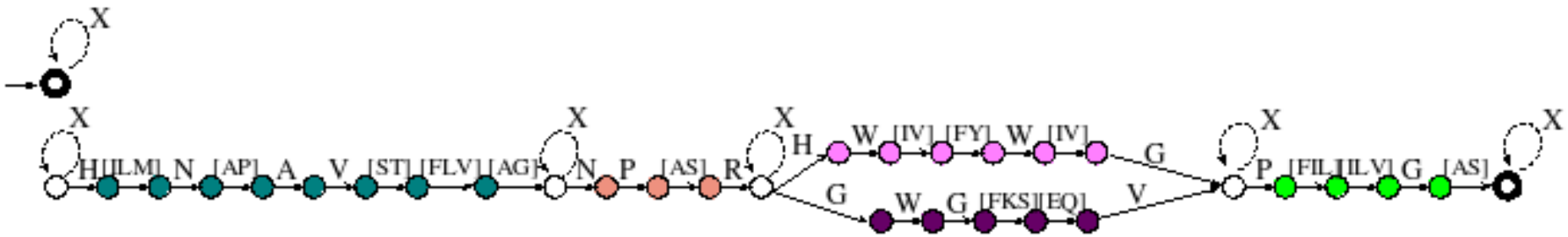


What if ?

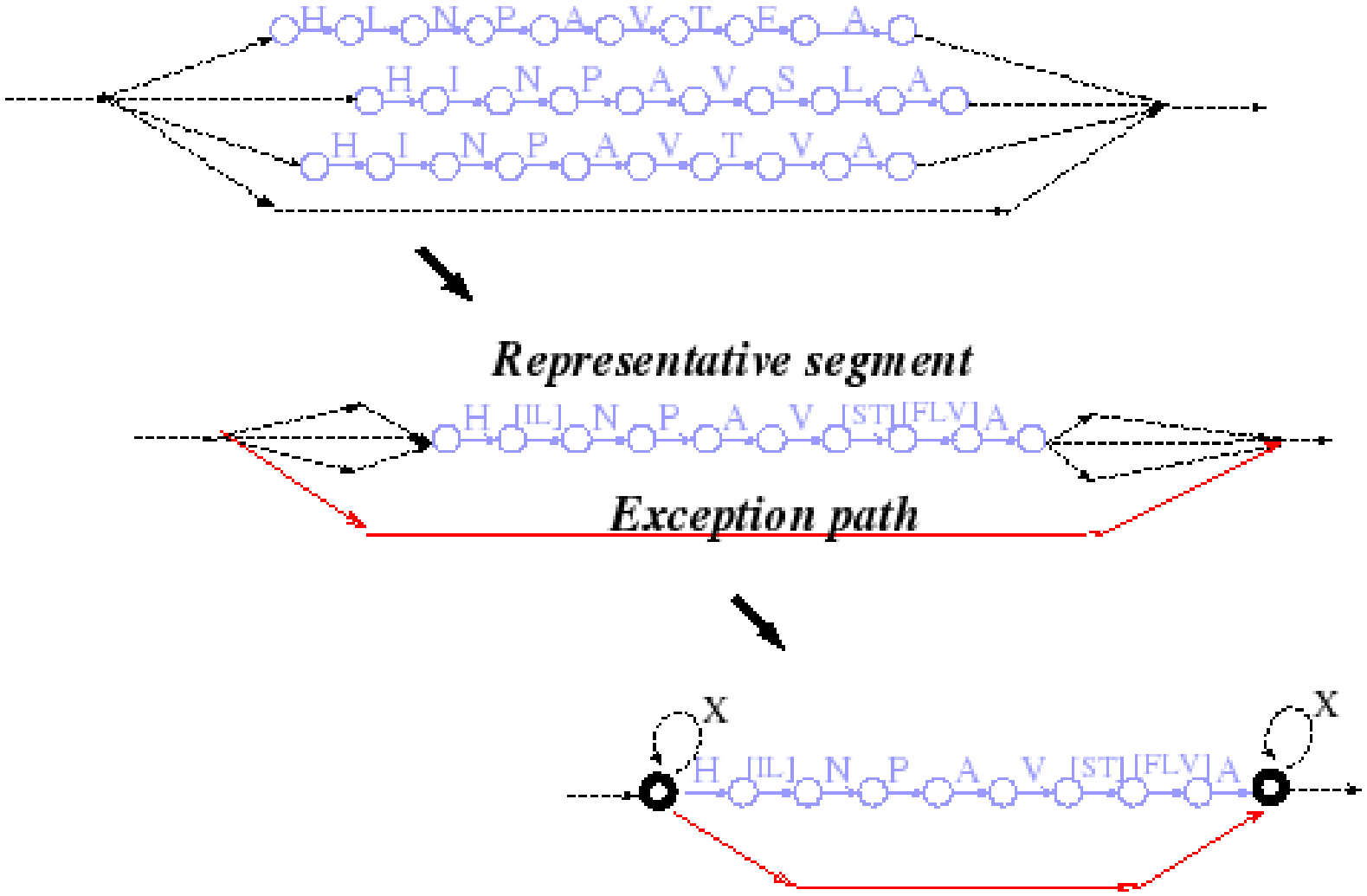
Characterization stage → Partial Local Multiple Alignment (PLMA) :



Merging PLMAs :

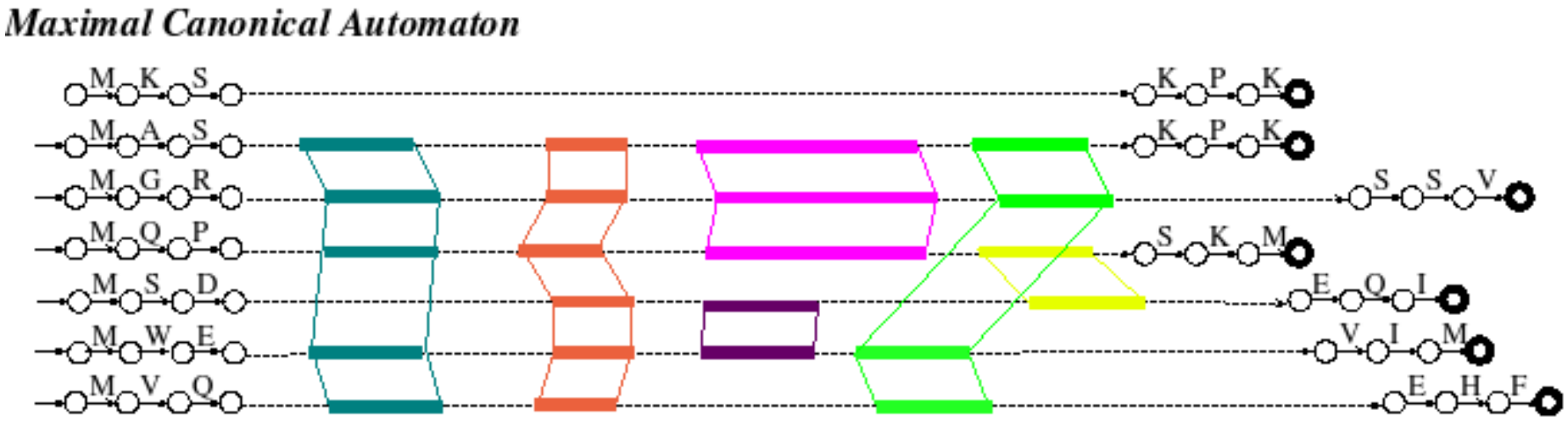


Exceptions

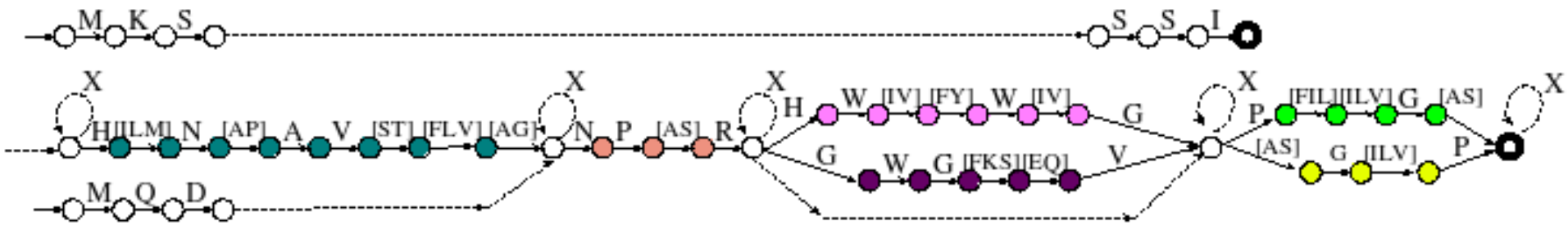

H L N P A V T E A
H I N P A V S L A
H I N P A V T V A
Representative segment
H [IL] N P A V [ST] [FLV] A
Exception path
X
H [IL] N P A V [ST] [FLV] A
X

Exceptions

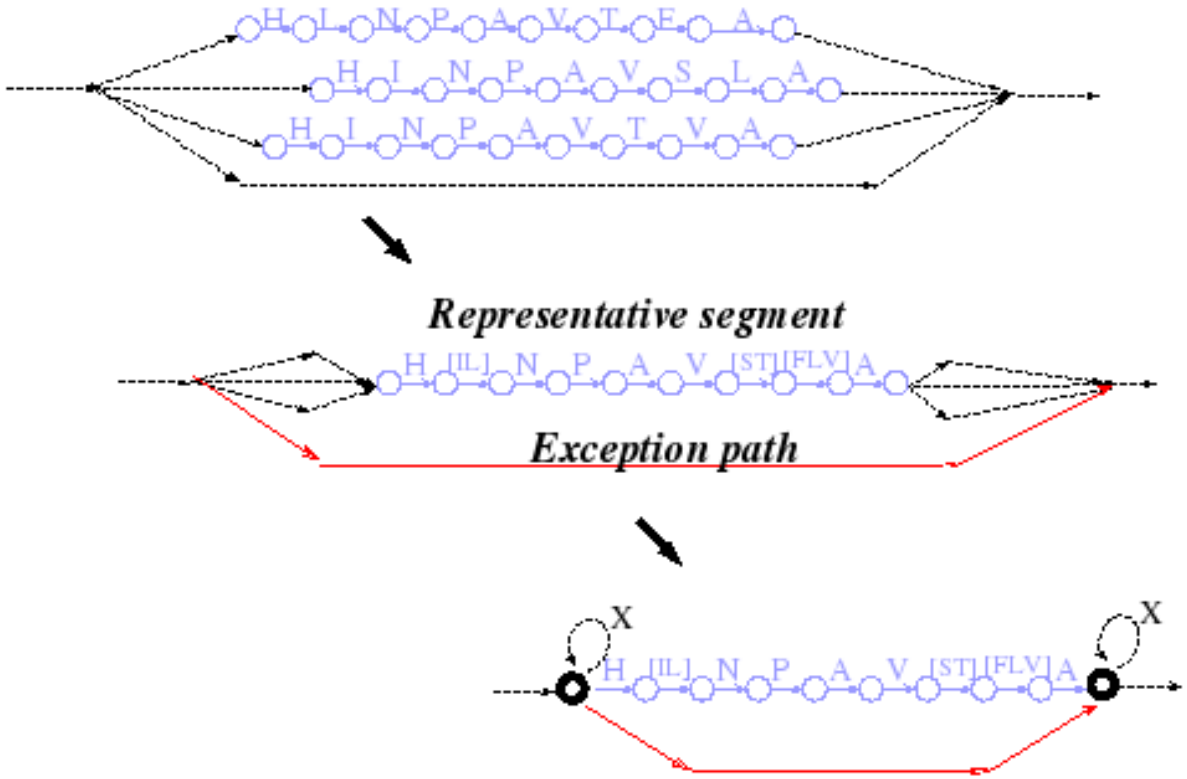
A more realistic result of the characterization stage :



Merging PLMAs, exception detection, gap generalisation :



Exceptions

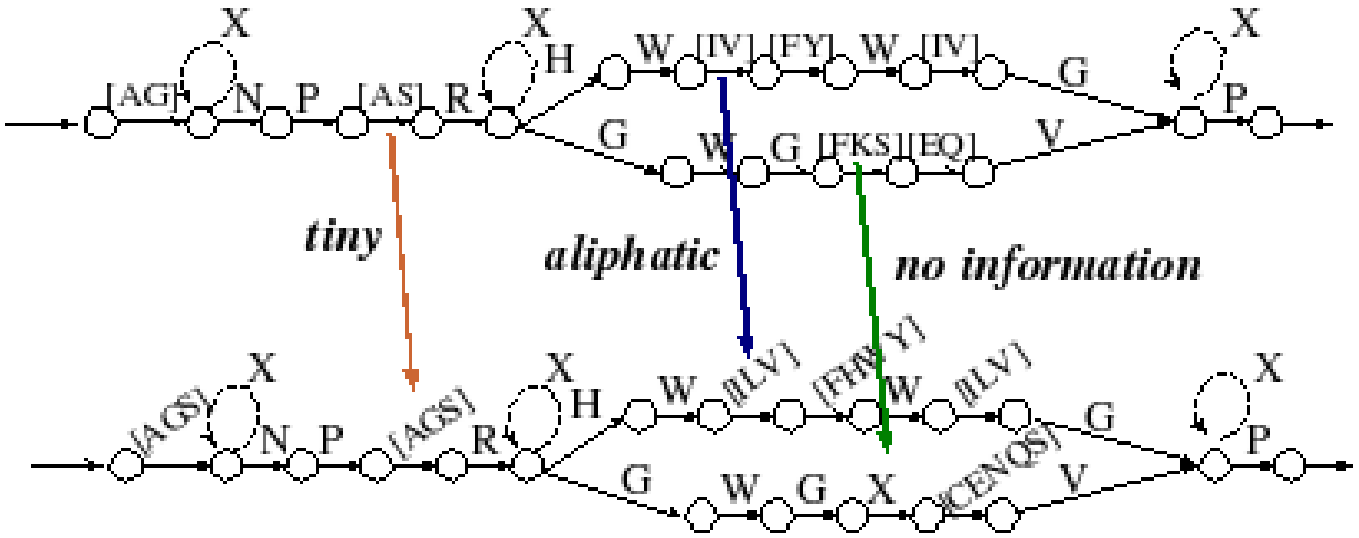


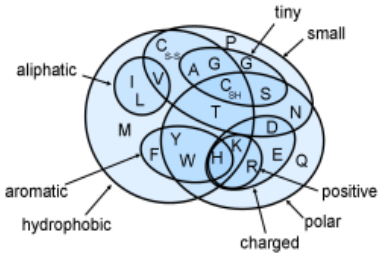
Detection

Transitive closure over accessibility graph between representative segments. Exceptions are edges between two representative segments s.t. a path of length ≥ 2 between them also exists.

Identification of informative positions

- Representative segments : average conservation of sub-sequences (BLOSUM)
- Identification of actual important *physico-chemical properties*

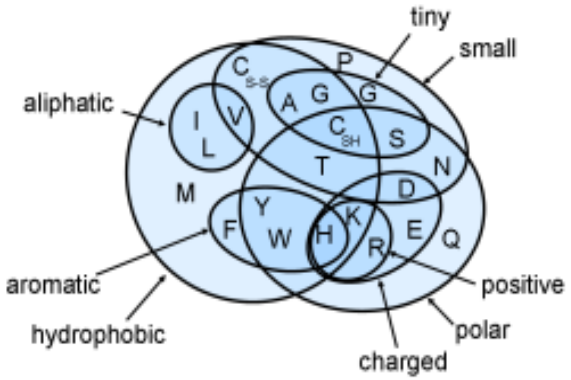




Likelihood ratio tests...

Testing expansion by likelihood ratio test

P : multiset of amino acid at a given position
G : smallest group containing aa of P
Σ : amino acids alphabet
$\lambda_G, \lambda_\Sigma$: thresholds



- Expansion $P \rightarrow G$

$$\text{Reject if } LR_{G/P} = \frac{L_G}{L_P} = \left(\frac{\sum_{a \in P} p_a}{\sum_{a \in G} p_a}\right)^n < \lambda_G$$

- or else Expansion $P \rightarrow \Sigma$

$$\text{Reject if } LR_{\Sigma/P} = \frac{L_G}{L_P} = (\sum_{a \in S} p_a)^n < \lambda_\Sigma$$

Demonstration

protomata-learner.genouest.org

Protomata Learner Parameters - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

http://protomata-learner.genouest.org/

Plate-forme Bio-informatique Genouest

Protomata-Learner version 0.07 (warning : Beta version)

Protomata-Learner documentation

Examples data

Description of your request: no_named_request

Email (optional)

Partial Local Multiple Alignements (PLMA)

Sequences

Advice : reduce the redundancy in your set of sequences, for example use the program decrease redundancy at Expasy.

Upload your sequences file (FASTA) Parcourir...

Or paste your sequences

Viral nucleic acid binding Clan

Carla_C4.fas

```
>Q02123|VNBP_POPMV
MVNMRKVLALMQVFRERYDHKCDFNFCDIAVSIVCRSELDFINEPGLSNYAKRRRARRLG
RCVRCFRVNPGFYFTKRCDGITCVPGISWNYDVEDYIKRGRVTGDRETPSTFHGYGYPVG
HKT
>P16654|VNBP_PVSP
MKAERLEMLLLCVYRLGYILPVDVCIKIISVAQVSVQGRSTYSCKRRARSIGRCWRCYRV
YPPVCNSKCDNRTCRPGISPNFKVVTFIRGWSN
>Q00572|VNBP_HELVS
MDKRNKANVVLSLCSMFASRGNCIPIPIVFNIYMRAFPKLVGRGTSTYARRRRARSILRC
ERCYRVYPPLPFSKKCDNRTCVPGISYNIKVADFIKWGVTEVIPHPGFNF
>P27336|VNBP_LSV
MSVWGAWKPNTPVGYKELKSSEIIDTQIMDEALKRRTTIVLCLLSAFPRDICRDILRRTS
SHIVGLGRSRYARRRRALQIGRCERCYRVYPPVCGSKCDNKTCRPGLSINTNVANYIDHG
VTEVIPWISPHRGQFYLRPK
>Q01687|VNBP_PVMG
MKDVTKVALLIARAMCTSSGTFVFELAFSIAECAGRPLGGGRSKYARRRRAISIARCHRC
YRLWPPTVFTTRCDNKYCVPGISYNVRVAQFIDEGVTEVIPSVINKRE
>P37992|VNBP_CVB
MDVIVKMLILRKFVEQGNVCPIHLCVDIYKRAFPRSVNKGRSSYARRRRALELGRCHRCY
RVYPPLFPEISRCDNRTCVPGISYNSKVRDYILWGVTEVIPHPGYNF
>P22625|VNBP_CLV
MRERKLRKQLEDLFKRFASVQHGHSDCINIIIAKIKSDQPGESKYARRRRAKSIARCPRC
ARVSPGFYFTTRCDGKTCRPGLSARPDLLEFIGIDLCVRSK
>O38024|O38024_9VIRU
MKCVTQTALLVARALYLHSGVFVFEIAYLISTRAGRPLGGGRSKYARRRRAIAAGRCHRC
YRLWPPTVFTIKCDNRSCVPGISYNVRVSASR
>O09491|O09491_9VIRU
MFSKTQFQRESMLIKQRTYRRLLRAIFKLHTNKNCVDVIDIIVNKIVCGRTGASKYARAR
RAKSIGRCPRCFRCAPGFYFTKNCDTKTCTPGISYNEKVKNFIVDGVTM
>O41484|O41484_9VIRU
MKADRLATLLLCVHRLGYVLPVEICVNIISLSAGPISRGRSTYGRKRRARSIGRCWRCYR
VYPPICNSKCDNRTCRPGISPNYNVVTFIRGWSN
```

u-----XEmacs: Carla_C4.fas (Fundamental)----All----------

Viral_NABP.fas

```
>Q04580|VNBP_SHVX
MHPHDLNLLCCLHFSKPSLPNDLKTLLFRACETSCKLNRRLLDNKPFQGTSKCAKRRRAK
RYNRCFDCGAYLYDDHVCKRFTSRSNSDCLSVIHQGPAKLYAEGAYRANSDAEQLIMNDM
LLIKSLKL
>Q67698|Q67698_9VIRU
MHTYDFNFLACLQFAQPNLPSDVRISIYMLSSASRKIGRKSQQNKPFTGTSKCAARRRAK
RYNRCFDCGALLNTDHVCKLFTSRASTDCLHVIREGPAKLYAERTFRKSSFAEQLILDDL
ELMKLYE
>Q67702|Q67702_9VIRU
MHPHDFNLLCCLHFSQPPLPKDLKEFIFSRCVDCCKLVRRTNQNKPFQGVSKCAKRRRAK
RYNRCFDCGAFLVDGHRCKVFVSKAHSDVLAVIHEGPAKLYAERSYRPNSDAAQLIENDI
LYIRALKL
>Q8QXW6|Q8QXW6_9VIRU
MHRYDFNLLACLHFSKPTLPNDVRIIIYNLTLNSRKLGRKAQQNKPFTGTSKCAARRRAK
RYGRCFDCGALLEANHVCKLFTTRSNSDCLSVIHEGPAKLYAERSFRKSSFAEQLMRNDL
QLMKLYK
>Q67694|Q67694_9VIRU
MIPQDFNILCCLHFAKPFIPQDLKAHLFFTCVNECKLVRIARENKPFLGTSKCAQRRRAK
RYNRCFECGAYLLDNHKCRIFVSRAQSDVLAVIHEGPAKLHAERTYRPNSDAALLIESDL
QYIKLFQNRKA
>Q67662|Q67662_9VIRU
MHPYNLNLLCCLHFSKPTLPLELRLYIYNLAAPNLLLGRKMLQNKPFQGTSKCAARRRAK
RYNRCFDCGGYLLNNHVCKQFPTRASTSCFNVIHEGPAKLYAEGAFRRDSFAEQLILNDL
ELMKLHNN
>Q8QXV8|Q8QXV8_9VIRU
MHPHDFNLLCCLHFSPNHLPKDLQILIFSRVTTECKLRRLSENKPFRGRSKFAMRRRAKR
YNRCFDCGAVLIDDHQCKVLTSRAQSDVLTVIREGPAKLYAERTYRPNSDAARLIEDDIL
YIKTLKL
>Q9QEE6|Q9QEE6_9VIRU
MEPHDQSGSTTRQLDEIRDRRGSQIRSVRLLPWRPFTRFPVCPSGTSPYSRGTHSQPSYV
RCQNCERARQWFRAHDGPRCLHQRPDYSRLQAPPDPFQHLNSFEPILLAALSVLRPLPRD
IQITIISCACDYFHSVRCASSRYLGSSRSAVKRRAARLNYCYKCGHPLYLNKPHTCRPGR
LCSASISERLALLREGPIRSLTENPINARAAHFLAHELLDPR
```

u-----XEmacs: Viral_NABP.fas (Fundamental)----All----------

CTV_P23.fas

```
>Q66245|Q66245_9CLOS
MDNTSGQTFVSVNLSDESNTATTDVEPVSSEADRLDFLQKMNPIIIDALIRKNSYQGARF
RARIIGVCVDCGRKHDKGLKTERKCKVNNTQSQNEVAHMLMHDPVKYLNKRKARAFSNAE
IFAIDLVMYTKERQLAIDLAAEREKTRLARRHPMRSPEETPEYYKFGRTAKAMLPDINAV
DVGDNEETSSEYPVSLSVSGGVLREHHFI
>Q9DJS8|Q9DJS8_9CLOS
MNDTSGQTFVSVNLSDESNTASTRVENVKSEADRLEFLRKMNPLIIDALVRKTNYQGARF
RARIIGVCVDCGRKHDKALKTERKCKVNNTQSQNEVAHMLMHDPVKYLNKRKARAFSNAE
MFAIELVLYTKEKQLAVDLAAEREKTRLARRHPMRSPEETPEHYKFGMTASAMLPELNAV
GVGDNEDTSSEYPVSLSVSGGVLREHHFI
>Q66258|Q66258_9CLOS
MDNPSGQTFISVNLSDESNTASTETKAVSSEADRLEFLRKMNPFIIDALIRKNSYQGARF
RARIIGVCVDCGRKHDKASRTERKCKVNNTQSQNEVAHMLMHDPVKYLNKRKARAFSNAE
MFAIDLVMHTKERQLAVDLAAEREKTRLARRHPMRSPEETPEHYKFGMTAKAMLPNINAV
DVGDNEDTSSEYPVSLSVSGGVLREHHLI
```

u--**-XEmacs: CTV_P23.fas (Fundamental)----All----------

protomata-learner.genouest.org

Protomata Learner Parameters - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

http://protomata-learner.genouest.org/

Description of your request: Viral_nucleic_acid_binding

Email (optional): goulven.kerbellec@irisa.fr

Partial Local Multiple Alignements (PLMA)

Sequences

Advice : reduce the redundancy in your set of sequences, for example use the program decrease redundancy at Expasy.

Upload your sequences file (FASTA)

Parcourir...

Or paste your sequences

```
>Q02123|VNBP_POPMV
MVNMRKVLALMQVFRERYDHKCDFNFCDIAVSIVCRSELDFINEPGLSNYAKRRRARRLG
RCVRCFRVNPGFYFTKRCDGITCVPGISWNYDVEDYIKRGRVTGDRETPSTFHGYGYPVG
HKT
>P16654|VNBP_PVSP
MKAERLEMLLLCVYRLGYILPVDVCIKIISVAQVSVQGRSTYSCKRRARSIGRCWRCYRV
YPPVCNSKCDNRTCRPGISPNFKVVTFIRGWSN
```

Sequence type : Proteins DNA DNA (with coding regions)

Paloma parameters

Fragments similarity threshold: 5

Consensus: weak

Minimum fragments size: 1

Maximal fragments size: 15

process is running

Protomata progression status - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

http://genoweb.univ-rennes1.fr/Serveur-GPO/lib/scripts/thanks_sge.php?&name=Protomata&file=Protor

Google

The program is running ...

Running the job Protomata (id 1276521)

Results of protomata ordered by quorum

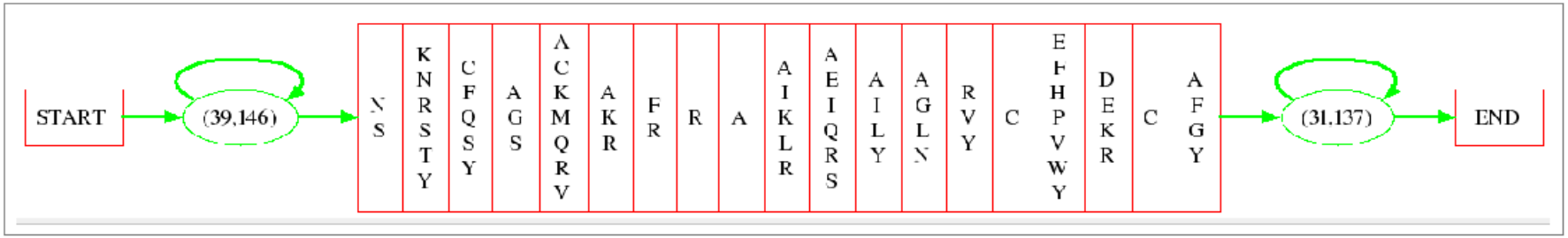
PatternDiscovery by Protomata - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

http://genoweb.univ-rennes1.fr/Serveur-GPO/outils/patternDiscovery/PROTOMATA/form_resultats.php?

Go to quorum >> 21 14 7

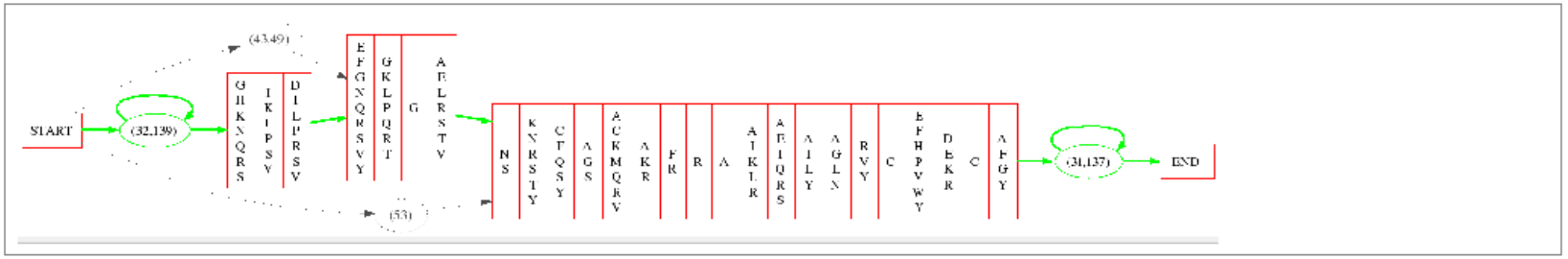
Protomaton quorum21



View protomaton : JAVA applet PNG format SVG format | *View alignment :* JAVA applet PNG format SVG format

Save protomaton (XML) Scan sequences with protomaton

Protomaton quorum14

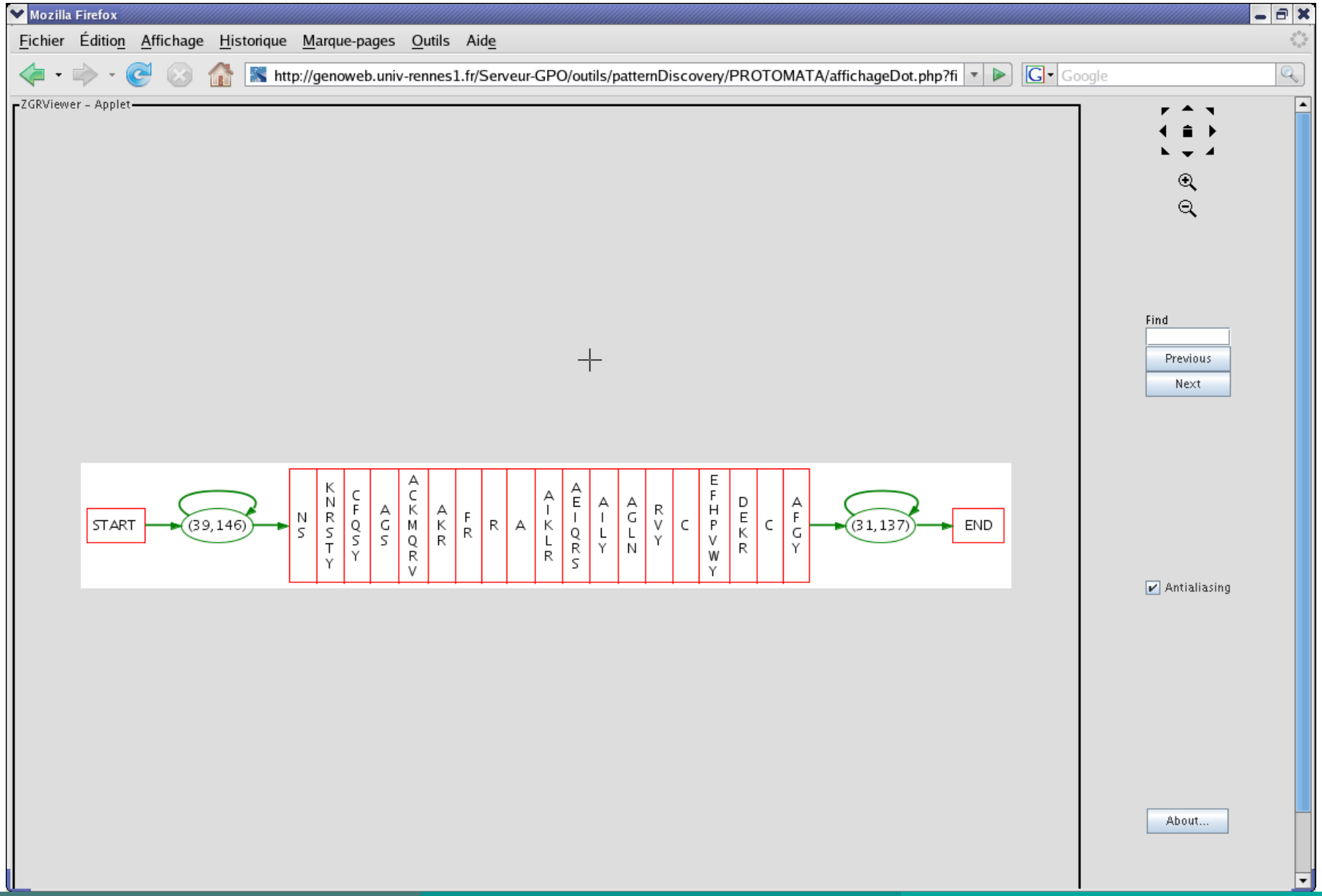


View protomaton : JAVA applet PNG format SVG format | *View alignment :* JAVA applet PNG format SVG format

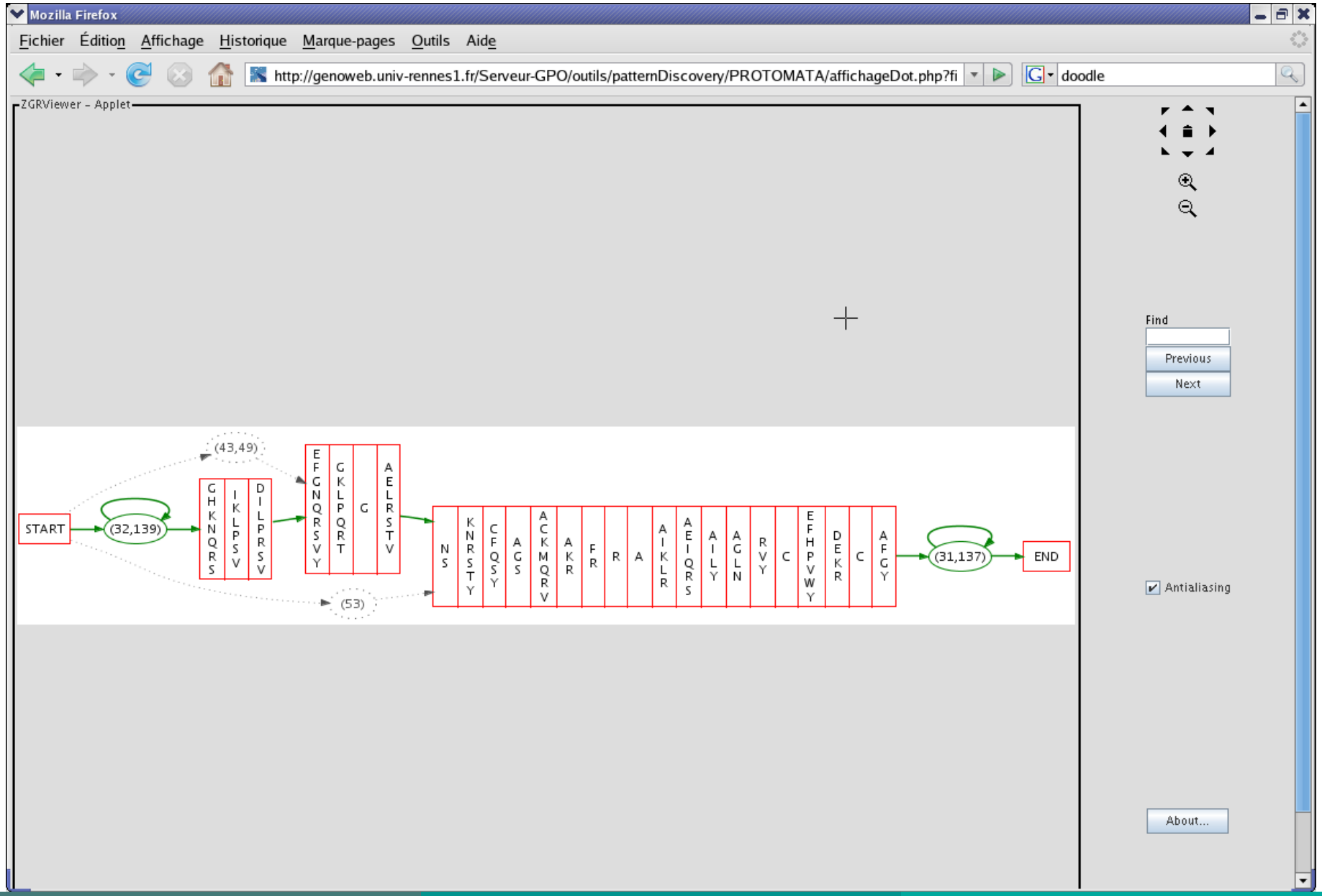
Save protomaton (XML) Scan sequences with protomaton

Protomaton quorum7

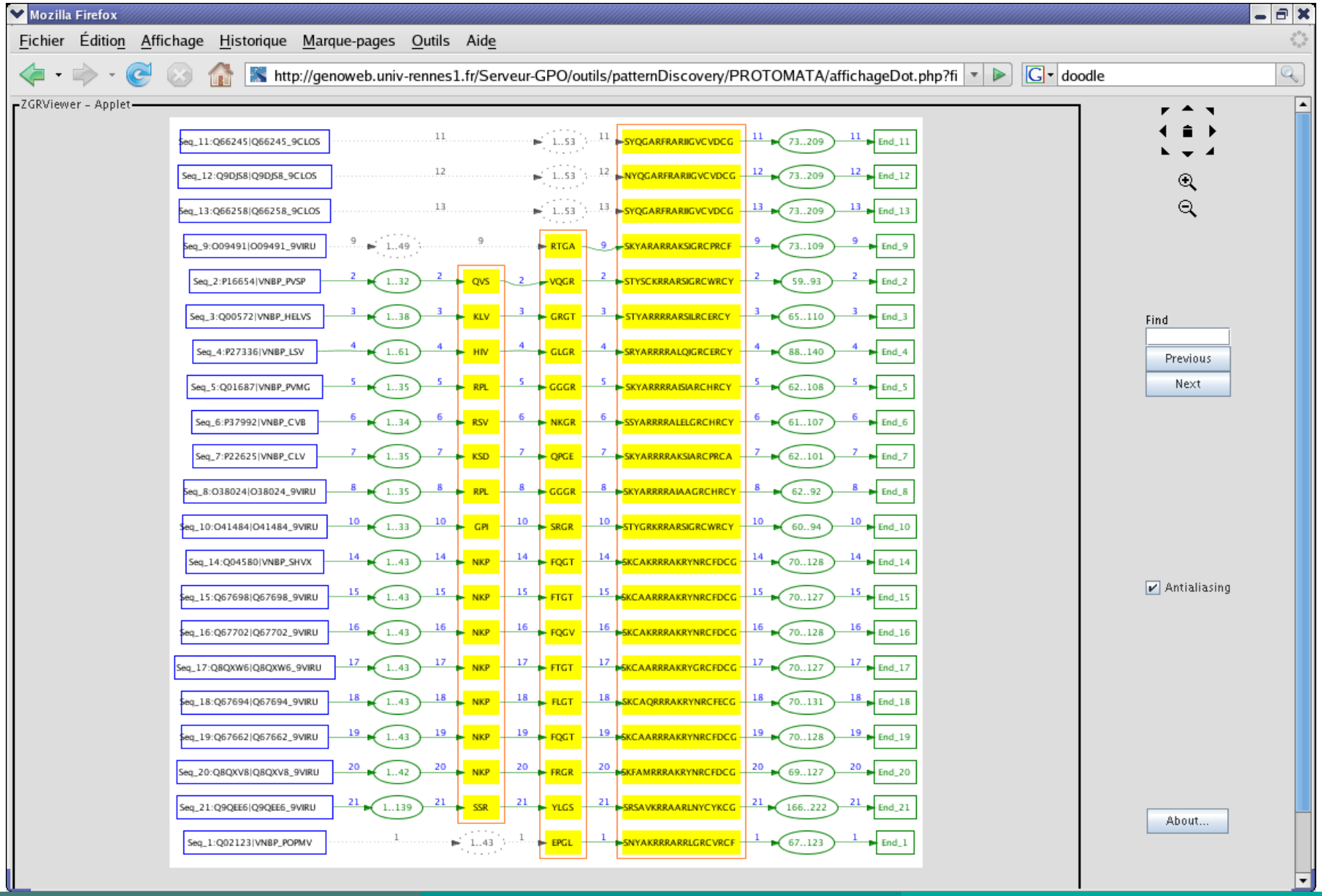
Quorum of 100%



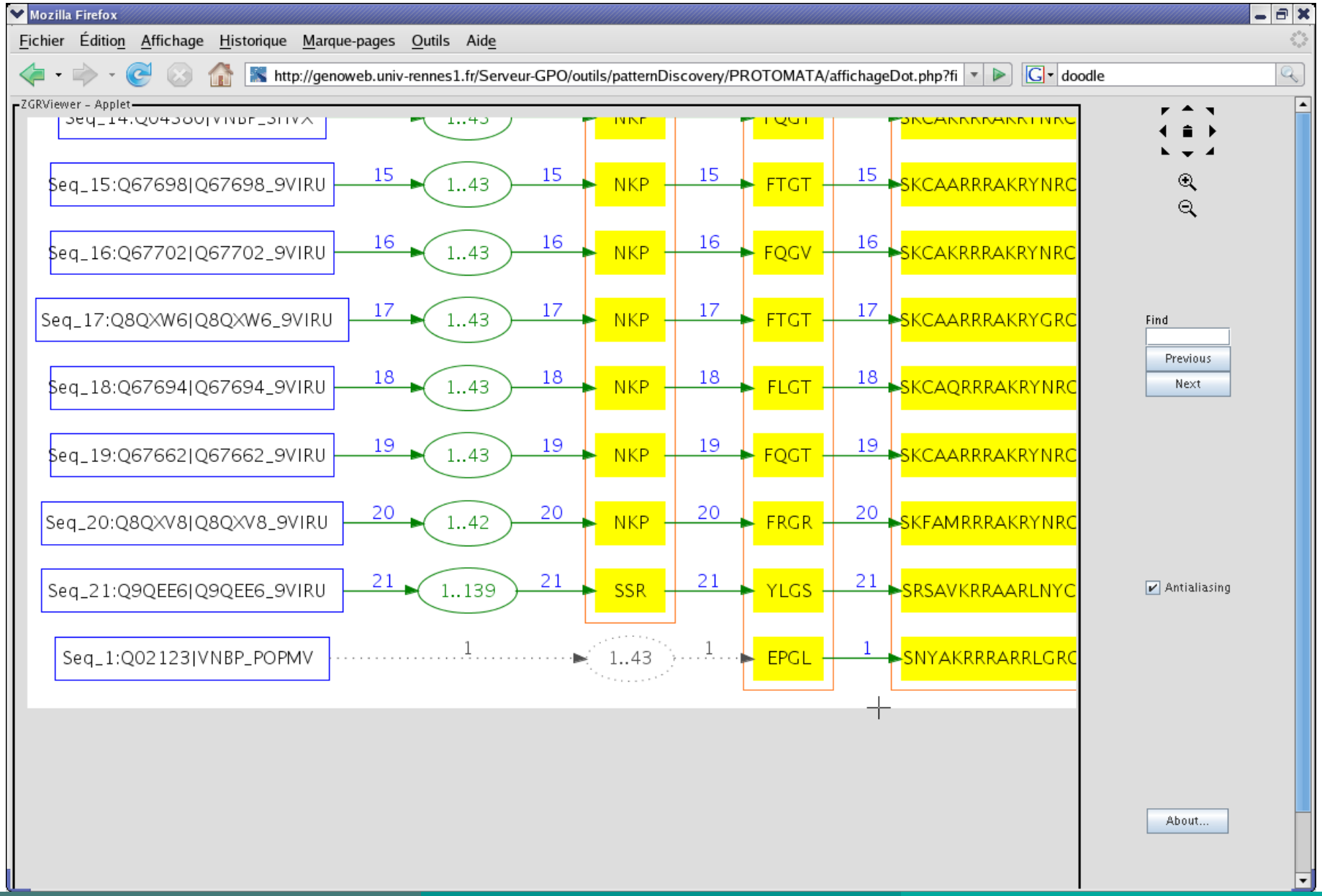
Quorum of 2/3


Mozilla Firefox
Fichier Édition Affichage Historique Marque-pages Outils Aide
http://genoweb.univ-rennes1.fr/Serveur-GPO/outils/patternDiscovery/PROTOMATA/affichageDot.php?fi
doodle
ZGRViewer - Applet
Find
Previous
Next
Antialiasing
About...
START
(32,139)
(43,49)
(53)
(31,137)
END

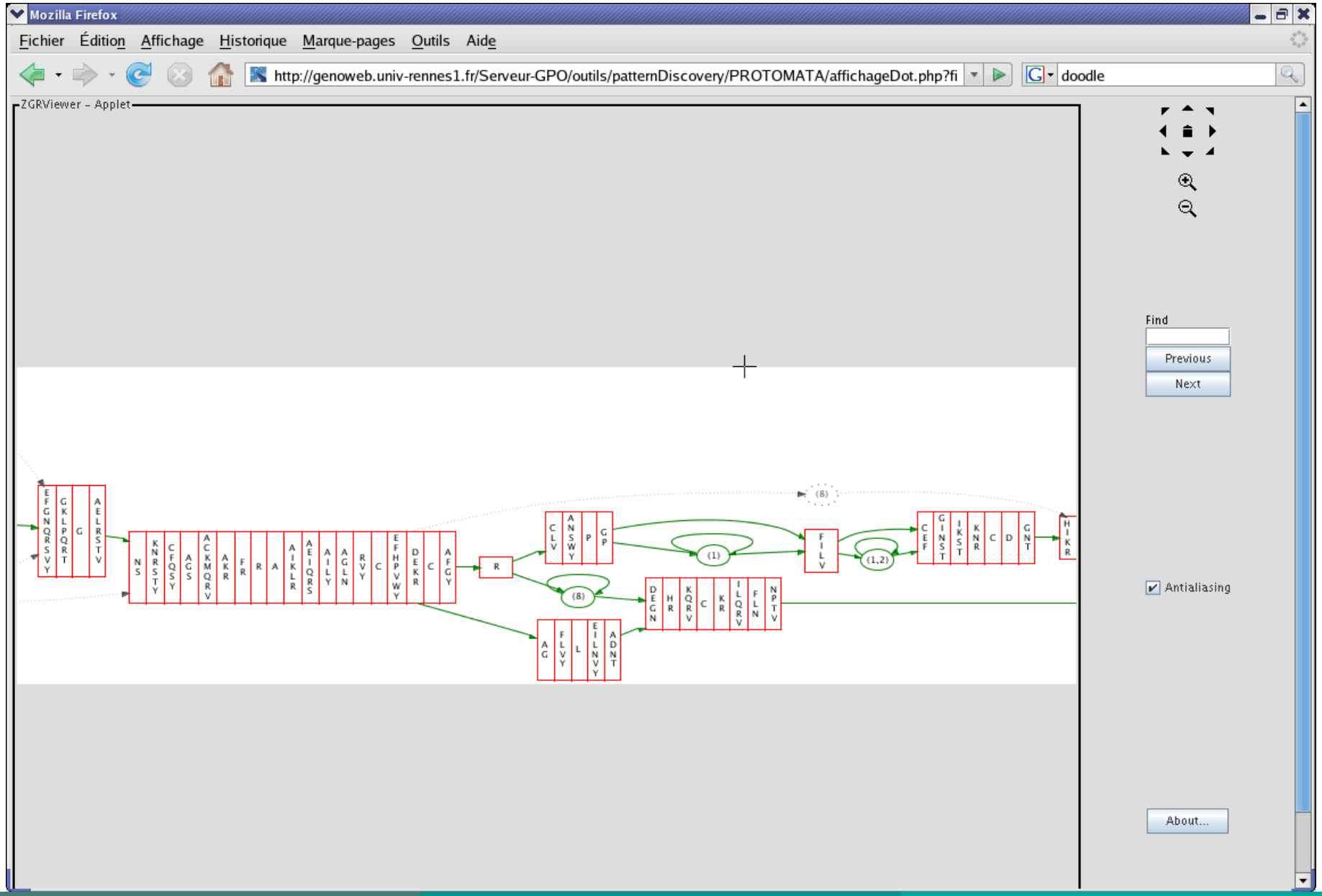
Partial Local Multiple Alignement view



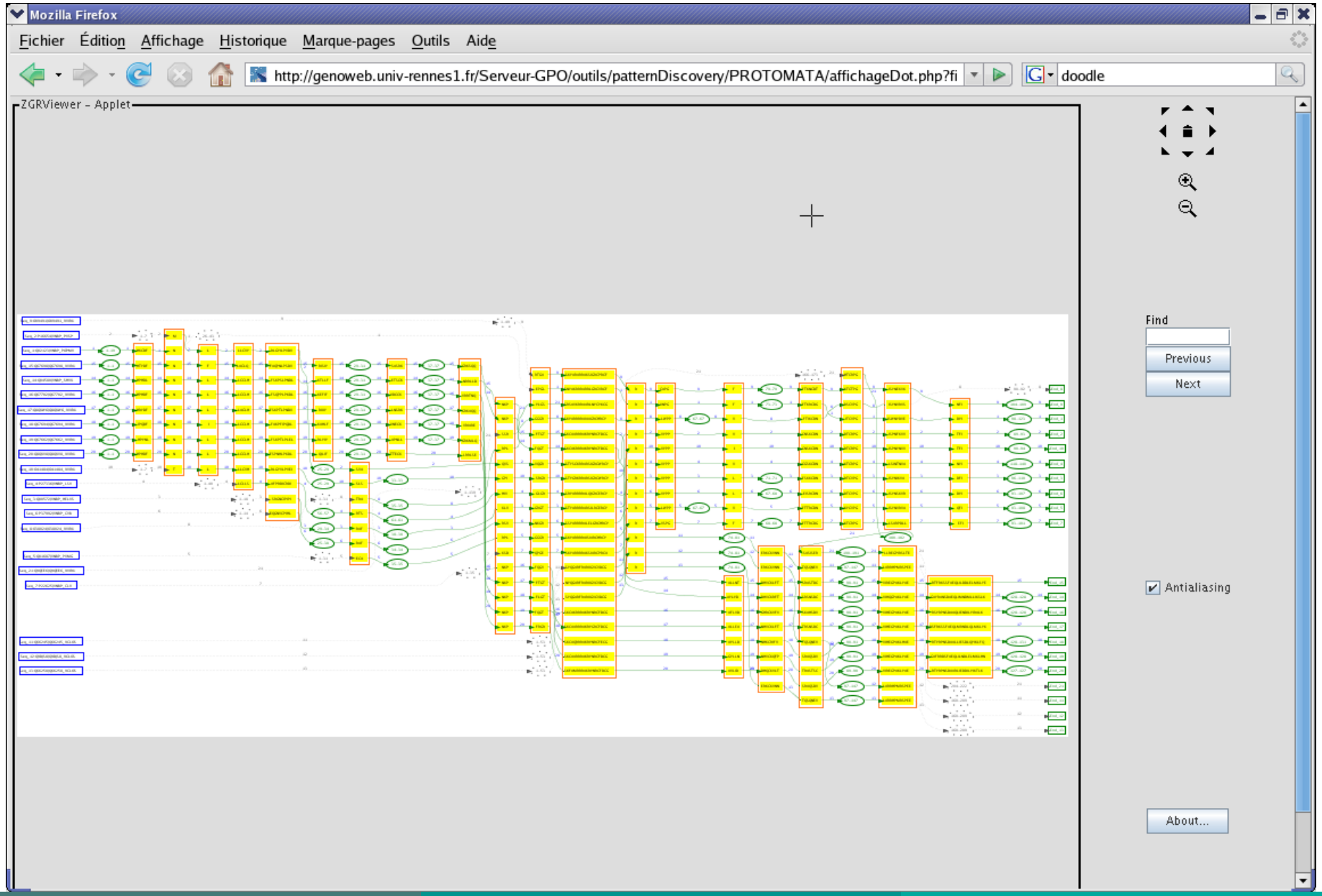
An exception



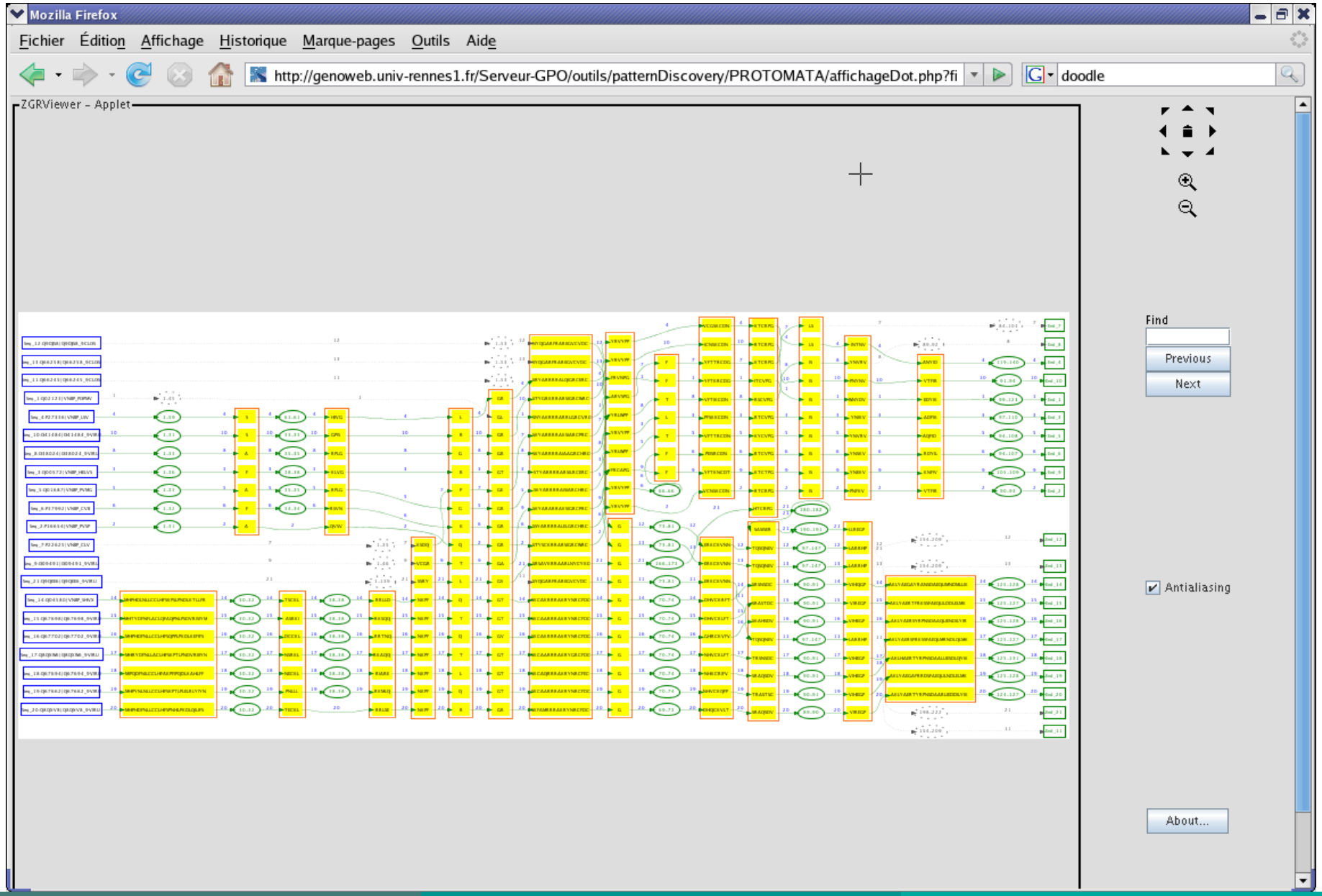
Quorum of 1/3


Mozilla Firefox
Fichier Édition Affichage Historique Marque-pages Outils Aide
http://genoweb.univ-rennes1.fr/Serveur-GPO/outils/patternDiscovery/PROTOMATA/affichageDot.php?fi
doodle
ZGRViewer - Applet
Find
Previous
Next
Antialiasing
About...

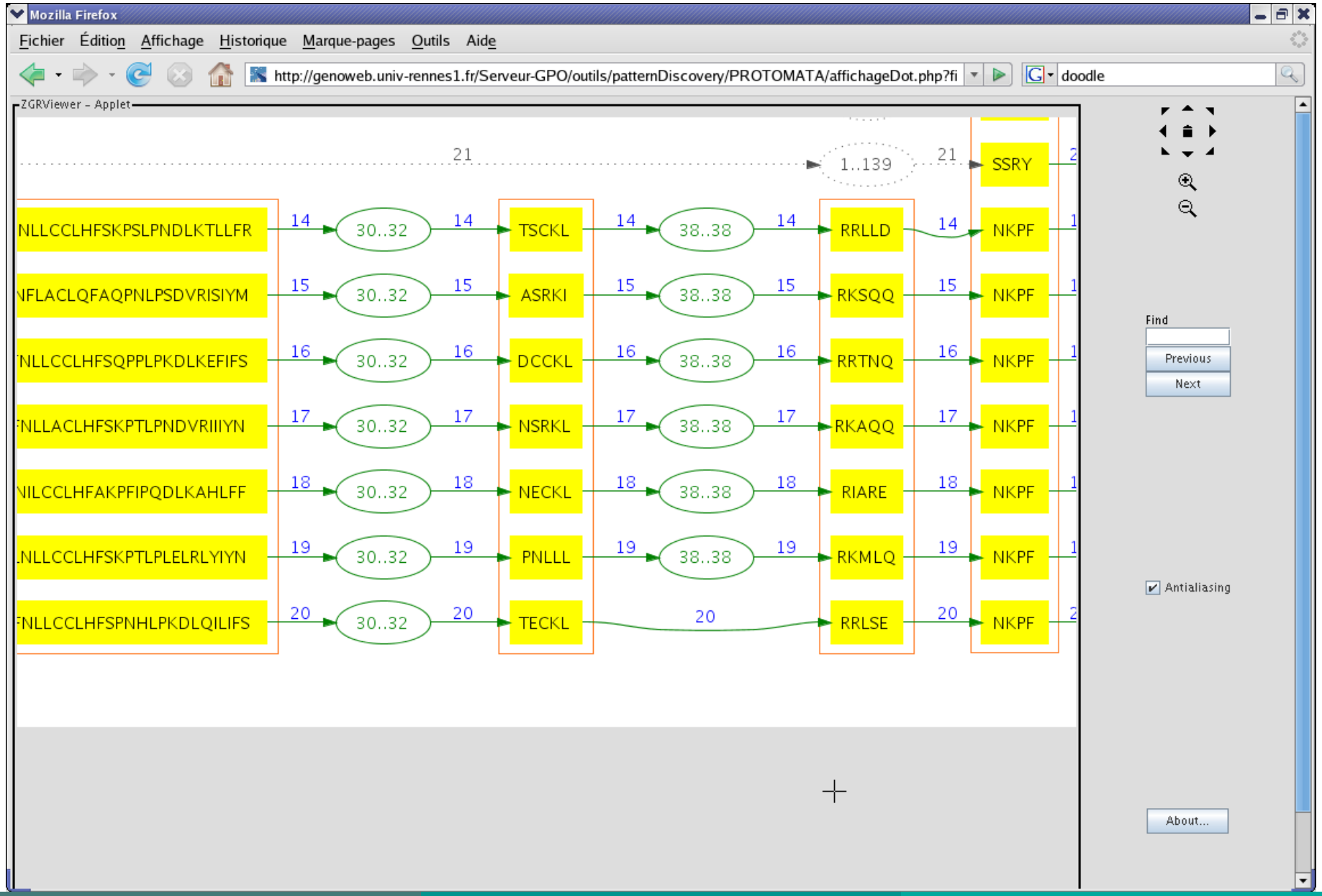
Alignment view of a weak concensus result



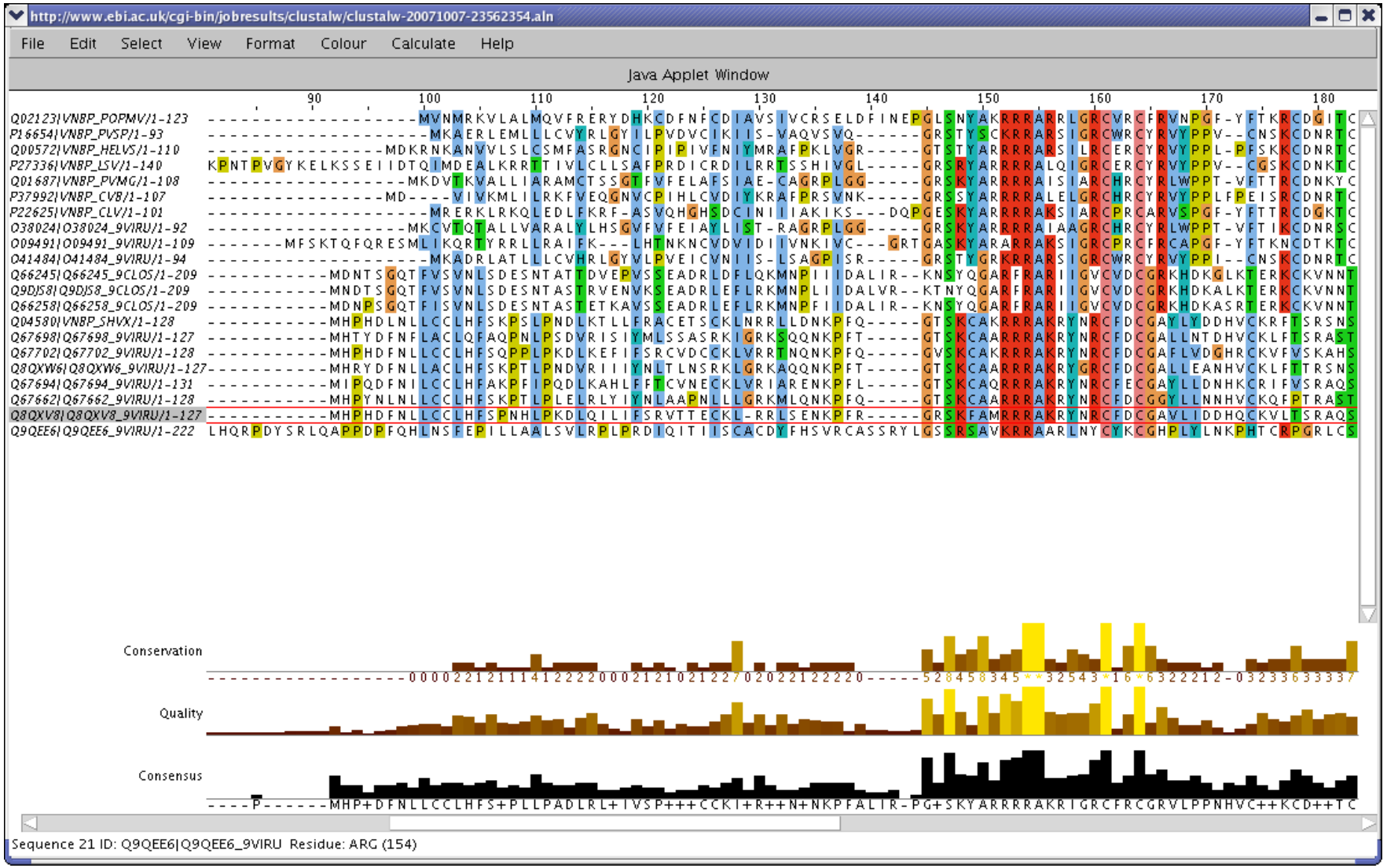
Alignment view of a strong concensus result



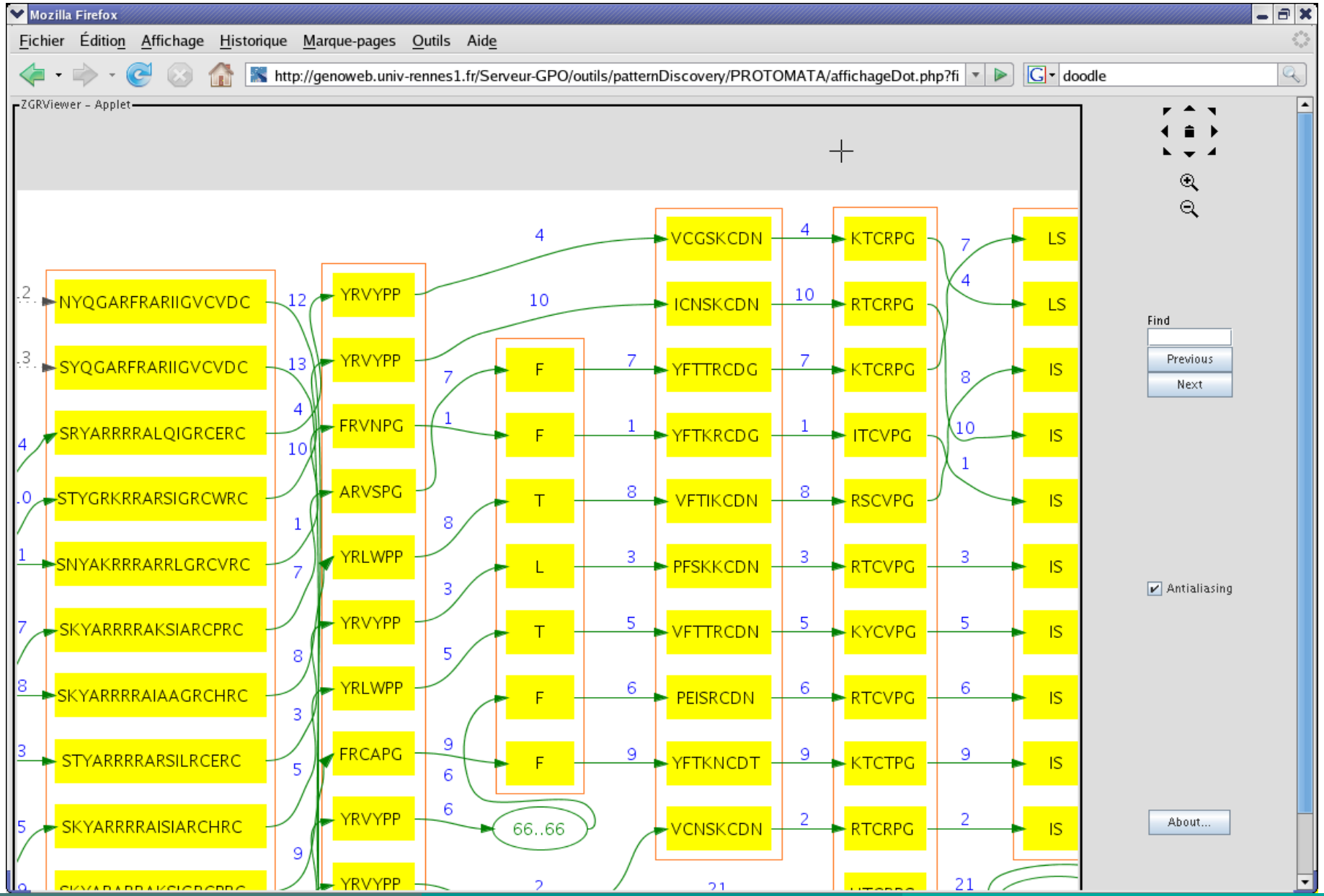
A deletion


Mozilla Firefox
Fichier Édition Affichage Historique Marque-pages Outils Aide
http://genoweb.univ-rennes1.fr/Serveur-GPO/outils/patternDiscovery/PROTOMATA/affichageDot.php?fi
doodle
ZGRViewer - Applet
21
1..139
21
SSRY
NLLCCLHFSKPSLPNDLKTLLFR
NFLACLQFAQPNLPSDVRISIYM
NLLCCLHFSQPPLPKDLKEFIFS
NLLACLHFSKPTLPNDVRIIIYN
NILCCLHFAKPFIPQDLKAHLFF
NLLCCLHFSKPTLPLELRLYIYN
NLLCCLHFSPNHLPKDLQILIFS
14 15 16 17 18 19 20
30..32
TSCKL ASRKI DCCKL NSRKL NECKL PNLLL TECKL
38..38
RRLLD RKSQQ RRTNQ RKAQQ RIARE RKMLQ RRLSE
NKPF
Find
Previous
Next
Antialiasing
About...

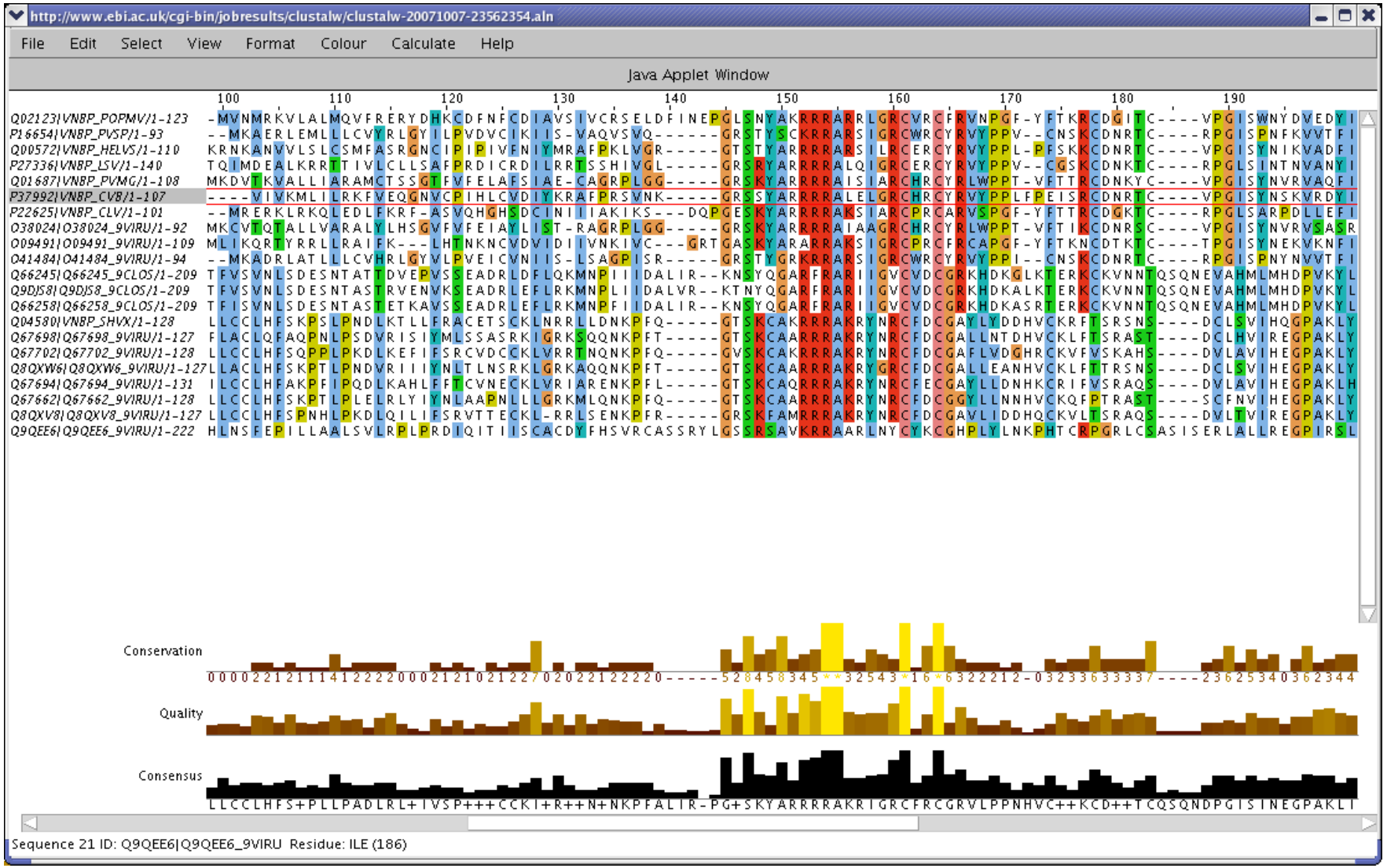
The same in Clustal-W



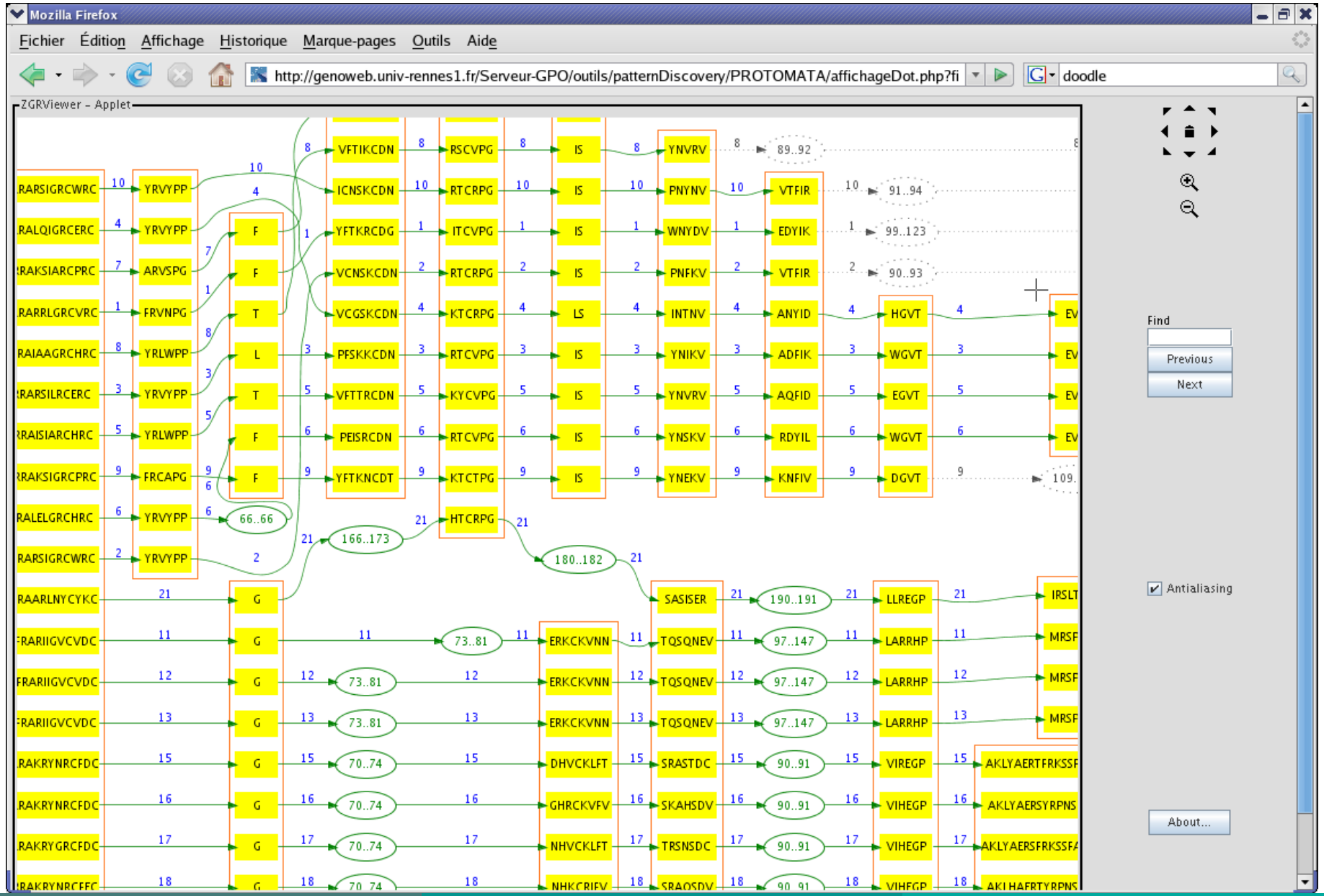
An insertion



The same in Clustal-W



Zinc finger motif



protomata-scan

PROTOSCAN - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

http://genoweb.univ-rennes1.fr/Serveur-GPO/outils/patternMatching/PROTOSCAN/index.php?file=/web,

-- PROTOSCAN --

-- BETA VERSION --

Your email (optional) goulven.kerbellec@irisa.fr

Your XML file of protomata

Scores matrix Select a default matrix Blosum62_plus4_inv.mat Upload your matrice file

The sequences

Search within a data base or a Personnal sequences

Data bank swissprot

Scan

Score threshold default WA threshold Define an other threshold

Enter your threshold 20

Start the scan

For any problem or any question don't hesitate to contact the Webmaster

scanning databases

Protoscan progression status - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

http://genoweb.univ-rennes1.fr/Serveur-GPO/lib/scripts/thanks_sge.php?&name=Protoscan&file=Protos

Google

The program is running ...

Running the job Protoscan (id 1276528)

results

WAPAM result - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils Aide

http://genoweb.univ-rennes1.fr/Serveur-GPO/lib/pattern/matching/parser_xml.php

Back | Results per page : 1500 | Next

Begin | Jump to : P14274|CB2A_SOLLC Chlo | End

Maximum sequences length : 30

Result 1 to 23 of 23

Num	Chromosome	Strand	F-/index/www-tmp/Protoscan4591810.wa				
			begin	end	cost	sequence	length
1	P14274|CB2A_SOLLC Chlorophyll a-b binding protein 1A, chloroplast precursor - Solanum lycopersicum (Tomato) (Lycopersicon esculentum)	plus	0	0	0	MAAAAMALSSPSFAG ... NNNAWAFATNFVPGK	265
2	P14275|CB2C_SOLLC Chlorophyll a-b binding protein 1C, chloroplast precursor - Solanum lycopersicum (Tomato) (Lycopersicon esculentum)	plus	0	0	0	MAAATMALSSPSFAG ... NNNAWAFATNFVPGK	265
3	P14276|CB2E_SOLLC Chlorophyll a-b binding protein 3A, chloroplast precursor - Solanum lycopersicum (Tomato) (Lycopersicon esculentum)	plus	0	0	0	MAASTMALSSSTFAG ... NNNAWAFATNFVPGK	267
4	P14277|CB2F_SOLLC Chlorophyll a-b binding protein 3B, chloroplast precursor - Solanum lycopersicum (Tomato) (Lycopersicon esculentum)	plus	0	0	0	MAASTMALSSSTFAG ... NNNAWAFATNFVPGK	267
5	P0C2W8|CO1A1_MAMAE Collagen alpha-1 (I) chain - Mammut americanum (American mastodon)	plus	0	0	0	GFSGLDGAKXXXXXX ... GLNGLPGPIGPPGPR	900
6	P0C2W2|CO1A1_TYREX Collagen alpha-1 (I) chain - Tyrannosaurus rex (Tyrant lizard king)	plus	0	0	0	GATGAPGIAGAPGFP ... XXXXXXGVVGLPGQR	570
7	P85154|CO1A2_MAMAE Collagen alpha-2(I) chain - Mammut americanum (American mastodon)	plus	0	0	0	GSDGSVGPVGPAGPN ... XXXXGPAGPSGPVGK	830
8	P85153|CO2A1_MAMAE Collagen alpha-1 (II) chain - Mammut americanum (American mastodon)	plus	0	0	0	GEPGGVGPIGPPGER ... GDQGASGPSGPAGPR	669
9	O48085|CYB_ERYTA Cytochrome b - Eryx tataricus (Tartar sand boa)	plus	0	0	0	MPHQQMLILFGLLPV ... NPIAGWIENNIMKDN	371
10	P16274|IFEA_HELPO Non-neuronal cytoplasmic intermediate filament protein A - Helix pomatia (Roman snail) (Edible snail)	plus	0	0	0	TSKISTTYEEEGRQS ... ADNEQIADMFSLGVG	551
11	O19816|MATK_ALLCA Maturase K - Allamanda cathartica (Yellow allamanda)	plus	0	0	0	MEAIQRYLQFDRSQQ ... WYLDITCINDLANQQ	505
12	P41413|PCSK5_RAT Proprotein convertase subtilisin/kexin type 5 precursor - Rattus norvegicus (Rat)	plus	0	0	0	MDWGWGSRCCRPGRR ... EDDELEYDDESYSYQ	1877
13	P22625|VNBP_CLV 11.6 kDa protein - Carnation latent virus (CLV)	plus	0	0	0	MRERKLRKQLEDLFK ... DLLEFIGIDLCVRSK	101
14	P37992|VNBP_CVB 12.6 kDa protein - Chrysanthemum virus B (CVB)	plus	0	0	0	MDVIVKMLILRKFVE ... LWGVTEVIPHPGYNF	107
15	Q00572|VNBP_HELVS 12.6 kDa protein - Helenium virus S (HelVS)	plus	0	0	0	MDKRNKANVVLSLCS ... KWGVTEVIPHPGFNF	110
16	P27336|VNBP_LSV 16 kDa protein - Lily symptomless virus (LSV)	plus	0	0	0	MSVWGAWKPNTPVGY ... PWISPHRGQFYLRPK	140
17	Q02123|VNBP_POPMV 14 kDa protein - Poplar mosaic virus (isolate ATCC Pv275) (PMV)	plus	0	0	0	MVNMRKVLALMQVFR ... PSTFHGYGYPVGHKT	123
18	Q01687|VNBP_PVMG 12 kDa protein - Potato virus M (strain German) (PVM)	plus	0	0	0	MKDVTKVALLIARAM ... EGVTEVIPSVINKRE	108
19	P16654|VNBP_PVSP 10.7 kDa protein - Potato virus S (strain Peruvian)	plus	0	0	0	MKAERLEMLLLCVYR ... SPNFKVVTFIRGWSN	93
20	Q04580|VNBP_SHVX 14.7 kDa protein - Shallot virus X (ShVX)	plus	0	0	0	MHPHDLNLLCCLHFS ... QLIMNDMLLIKSLKL	128
21	Q6PSC4|MATK_DOLLA Maturase K - Dolichos lab lab (Field bean) (Lablab purpureus)	plus	0	0	2	MEQYQAYLELRRSRY ... WYLDIFFRNDFVNHF	504
22	Q71ML1|MATK_CHIUM Maturase K - Chimaphila umbellata (Pipsissewa)	plus	0	0	10	MEEFKRNLELDRSQX ... WYLDIICINDLSNHE	504
23	P17530|VNBP_PVMR 10.7 kDa protein - Potato virus M (strain Russian) (PVM)	plus	0	0	19	MKDVTKVALLIARAM ... EGVTEVIPSVINKRE	108

Back | Results per page : 1500 | Next

Begin | Jump to : P14274|CB2A_SOLLC Chlo | End

Maximum sequences length : 30

Result 1 to 23 of 23

Conclusion

Conclusion

- New sequence-driven algorithm
- Bridging the gap between Multiple Sequences Alignements and Pattern Discovery and beyond ...
- Web interface
- Let's play with all kind of data !

Acknowledgments

- Région Bretagne
- Inria/Irisa
- Symbiose Team
- Plateforme Genouest (Laetitia Guillot)
- Christian Delamarche (MIP) and Thierry Guillaudeux (TNF)
- Boris Idmont, Daniel Fredouille, Thi Hong Hanh Hoang, Laetitia Guillot, François Coste and Jacques Nicolas