BIN 506: Protein & DNA Sequence Analysis

Assignment 04

By: Taha Ahmad

Student ID: 2546125

Instructor: Dr Yesim Aydin Son

1) **Discuss the following: Why does using protein sequences result in more accurate sequence alignments? Think about biological reasons, statistical reasons, and computational reasons.**

Answer:

There are a number of biological, statistical and computational reasons for protein sequences resulting in more accurate sequence alignment,

Biological Factors:

Protein sequences are under stronger functional constraints than DNA. Protein alignments highlight the conserved functional domains or motifs like enzyme active sites which are critical for biological activity therefore proteins evolve slower than DNA as natural selection preserves the structure of a functional protein, for example non synonymous mutations which alter the amino acid sequence of a protein are often deleterious and are selected against which leads to slower evolutionary rates. On the other hand synonymous mutations do not affect the amino acid sequences and consequently the proteins. These are common in DNA but irrelevant in protein alignment.

Statistical reasons:

Proteins have 20 amino acids compared to 4 nucleotides in DNA, this lowers the probability of random matches thus reducing false positive alignments. DNA matrices consist of very simple DNA match/ mismatch scores. For protein alignments we have more complicated models/ substitutions matrices like BLOSUM and PAM which incorporate empirical data on evolutionary likelihood and biochemical similarities as well, all of which improves the alignment accuracy. Furthermore each amino acid carries more evolutionary information due to diversity and biochemical activity, this enables us to better identify between homologous and random matches.

Computational Reasons:

An amino acid which is the building block of proteins is made up of 3 nucleotides per 1 amino acid. This means that a protein sequence is 3 times shorter than their coding dna counterparts, this simplifies alignment algorithms by reducing the computational complexity. Unlike a DNA sequence protein alignments are not affected by frameshifts or ambiguous reading frames in dna. The slower evolution and high information density allows protein alignments to detect distant evolutionary relationships better than DNA alignments especially for sequences with low nucleotide similarity but conserved protein function.

2) **Answer the following questions about PAM and BLOSUM matrices**

a) **What kind of biological information are they built upon?**

PAM which stands for point accepted mutations are built upon evolutionary models of accepted mutations only (mutations which exist through evolution), it is built on global alignments of protein with high similarity. PAM is derived from closely related protein sequences which allows us to see how amino acids change over time. A PAM matrix gives the probability of one amino acid mutating into another during evolution.

BlOSUM which stands for block substitution matrices uses local alignments instead of global and also unlike PAM it does not assume any evolutionary path. The matrix is built using conserved regions from an array of protein families which are divergent. It measures how one amino acid is substituted for another in aligned blocks that do not contain gaps, it is derived from protein sequences grouped by percentage identity for example BLOSUM62 are sequences that are 62% identical and BLOSUM95 are those which are 95% identical.

In essence pam emphasises on evolutionary time whereas BLOSUM emphasises on sequence diversity.

b) **Explain the difference between PAM250 and PAM500 and the difference between BLOSUM62 and BLOSUM95.**

PAM 250 vs PAM 500:

PAM250 indicates 250% average accepted mutations per site and is suitable for moderately divergent sequences. On the other hand PAM500 models 500% accepted mutations thus indicating longer evolutionary divergence. PAM500 is designed for highly divergent sequences
The PAM numbers increases with evolutionary distance

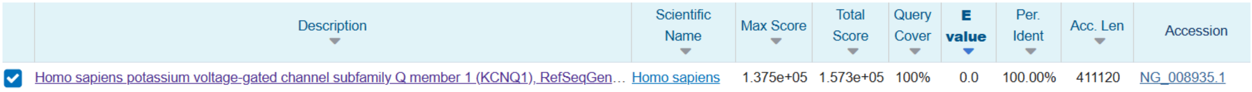
BLOSUM62 vs BLOSUM95:

BlOSUM62 is built from clusters with less than 62% identity and is optimised for general-purpose alignment of moderately divergent sequences. On the other hand BLOSUM95 uses clusters with less than 95% identity which means that only closely related sequences are included and is best for aligning very similar sequences
BLOSUM numbers, unlike PAM, decrease with evolutionary distance.

3) **Use “unknown.fasta” and search for the source using NCBI BLAST, UCSC BLAT, and Ensembl BLAST/BLAT.**

a) **Share the list of the results. Write down the name and the type (gene, transcript, etc.) of the source.**

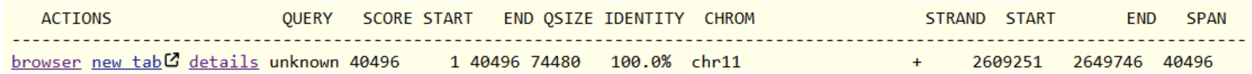
NCBI BLASTN Hit:



	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Homo sapiens potassium voltage-gated channel subfamily Q member 1 (KCNQ1), RefSeqGen...	Homo sapiens	1.375e+05	1.573e+05	100%	0.0	100.00%	411120	NG_008935.1

Name: KCNQ1
Type: Gene
Description: Homo sapiens potassium voltage-gated channel subfamily Q member 1 (KCNQ1), RefSeqGene (LRG_287) on chromosome 11

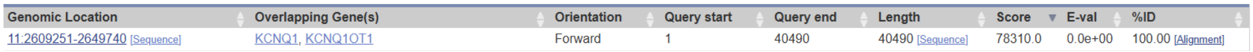
UCSC BLAT Hit:



```
   ACTIONS                  QUERY   SCORE START   END QSIZE IDENTITY  CHROM                   STRAND  START       END   SPAN
---------------------------------------------------------------------------------------------------------------------------
browser new tab details unknown 40496       1 40496 74480    100.0%  chr11                      +      2609251    2649746  40496
```

Name: KCNQ1
Type : Gene
Details: The sequence aligns to the genomic region of chr11:2,609,251-2,649,746, which corresponds to the KCNQ1 gene locus.

ENSEMBL BLAT Hit:

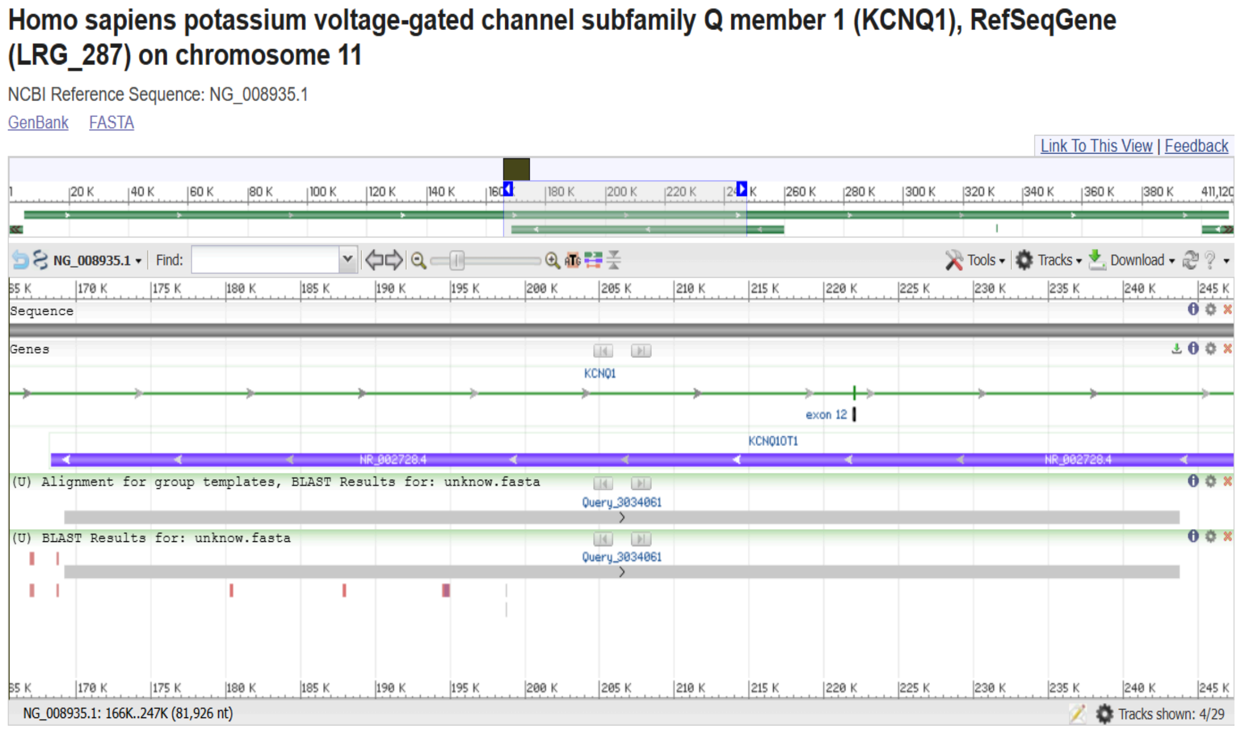


Genomic Location	Overlapping Gene(s)	Orientation	Query start	Query end	Length	Score	E-val	%ID
11:2609251-2649740 [Sequence]	KCNQ1, KCNQ1OT1	Forward	1	40490	40490 [Sequence]	78310.0	0.0e+00	100.00 [Alignment]

Name: KCNQ1
Type: Gene
Details: although we can see KCNQ1OT1 along with KCNQ1 as part of the overlapping genes however KCNQ1OT1 is a non coding RNA therefore we ignore it.

b) **Select the top scoring hit (both in terms of E-Value and score). Click on the link that will direct you to the genome browser on each database. Share a screenshot from each genome browser.**

NCBI BLASTN:



Homo sapiens potassium voltage-gated channel subfamily Q member 1 (KCNQ1), RefSeqGene (LRG_287) on chromosome 11

NCBI Reference Sequence: NG_008935.1

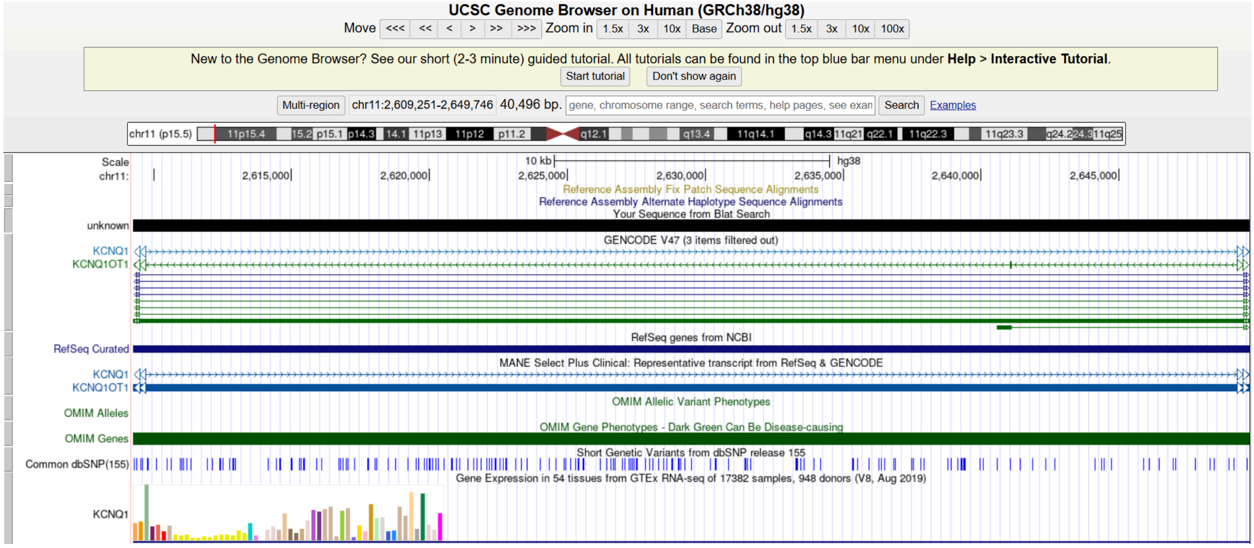
GenBank FASTA

Link To This View | Feedback

NG_008935.1: 166K..247K (81,926 nt)

Tracks shown: 4/29

UCSC BLAT:



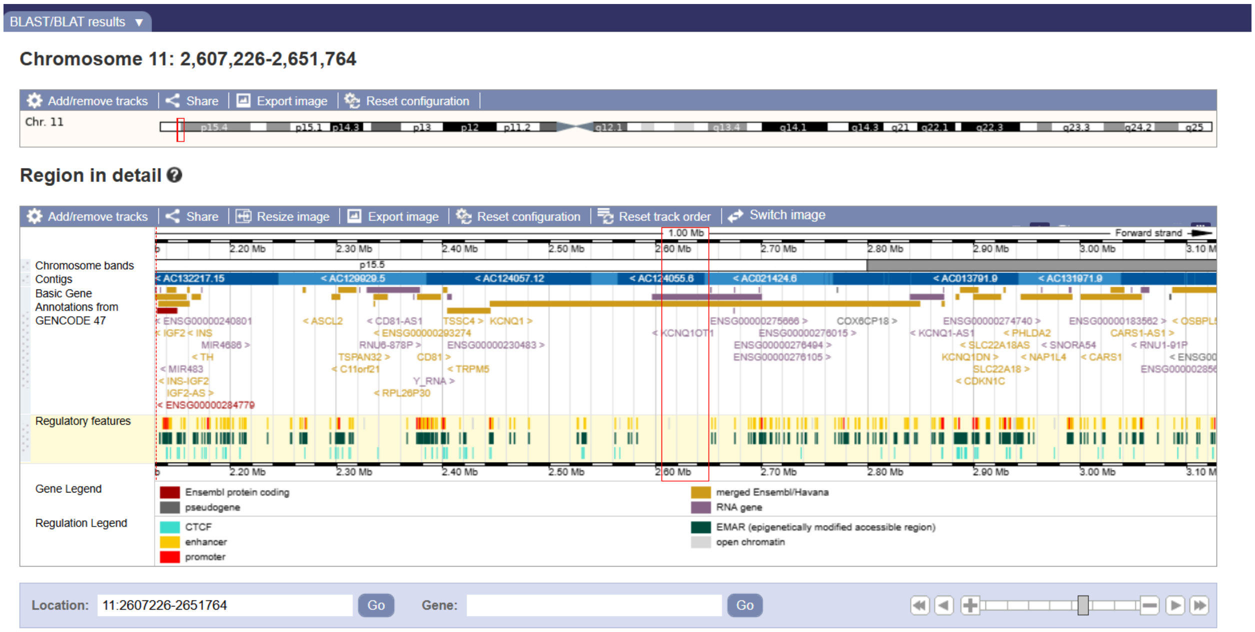
UCSC Genome Browser on Human (GRCh38/hg38)

New to the Genome Browser? See our short (2-3 minute) guided tutorial. All tutorials can be found in the top blue bar menu under **Help > Interactive Tutorial**.

chr11:2,609,251-2,649,746 40,496 bp.

Move <<< << < > >> >>> Zoom in 1.5x 3x 10x Base Zoom out 1.5x 3x 10x 100x
Start tutorial
Don't show again
Multi-region
gene, chromosome range, search terms, help pages, see exan
Search
Examples
chr11 (p15.5)
11p15.4
15.2
p15.1
p14.3
14.1
11p13
11p12
p11.2
q12.1
q13.4
11q14.1
q14.3
11q21
q22.1
11q22.3
11q23.3
q24.2
24.3
11q25
Scale
chr11:
10 kb
hg38
2,615,000
2,620,000
2,625,000
2,630,000
2,635,000
2,640,000
2,645,000
Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments
Your Sequence from Blat Search
unknown
GENCODE V47 (3 items filtered out)
KCNQ1
KCNQ1OT1
RefSeq genes from NCBI
RefSeq Curated
MANE Select Plus Clinical: Representative transcript from RefSeq & GENCODE
KCNQ1
KCNQ1OT1
OMIM Allelic Variant Phenotypes
OMIM Alleles
OMIM Gene Phenotypes - Dark Green Can Be Disease-causing
OMIM Genes
Short Genetic Variants from dbSNP release 155
Common dbSNP(155)
Gene Expression in 54 tissues from GTEx RNA-seq of 17382 samples, 948 donors (V8, Aug 2019)
KCNQ1

ENSEMBL BLAST/BLAT :



Chromosome 11: 2,607,226-2,651,764

Region in detail

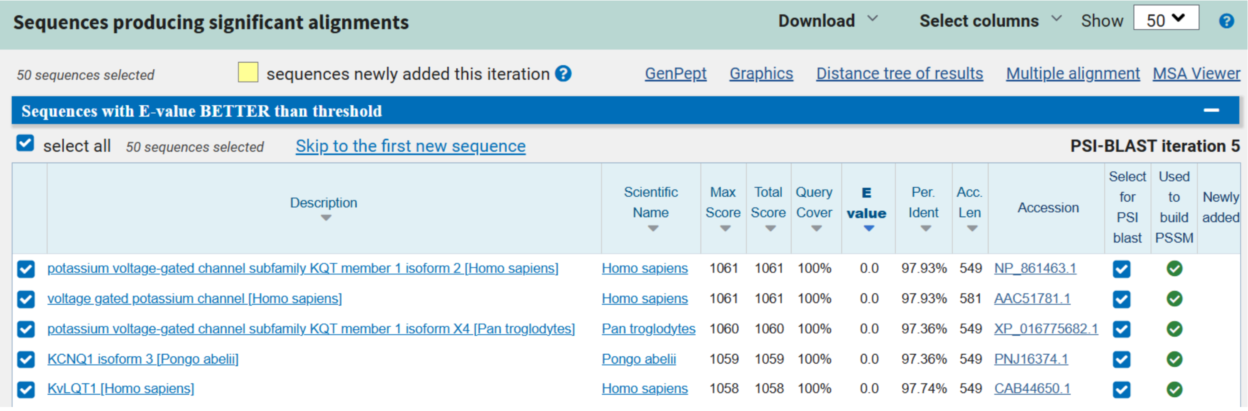
Location: 11:2607226-2651764

4) Use “protein.fasta” and run a search using blastp and psi-blast.

- **For both results, set the ‘Show’ value to 50 and make sure to keep it that way.**
- **Set ‘Number of sequences’ to 50 in psi-blast results and make 5 iterations.**
- **Make sure the resulting hits are sorted by the E-value.**
- **Answer the following questions by comparing the blastp result to the 5th iteration of psi-blast.**

a) Are the first 5 hits the same in both tools?

PSI-BLAST:



Sequences producing significant alignments

50 sequences selected

sequences newly added this iteration

Sequences with E-value BETTER than threshold

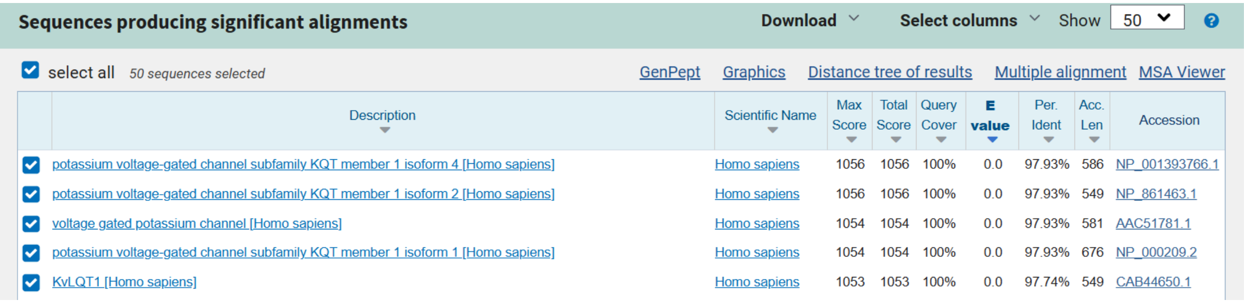
☑ select all 50 sequences selected

Skip to the first new sequence

PSI-BLAST iteration 5

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
☑	potassium voltage-gated channel subfamily KQT member 1 isoform 2 [Homo sapiens]	Homo sapiens	1061	1061	100%	0.0	97.93%	549	NP_861463.1	☑	✓	
☑	voltage gated potassium channel [Homo sapiens]	Homo sapiens	1061	1061	100%	0.0	97.93%	581	AAC51781.1	☑	✓	
☑	potassium voltage-gated channel subfamily KQT member 1 isoform X4 [Pan troglodytes]	Pan troglodytes	1060	1060	100%	0.0	97.36%	549	XP_016775682.1	☑	✓	
☑	KCNQ1 isoform 3 [Pongo abelii]	Pongo abelii	1059	1059	100%	0.0	97.36%	549	PNJ16374.1	☑	✓	
☑	KvLQT1 [Homo sapiens]	Homo sapiens	1058	1058	100%	0.0	97.74%	549	CAB44650.1	☑	✓	

BLASTP:



Sequences producing significant alignments

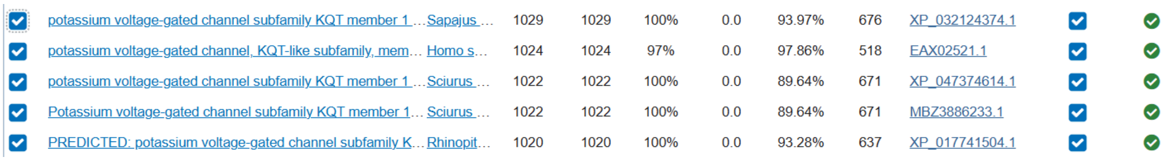
☑ select all 50 sequences selected

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	potassium voltage-gated channel subfamily KQT member 1 isoform 4 [Homo sapiens]	Homo sapiens	1056	1056	100%	0.0	97.93%	586	NP_001393766.1
☑	potassium voltage-gated channel subfamily KQT member 1 isoform 2 [Homo sapiens]	Homo sapiens	1056	1056	100%	0.0	97.93%	549	NP_861463.1
☑	voltage gated potassium channel [Homo sapiens]	Homo sapiens	1054	1054	100%	0.0	97.93%	581	AAC51781.1
☑	potassium voltage-gated channel subfamily KQT member 1 isoform 1 [Homo sapiens]	Homo sapiens	1054	1054	100%	0.0	97.93%	676	NP_000209.2
☑	KvLQT1 [Homo sapiens]	Homo sapiens	1053	1053	100%	0.0	97.74%	549	CAB44650.1

No, the first 5 hits in both tools are different as we can observe from both figures which show the results of BLASTP and PSI-BLAST. By only looking at the scientific names we can see that all of the 5 hits of BLASTP are from HomoSapiens taxon group while PSI-BLAST gives us 3 homosapiens and 2 form Pongo Abelii taxa.

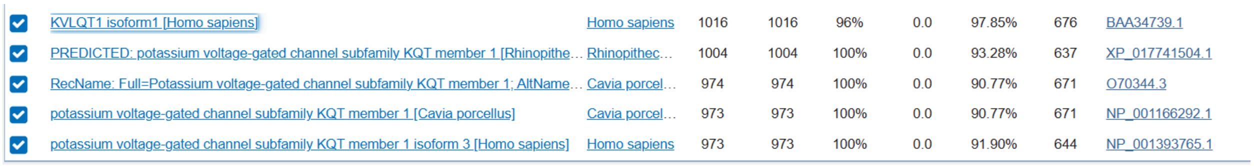
b) **Select the last 5 hits in both tools and compare their scores, E-values, and query covers. Submit a screenshot of ‘Graphical Summary’ with an explanation for both graphs for the last 5 hits. What do the graphs tell? What is the difference between the two graphs?**

PSI-BLAST:



☑	potassium voltage-gated channel subfamily KQT member 1 ...	Sapajus ...	1029	1029	100%	0.0	93.97%	676	XP_032124374.1	☑	✓
☑	potassium voltage-gated channel, KQT-like subfamily, mem...	Homo s...	1024	1024	97%	0.0	97.86%	518	EAX02521.1	☑	✓
☑	potassium voltage-gated channel subfamily KQT member 1 ...	Sciurus ...	1022	1022	100%	0.0	89.64%	671	XP_047374614.1	☑	✓
☑	Potassium voltage-gated channel subfamily KQT member 1...	Sciurus ...	1022	1022	100%	0.0	89.64%	671	MBZ3886233.1	☑	✓
☑	PREDICTED: potassium voltage-gated channel subfamily K...	Rhinopit...	1020	1020	100%	0.0	93.28%	637	XP_017741504.1	☑	✓

BLASTP:



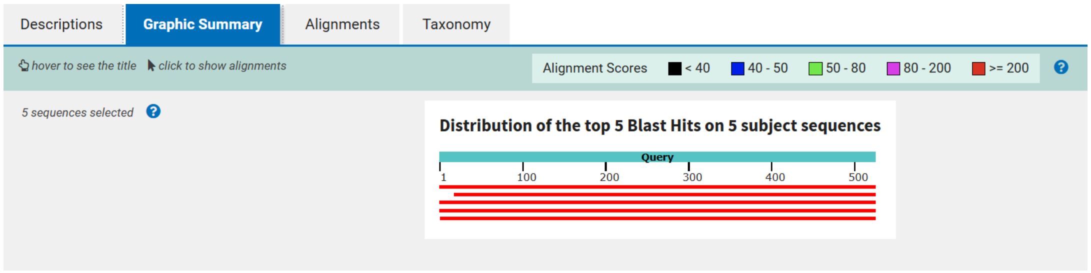
☑	KVLQT1 isoform1 [Homo sapiens]	Homo sapiens	1016	1016	96%	0.0	97.85%	676	BAA34739.1
☑	PREDICTED: potassium voltage-gated channel subfamily KQT member 1 [Rhinopithe...	Rhinopithec...	1004	1004	100%	0.0	93.28%	637	XP_017741504.1
☑	RecName: Full=Potassium voltage-gated channel subfamily KQT member 1; AltName...	Cavia porcel...	974	974	100%	0.0	90.77%	671	O70344.3
☑	potassium voltage-gated channel subfamily KQT member 1 [Cavia porcellus]	Cavia porcel...	973	973	100%	0.0	90.77%	671	NP_001166292.1
☑	potassium voltage-gated channel subfamily KQT member 1 isoform 3 [Homo sapiens]	Homo sapiens	973	973	100%	0.0	91.90%	644	NP_001393765.1

As we can observe in the figures of blastp and psi-blast hits above the last 5 results are different to each other as well.

We see that for all the hits the max and min score are equal, a 100% query cover and the E-Value also being 0.0 for all hits, as for psi-blast the max and min are also equal with a 100% query cover and 0.0 E value for all results just like the blastp however the max and min score of psi-blast are greater than those of blastp (man and min score of psi-blast > 120 & max and min score of blastp > 970 but < 1016 with 1016 being the value of 5th last blastp hit). For both tools the min and max score decreases as we go down the list. The corresponding Accession numbers also differ in both the tools.

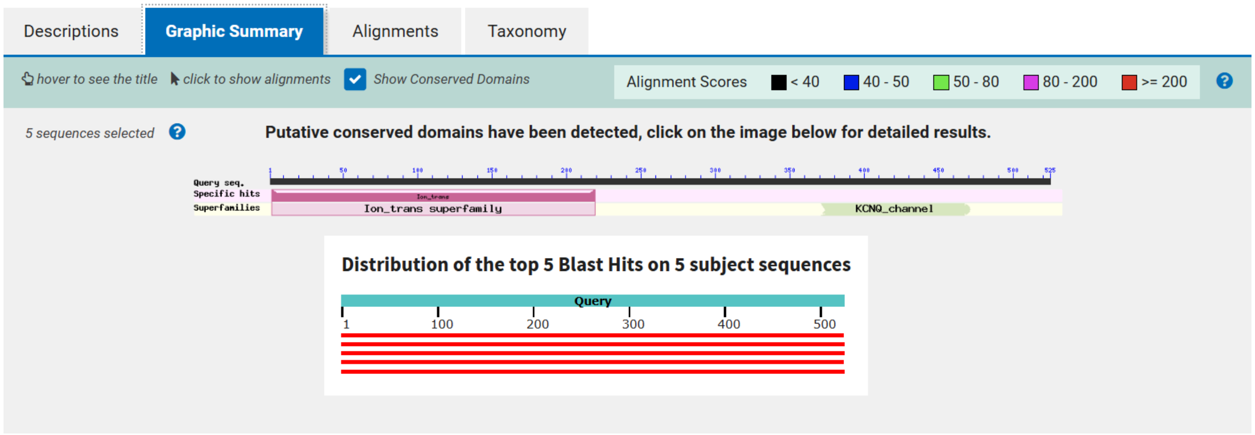
Graphical Summary:

PSI-BLAST:



Distribution of the top 5 Blast Hits on 5 subject sequences

BLASTP:



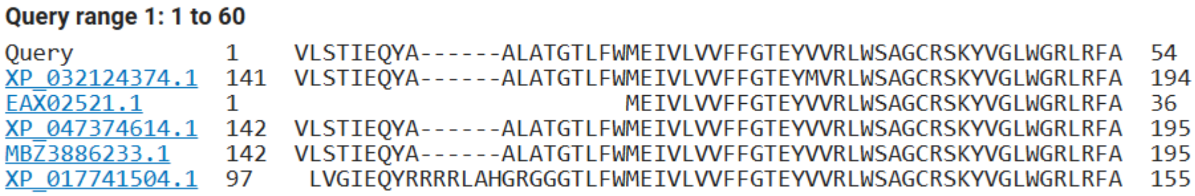
Although all the alignment scores are > 200 in both the graphs. We can clearly observe a key difference between the graphical summary of the last 5 hits of blastp and psi-blast.

The PSI-BLASTP output displays five red lines representing sequence alignments on a scale of 1-500 with one noticeably shorter line (let's say 20-500 instead of 1-500). This shorter line is a key indicator of PSI-BLASTP iterative approach, which refines its search across multiple iterations to detect distant homologs. The partial alignment (shorter line) reflects its sensitivity to sequences with gaps, mismatches, or insertions, shows the algorithms ability to uncover evolutionary connections that might not be immediately obvious

In contrast, the BLASTP output shows five equal red lines, each spanning the full length of the scale (1-500), indicating precise, full-length matches between the query sequence and database hits. This highlights BLASTP's focus on high similarity alignments, where sequences are closely related with minimal gaps or differences

c) **Again, select the last 5 hits, go to the 'Alignments' tab, and change the 'Alignment View' to 'Flat query-anchored with letters for identities'. Explain the differences you see with a few sentences. You may take hints from the graphs from 4b.**

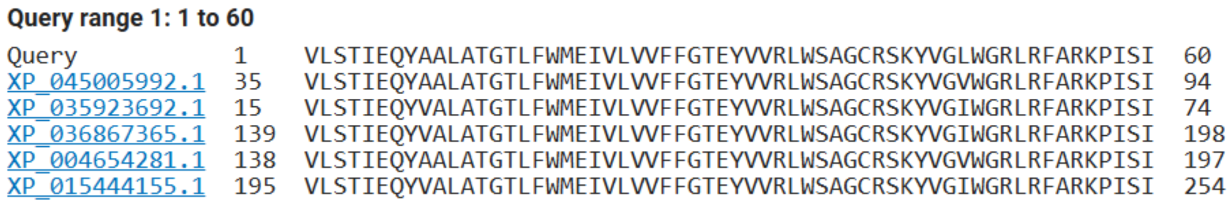
PSI-BLAST:



Query range 1: 1 to 60

```
Query           1    VLSTIEQYA------ALATGTLFWMEIVLVVFFGTEYVVRLWSAGCRSKYVGLWGRLRFA  54
XP_032124374.1  141  VLSTIEQYA------ALATGTLFWMEIVLVVFFGTEYMVRLWSAGCRSKYVGLWGRLRFA  194
EAX02521.1      1                            MEIVLVVFFGTEYVVRLWSAGCRSKYVGLWGRLRFA  36
XP_047374614.1  142  VLSTIEQYA------ALATGTLFWMEIVLVVFFGTEYVVRLWSAGCRSKYVGLWGRLRFA  195
MBZ3886233.1    142  VLSTIEQYA------ALATGTLFWMEIVLVVFFGTEYVVRLWSAGCRSKYVGLWGRLRFA  195
XP_017741504.1  97    LVGIEQYRRRRLAHGRGGGTLFWMEIVLVVFFGTEYVVRLWSAGCRSKYVGLWGRLRFA  155
```

BLASTP:



Query range 1: 1 to 60

```
Query           1    VLSTIEQYAALATGTLFWMEIVLVVFFGTEYVVRLWSAGCRSKYVGLWGRLRFARKPISI  60
XP_045005992.1  35   VLSTIEQYAALATGTLFWMEIVLVVFFGTEYVVRLWSAGCRSKYVGVWGRLRFARKPISI  94
XP_035923692.1  15   VLSTIEQYVALATGTLFWMEIVLVVFFGTEYVVRLWSAGCRSKYVGIWGRLRFARKPISI  74
XP_036867365.1  139  VLSTIEQYVALATGTLFWMEIVLVVFFGTEYVVRLWSAGCRSKYVGIWGRLRFARKPISI  198
XP_004654281.1  138  VLSTIEQYAALATGTLFWMEIVLVVFFGTEYVVRLWSAGCRSKYVGVWGRLRFARKPISI  197
XP_015444155.1  195  VLSTIEQYVALATGTLFWMEIVLVVFFGTEYVVRLWSAGCRSKYVGIWGRLRFARKPISI  254
```

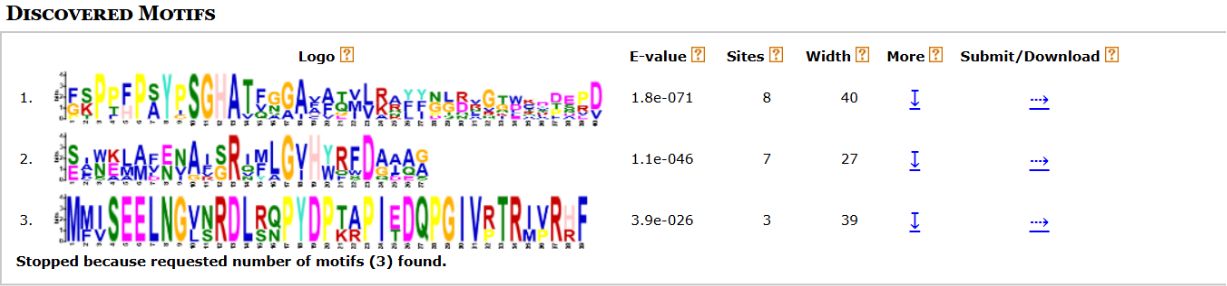
Explanation:

In the ‘Flat query-anchored with letters for identities’ view, the BLASTP alignments for the last 5 hits (XP_0450089592.1 to XP_015444155.1) show near-perfect matches to the query (VLSTIEQYAA...), with almost all positions as letters, indicating high identity and no gaps or mismatches in the aligned region (1-54). In contrast, the PSI-BLASTP alignments (XP_032124374.1 to XP_017741504.1) display more variability, with gaps (dashes like VLSTIEQYA-----ALATGTLFM...) and mismatches, showing these hits are less identical and include sequences with evolutionary divergence. This aligns with the graphs from 4b, where PSI-BLASTP had a shorter line for one hit, reflecting its iterative sensitivity to distant homologs, while BLASTP equal lines indicated consistent, full-length matches.

The graphs from 4b support this: PSI-BLASTP’s shorter line (eg 1-450) hinted at partial alignments, which we now see as gaps and mismatches in the alignment view. BLASTP’s equal lines (all 1-500) predicted its perfect alignments, confirmed by the near-identical sequences here. This further defines PSI-BLASTP focus on evolutionary relationships as compared to BLASTP’s emphasis on exact similarity.

5) Submit the msa.txt file to MEME Suite.

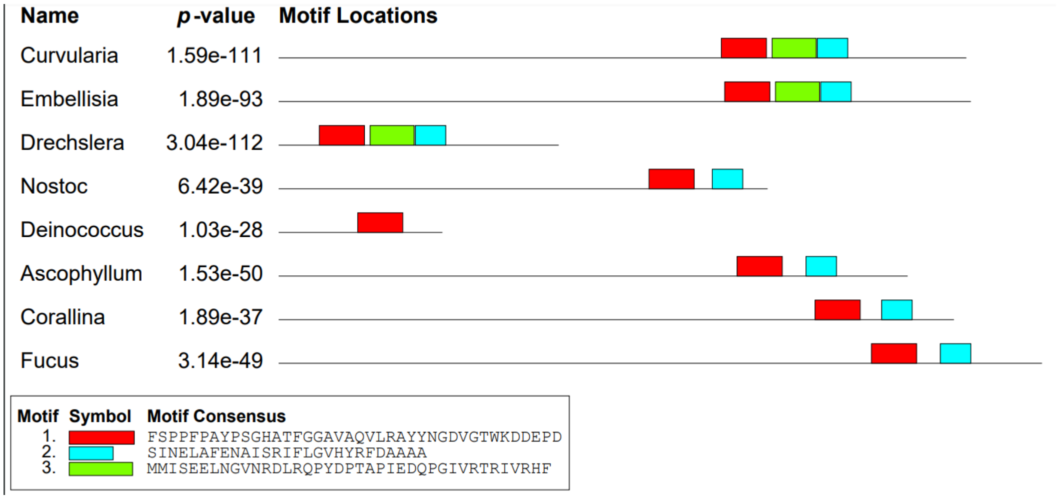
a) **How many motifs are these sequences sharing? How are these motifs positioned on the sequences? Add a screenshot of the graphical view of the motifs.**



DISCOVERED MOTIFS

	Logo	E-value	Sites	Width	More	Submit/Download
1.		1.8e-071	8	40	↧	⇢
2.		1.1e-046	7	27	↧	⇢
3.		3.9e-026	3	39	↧	⇢

Stopped because requested number of motifs (3) found.

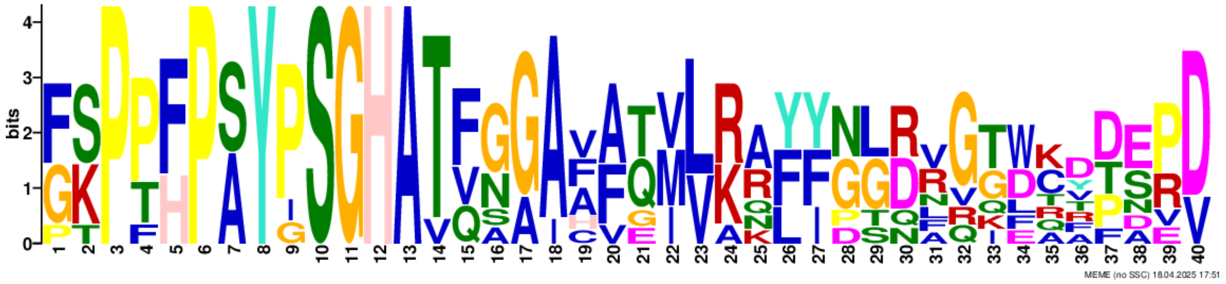


Name	p-value	Motif Locations
Curvularia	1.59e-111	
Embellisia	1.89e-93	
Drechslera	3.04e-112	
Nostoc	6.42e-39	
Deinococcus	1.03e-28	
Ascophyllum	1.53e-50	
Corallina	1.89e-37	
Fucus	3.14e-49	

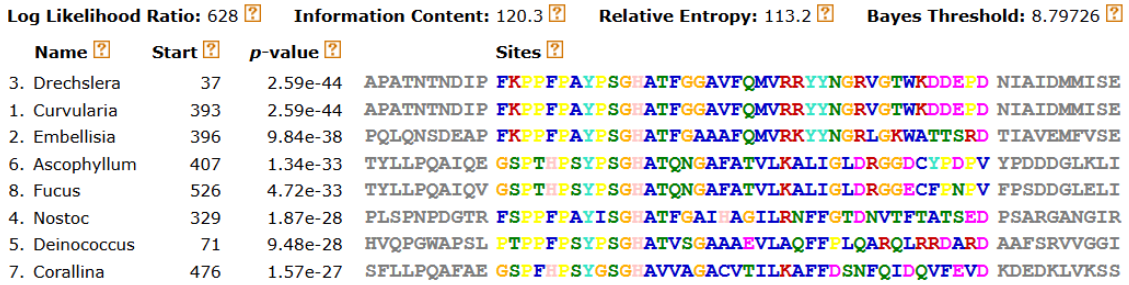
Motif	Symbol	Motif Consensus
1.	🟥	FSPPFPAYPSGHATFGGAVAQVLRAYYNGDVGTWKDDEPD
2.	🟦	SINELAFENAISRIFLGVHYRFDAAAA
3.	🟩	MMISEELNGVNRDLRQPYDPTAPIEDQPGIVRTRIVRHF

Motifs shared = 3 (default parameter in MEME Suite).
Motifs have a general pattern in the order of red, green and blue in order
Motif 1 (red): Found in all 8 sequences (Curvularia, Embellisia, Drechslera, Nostoc, Deinococcus, Ascophyllum, Coralina, Fucus).
Motif 2 (green): Found in 3 sequences (Curvularia, Embellisia, Drechslera)
Motif 3 (blue): Found in 7 sequences (Curvularia, Embellisia, Drechslera, Ascophyllum, Nostoc, Coralina, Fucus)

b) **Select the most significant motif and add 1) the seq logo, 2) the alignment, and 3) scores for that motif. What does the seq logo and size of the letters represent? Explain using the screenshots you provided.**



Motif 1 has the lowest E-value (1.8e-071), making it the most significant. It appears in 8 sites (all sequences) with a width of 40.



Log Likelihood Ratio: 628 ? **Information Content:** 120.3 ? **Relative Entropy:** 113.2 ? **Bayes Threshold:** 8.79726 ?

Name ?	**Start** ?	***p*-value** ?	**Sites** ?
3. Drechslera	37	2.59e-44	APATNTNDIP FKPPFPAYPSGHATFGGAVFQMVRRYYNGRVGTWKDDEPD NIAIDMMISE
1. Curvularia	393	2.59e-44	APATNTNDIP FKPPFPAYPSGHATFGGAVFQMVRRYYNGRVGTWKDDEPD NIAIDMMISE
2. Embellisia	396	9.84e-38	PQLQNSDEAP FKPPFPAYPSGHATFGAAAFQMVRKYYNGRLGKWATTSRD TIAVEMFVSE
6. Ascophyllum	407	1.34e-33	TYLLPQAIQE GSPTHPSYPSGHATQNGAFATVLKALIGLDRGGDCYPDPV YPDDDGLKLI
8. Fucus	526	4.72e-33	TYLLPQAIQV GSPTHPSYPSGHATQNGAFATVLKALIGLDRGGECFPNPV FPSDDGLELI
4. Nostoc	329	1.87e-28	PLSPNPDGTR FSPPFPAYISGHATFGAIHAGILRNFFGTDNVTFTATSED PSARGANGIR
5. Deinococcus	71	9.48e-28	HVQPGWAPSL PTPPFPSYPSGHATVSGAAAEVLAQFFPLQARQLRRDARD AAFSRVVGGI
7. Corallina	476	1.57e-27	SFLLPQAFAE GSPFHPSYGSGHAVVAGACVTILKAFFDSNFQIDQVFEVD KDEDKLVKSS

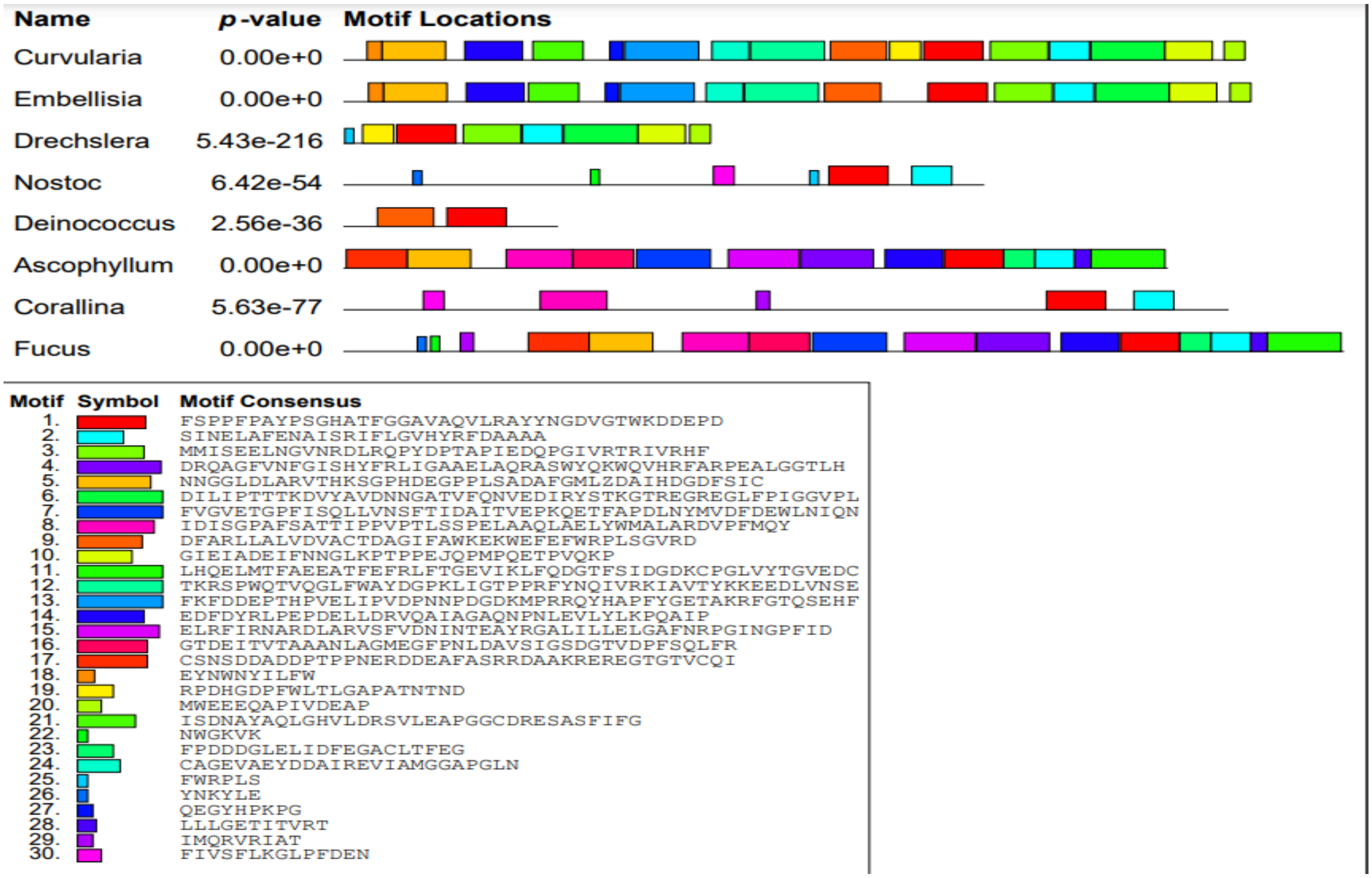
REMEMBER TO LABEL THE ALIGNMENT AND SCORE OF MOTIF 1 IN THE ABOVE PICTURE

The sequence logo for Motif 1 represents the consensus sequence across 8 sites, the x-axis shows positions 1 to 40 and the y-axis indicate information content in bits, reflecting conservation at each position. The height of the letter shows how conserved a position is, therefore the letters are taller, for example at positions 3–13 (PPYSGHA), indicating high conservation where nearly all sequences share the same amino acid, such as "P" at position 3 and 6. Shorter stacks with multiple letters, like at positions 26–40 (e.g V, E, N), show lower conservation with more variability across sequences. Within each stack, the size of individual letters reflects their frequency therefore a taller letter like "S" at position 10 means it's more frequent at that position.

c) **Change the parameters to find the maximum possible number of motifs. How many of them are significant? Submit a screenshot of the graphical view of the motifs.**

We adjusted the meme suite parameter settings to search for a maximum of 30 motifs instead of our previous 3, the output identified all 30 motifs. The first 18 motifs are colorful, indicating they are statistically significant (with E-values ranging from 1.8e-071 for Motif 1 to 2.3e-002 for Motif 18), while the last 12 motifs (19 to 30) are dimmed and not colorful, suggesting they are not significant as their E-values (e.g, 8.3e-002 for Motif 19 to 6.4e+003 for Motif 30) are higher and likely above the default significance threshold. Therefore out of the 30 motifs 18 are significant. (Graphical view of motifs given below)

Graphical view of the motifs:



Name	*p*-value	Motif Locations
Curvularia	0.00e+0	
Embellisia	0.00e+0	
Drechslera	5.43e-216	
Nostoc	6.42e-54	
Deinococcus	2.56e-36	
Ascophyllum	0.00e+0	
Corallina	5.63e-77	
Fucus	0.00e+0	

Motif	Symbol	Motif Consensus
1.	▬	FSPPFPAYPSGHATFGGAVAQVLRAYYNGDVGTWKDDEPD
2.	▬	SINELAFENAISRIFLGVHYRFDAAAA
3.	▬	MMISEELNGVNRDLRQPYDPTAPIEDQPGIVRTRIVRHF
4.	▬	DRQAGFVNFGISHYFRLIGAAELAQRASWYQKWQVHRFARPEALGGTLH
5.	▬	NNGGLDLARVTHKSGPHDEGPPLSADAFGMLZDAIHDGDFSIC
6.	▬	DILIPTTTKDVYAVDNNGATVFQNVEDIRYSTKGTREGREGLFPIGGVPL
7.	▬	FVGVETGPFISQLLVNSFTIDAITVEPKQETFAPDLNYMVDFDEWLNIQN
8.	▬	IDISGPAFSATTIPPVPTLSSPELAAQLAELYWMALARDVPFMQY
9.	▬	DFARLLALVDVACTDAGIFAWKEKWEFEFWRPLSGVRD
10.	▬	GIEIADEIFNNGLKPTPPEJQPMPQETPVQKP
11.	▬	LHQELMTFAEEATFEFRLFTGEVIKLFQDGTFSIDGDKCPGLVYTGVEDC
12.	▬	TKRSPWQTVQGLFWAYDGPKLIGTPPRFYNQIVRKIAVTYKKEEDLVNSE
13.	▬	FKFDDEPTHPVELIPVDPNNPDGDKMPRRQYHAPFYGETAKRFGTQSEHF
14.	▬	EDFDYRLPEPDELLDRVQAIAGAQNPNLEVLYLKPQAIP
15.	▬	ELRFIRNARDLARVSFVDNINTEAYRGALILLELGAFNRPGINGPFID
16.	▬	GTDEITVTAAANLAGMEGFPNLDAVSIGSDGTVDPFSQLFR
17.	▬	CSNSDDADDPTPPNERDDEAFASRRDAAKREREGTGTVCQI
18.	▬	EYNWNYILFW
19.	▬	RPDHGDPFWLTLGAPATNTND
20.	▬	MWEEEQAPIVDEAP
21.	▬	ISDNAYAQLGHVLDRSVLEAPGGCDRESASFIFG
22.	▬	NWGKVK
23.	▬	FPDDDGLELIDFEGACLTFEG
24.	▬	CAGEVAEYDDAIREVIAMGGAPGLN
25.	▬	FWRPLS
26.	▬	YNKYLE
27.	▬	QEGYHPKPG
28.	▬	LLLGETITVRT
29.	▬	IMQRVRIAT
30.	▬	FIVSFLKGLPFDEN