

## assignment 7 Rebekka Trenkle

model sequence: human langerin  
uniprot Q9UJ71

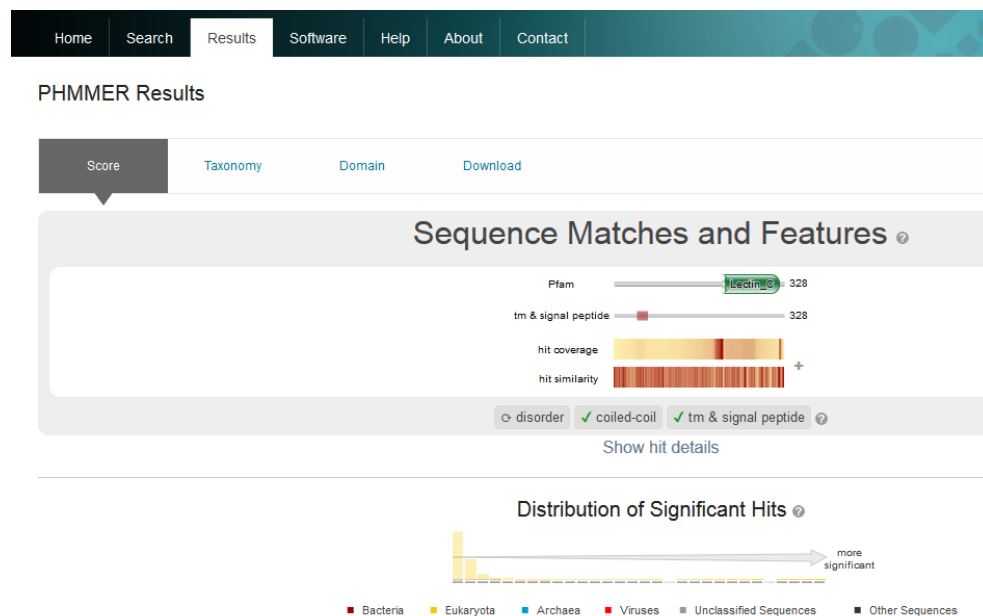
<https://www.uniprot.org/uniprot/Q9UJ71#sequences>

```
>sp|Q9UJ71|CLC4K_HUMAN C-type lectin domain family 4 member K
OS=Homo sapiens OX=9606 GN=CD207 PE=1 SV=2
MTVEKEAPDAHFTVDKQNISLWPREPPPKSGPSLVPGKTPTVRAALICLTLVLVASVLLQ
AVLYPRFMGTISDVKTNVQLLKGRVDNISTLDSEIKKNSDGMEAAGVQIQMVNESLGYVR
SQFLKLKTSVEKANAQIQILTRSWEEVSTLNAQIPELKSDLEKASALNTKIRALQGSLEN
MSKLLKRQNDILQVVSQGWKYFKGNFYFSLIPKTWYSAEQFCVSRNSHLTSVTSESEQE
FLYKTAGGLIYWIGLTKAGMEGDWSWVDDTPFNKVQSVRFWIPGEPNAGNNEHCGNIKA
PSLQAWNDAPCDKTFLFICKRPYVPSEP
```

### 1 phmmer

phmmer aligns the given protein sequence to a protein sequence data base. Procedure:

- select phmmer
- upload data, submit
- the result shows the sequence features found by the algorithm




- click on domains: shows match with query architectures. Highlighted in red are the exact matches, here Lectin C

## Domain Architectures








**5889**  
SEQUENCES

**Exact match with query architecture: Lectin\_C, example: CLC4K\_HUMAN**

Sequence Features  328

[Hide All](#)

---

<a href="#">CLC4K_HUMAN</a> Homo sapiens	Sequence Features  328
<a href="#">G3QPX8_GORGO</a> Gorilla gorilla gorilla	Sequence Features  328
<a href="#">H2R8S3_PANTR</a> Pan troglodytes	Sequence Features  328
<a href="#">A0A2R8ZP90_PANPA</a> Pan paniscus	Sequence Features  328
<a href="#">G1RGW2_NOMLE</a> Nomascus leucogenys	Sequence Features  328
<a href="#">A0A2K5IFY7_COLAP</a> Colobus angolensis palliatus	Sequence Features  328
<a href="#">A0A2K6AMD9_MACNE</a> Macaca nemestrina	Sequence Features  328

- going back to scores (after selecting the exact matches): shows the scores and hit positions of the aligned sequences on the right for every found sequence
- one can take a closer look at the aligned sequences, e.g. with human and wolf query sequence:
- under "taxonomy" one can look at the evolutionary tree and select the desired branch, here done as an example for human
- with that selection one can look again at the aligned sequences and the the hit positions
- the results in github are only for this selection

## 2 hmmscan

hmmscan queries a protein sequence against a collection of profile-HMM.

(show all) alignments Your search took: 5.27 secs showing rows 1 - 100 of 5889

Target 321 REVERSED 328

- the result gives the closest related protein sequence

### 3 hmmsearch

hmmsearch scans the aligned sequences/the profile HMM against a lot of sequences

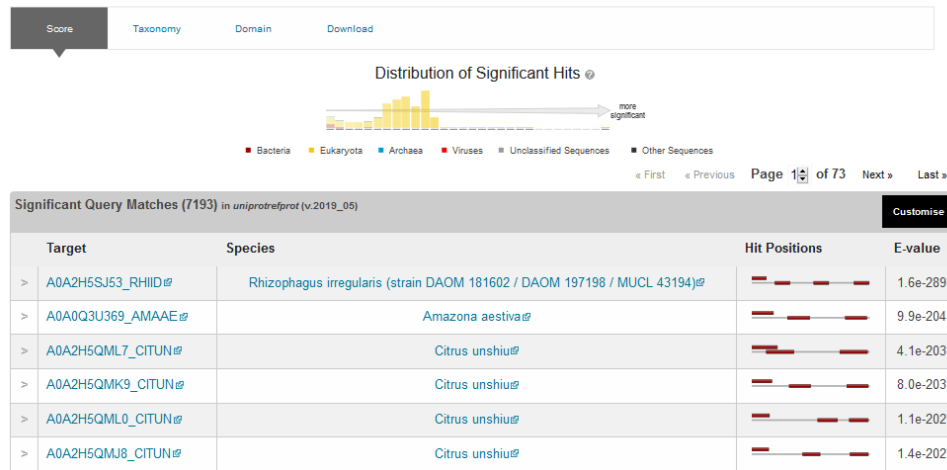
- example data used because the clustal omega results couldn't be read as input





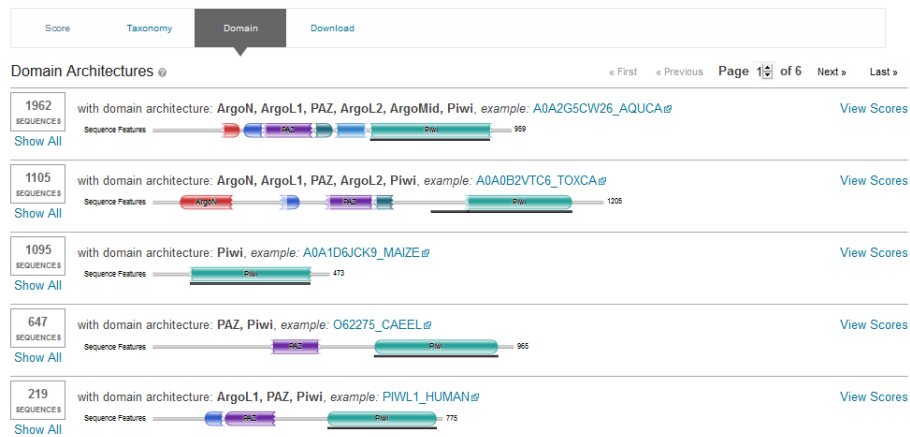
## HMMSEARCH Results

Search Again



## HMMSEARCH Results

Search Again



- start next iterations
- results show the sequence matches and features
- in addition, matches that were found (or lost) in every iteration step are highlighted
- the alignments, hit positions and so forth can be looked at individually for each iteration

Search Again

Jackhmmer Summary					
Iteration	Results	Hits			
		New	Lost	Dropped	Total
1	4177FA58-912C-11E9-8062-6E99F75AEC3D.1	+10516	-	-	10516
2	4177FA58-912C-11E9-8062-6E99F75AEC3D.2	+13829	-1	-	24344
3	4177FA58-912C-11E9-8062-6E99F75AEC3D.3	+19735	-19	83	43977

Start Next Iteration

Significant Query Matches (43977) in uniprotrefprot (v.2019_05)												Customise
Target			Species			Hit Positions			E-value			
> A0A226NJF9_CALSU			Callipepla squamata			<div></div>			0.0e+00			<input checked="" type="checkbox"/>
v A0A226PYY6_COLVI			Colinus virginianus			<div></div>			0.0e+00			<input checked="" type="checkbox"/>
Query		Target Envelope		Target Alignment		Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	E-value	
start	end	start	end	start	end						Ind.	Cond.
248	321	175	264	194	260	3.31	0.88	40.3 (27)	74.6 (50)	55.6	1.0e-11	2.3e-14
<div>Query 248 gssyWIGledlgtegswnkvdGspIspvssjt nWapgcPnnaggncDvalssstgkndvdGseKlprlQeK 321 + +W+GI ++ ++sw+W+ Gsp+ y nWapg P+ + C+ +++ +gkW +v C++k1+1C+k Target 194 SAKFWIGLVRK-LDSSNWTGSGPFR-----VINWAPGNPSEF-LIQMGTFQGR-WGKWEIVACDQKLGVIQK 260 PP 468*****97.789*****8....*****86.4899****997.6*****9</div>												
Query		Target Envelope		Target Alignment		Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	E-value	
start	end	start	end	start	end						Ind.	Cond.
194	321	261	409	281	406	5.24	0.93	34.9 (44)	77.0 (97)	142.8	3.0e-38	6.9e-41
<div>Query 194 sCpgeWlfggsCYkfstekltNeeAraaGslgnhLstseEldflrallg...gssyWIGledlgtegswnkvdG 270 +Cp++W+ + g+CY+++++k W +A+++C+++g+L s+++ eE +f+ s+l g + + WlGld + +W+dG+ Target 281 WAKFWIGLVRK-LDSSNWTGSGPFR-----VINWAPGNPSEF-LIQMGTFQGR-WGKWEIVACDQKLGVIQK 360 <a href="#">[show all]</a> alignments ***** Your search took: 69 secs ***** showing rows 1 - 100 of 96046</div>												