STRUCTURA BIOLOGICĂ.
§2.4. UTILIZAREA PROGRAMELOR DEDICATE:IDENTIFICAREA PATTERN-URILOR

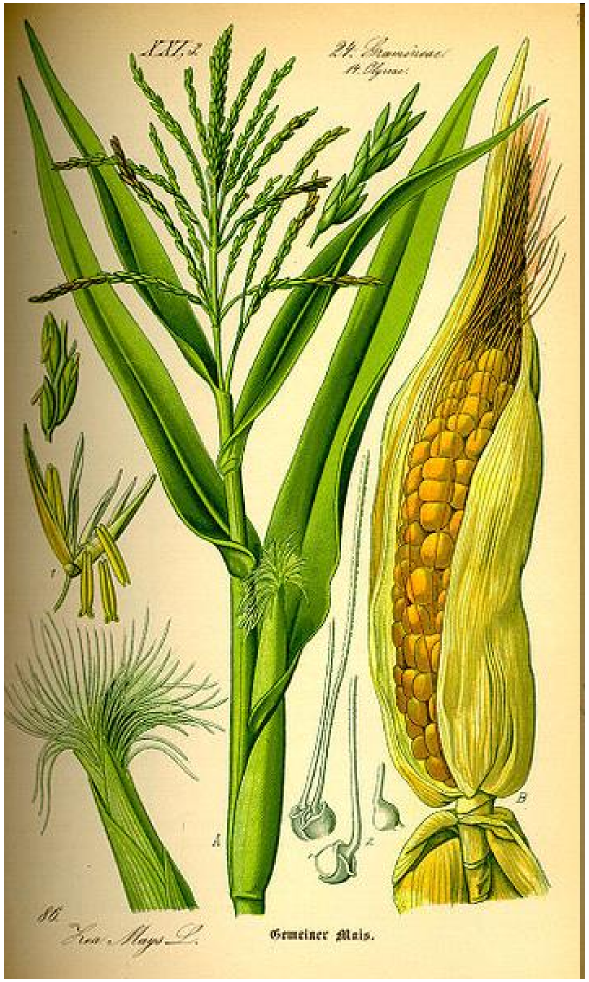
Sorana D. BOLBOACĂ

Despre …

- IDENTIFICAREA PATTERN-URILOR:
    - UTILIZAREA PROGRAMULUI CLC

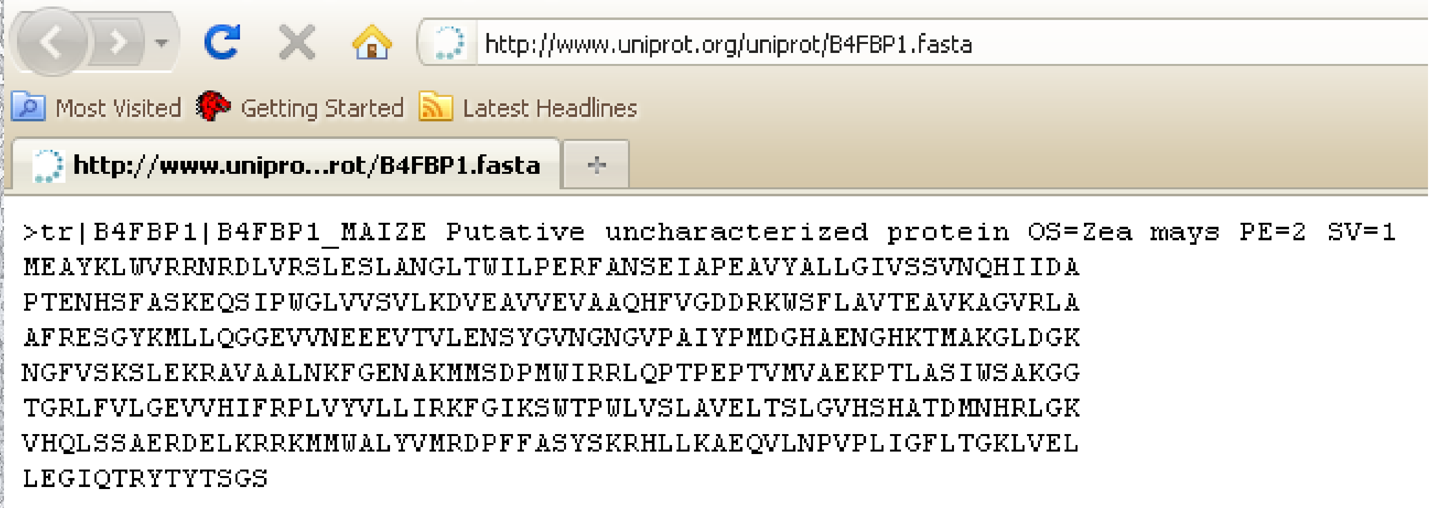
PROBLEMA: AMINOACIZI

- **Zea mays (Maize)**
  - Proteină necaracterizată
  - Dimensiune: 437 aa
  - Nivel: transcriere
  - Denumire: B4FBP1_MAIZE

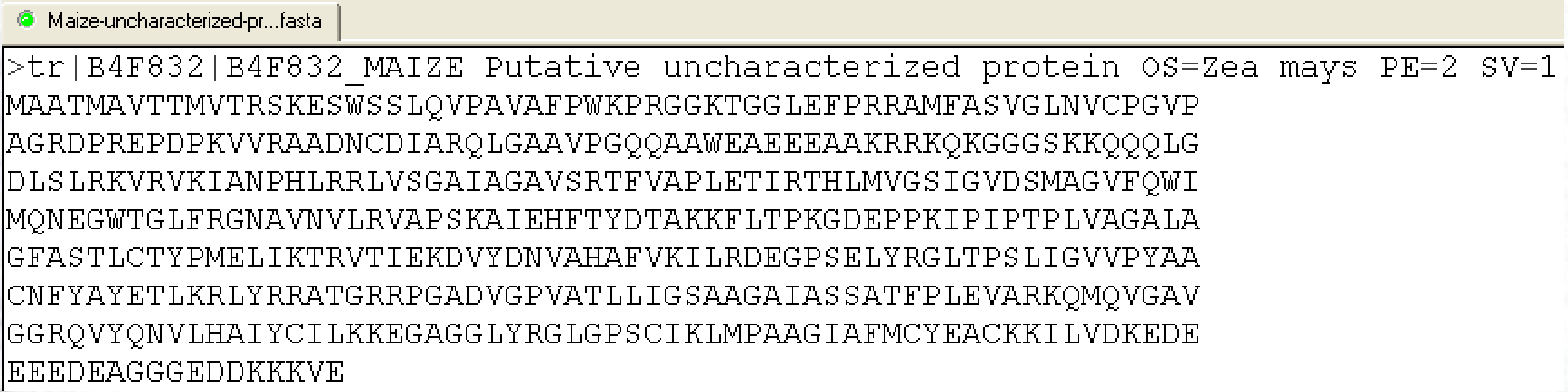


IDENTIFICAREA ŞI SALVAREA SECVENŢEI

- http://www.uniprot.org/uniprot/B4FBP1.fasta

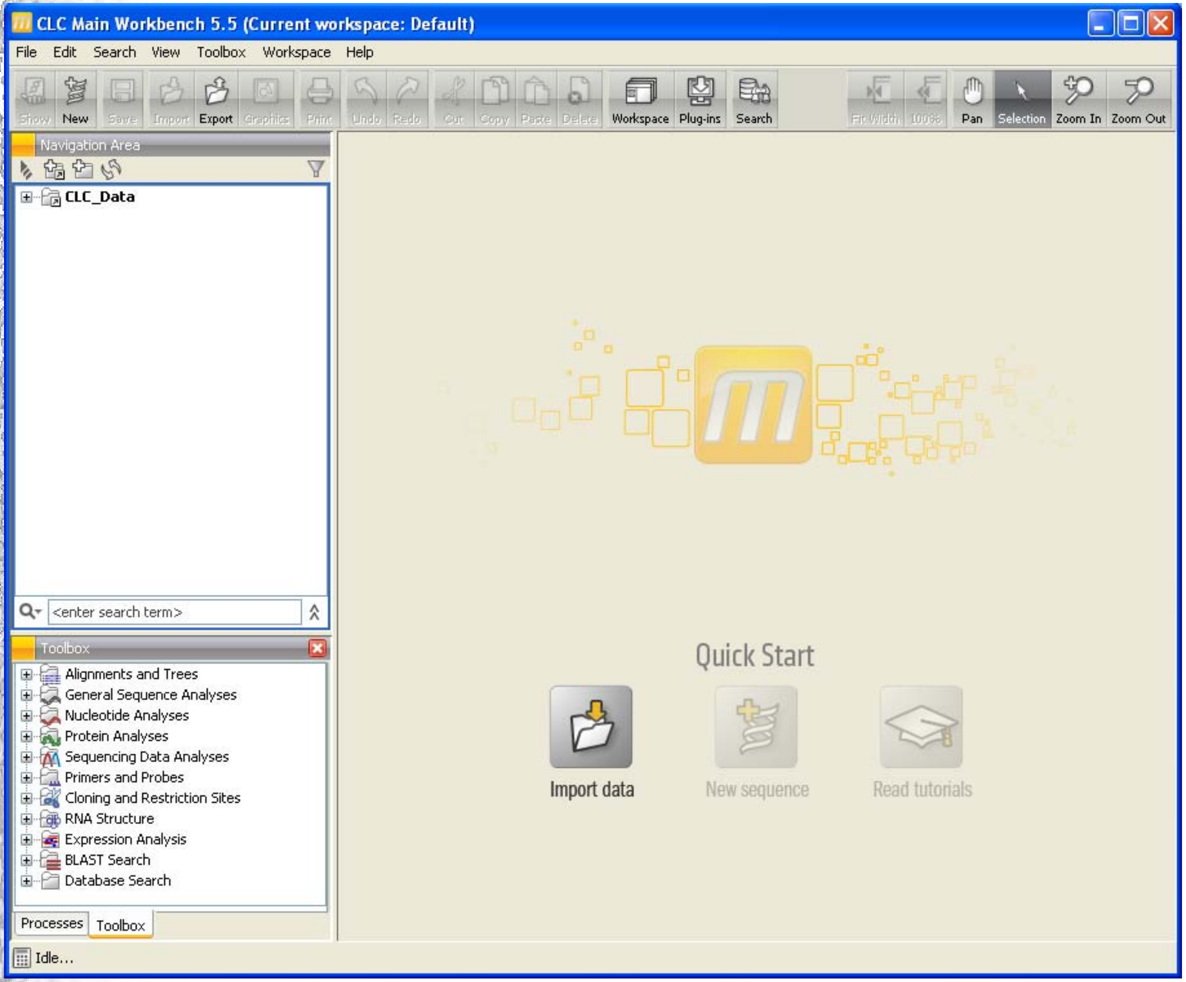


```
>tr|B4FBP1|B4FBP1_MAIZE Putative uncharacterized protein OS=Zea mays PE=2 SV=1
MEAYKLWVRRNRDLVRSLESLANGLTWILPERFANSEIAPEAVYALLGIVSSVNQHIIDA
PTENHSFASKEQSIPWGLVVSVLKDVEAVVEVAAQHFVGDDRKWSFLAVTEAVKAGVRLA
AFRESGYKMLLQGGEVVNEEEVTVLENSYGVNGNGVPAIYPMDGHAENGHKTMAKGLDGK
NGFVSKSLEKRAVAALNKFGENAKMMSDPMWIRRLQPTPEPTVMVAEKPTLASIWSAKGG
TGRLFVLGEVVHIFRPLVYVLLIRKFGIKSWTPWLVSLAVELTSLGVHSHATDMNHRLGK
VHQLSSAERDELKRRKMMWALYVMRDPFFASYSKRHLLKAEQVLNPVPLIGFLTGKLVEL
LEGIQTRYTYTSGS
```

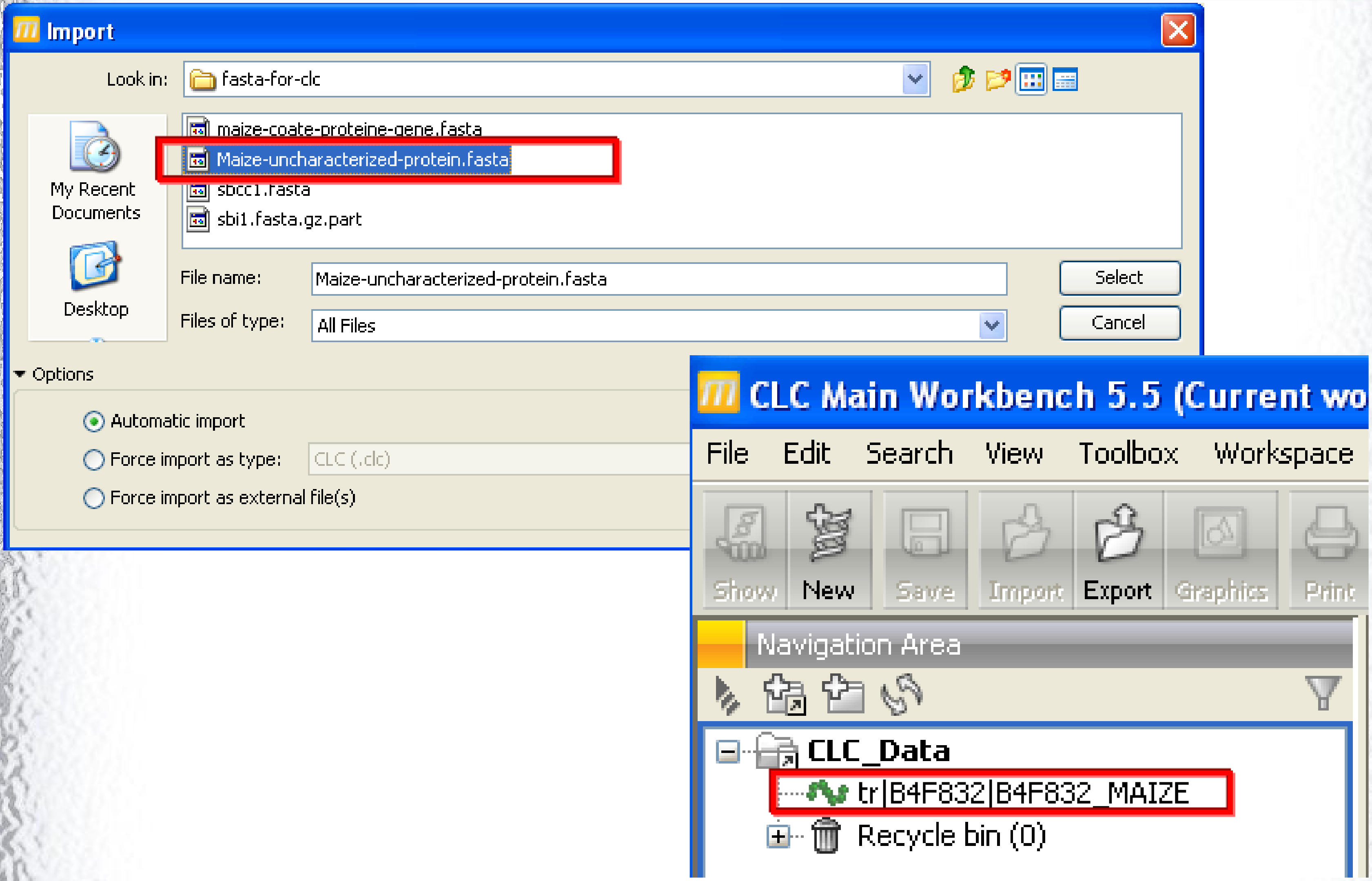


```
>tr|B4F832|B4F832_MAIZE Putative uncharacterized protein OS=Zea mays PE=2 SV=1
MAATMAVTTMVTRSKESWSSLQVPAVAFPWKPRGGKTGGLEFPRRAMFASVGLNVCPGVP
AGRDPREPDPKVVRAADNCDIARQLGAAVPGQQAAWEAEEEAAKRRKQKGGGSKKQQQLG
DLSLRKVRVKIANPHLRRLVSGAIAGAVSRTFVAPLETIRTHLMVGSIGVDSMAGVFQWI
MQNEGWTGLFRGNAVNVLRVAPSKAIEHFTYDTAKKFLTPKGDEPPKIPIPTPLVAGALA
GFASTLCTYPMELIKTRVTIEKDVYDNVAHAFVKILRDEGPSELYRGLTPSLIGVVPYAA
CNFYAYETLKRLYRRATGRRPGADVGPVATLLIGSAAGAIASSATFPLEVARKQMQVGAV
GGRQVYQNVLHAIYCILKKEGAGGLYRGLGPSCIKLMPAAGIAFMCYEACKKILVDKEDE
EEEDEAGGGEDDKKKVE
```

IMPORTUL FIŞIERULUI FASTA ÎN CLC



IMPORTUL FIŞIERULUI FASTA ÎN CLC



Look in: fasta-for-clc

- maize-coate-proteine-gene.fasta
- Maize-uncharacterized-protein.fasta
- sbcc1.fasta
- sbi1.fasta.gz.part

File name: Maize-uncharacterized-protein.fasta

Files of type: All Files

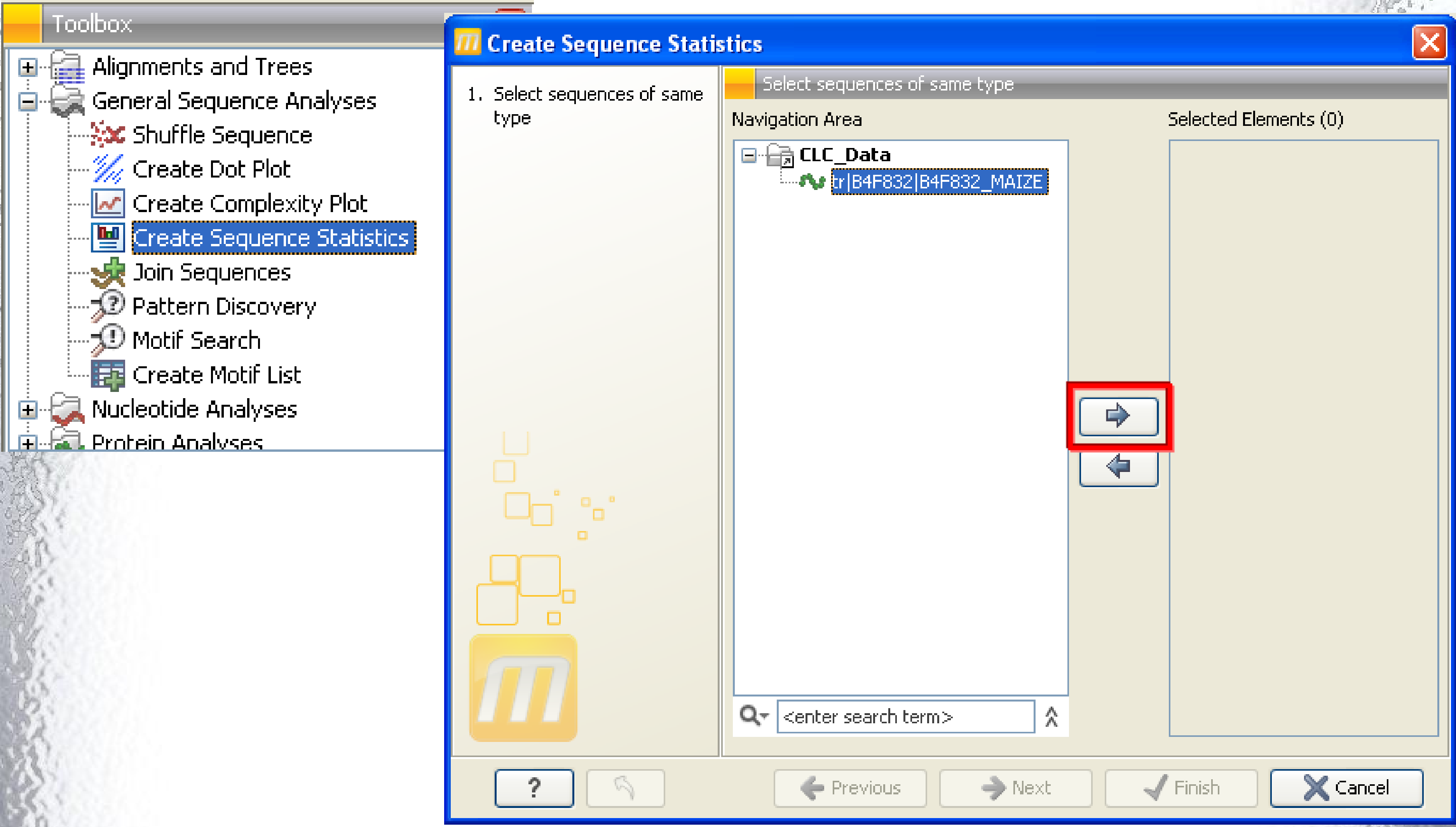
● Automatic import

○ Force import as type: CLC (.clc)

○ Force import as external file(s)

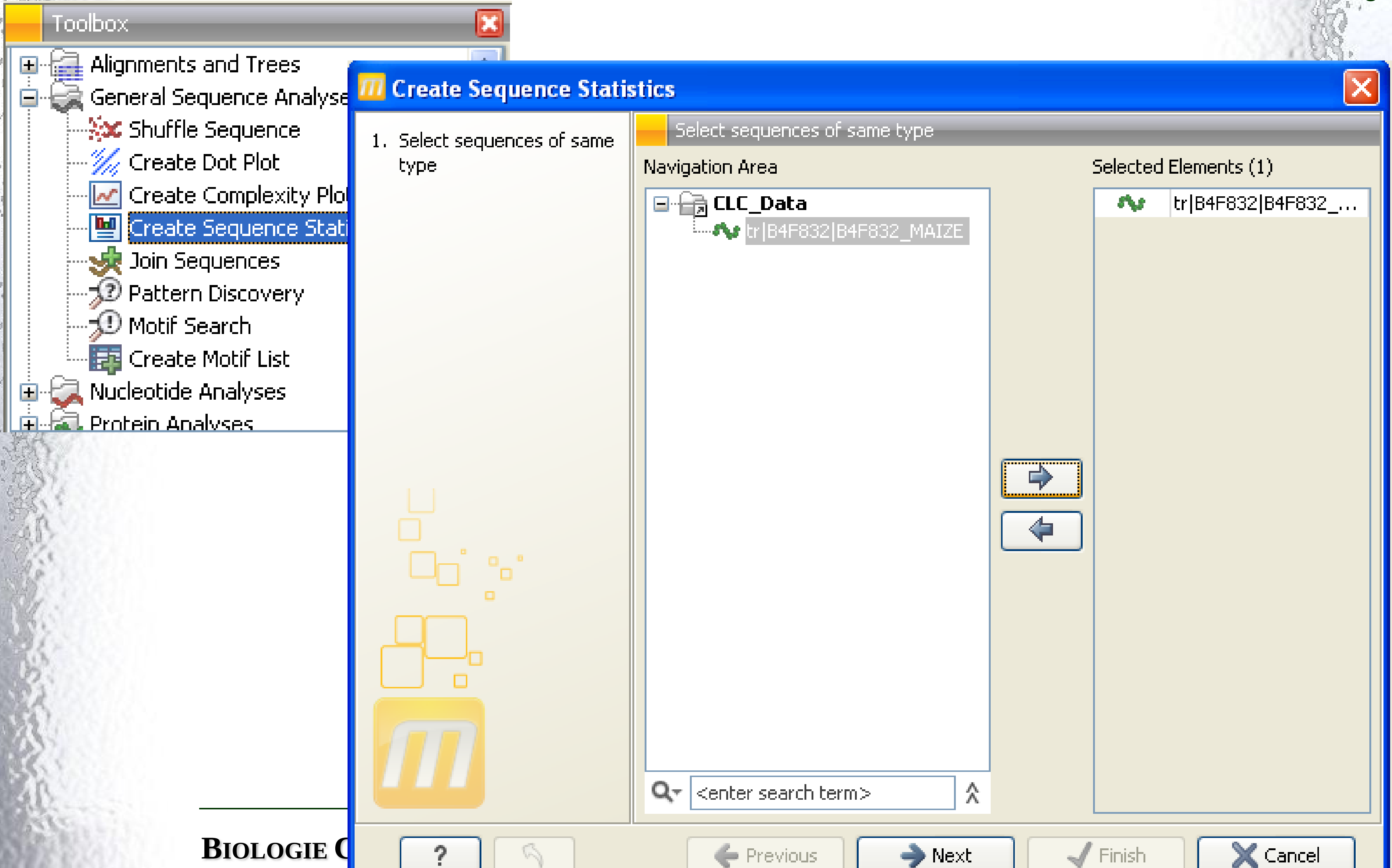
CLC_Data
- tr|B4F832|B4F832_MAIZE
- Recycle bin (0)

STATISTICA SECVENȚEI



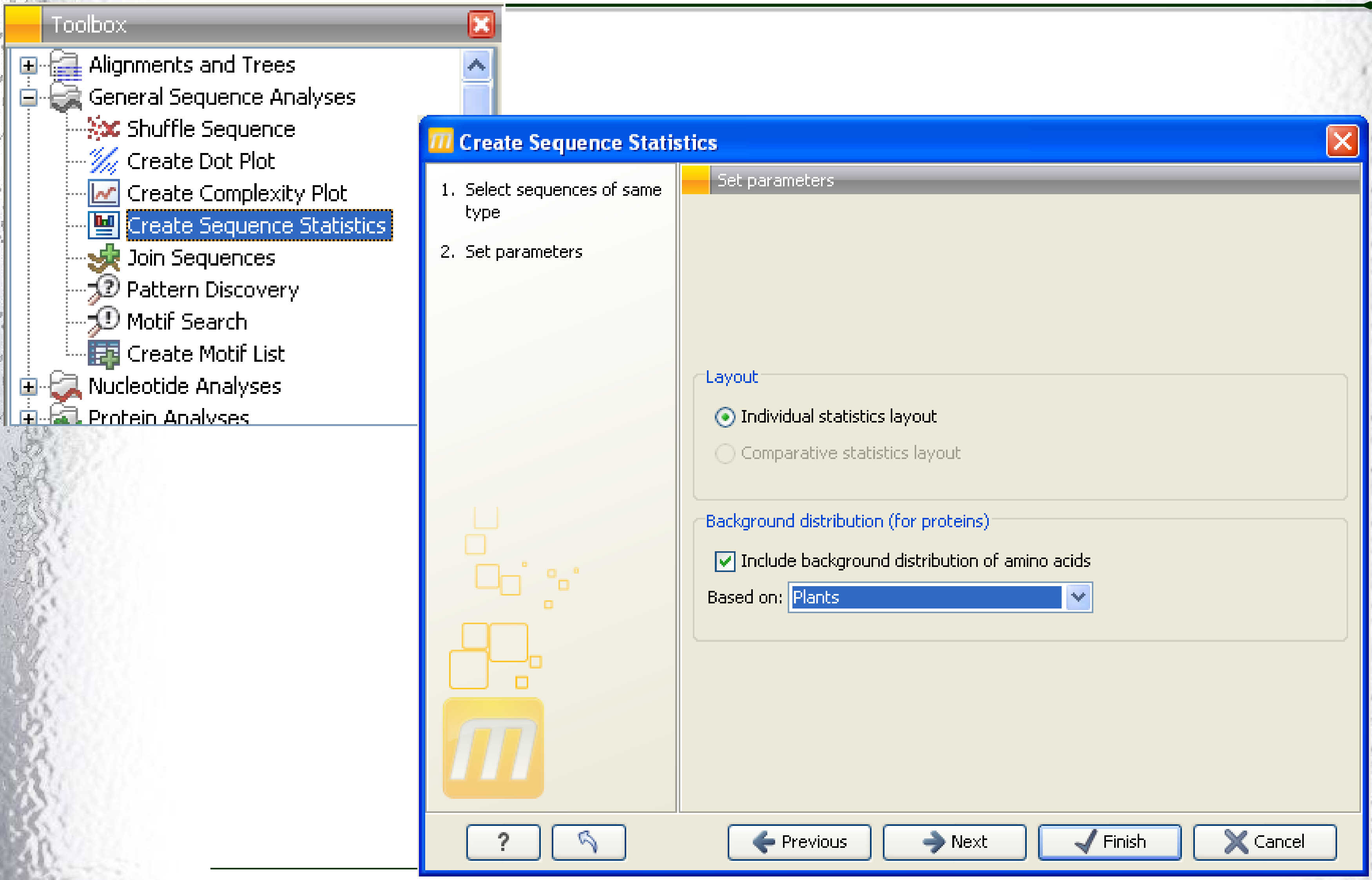
1. Select sequences of same type

STATISTICA SECVENȚEI



1. Select sequences of same type

STATISTICA SECVENȚEI



Create Sequence Statistics

1. Select sequences of same type
2. Set parameters

Set parameters

Layout

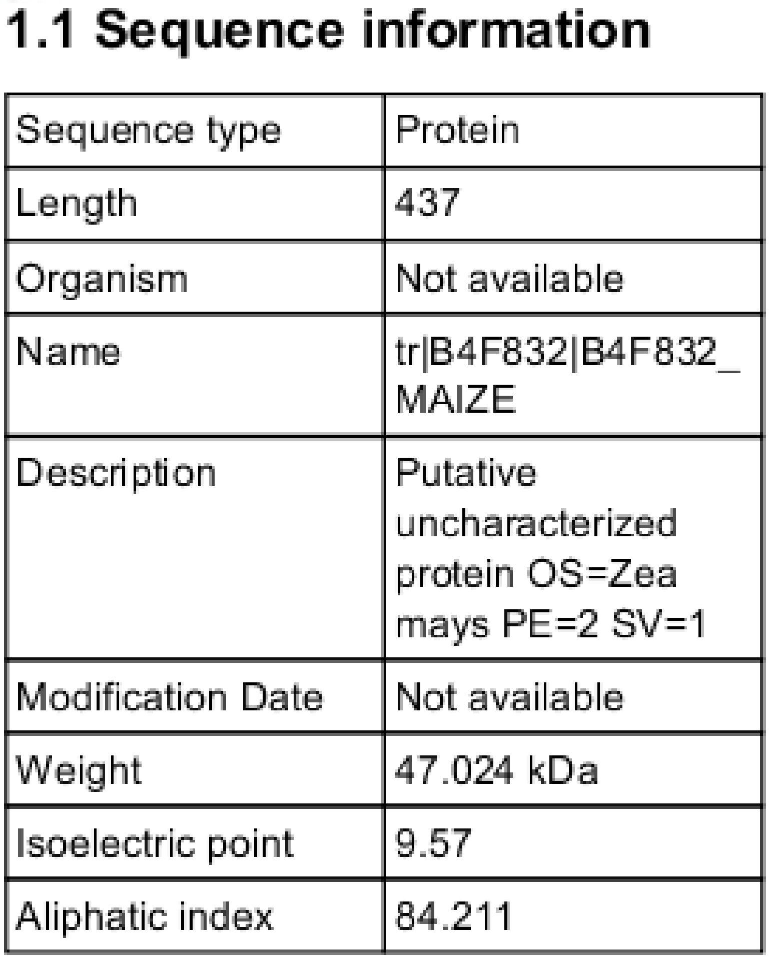
● Individual statistics layout
○ Comparative statistics layout

Background distribution (for proteins)

☑ Include background distribution of amino acids

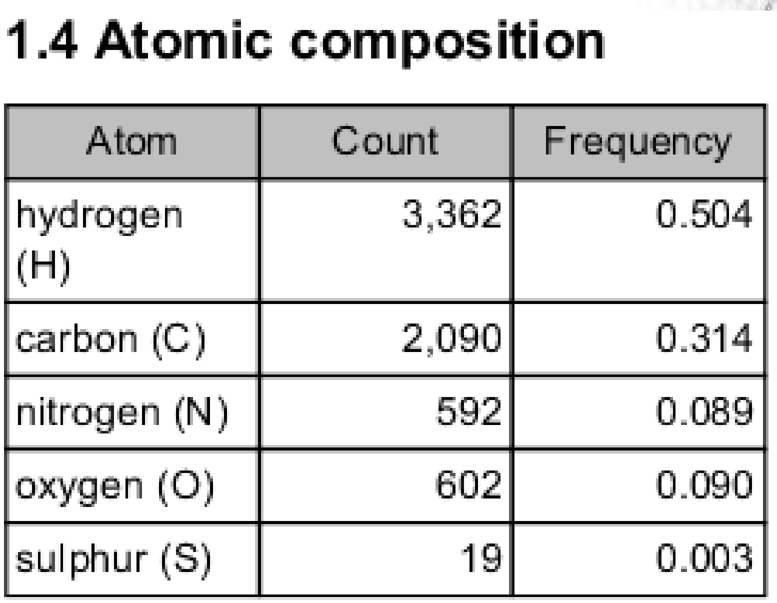
Based on: Plants

STATISTICA SECVENȚEI



1.1 Sequence information

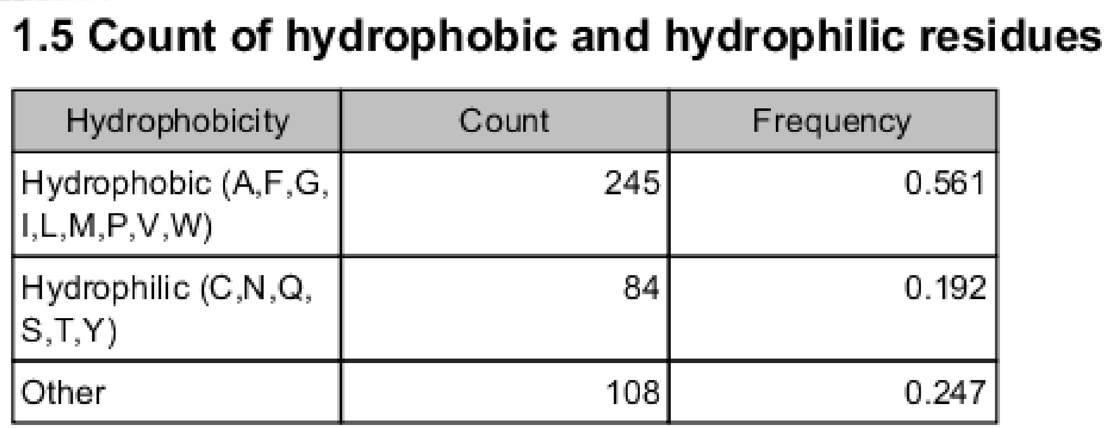
Sequence type	Protein
Length	437
Organism	Not available
Name	tr|B4F832|B4F832_MAIZE
Description	Putative uncharacterized protein OS=Zea mays PE=2 SV=1
Modification Date	Not available
Weight	47.024 kDa
Isoelectric point	9.57
Aliphatic index	84.211



1.4 Atomic composition

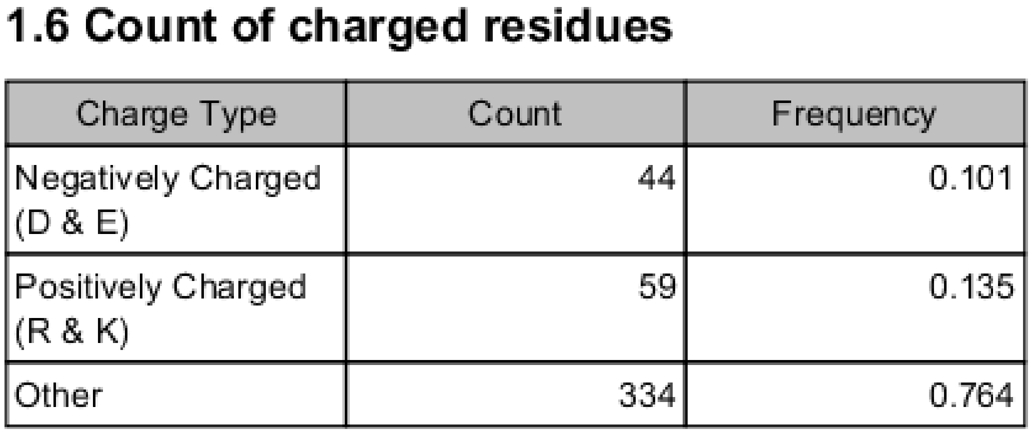
Atom	Count	Frequency
hydrogen (H)	3,362	0.504
carbon (C)	2,090	0.314
nitrogen (N)	592	0.089
oxygen (O)	602	0.090
sulphur (S)	19	0.003

STATISTICA SECVENŢEI



1.5 Count of hydrophobic and hydrophilic residues

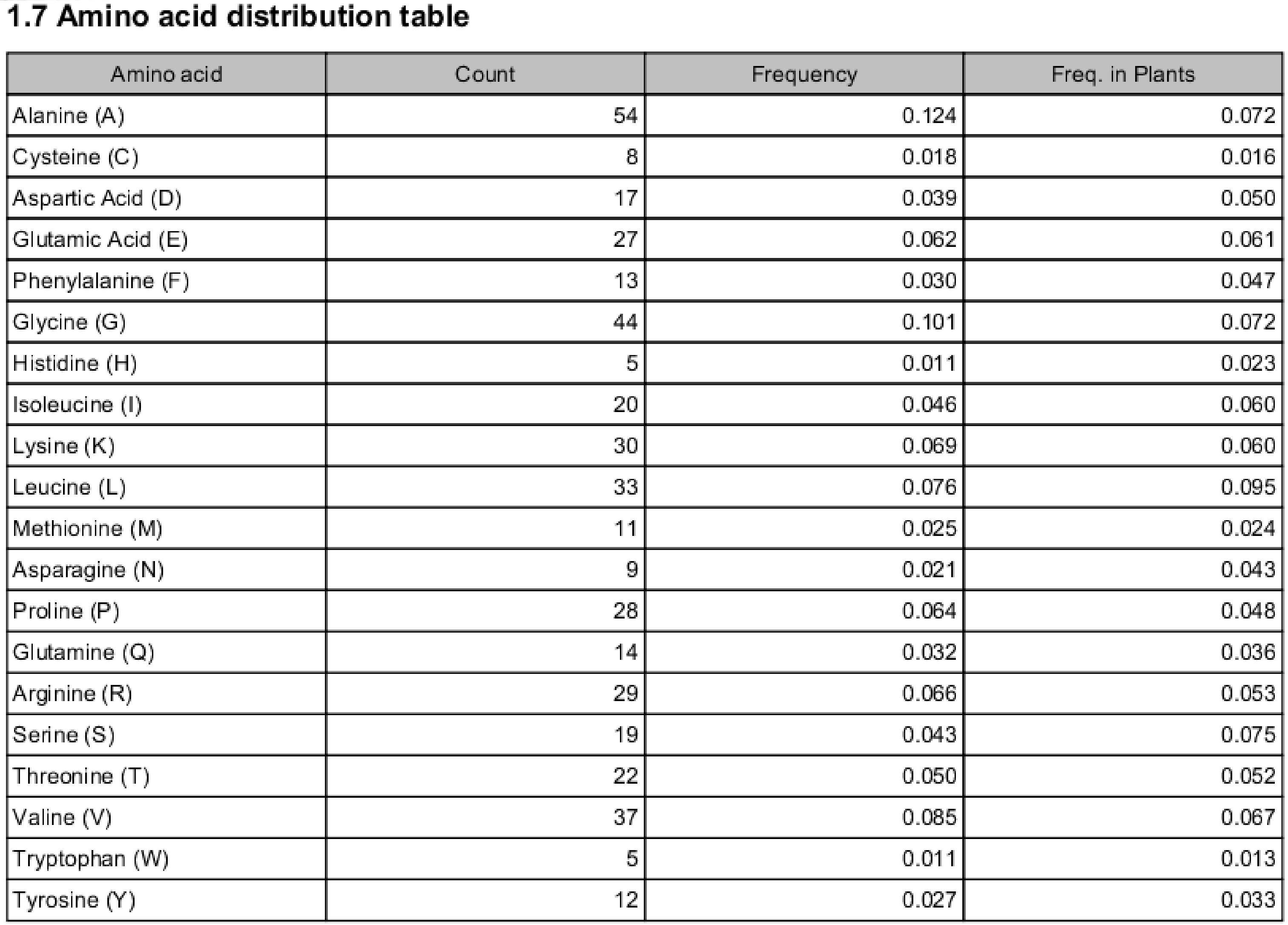
Hydrophobicity	Count	Frequency
Hydrophobic (A,F,G, I,L,M,P,V,W)	245	0.561
Hydrophilic (C,N,Q, S,T,Y)	84	0.192
Other	108	0.247



1.6 Count of charged residues

Charge Type	Count	Frequency
Negatively Charged (D & E)	44	0.101
Positively Charged (R & K)	59	0.135
Other	334	0.764

STATISTICA SECVENŢEI

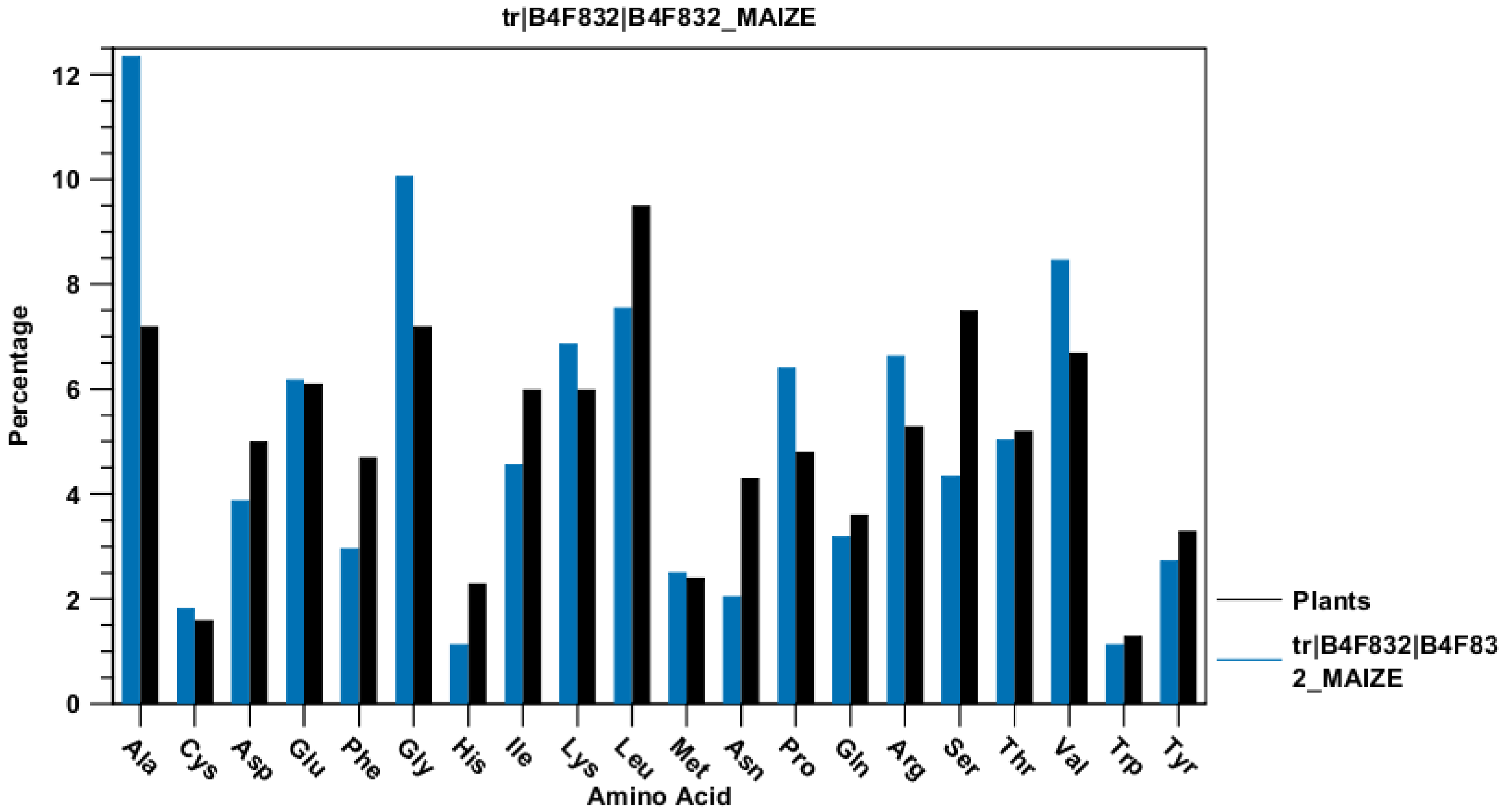


1.7 Amino acid distribution table

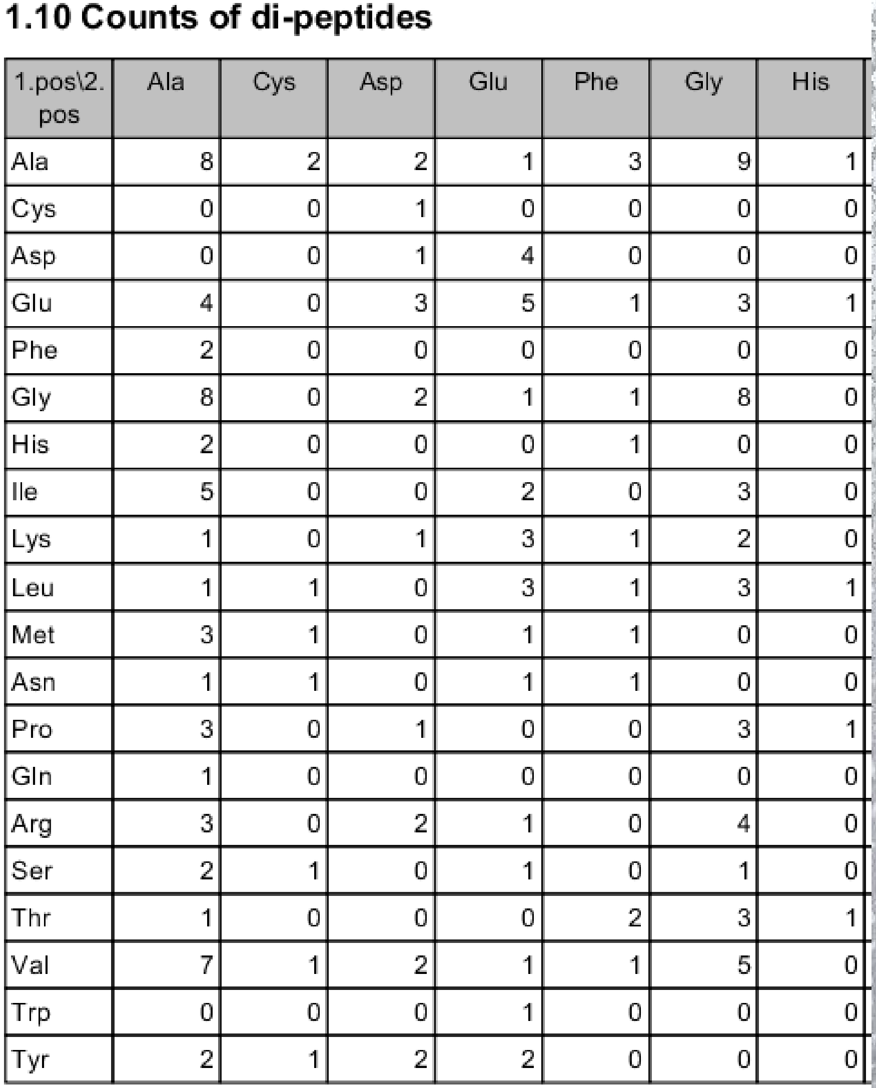
Amino acid	Count	Frequency	Freq. in Plants
Alanine (A)	54	0.124	0.072
Cysteine (C)	8	0.018	0.016
Aspartic Acid (D)	17	0.039	0.050
Glutamic Acid (E)	27	0.062	0.061
Phenylalanine (F)	13	0.030	0.047
Glycine (G)	44	0.101	0.072
Histidine (H)	5	0.011	0.023
Isoleucine (I)	20	0.046	0.060
Lysine (K)	30	0.069	0.060
Leucine (L)	33	0.076	0.095
Methionine (M)	11	0.025	0.024
Asparagine (N)	9	0.021	0.043
Proline (P)	28	0.064	0.048
Glutamine (Q)	14	0.032	0.036
Arginine (R)	29	0.066	0.053
Serine (S)	19	0.043	0.075
Threonine (T)	22	0.050	0.052
Valine (V)	37	0.085	0.067
Tryptophan (W)	5	0.011	0.013
Tyrosine (Y)	12	0.027	0.033

STATISTICA SECVENȚEI

1.8 Amino acid distribution histogram



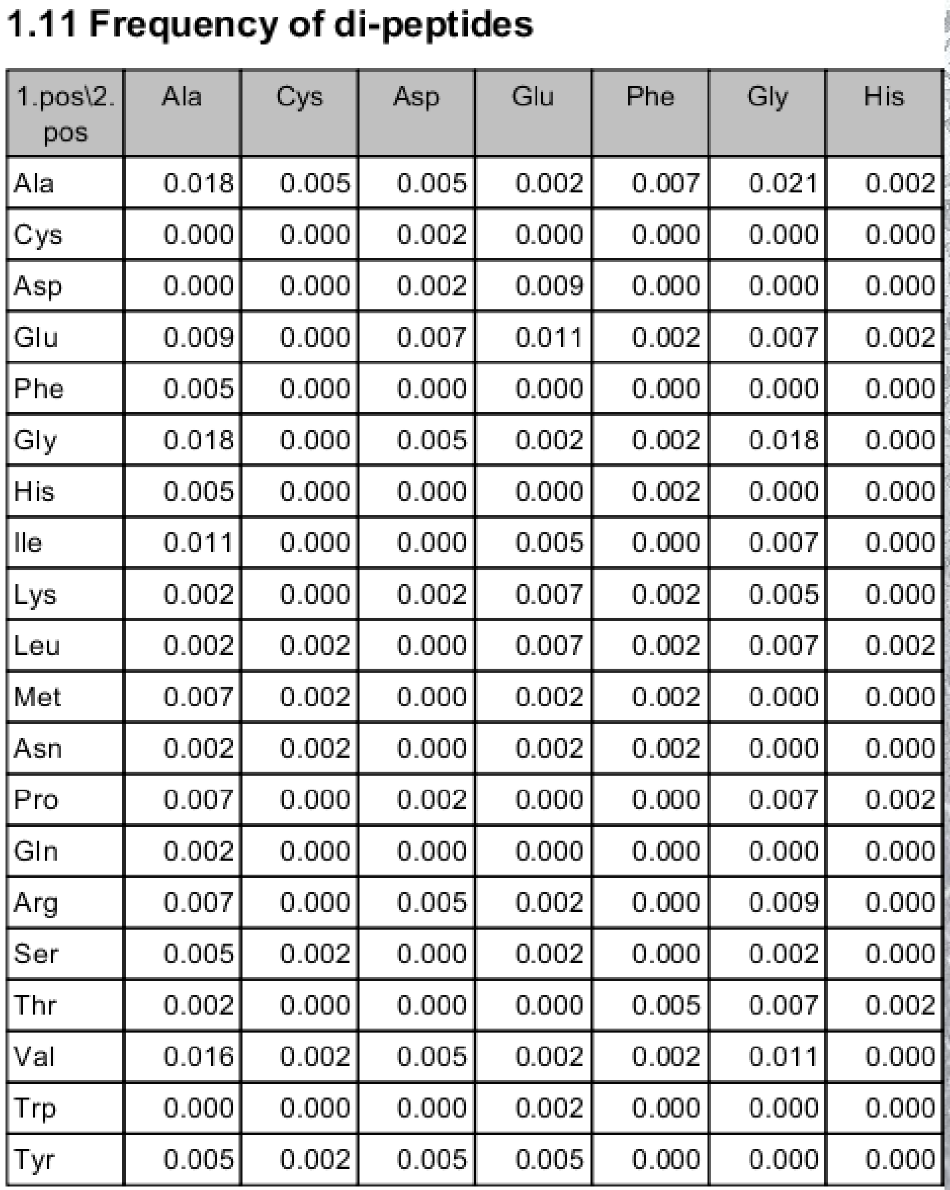
STATISTICA SECVENȚEI



1.10 Counts of di-peptides

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

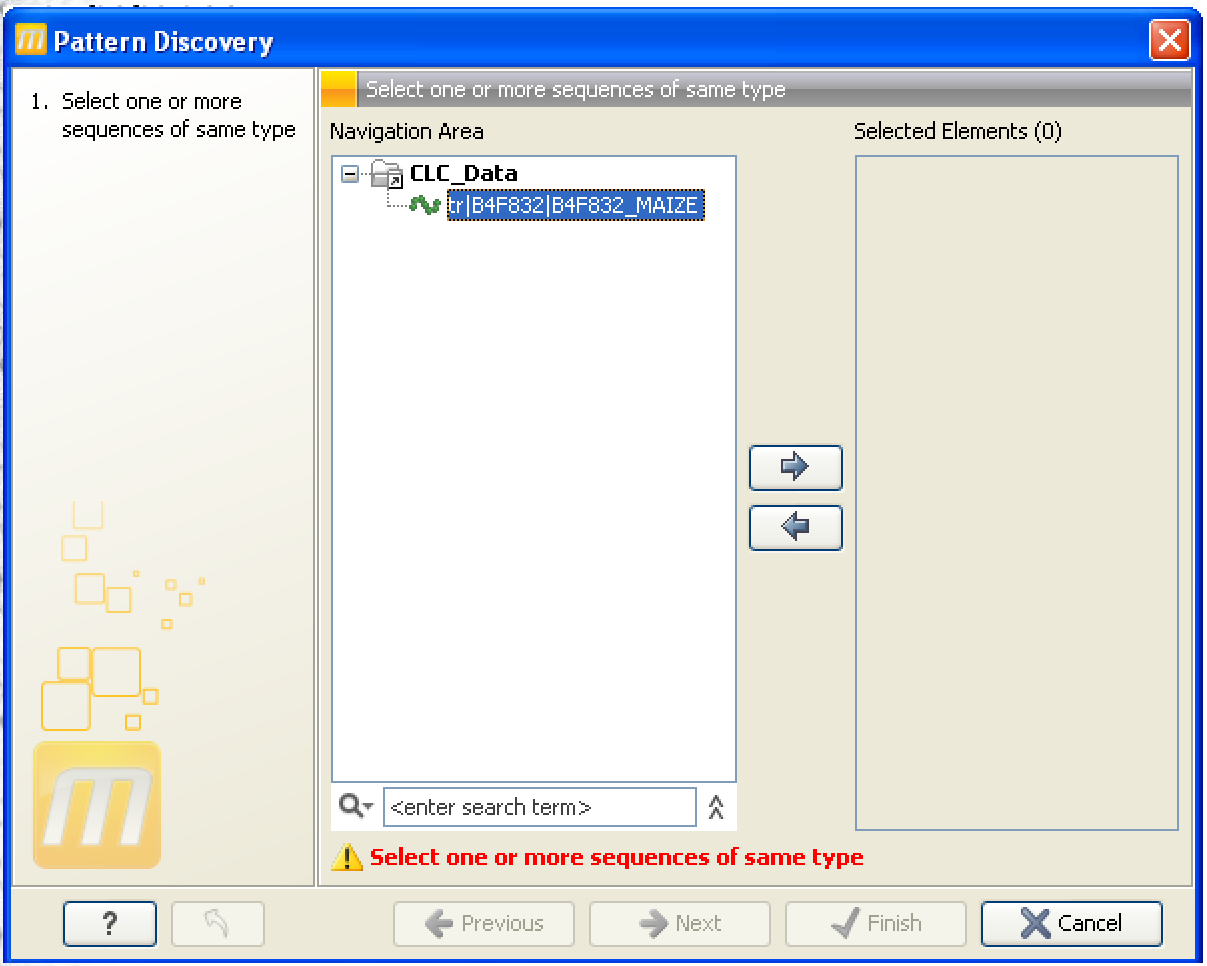
STATISTICA SECVENȚEI



1.11 Frequency of di-peptides

1.pos\2. pos	Ala	Cys	Asp	Glu	Phe	Gly	His
Ala	0.018	0.005	0.005	0.002	0.007	0.021	0.002
Cys	0.000	0.000	0.002	0.000	0.000	0.000	0.000
Asp	0.000	0.000	0.002	0.009	0.000	0.000	0.000
Glu	0.009	0.000	0.007	0.011	0.002	0.007	0.002
Phe	0.005	0.000	0.000	0.000	0.000	0.000	0.000
Gly	0.018	0.000	0.005	0.002	0.002	0.018	0.000
His	0.005	0.000	0.000	0.000	0.002	0.000	0.000
Ile	0.011	0.000	0.000	0.005	0.000	0.007	0.000
Lys	0.002	0.000	0.002	0.007	0.002	0.005	0.000
Leu	0.002	0.002	0.000	0.007	0.002	0.007	0.002
Met	0.007	0.002	0.000	0.002	0.002	0.000	0.000
Asn	0.002	0.002	0.000	0.002	0.002	0.000	0.000
Pro	0.007	0.000	0.002	0.000	0.000	0.007	0.002
Gln	0.002	0.000	0.000	0.000	0.000	0.000	0.000
Arg	0.007	0.000	0.005	0.002	0.000	0.009	0.000
Ser	0.005	0.002	0.000	0.002	0.000	0.002	0.000
Thr	0.002	0.000	0.000	0.000	0.005	0.007	0.002
Val	0.016	0.002	0.005	0.002	0.002	0.011	0.000
Trp	0.000	0.000	0.000	0.002	0.000	0.000	0.000
Tyr	0.005	0.002	0.005	0.005	0.000	0.000	0.000

IDENTIFICAREA PATTERN-URILOR ÎN CLC



Pattern Discovery

1. Select one or more sequences of same type

Select one or more sequences of same type

Navigation Area

Selected Elements (0)

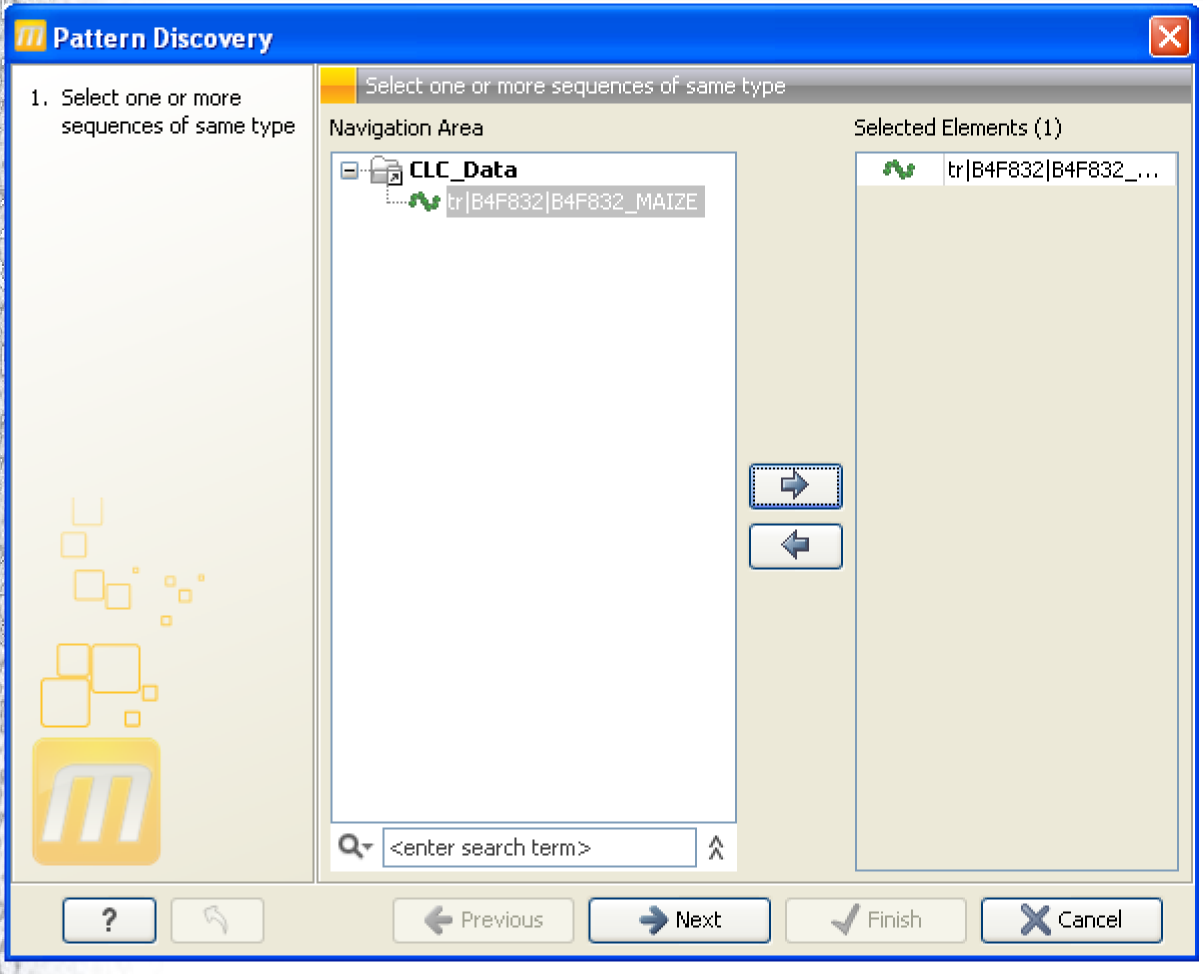
CLC_Data

tr|B4F832|B4F832_MAIZE

<enter search term>

Select one or more sequences of same type

IDENTIFICAREA PATTERN-URILOR ÎN CLC



Pattern Discovery

1. Select one or more sequences of same type

Select one or more sequences of same type

Navigation Area

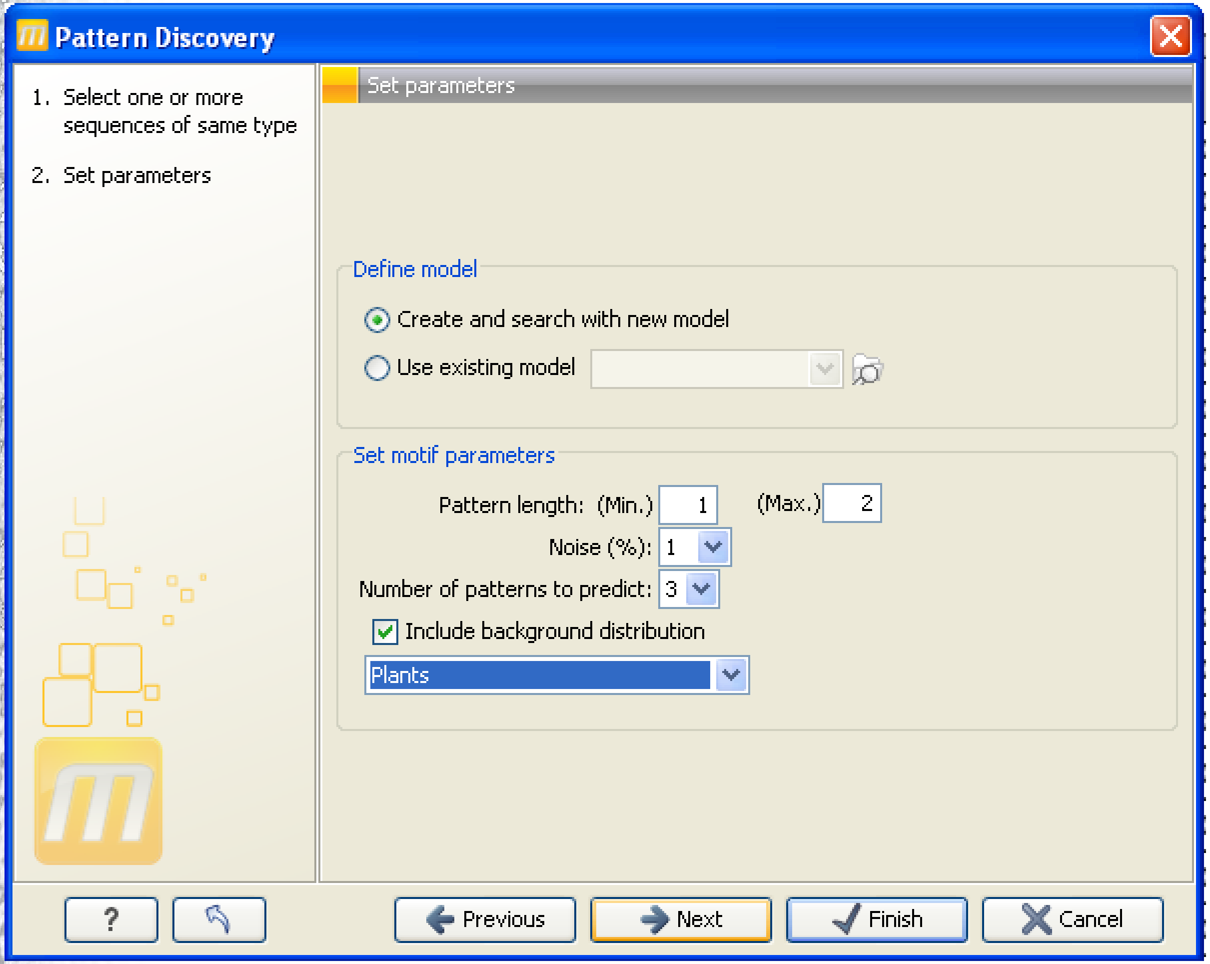
CLC_Data

tr|B4F832|B4F832_MAIZE

Selected Elements (1)

tr|B4F832|B4F832_...

IDENTIFICAREA PATTERN-URILOR ÎN CLC



1. Select one or more sequences of same type
2. Set parameters

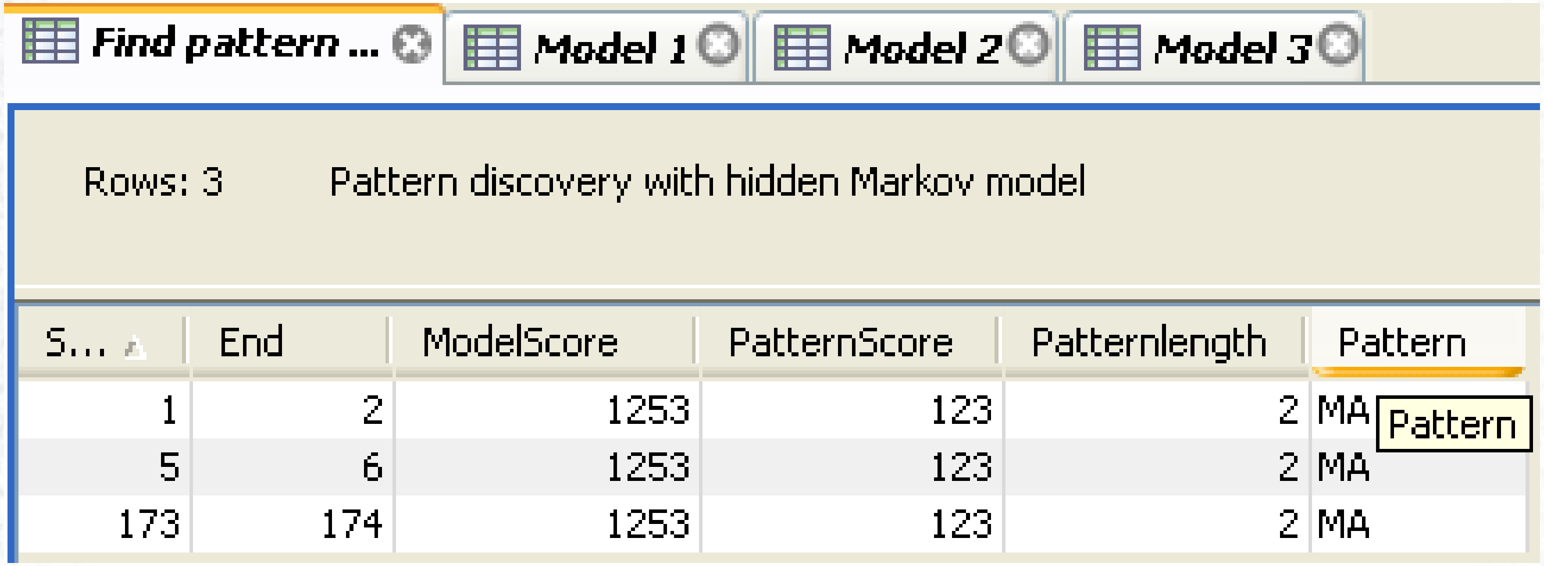
Define model

● Create and search with new model
○ Use existing model

Set motif parameters

Pattern length: (Min.) 1 (Max.) 2
Noise (%): 1
Number of patterns to predict: 3
☑ Include background distribution
Plants

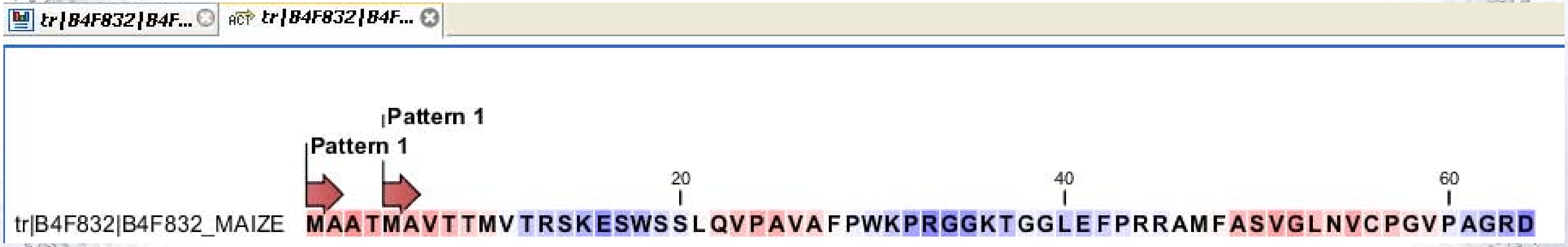
IDENTIFICAREA PATTERN-URILOR ÎN CLC



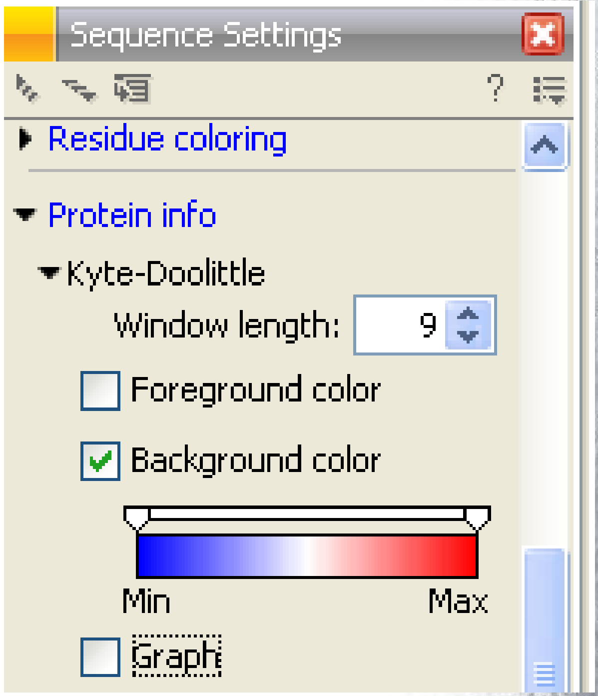
Rows: 3 Pattern discovery with hidden Markov model

S...	End	ModelScore	PatternScore	Patternlength	Pattern
1	2	1253	123	2	MA
5	6	1253	123	2	MA
173	174	1253	123	2	MA

IDENTIFICAREA PATTERN-URILOR ÎN CLC



tr|B4F832|B4F832_MAIZE MAATMAVTTMVTRSKESWSSLQVPAVAFPWKPRGGKTGGLEFPRRAMFASVGLNVCPGVPAGRD



Kyte-Doolittle

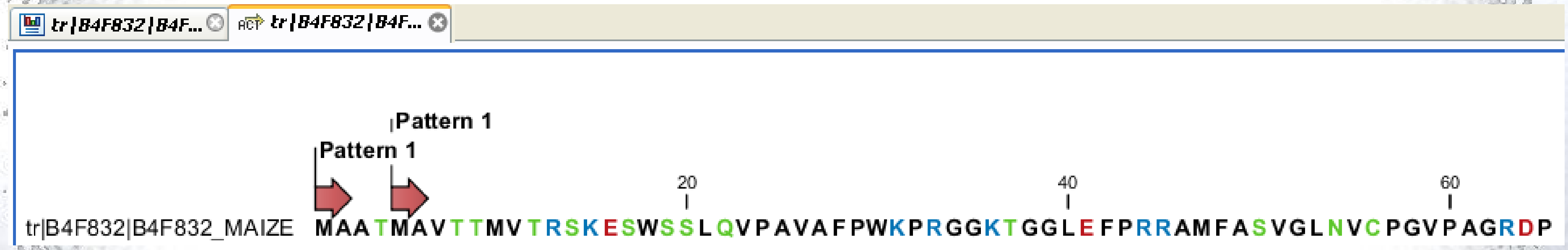
Window length: 9

☐ Foreground color

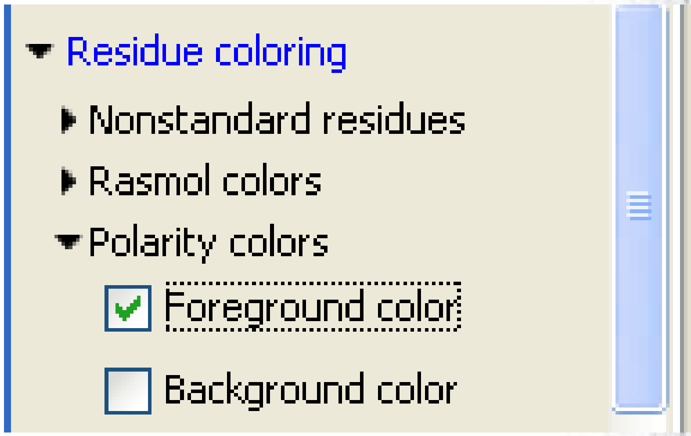
☑ Background color

☐ Graph

IDENTIFICAREA PATTERN-URILOR ÎN CLC

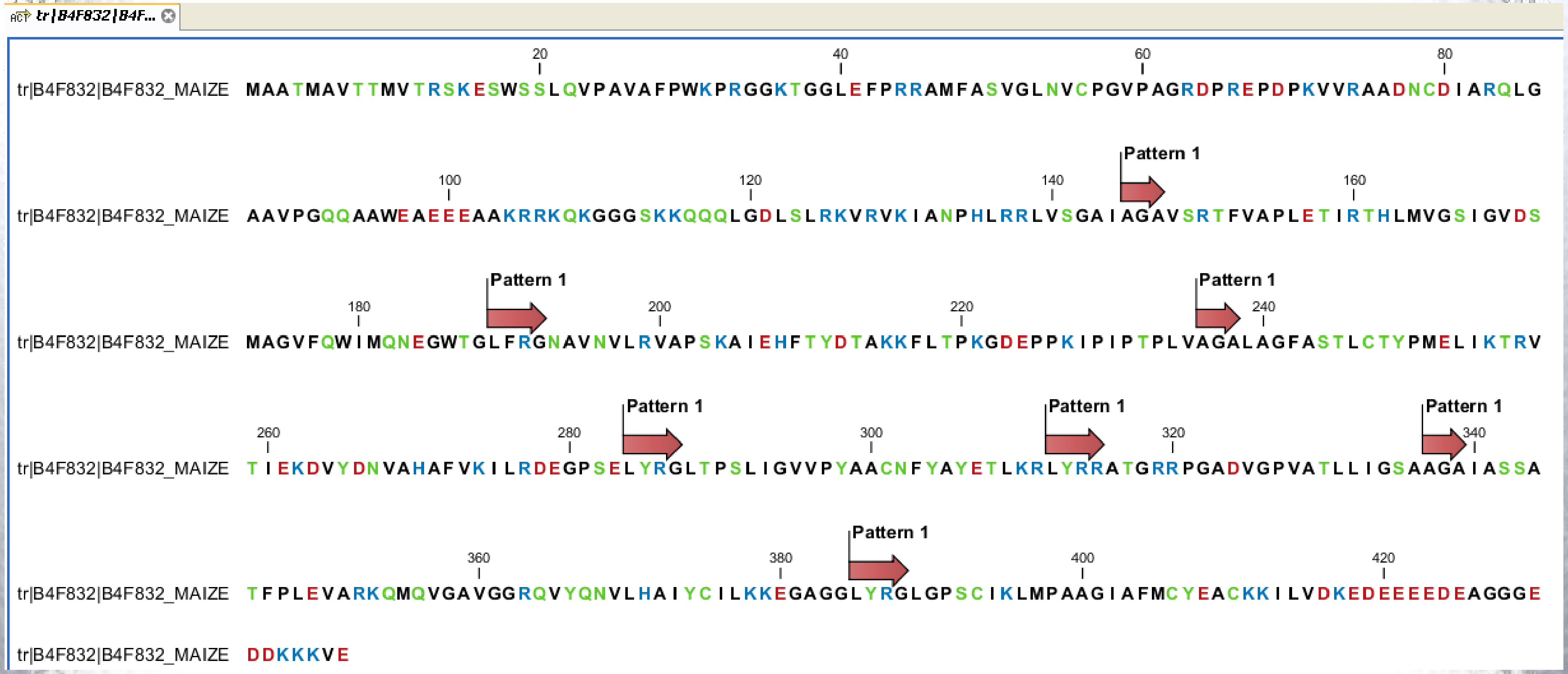


tr|B4F832|B4F832_MAIZE MAATMAVTTMVTRSKESWSSLQVPAVAFPWKPRGGKTGGLEFPRRAMFASVGLNVCPGVPAGRDP



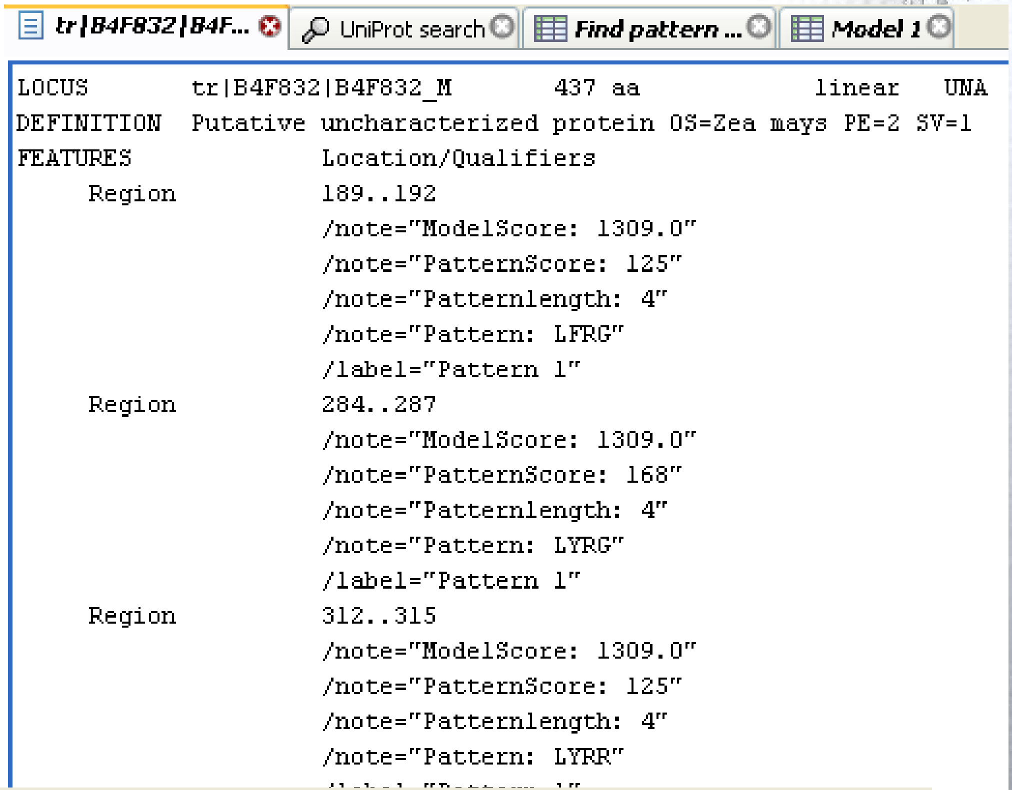
- Residue coloring
  - Nonstandard residues
  - Rasmol colors
  - Polarity colors
    - ☑ Foreground color
    - ☐ Background color

IDENTIFICAREA PATTERN-URILOR ÎN CLC

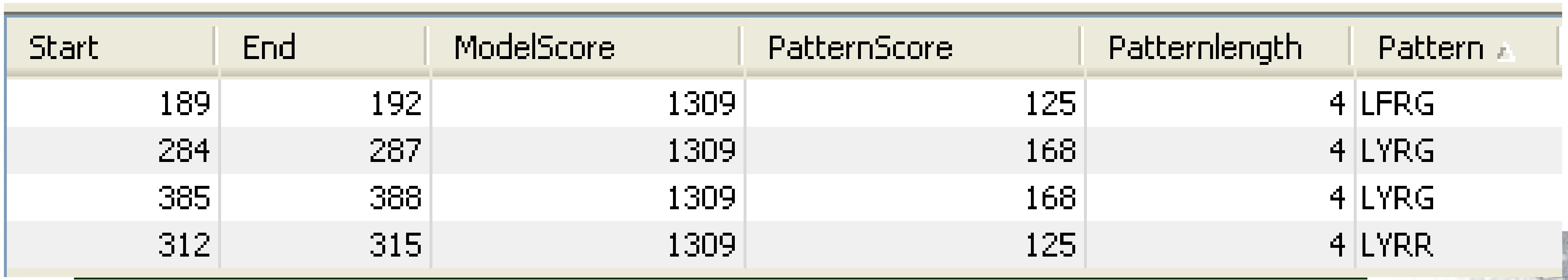


```
tr|B4F832|B4F832_MAIZE   MAATMAVTTMVTRSKESWSSLQVPAVAFPWKPRGGKTGGLEFPRRAMFASVGLNVCPGVPAGRDPREPDPKVVRAADNCDIARQLG
tr|B4F832|B4F832_MAIZE   AAVPGQQAAWEAEEEAAKRRKQKGGGSKKQQQLGDLSLRKVRVKIANPHLRRLVSGAIAGAVSRTFVAPLETIRTHLMVGSIGVDS
tr|B4F832|B4F832_MAIZE   MAGVFQWIMQNEGWTGLFRGNAVNVLRVAPSKAIEHFTYDTAKKFLTPKGDEPPKIPIPTPLVAGALAGFASTLCTYPMELIKTRV
tr|B4F832|B4F832_MAIZE   TIEKDVYDNVAHAFVKILRDEGPSELYRGLTPSLIGVVPYAACNFYAYETLKRLYRRATGRRPGADVGPVATLLIGSAAGAIASSA
tr|B4F832|B4F832_MAIZE   TFPLEVARKQMQVGAVGGRQVYQNVLHAIYCILKKEGAGGLYRGLGPSCIKLMPAAGIAFMCYEACKKILVDKEDEEEEDEAGGGE
tr|B4F832|B4F832_MAIZE   DDKKKVE
```

IDENTIFICAREA PATTERN-URILOR ÎN CLC

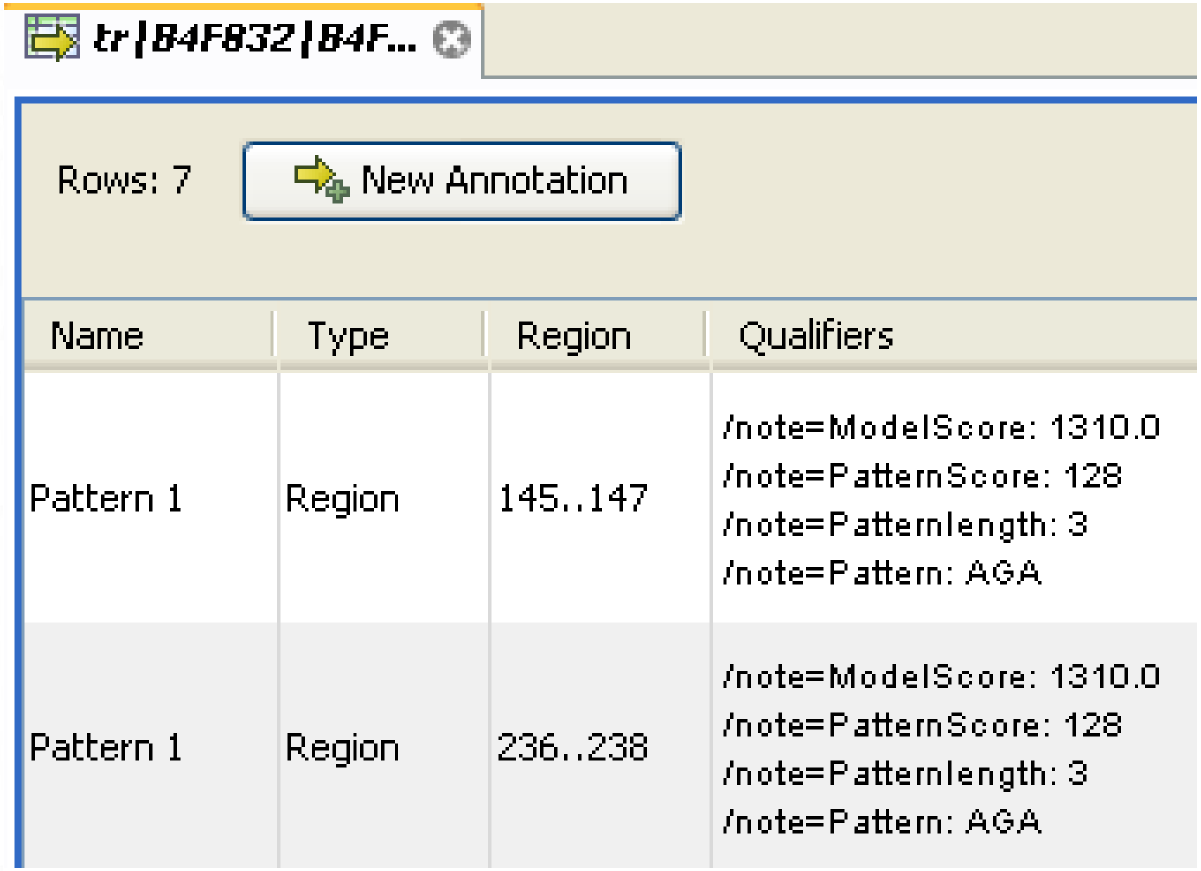


```
LOCUS       tr|B4F832|B4F832_M       437 aa            linear   UNA
DEFINITION  Putative uncharacterized protein OS=Zea mays PE=2 SV=1
FEATURES             Location/Qualifiers
     Region          189..192
                     /note="ModelScore: 1309.0"
                     /note="PatternScore: 125"
                     /note="Patternlength: 4"
                     /note="Pattern: LFRG"
                     /label="Pattern 1"
     Region          284..287
                     /note="ModelScore: 1309.0"
                     /note="PatternScore: 168"
                     /note="Patternlength: 4"
                     /note="Pattern: LYRG"
                     /label="Pattern 1"
     Region          312..315
                     /note="ModelScore: 1309.0"
                     /note="PatternScore: 125"
                     /note="Patternlength: 4"
                     /note="Pattern: LYRR"
```



Start	End	ModelScore	PatternScore	Patternlength	Pattern
189	192	1309	125	4	LFRG
284	287	1309	168	4	LYRG
385	388	1309	168	4	LYRG
312	315	1309	125	4	LYRR

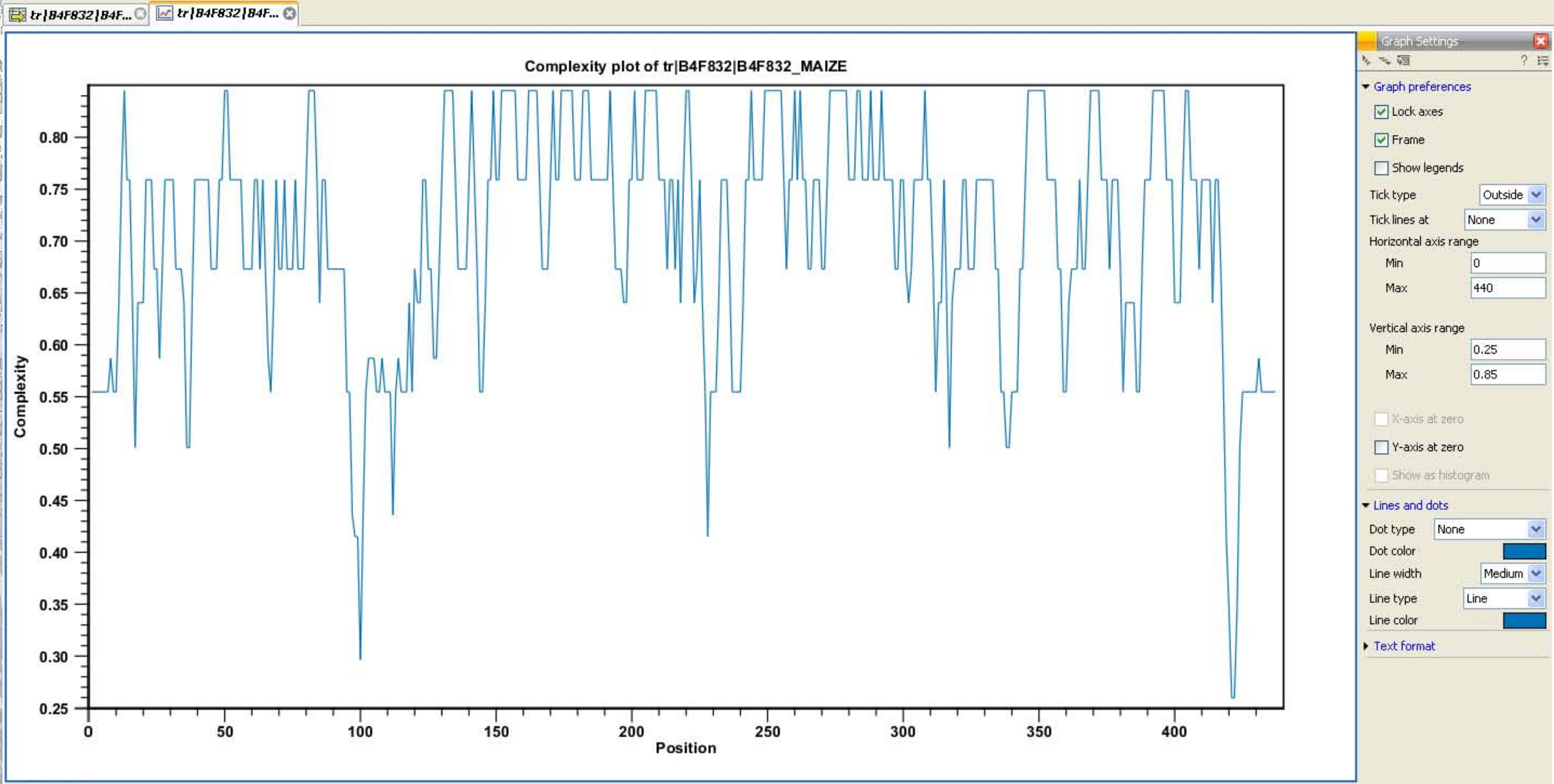
CLC REZULTATE



Rows: 7

Name	Type	Region	Qualifiers
Pattern 1	Region	145..147	/note=ModelScore: 1310.0 /note=PatternScore: 128 /note=Patternlength: 3 /note=Pattern: AGA
Pattern 1	Region	236..238	/note=ModelScore: 1310.0 /note=PatternScore: 128 /note=Patternlength: 3 /note=Pattern: AGA

CLC REZULTATE: COMPELXITY PLOT



Graph preferences

☑ Lock axes

☑ Frame

☐ Show legends

Tick type: Outside

Tick lines at: None

Horizontal axis range

Min: 0

Max: 440

Vertical axis range

Min: 0.25

Max: 0.85

☐ X-axis at zero

☐ Y-axis at zero

☐ Show as histogram

Lines and dots

Dot type: None

Dot color

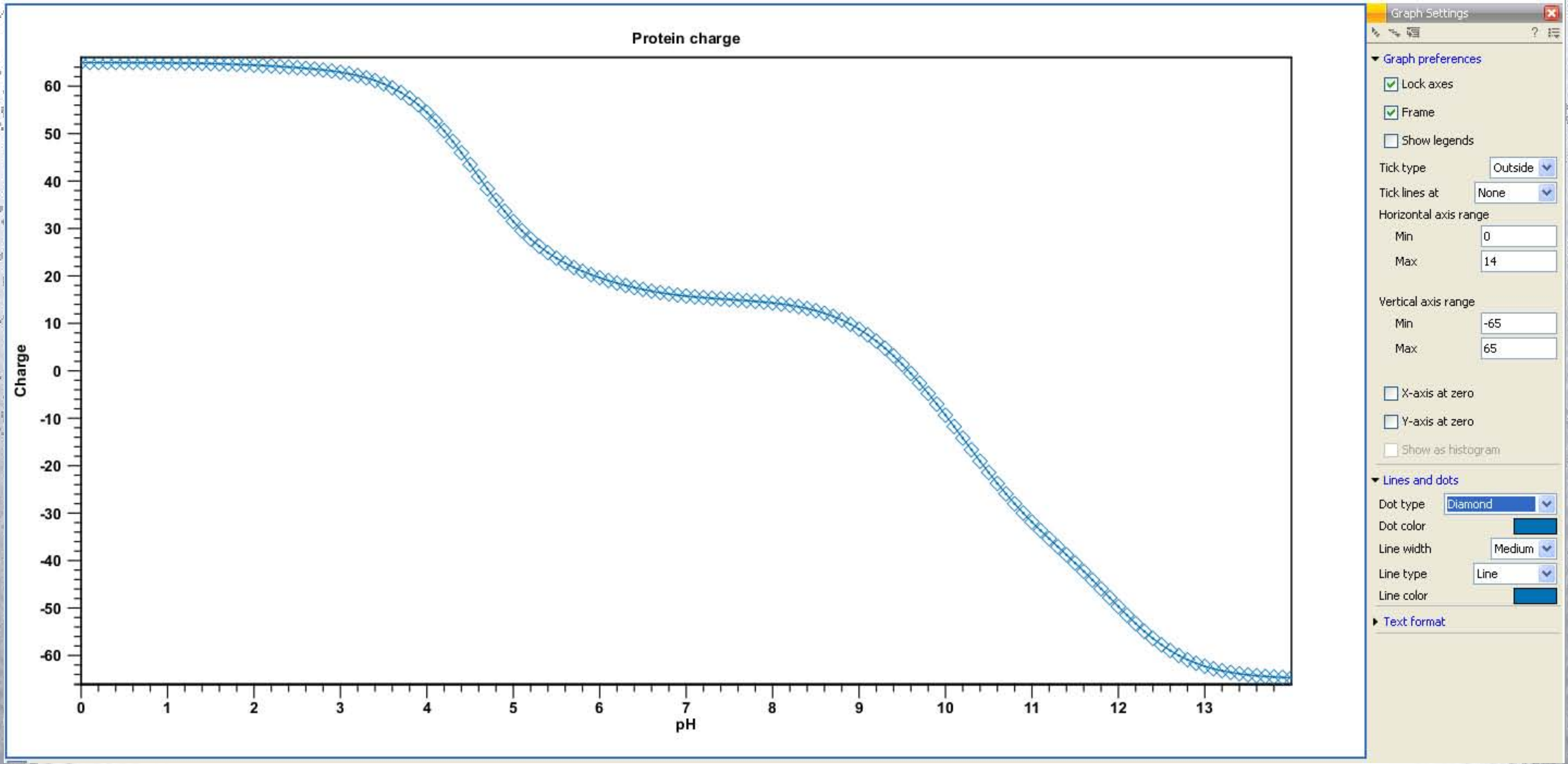
Line width: Medium

Line type: Line

Line color

Text format

CLC REZULTATE: PROTEINE CHARGE PLOT



Graph preferences

☑ Lock axes

☑ Frame

☐ Show legends

Tick type: Outside

Tick lines at: None

Horizontal axis range

Min: 0

Max: 14

Vertical axis range

Min: -65

Max: 65

☐ X-axis at zero

☐ Y-axis at zero

☐ Show as histogram

Lines and dots

Dot type: Diamond

Dot color

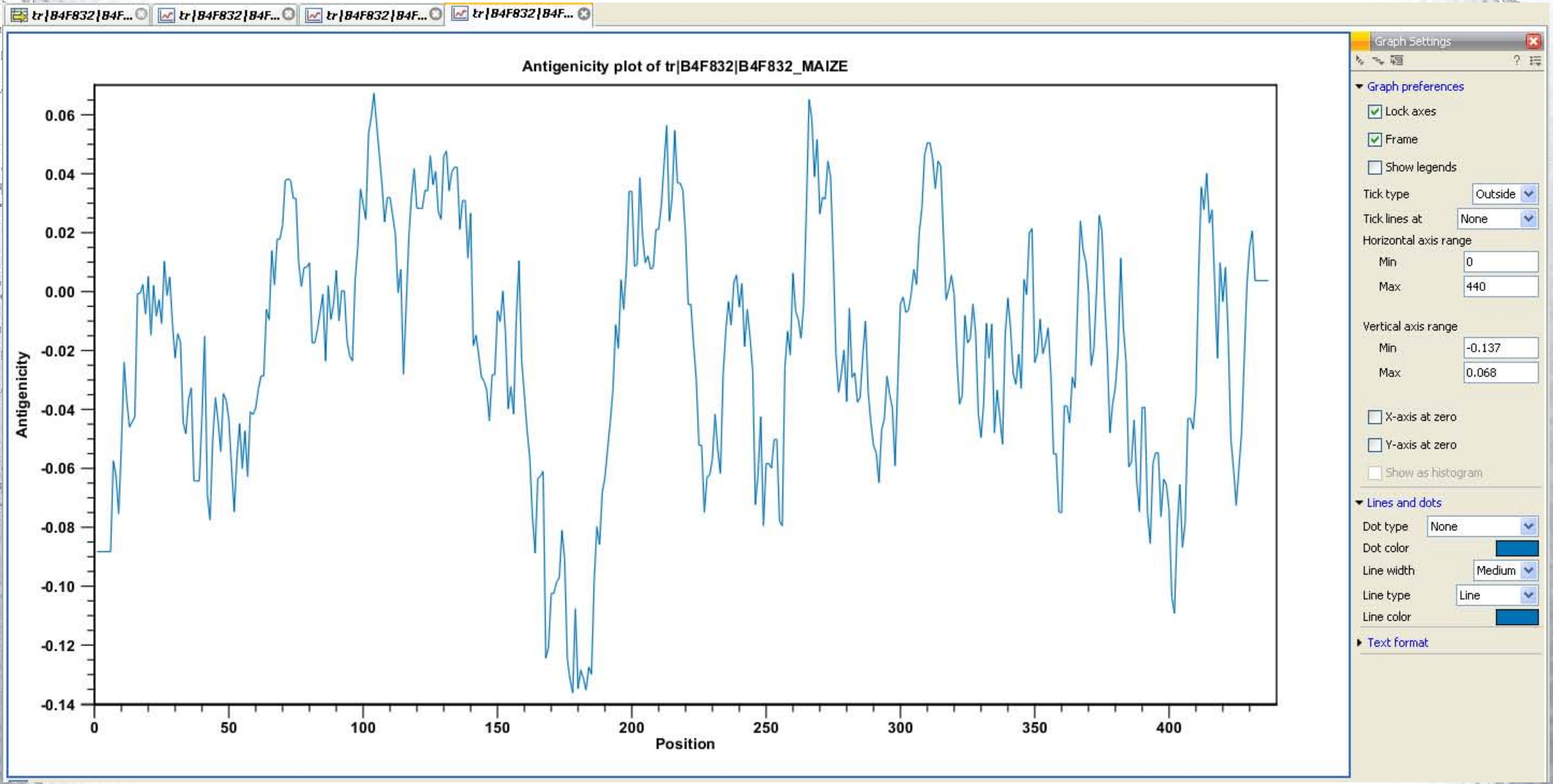
Line width: Medium

Line type: Line

Line color

Text format

CLC Rezultate: Antigenicity Plot



Graph preferences

- ☑ Lock axes
- ☑ Frame
- ☐ Show legends

Tick type: Outside

Tick lines at: None

Horizontal axis range
- Min: 0
- Max: 440

Vertical axis range
- Min: -0.137
- Max: 0.068

- ☐ X-axis at zero
- ☐ Y-axis at zero
- ☐ Show as histogram

Lines and dots

Dot type: None

Dot color: 🟦

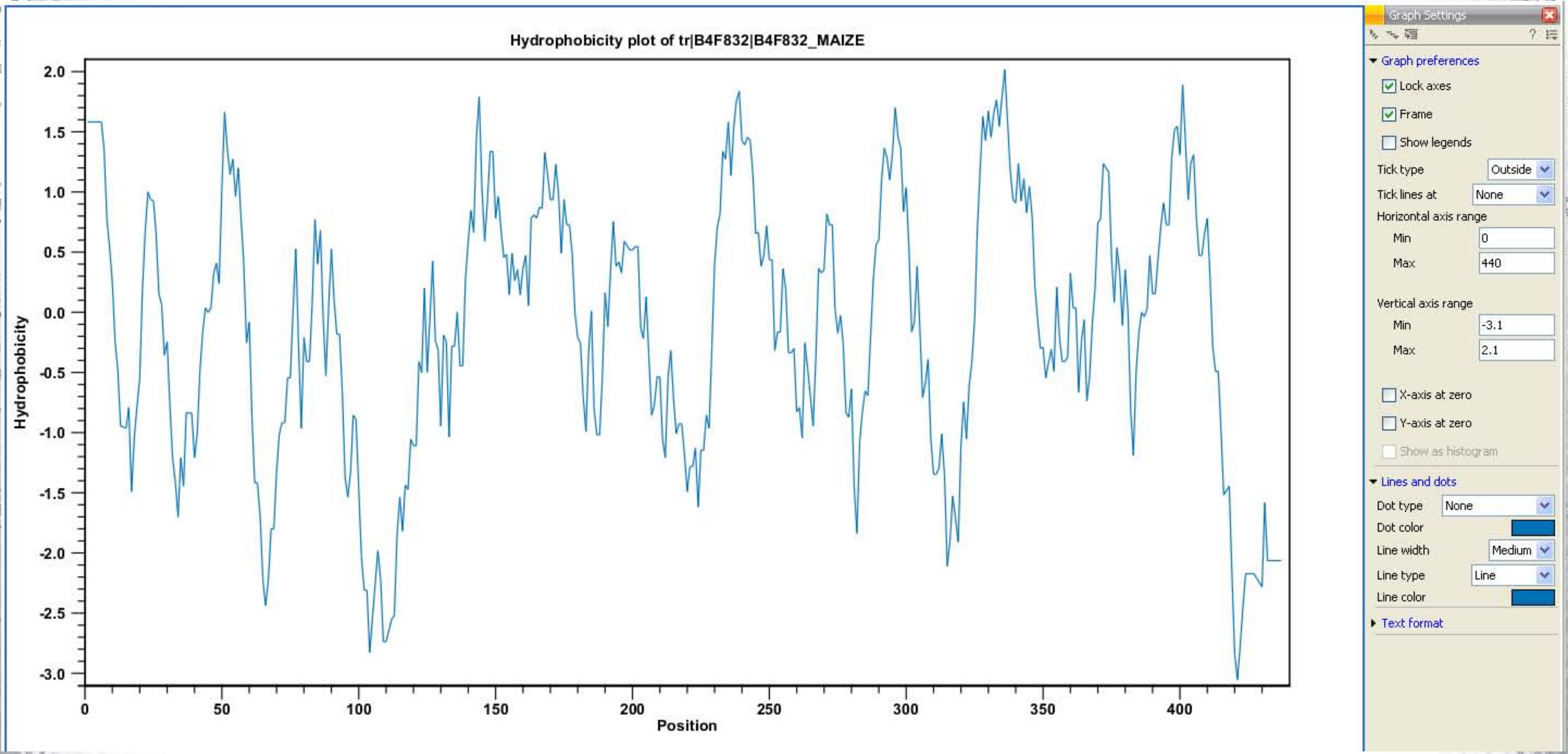
Line width: Medium

Line type: Line

Line color: 🟦

Text format

CLC REZULTATE: HYDROPHOBICITY PLOT



Graph preferences

☑ Lock axes

☑ Frame

☐ Show legends

Tick type: Outside

Tick lines at: None

Horizontal axis range
Min: 0
Max: 440

Vertical axis range
Min: -3.1
Max: 2.1

☐ X-axis at zero

☐ Y-axis at zero

☐ Show as histogram

Lines and dots

Dot type: None

Dot color: 🟦

Line width: Medium

Line type: Line

Line color: 🟦

Text format

PROBLEMA: NUCLEOTIDE

- Blackberry chlorotic ringspot virus
  - Acronim: BCRV
- ARN 1
  - mRNA
  - 3478 bp

IDENTIFICAREA SECVENȚEI

- http://www.ncbi.nlm.nih.gov/nuccore/71836079

- Jones AT, McGavinWJ, GeppV, Zimmerman MT, Scott SW. Purification and properties of blackberry chlorotic ringspot, a new virus species in subgroup 1 of the genus Ilarvirus found naturally infecting blackberry in the UK. Ann Appl Biol 2006;149(2):125-35.

Identificăm secvența fasta

Blackberr otic ringspot virus segment RNA 1, complete sequence

GenBank: D 91193.1

```
LOCUS       DQ091193                3478 bp    mRNA    linear   VRL 30-MAR-2010
DEFINITION  Blackberry chlorotic ringspot virus segment RNA 1, complete
            sequence.
ACCESSION   DQ091193
VERSION     DQ091193.1  GI:71836079
KEYWORDS    .
SOURCE      Blackberry chlorotic ringspot virus
  ORGANISM  Blackberry chlorotic ringspot virus
```

SALVAREA ÎN FORMAT FASTA

Search: Nucleotide

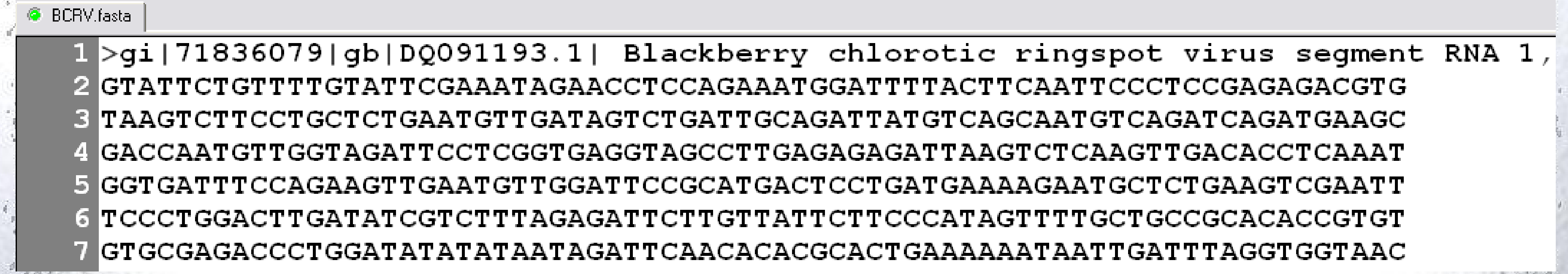
Display Settings: FASTA

Blackberry chlorotic ringspot virus segment RNA 1, complete sequence

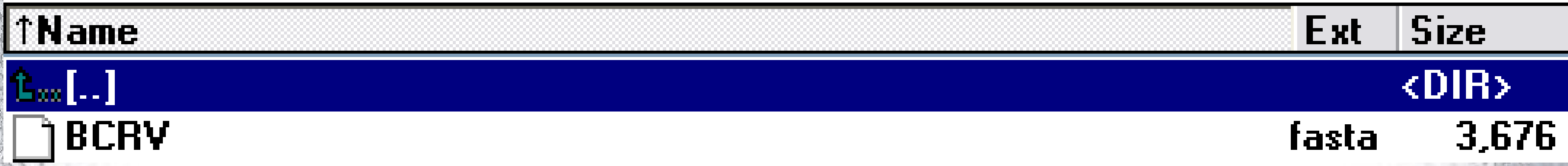
GenBank: DQ091193.1

```
>gi|71836079|gb|DQ091193.1| Blackberry chlorotic ringspot virus segment RNA 1,
complete sequence
GTATTCTGTTTTGTATTCGAAATAGAACCTCCAGAAATGGATTTTACTTCAATTCCCTCCGAGAGACGTG
TAAGTCTTCCTGCTCTGAATGTTGATAGTCTGATTGCAGATTATGTCAGCAATGTCAGATCAGATGAAGC
GACCAATGTTGGTAGATTCCTCGGTGAGGTAGCCTTGAGAGAGATTAAGTCTCAAGTTGACACCTCAAAT
GGTGATTTCCAGAAGTTGAATGTTGGATTCCGCATGACTCCTGATGAAAAGAATGCTCTGAAGTCGAATT
```

SALVAREA ÎN FORMAT FASTA

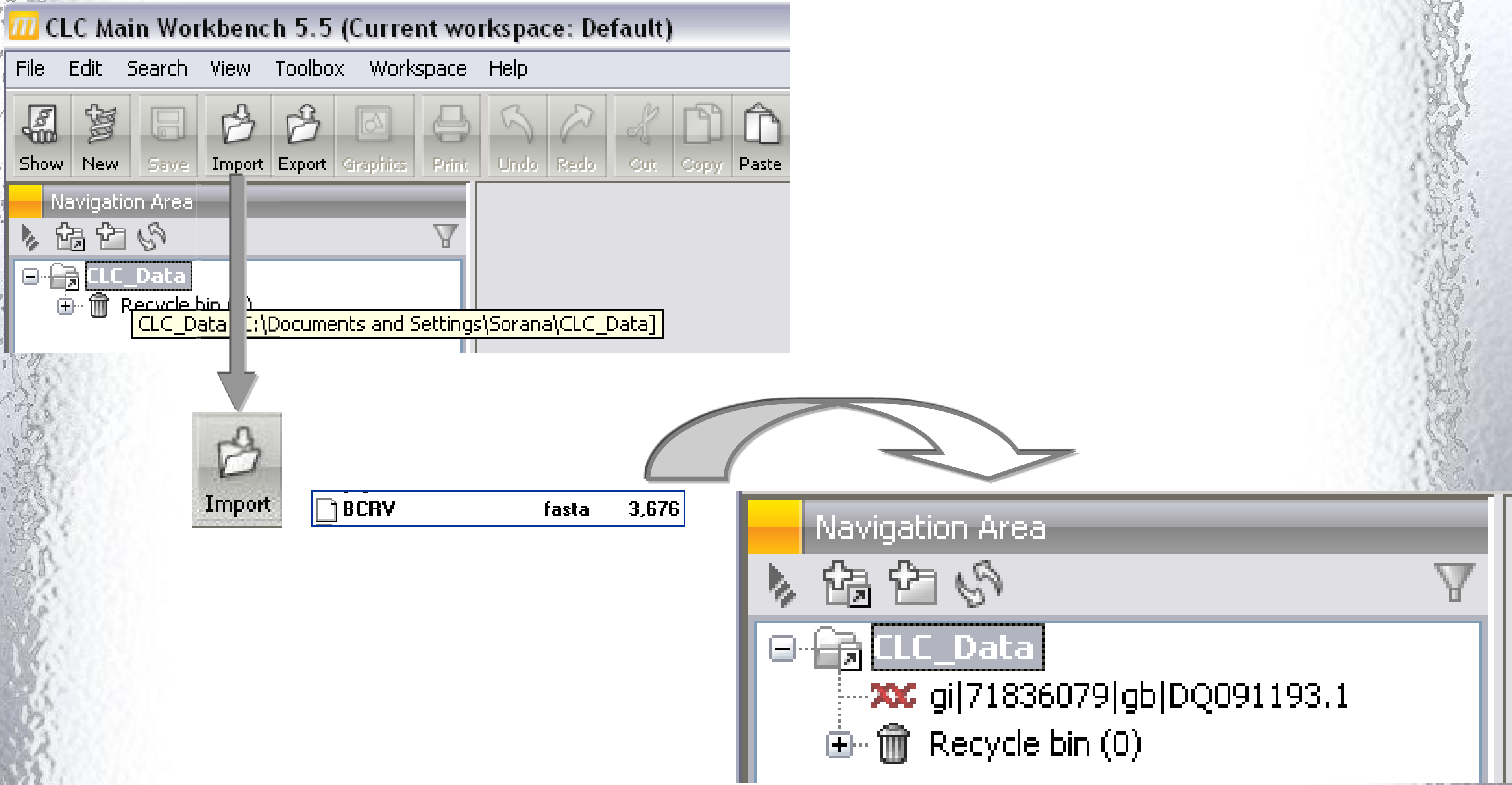


```
>gi|71836079|gb|DQ091193.1| Blackberry chlorotic ringspot virus segment RNA 1,
GTATTCTGTTTTGTATTCGAAATAGAACCTCCAGAAATGGATTTTACTTCAATTCCCTCCGAGAGACGTG
TAAGTCTTCCTGCTCTGAATGTTGATAGTCTGATTGCAGATTATGTCAGCAATGTCAGATCAGATGAAGC
GACCAATGTTGGTAGATTCCTCGGTGAGGTAGCCTTGAGAGAGATTAAGTCTCAAGTTGACACCTCAAAT
GGTGATTTCCAGAAGTTGAATGTTGGATTCCGCATGACTCCTGATGAAAAGAATGCTCTGAAGTCGAATT
TCCCTGGACTTGATATCGTCTTTAGAGATTCTTGTTATTCTTCCCATAGTTTTGCTGCCGCACACCGTGT
GTGCGAGACCCTGGATATATATAATAGATTCAACACACGCACTGAAAAAATAATTGATTTAGGTGGTAAC
```



↑Name	Ext	Size
[..]		<DIR>
BCRV	fasta	3,676

IMPORTUL FIŞIERULUI FASTA ÎN CLC

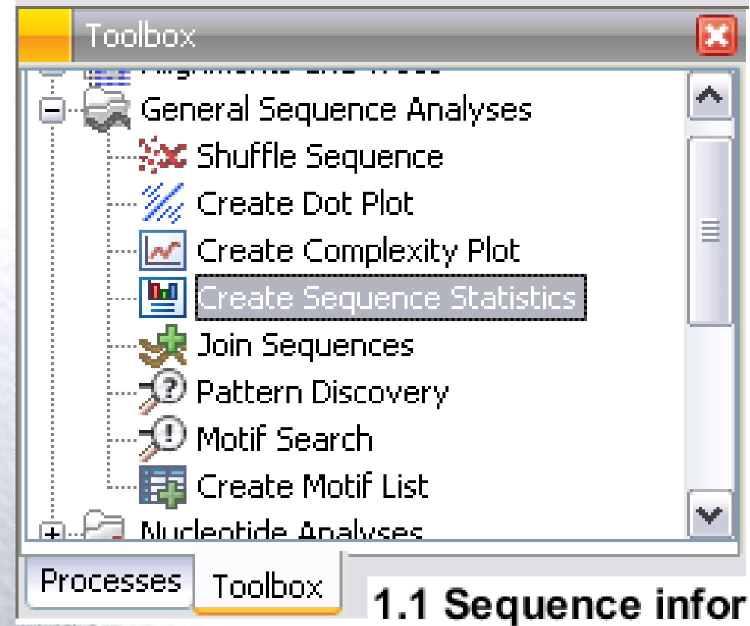


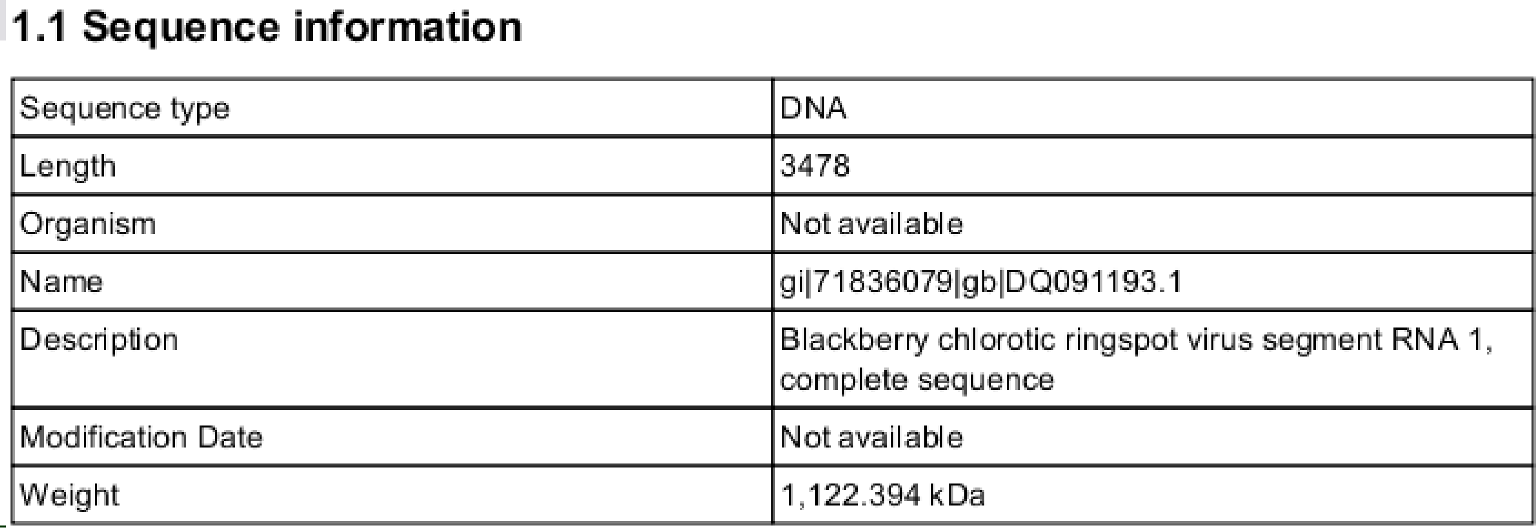
CLC_Data [C:\Documents and Settings\Sorana\CLC_Data]

BCRV fasta 3,676

CLC_Data
gi|71836079|gb|DQ091193.1
Recycle bin (0)

CLC: ANALIZA SECVENŢEI

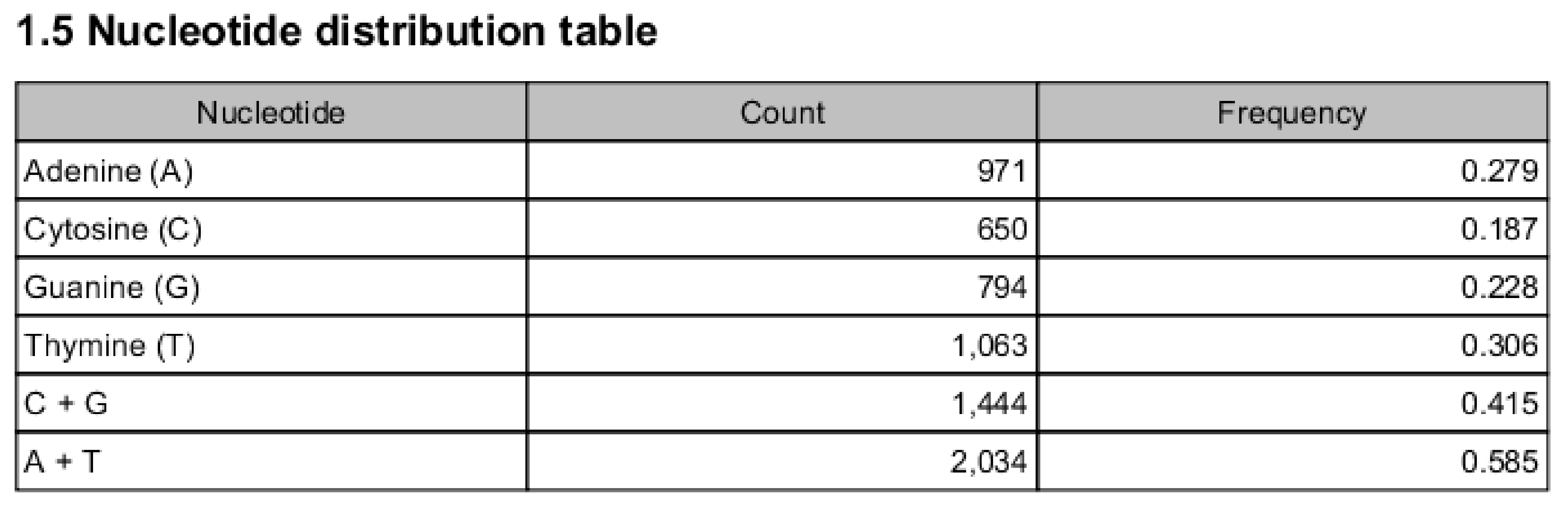




1.1 Sequence information

Sequence type	DNA
Length	3478
Organism	Not available
Name	gi|71836079|gb|DQ091193.1
Description	Blackberry chlorotic ringspot virus segment RNA 1, complete sequence
Modification Date	Not available
Weight	1,122.394 kDa

CLC: ANALIZA SECVENȚEI

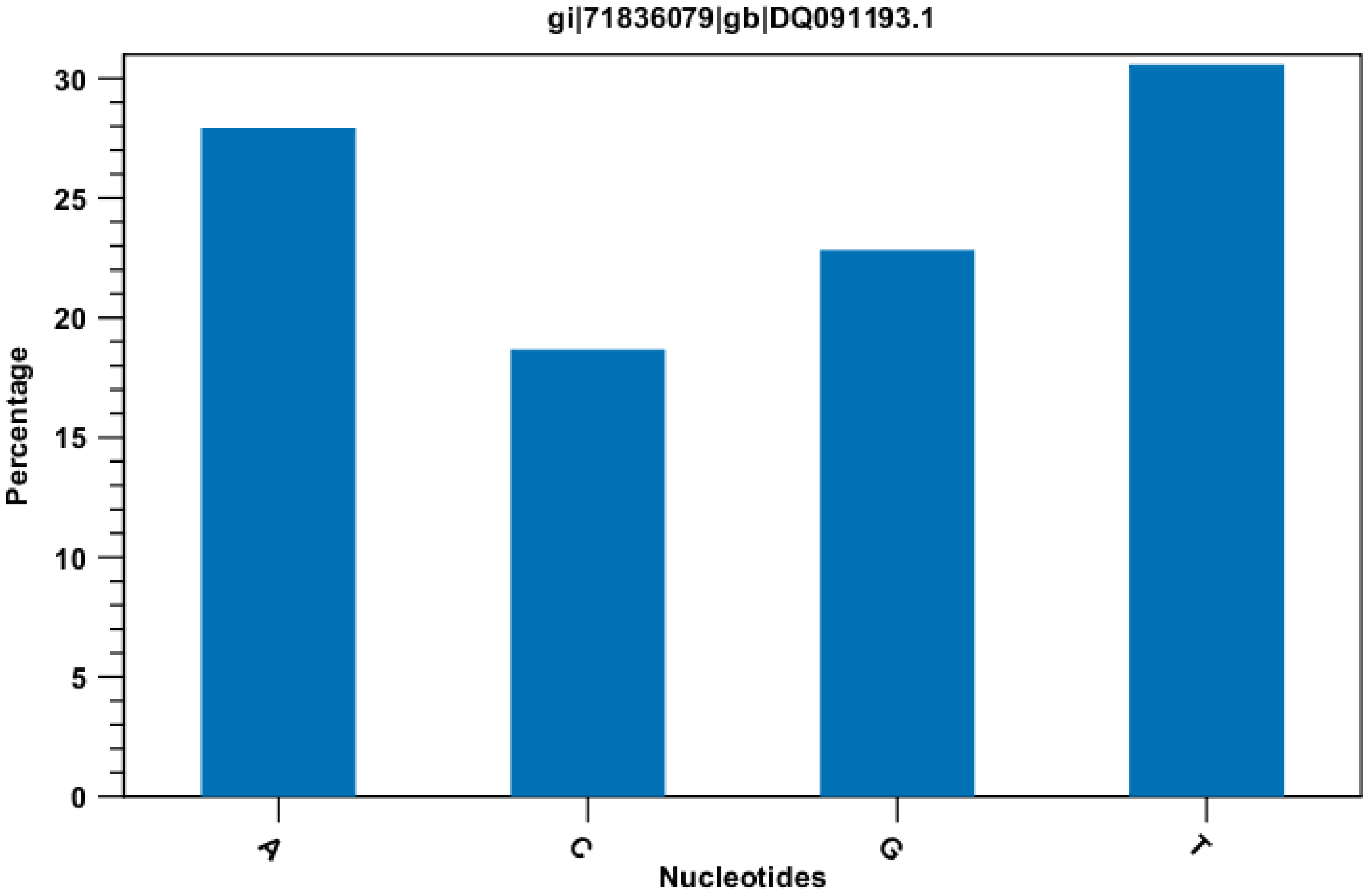


1.5 Nucleotide distribution table

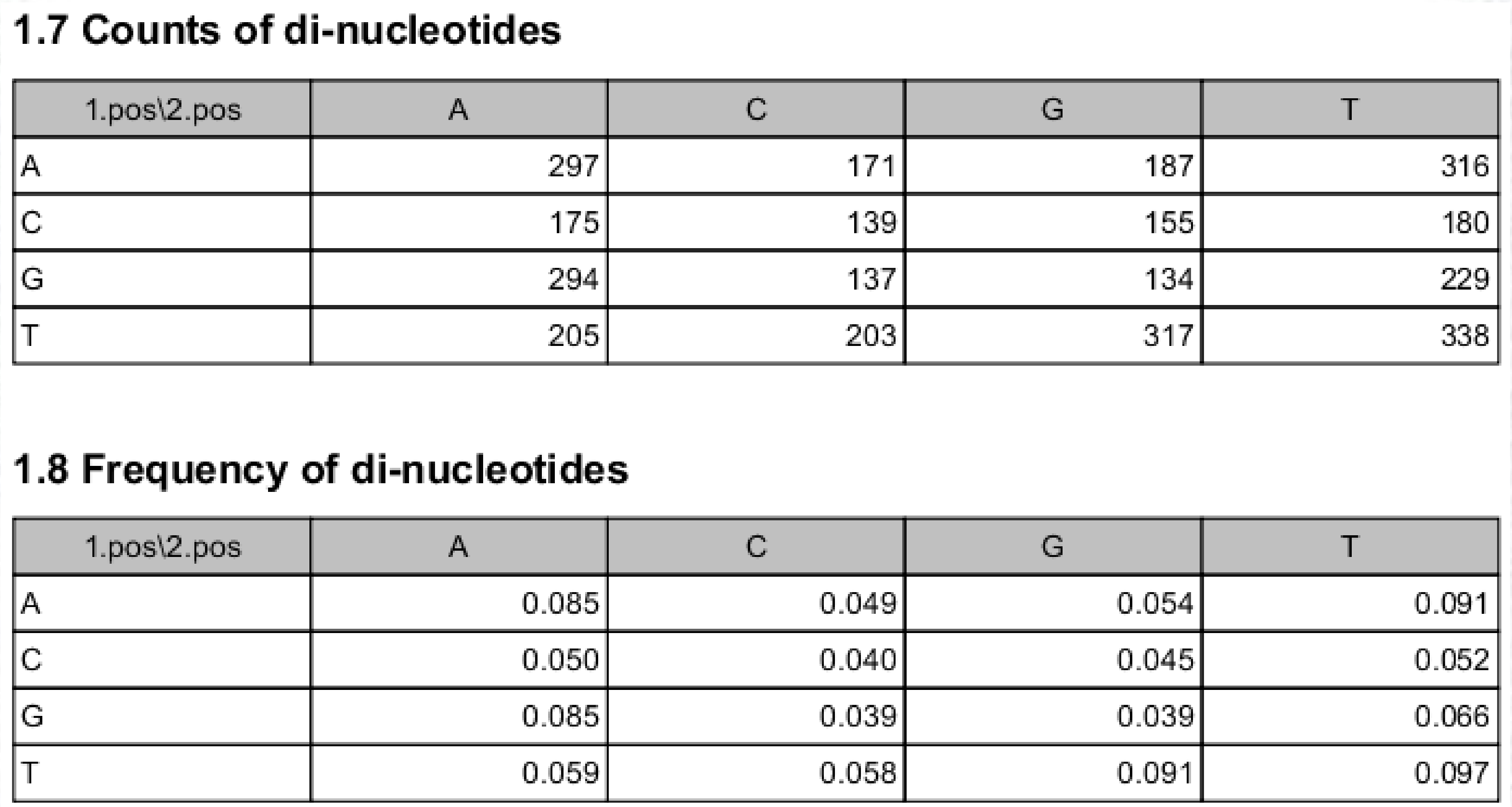
Nucleotide	Count	Frequency
Adenine (A)	971	0.279
Cytosine (C)	650	0.187
Guanine (G)	794	0.228
Thymine (T)	1,063	0.306
C + G	1,444	0.415
A + T	2,034	0.585

CLC: ANALIZA SECVENȚEI

1.6 Nucleotide distribution histogram



CLC: ANALIZA SECVENŢEI



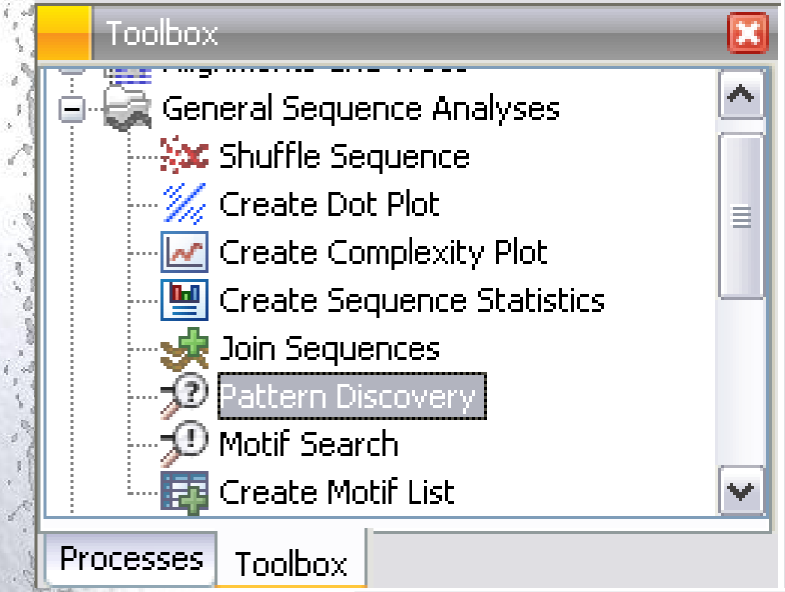
1.7 Counts of di-nucleotides

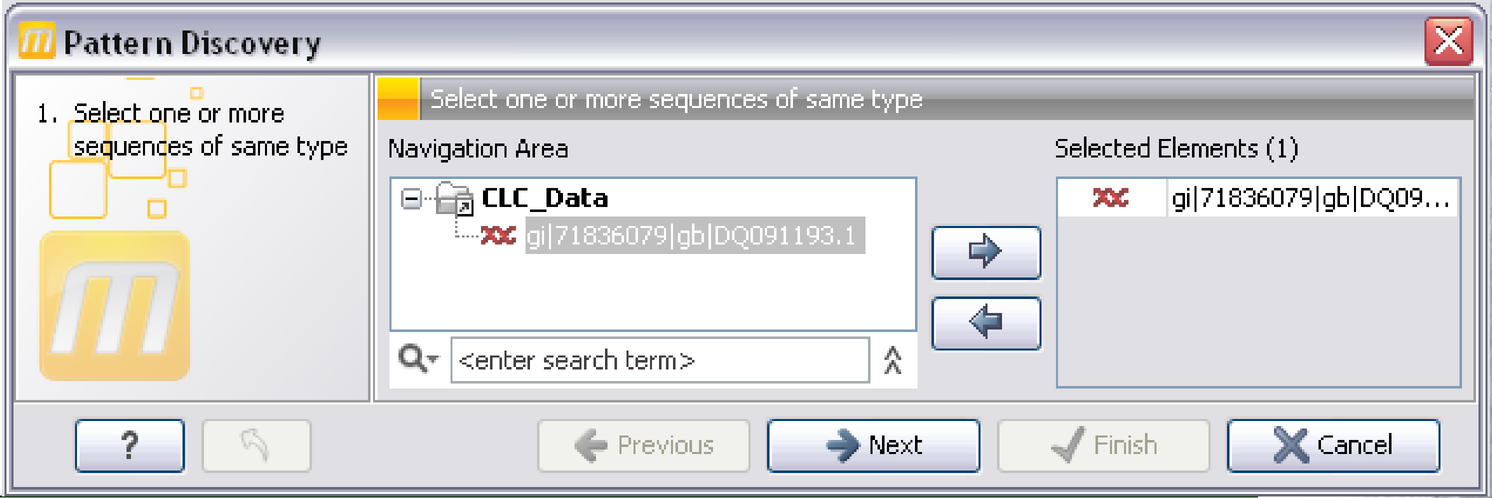
1.pos\2.pos	A	C	G	T
A	297	171	187	316
C	175	139	155	180
G	294	137	134	229
T	205	203	317	338

1.8 Frequency of di-nucleotides

1.pos\2.pos	A	C	G	T
A	0.085	0.049	0.054	0.091
C	0.050	0.040	0.045	0.052
G	0.085	0.039	0.039	0.066
T	0.059	0.058	0.091	0.097

IDENTIFICAREA PATTERN-URILOR ÎN CLC



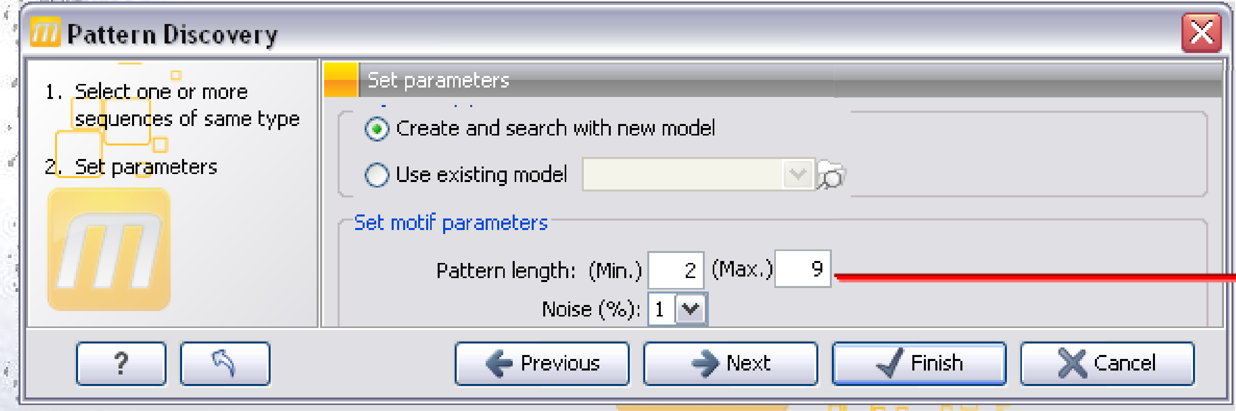


1. Select one or more sequences of same type

gi|71836079|gb|DQ091193.1

gi|71836079|gb|DQ09...

IDENTIFICAREA PATTERN-URILOR ÎN CLC



1. Select one or more sequences of same type
2. Set parameters

● Create and search with new model

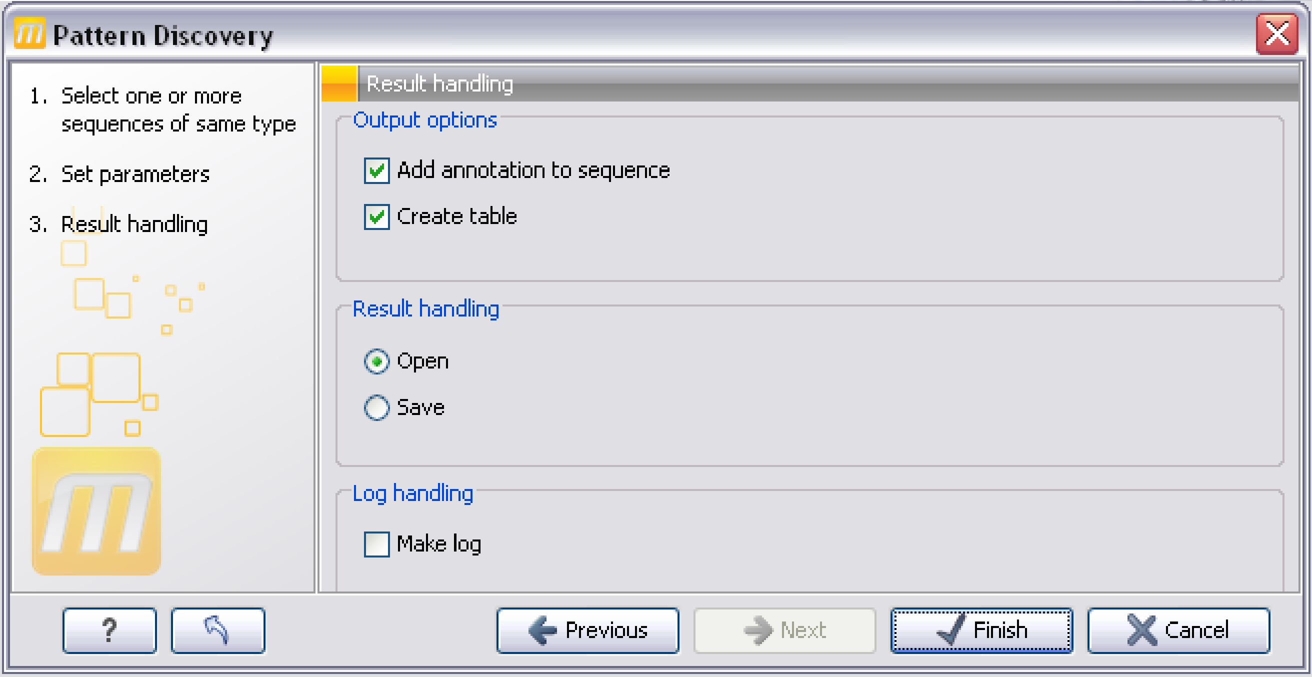
○ Use existing model

Set motif parameters

Pattern length: (Min.) 2 (Max.) 9

Noise (%): 1

Pentru a identifica secvenţe de 2 nucleotide min = 1 şi max = 2



1. Select one or more sequences of same type
2. Set parameters
3. Result handling

Output options

☑ Add annotation to sequence

☑ Create table

Result handling

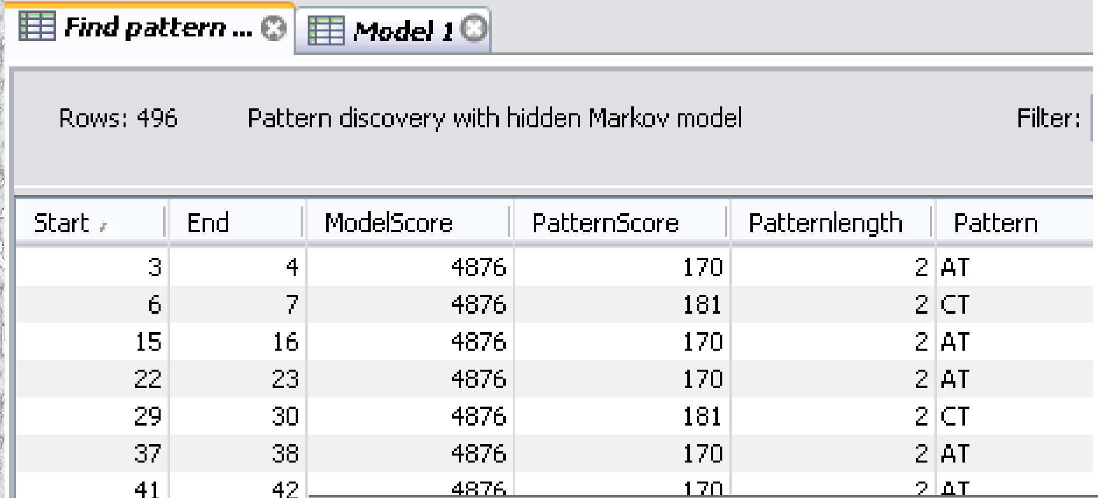
● Open

○ Save

Log handling

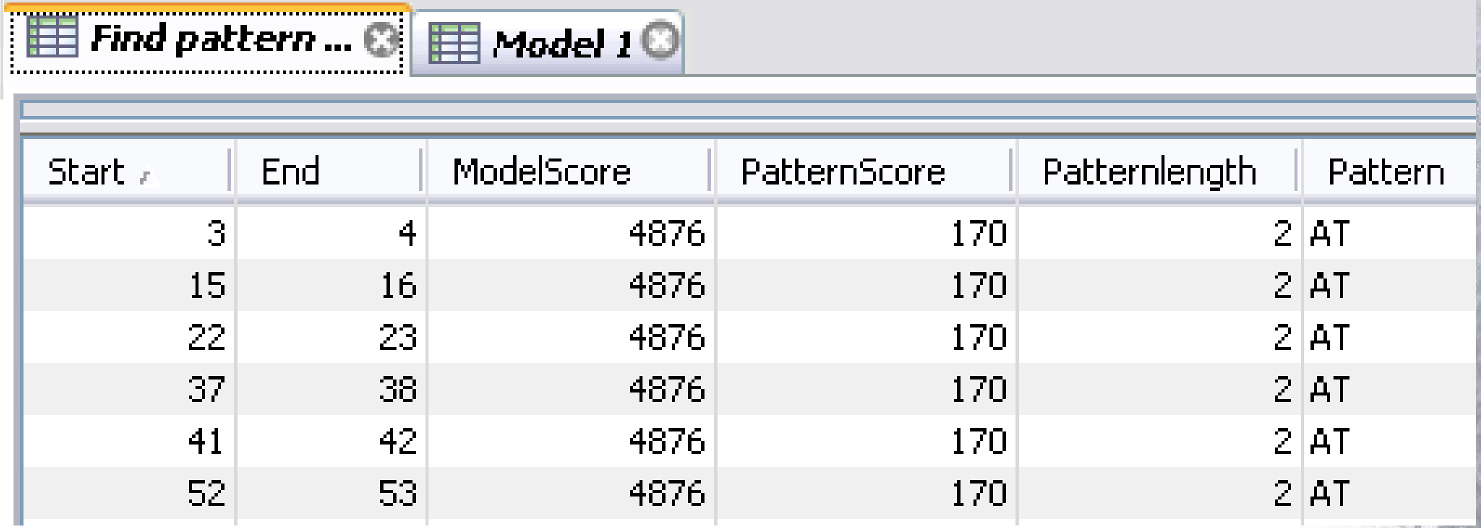
☐ Make log

CLC REZULTATE: PATTERN 2 NUCLEOTIDE



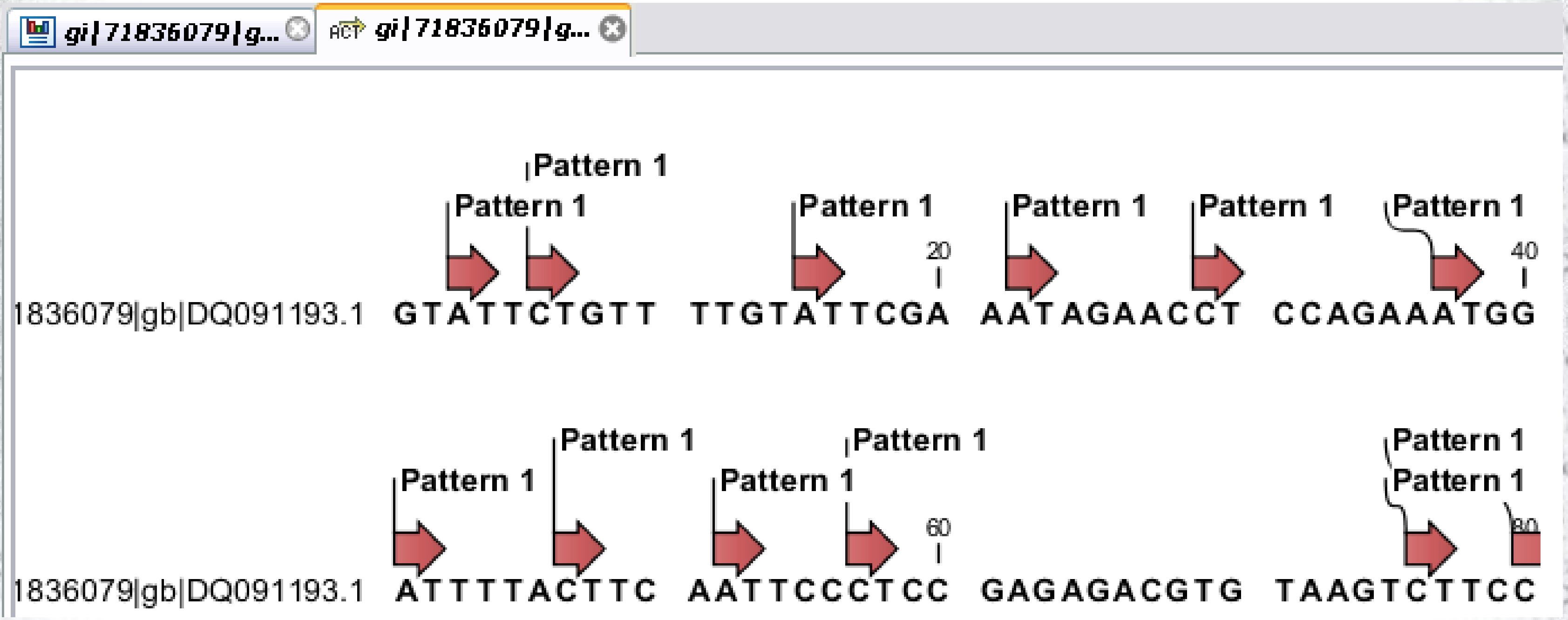
Rows: 496 Pattern discovery with hidden Markov model

Start	End	ModelScore	PatternScore	Patternlength	Pattern
3	4	4876	170	2	AT
6	7	4876	181	2	CT
15	16	4876	170	2	AT
22	23	4876	170	2	AT
29	30	4876	181	2	CT
37	38	4876	170	2	AT
41	42	4876	170	2	AT
47	48				
52	53				



Start	End	ModelScore	PatternScore	Patternlength	Pattern
3	4	4876	170	2	AT
15	16	4876	170	2	AT
22	23	4876	170	2	AT
37	38	4876	170	2	AT
41	42	4876	170	2	AT
52	53	4876	170	2	AT

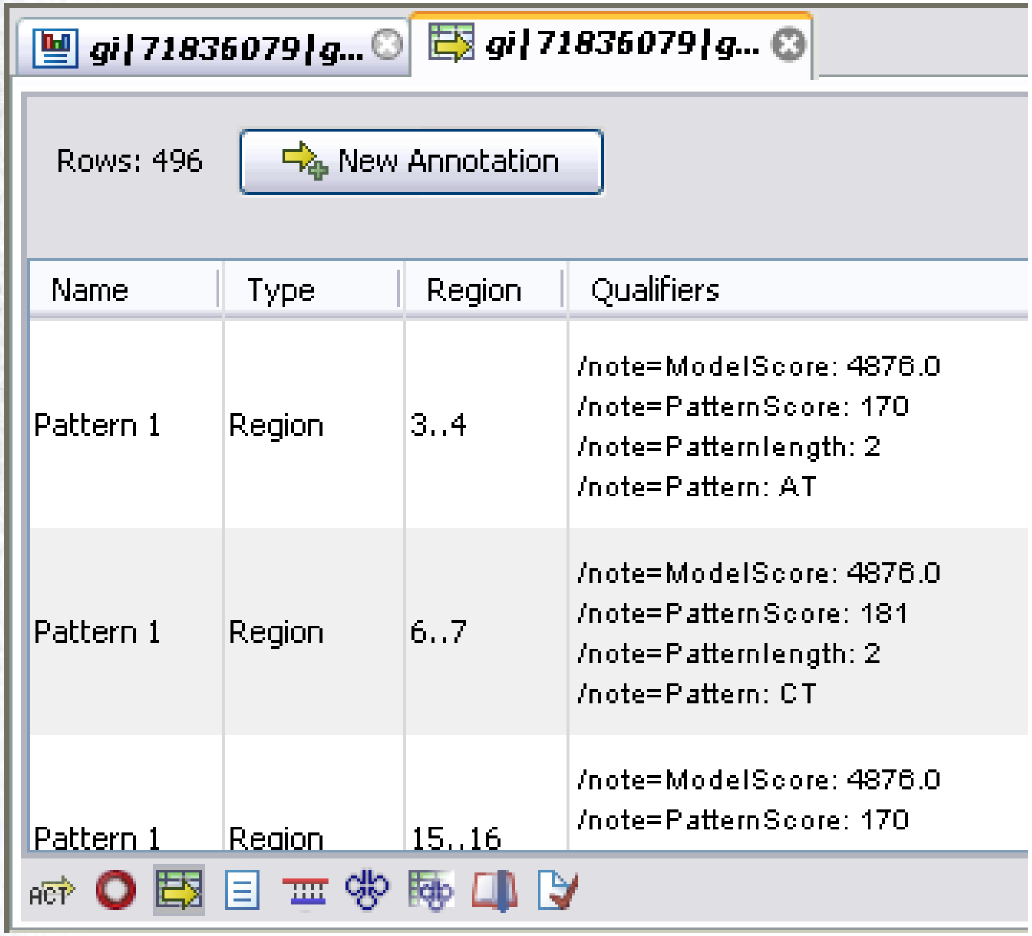
CLC REZULTATE: PATTERN 2 NUCLEOTIDE



1836079|gb|DQ091193.1 GTATTCTGTT TTGTATTCGA AATAGAACCT CCAGAAATGG

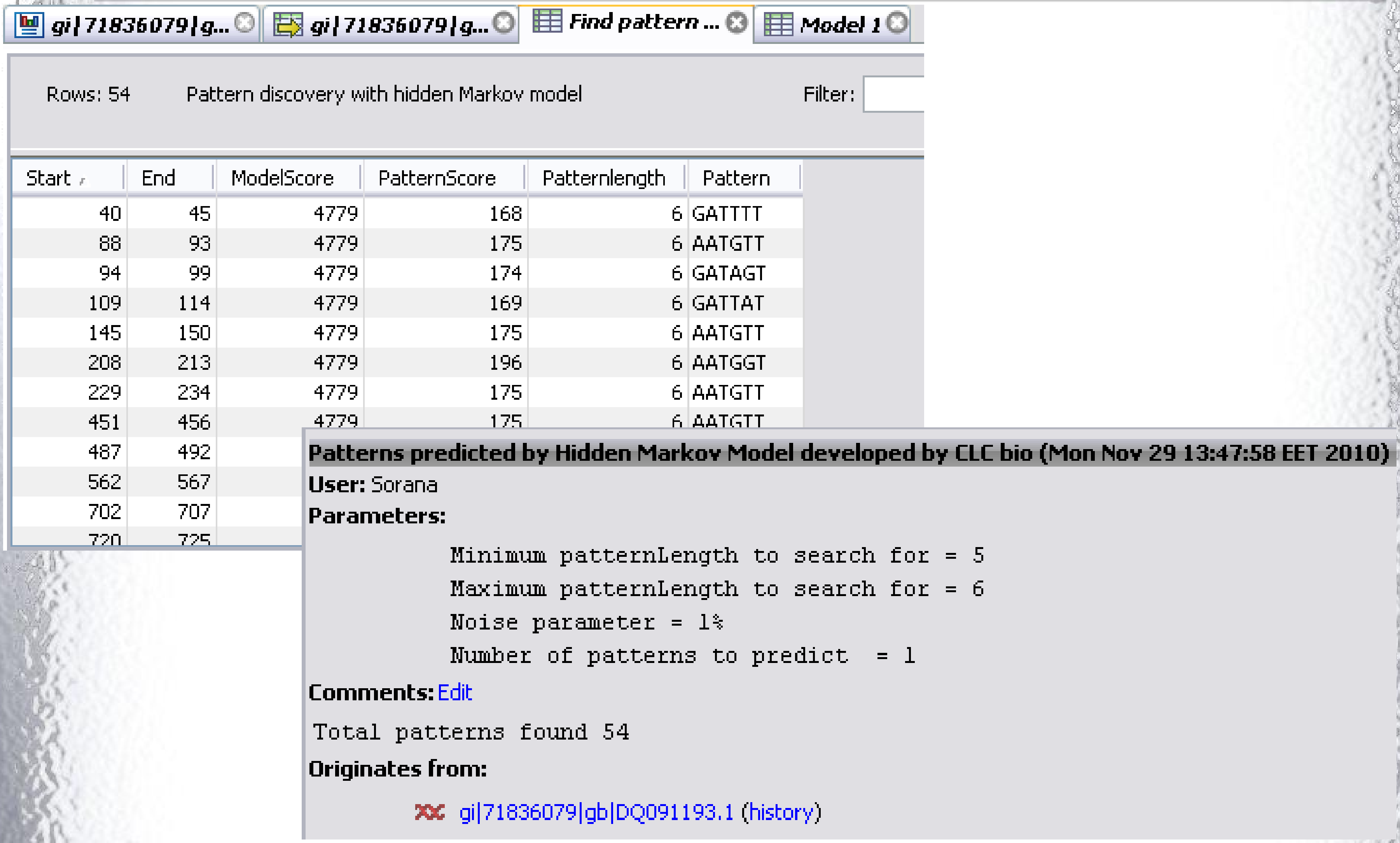
1836079|gb|DQ091193.1 ATTTTACTTC AATTCCCTCC GAGAGACGTG TAAGTCTTCC

CLC REZULTATE: PATTERN 2 NUCLEOTIDE



Name	Type	Region	Qualifiers
Pattern 1	Region	3..4	/note=ModelScore: 4876.0 /note=PatternScore: 170 /note=Patternlength: 2 /note=Pattern: AT
Pattern 1	Region	6..7	/note=ModelScore: 4876.0 /note=PatternScore: 181 /note=Patternlength: 2 /note=Pattern: CT
Pattern 1	Region	15..16	/note=ModelScore: 4876.0 /note=PatternScore: 170

CLC REZULTATE: PATTERN 6 NUCLEOTIDE



Rows: 54 Pattern discovery with hidden Markov model

Start	End	ModelScore	PatternScore	Patternlength	Pattern
40	45	4779	168	6	GATTTT
88	93	4779	175	6	AATGTT
94	99	4779	174	6	GATAGT
109	114	4779	169	6	GATTAT
145	150	4779	175	6	AATGTT
208	213	4779	196	6	AATGGT
229	234	4779	175	6	AATGTT
451	456	4779	175	6	AATGTT
487	492				
562	567				
702	707				
720	725				

Patterns predicted by Hidden Markov Model developed by CLC bio (Mon Nov 29 13:47:58 EET 2010)

User: Sorana

Parameters:

```
Minimum patternLength to search for = 5
Maximum patternLength to search for = 6
Noise parameter = 1%
Number of patterns to predict  = 1
```

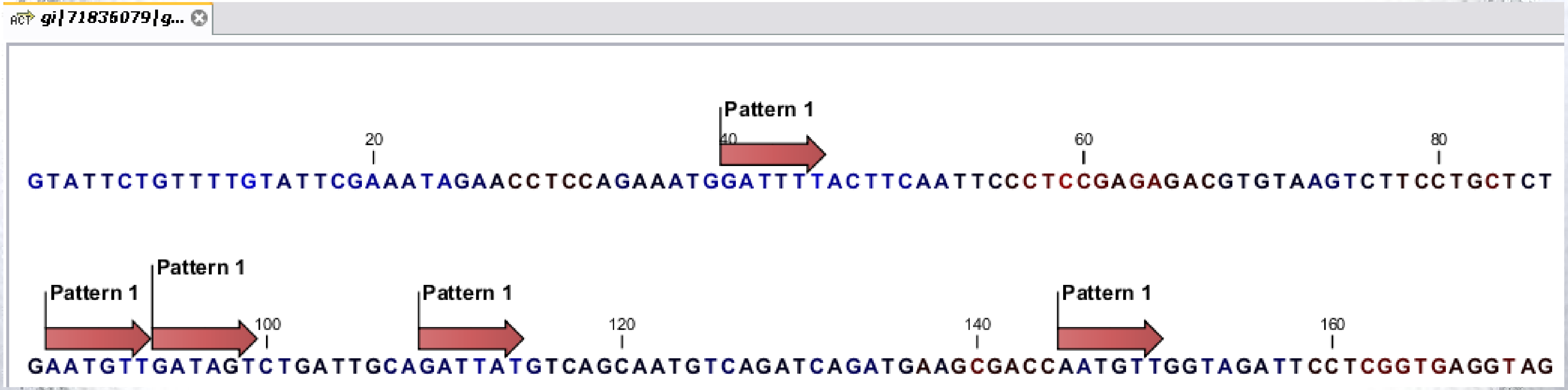
Comments: Edit

Total patterns found 54

Originates from:

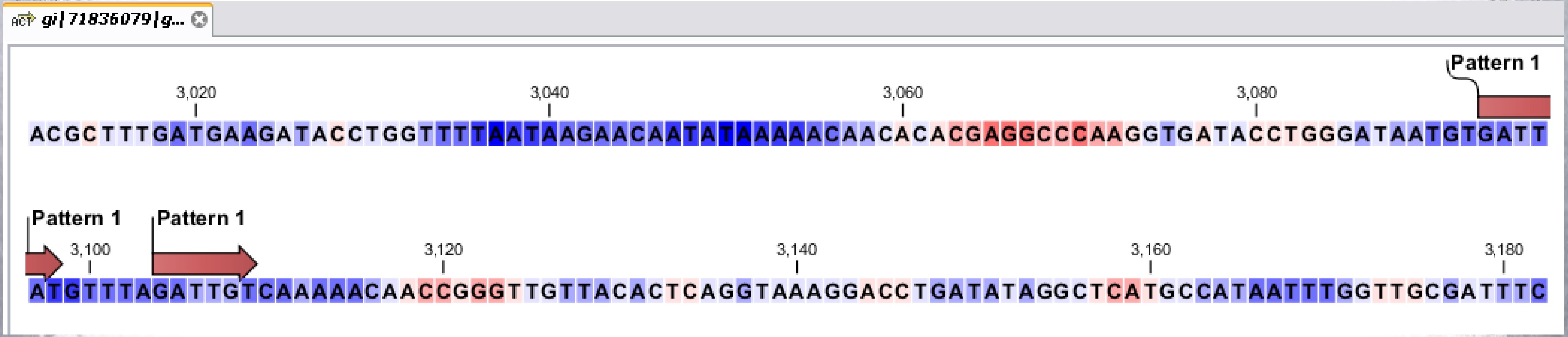
gi|71836079|gb|DQ091193.1 (history)

CLC REZULTATE: PATTERN 6 NUCLEOTIDE



GTATTCTGTTTTGTATTCGAAATAGAACCTCCAGAAATGGATTTTACTTCAATTCCCTCCGAGAGACGTGTAAGTCTTCCTGCTCT

GAATGTTGATAGTCTGATTGCAGATTATGTCAGCAATGTCAGATCAGATGAAGCGACCAATGTTGGTAGATTCCTCGGTGAGGTAG

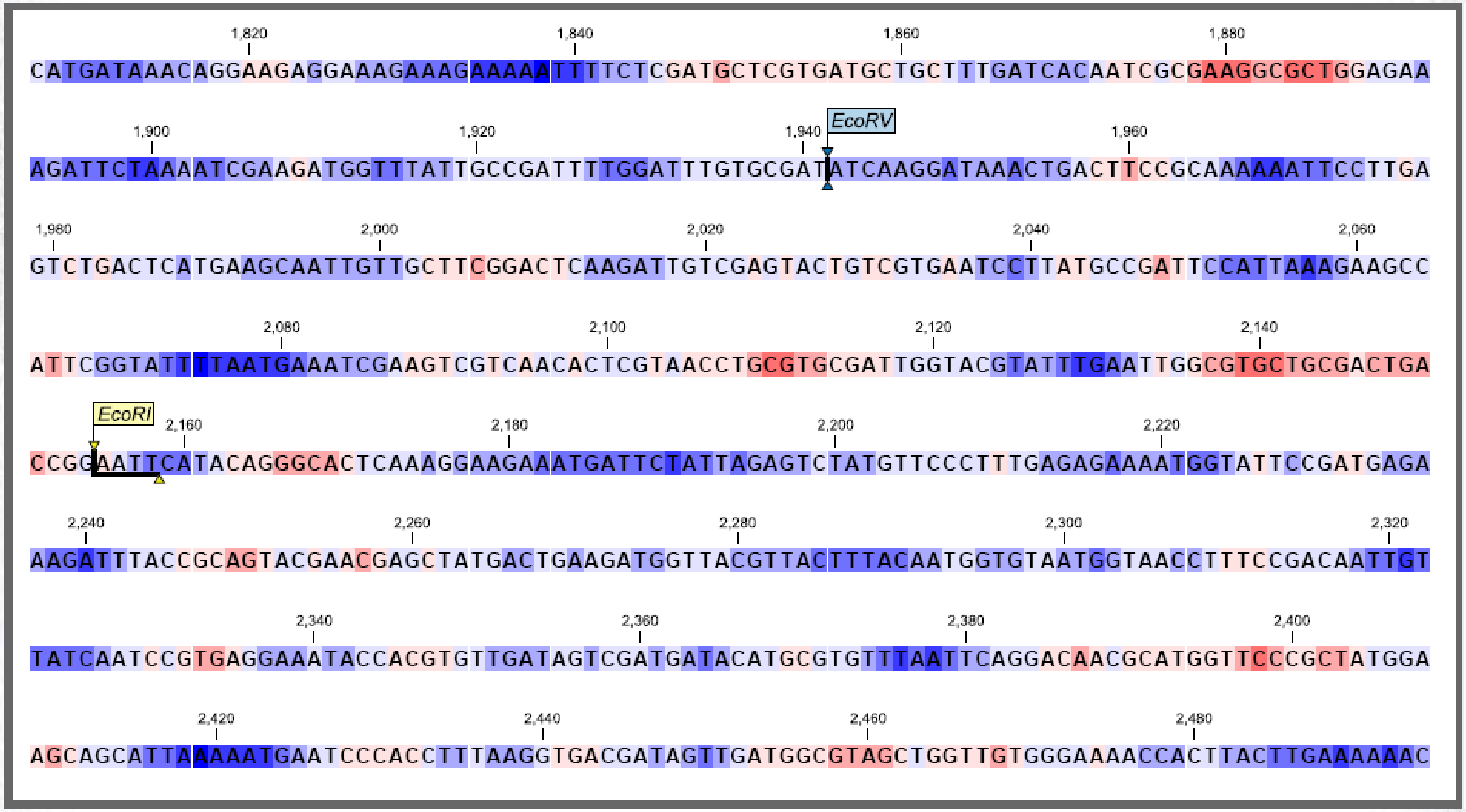


ACGCTTTGATGAAGATACCTGGTTTTAATAAGAACAATATAAAAACAACACACGAGGCCCAAGGTGATACCTGGGATAATGTGATT

ATGTTTAGATTGTCAAAAACAACCGGGTTGTTACACTCAGGTAAAGGACCTGATATAGGCTCATGCCATAATTTGGTTGCGATTTC

CLC REZULTATE: GRAPHICS

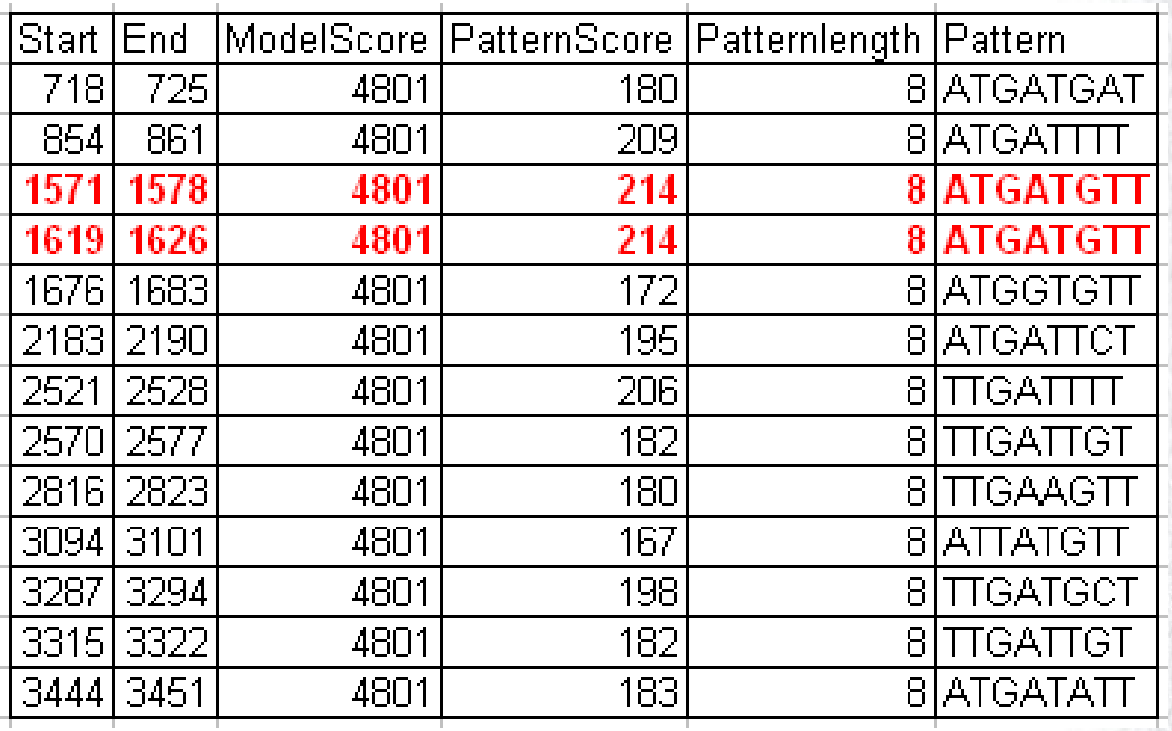
- Aria vizibilă sau întreaga secvenţă



```
CATGATAAACAGGAAGAGGAAAGAAAGAAAAATTTTCTCGATGCTCGTGATGCTGCTTTGATCACAATCGCGAAGGCGCTGGAGAA
AGATTCTAAAATCGAAGATGGTTTATTGCCGATTTTGGATTTGTGCGATATCAAGGATAAACTGACTTCCGCAAAAAATTCCTTGA
GTCTGACTCATGAAGCAATTGTTGCTTCGGACTCAAGATTGTCGAGTACTGTCGTGAATCCTTATGCCGATTCCATTAAAGAAGCC
ATTCGGTATTTTAATGAAATCGAAGTCGTCAACACTCGTAACCTGCGTGCGATTGGTACGTATTTGAATTGGCGTGCTGCGACTGA
CCGGAATTCATACAGGGCACTCAAAGGAAGAAATGATTCTATTAGAGTCTATGTTCCCTTTGAGAGAAAATGGTATTCCGATGAGA
AAGATTTACCGCAGTACGAACGAGCTATGACTGAAGATGGTTACGTTACTTTACAATGGTGTAATGGTAACCTTTCCGACAATTGT
TATCAATCCGTGAGGAAATACCACGTGTTGATAGTCGATGATACATGCGTGTTTAATTCAGGACAACGCATGGTTCCCGCTATGGA
AGCAGCATTAAAAATGAATCCCACCTTTAAGGTGACGATAGTTGATGGCGTAGCTGGTTGTGGGAAAACCACTTACTTGAAAAAAC
```

CLC REZULTATE: DACĂ ŞTIM CE CĂUTĂM

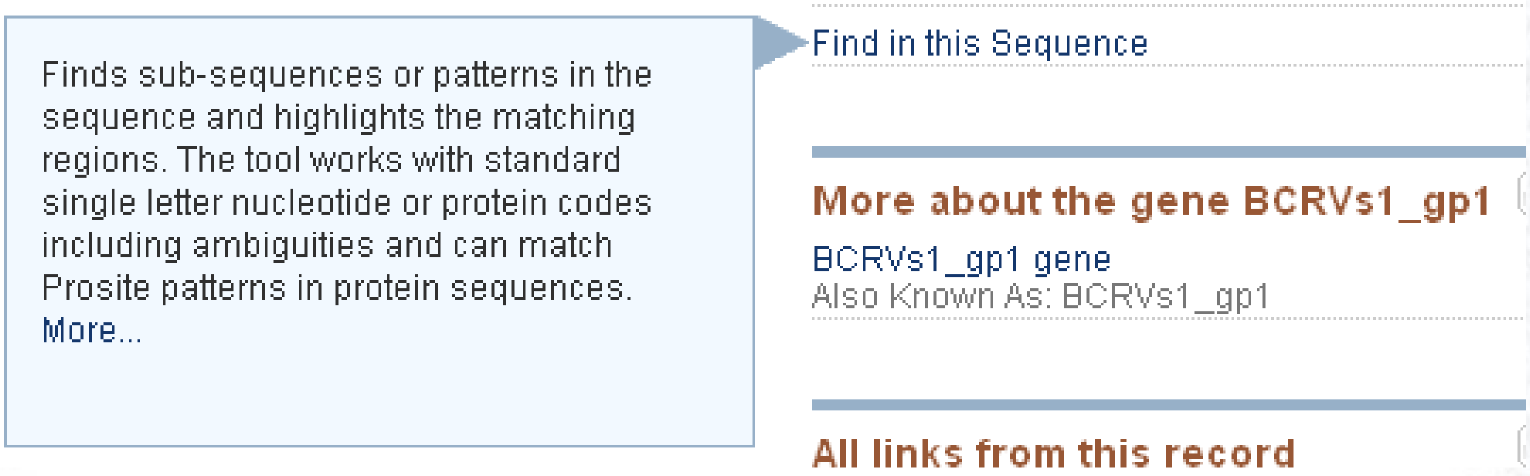
- Căutăm secvenţa: ATGATGTT (7 nucleotide)



Start	End	ModelScore	PatternScore	Patternlength	Pattern
718	725	4801	180	8	ATGATGAT
854	861	4801	209	8	ATGATTTT
1571	**1578**	**4801**	**214**	**8**	**ATGATGTT**
1619	**1626**	**4801**	**214**	**8**	**ATGATGTT**
1676	1683	4801	172	8	ATGGTGTT
2183	2190	4801	195	8	ATGATTCT
2521	2528	4801	206	8	TTGATTTT
2570	2577	4801	182	8	TTGATTGT
2816	2823	4801	180	8	TTGAAGTT
3094	3101	4801	167	8	ATTATGTT
3287	3294	4801	198	8	TTGATGCT
3315	3322	4801	182	8	TTGATTGT
3444	3451	4801	183	8	ATGATATT

CLC Rezultate: Dacă Ştim ce Căutăm

http://www.ncbi.nlm.nih.gov/nuccore/71836079?report=fasta



Finds sub-sequences or patterns in the sequence and highlights the matching regions. The tool works with standard single letter nucleotide or protein codes including ambiguities and can match Prosite patterns in protein sequences. More...

Find in this Sequence

More about the gene BCRVs1_gp1

BCRVs1_gp1 gene
Also Known As: BCRVs1_gp1

All links from this record

ATGATGTT

CLC REZULTATE: DACĂ ŞTIM CE CĂUTĂM

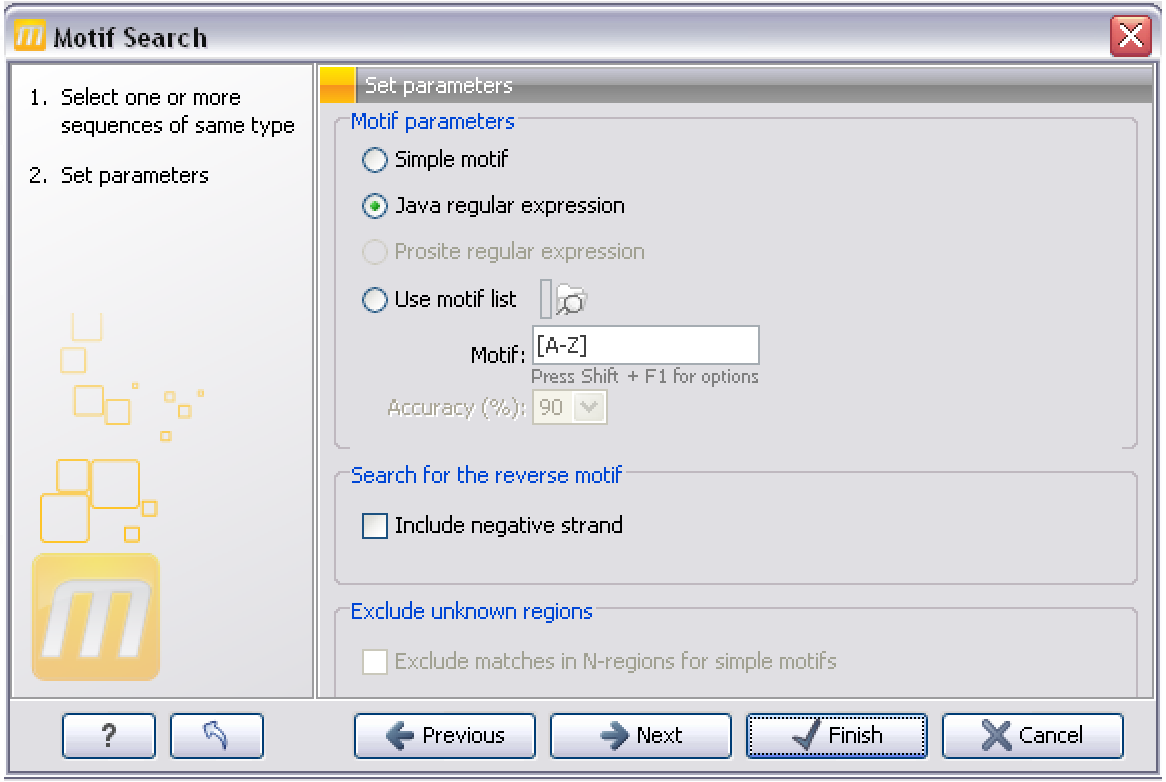
www.ncbi.nlm.nih.gov/nuccore/71836079?report=fasta

Display Settings: FASTA

Blackberry chlorotic ringspot virus segment RNA 1, complete sequence

TTGGTCGGAAAACCGTCACAAGAATCCGGG**ATGATGTT**TCAATTGTACACCCGAGAAATTTGCTTAAGGA
GATCATTT**ATGATGTT**AAGCGCTTATTTGGTCTCACTCTTTCTGGTTCCGACTATAATTTAGTCGATGGT

CLC REZULTATE: SECVENȚE MOTIFS



Motif Search

1. Select one or more sequences of same type
2. Set parameters

Set parameters

Motif parameters

○ Simple motif

● Java regular expression

○ Prosite regular expression

○ Use motif list

Motif: [A-Z]

Press Shift + F1 for options

Accuracy (%): 90

Search for the reverse motif

☐ Include negative strand

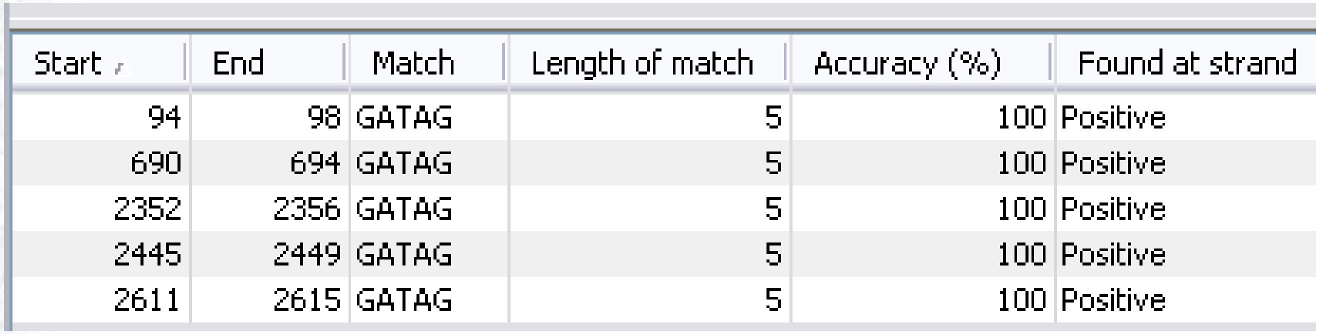
Exclude unknown regions

☐ Exclude matches in N-regions for simple motifs

CLC REZULTATE: SECVENŢE MOTIFS

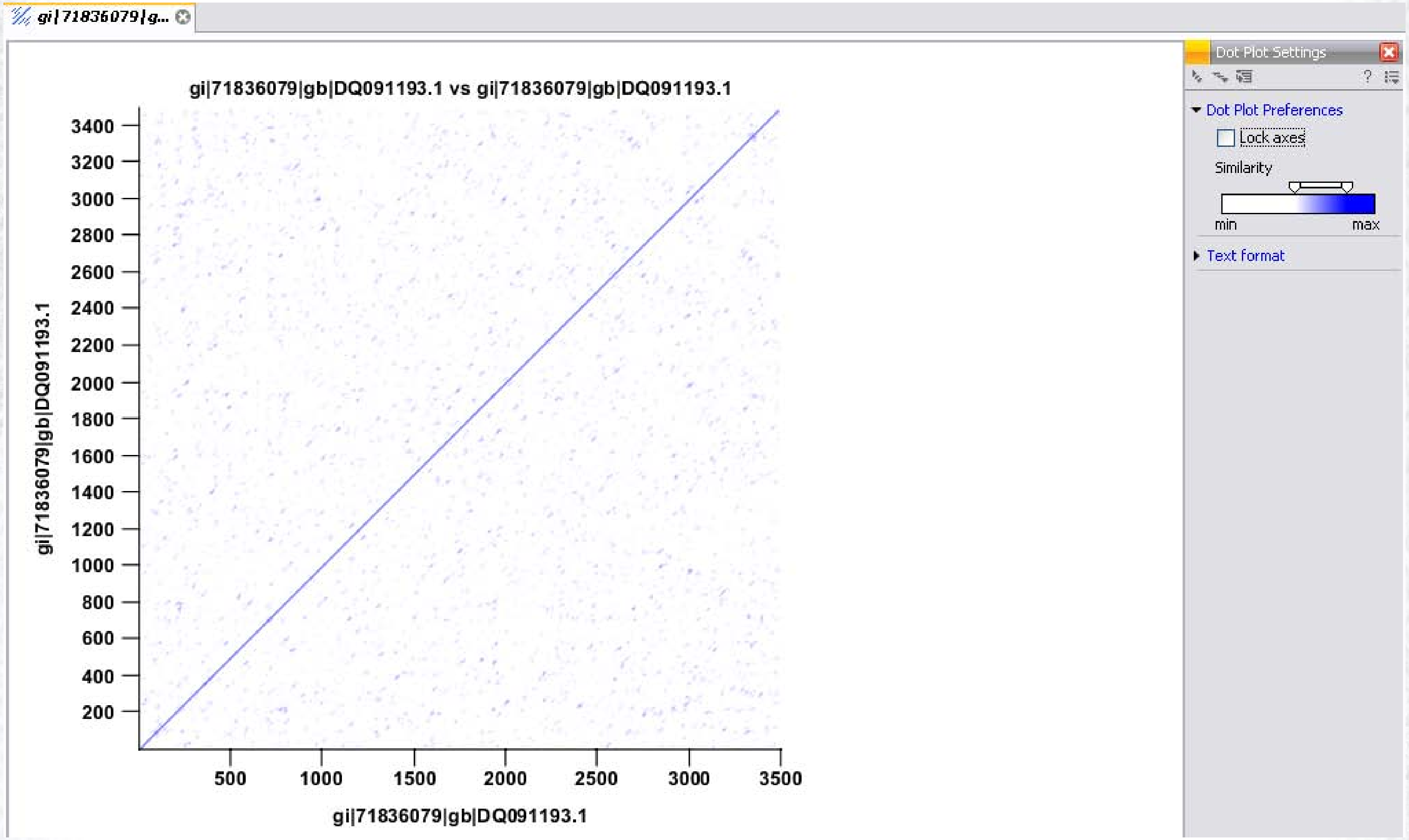
- Cerinţe:
  - Utilizatorul are cunoştinţe despre secvenţa motif pe care doreşte să o identifice
  - Secvenţa este definită de utilizator – algoritmul caută secvenţa prin identificarea pattern-urilor identice sau degenerate

CLC REZULTATE: SECVENȚE MOTIFS



Start	End	Match	Length of match	Accuracy (%)	Found at strand
94	98	GATAG	5	100	Positive
690	694	GATAG	5	100	Positive
2352	2356	GATAG	5	100	Positive
2445	2449	GATAG	5	100	Positive
2611	2615	GATAG	5	100	Positive

CLC REZULTATE: CREATE DOT PLOT



☐ Lock axes