

# PHMMer

Zuerst musste ich für den Protein-hidden-markov-modeler (phmmer) eine Proteinsequenz finden, die ich von [http://pathblast.org/docs/sequence\\_format.html](http://pathblast.org/docs/sequence_format.html) nahm. Wie für alle anderen Beispiele auch, musste ich hierzu eine fasta-Datei finden. Dies war ein Protein aus dem Haushuhn, was zur Abwehr gegen Gramnegative Bakterien wirkt. Diese wurde dann mit dem Online-tool mit bekannten Proteinsequenzen verglichen und mehrere Listen mit Ergebnissen wurden ausgegeben. Die erste namens "score" beschrieb die Übereinstimmung mit bekannten Proteinsequenzen auf Aminosäuren genau, geordnet nach statistischer Übereinstimmung, die namens Domain nach Übereinstimmung im Bezug auf bekannte Domänen in Proteinen und die namens "Taxonomy" geordnet nach Verwandtschaftsbezug.

PHMMER Results

Search Again

ScoreTaxonomyDomainDownload

Distribution of Significant Hits

BacteriaEukaryotaArchaeaVirusesUnclassified SequencesOther Sequences

Did you know? Clicking the button customise, in the table header below, gives you the opportunity toggle up to twelve columns of data in this table. hide this X  
Also, have a look at the new Cross-references column, showing references to other resources at the EBI.

« First« PreviousPage 1 of 81Next»Last »

Significant Query Matches (7782) in uniprotrefprot (v.2019\_05)Customise

Target	Species	E-value
> A0A226PX63_COLVI	Colinus virginianus	0.0e+00
> A0A226NIP0_CALSU	Callipepla squamata	0.0e+00
> A0A218UPK0_9PASE	Lonchura striata domestica	4.2e-221
> A0A218UQ52_9PASE	Lonchura striata domestica	9.8e-191
> A0A226NI85_CALSU	Callipepla squamata	7.0e-187
> A0A2U4BFG2_TURTR	Tursiops truncatus	2.4e-166
> OVALX_CHICK	Gallus gallus	2.2e-151

Significant Query Matches (7782) in uniprotrefprot (v.2019\_05)Customise

Target	Species	E-value
v A0A226PX63_COLVI	Colinus virginianus	0.0e+00

Query	Target Envelope	Target Alignment	Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	E-value				
start	end	start	end	start	end			Ind.	Cond.			
1	227	188	417	188	414	1.02	0.99	75.8 (172)	88.1 (200)	373.5	7.9e-109	1.5e-112

Query

1 qikdlivssstgldtllvlnelyfkgwktafnaedtrempfhvtkgeskpvqgmncnnsfnvatlpaeckmlelpfa80

Target

188 QIKDLLVPSSVDFTGMVFINTIIFKGIWKTAFNTEDTREMPFSMTKQESKPVQGMCLNDFNMATLPAEKMRILELPPA267

PP

89\*\*\*\*\*

Query

81 sgdlmlvllpdevsclerlektinfekltawtntptmekrrvkylpkmkieekynltsvmlalgmtldfipsanltgi160

Target

268 SGGLSMLVLLPDEVSGLEQIPKAINFEKLAEWSTNAMEKRSKMVILPRMKIEEKYNLTSTIMALGMTDLFSPRSANLTGI347

PP

\*\*\*\*\*

Query

161 Sseelklsavhgafmelsedglaagstgvleikhsesefradhpfllfikhnptnltivylfg227

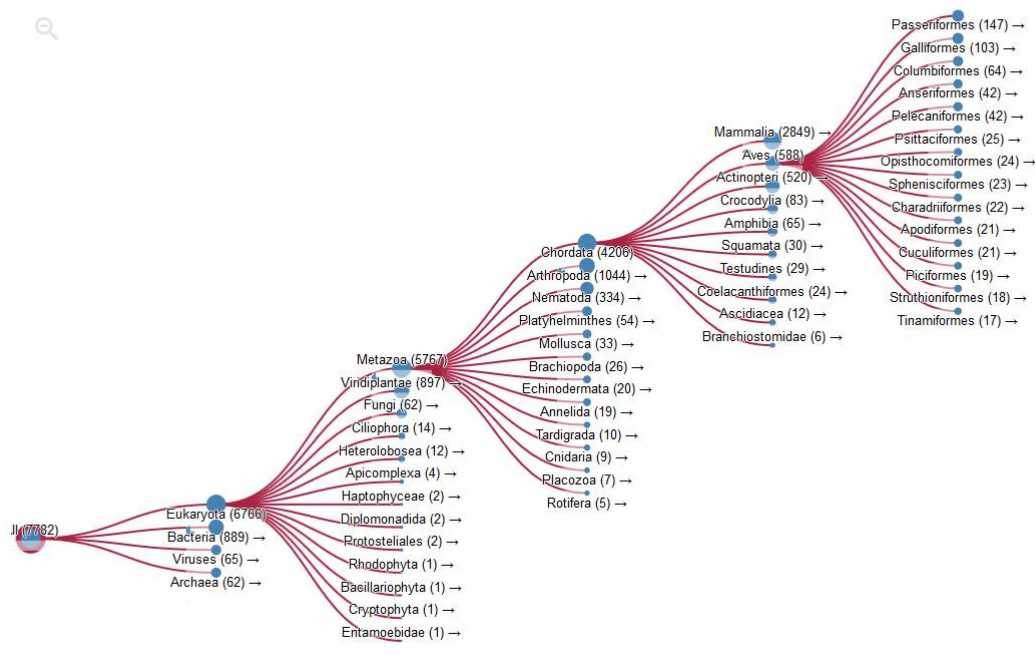
Target

348 SSVENLWISDAVHGAFMEVNEEGTEAAGSTGAIGNIKHSVEFEPRADHFFLLIRYNPTNVILPFD414

PP

\*\*\*\*\*g6

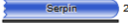

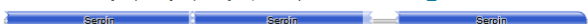
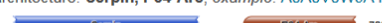
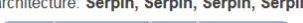
## Taxonomic distribution of all search hits

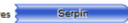
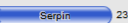

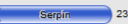






Wie zu erwarten gab es eine genaue Übereinstimmung mit einer Sequenz von genau diesem Hühnerprotein.

### Domain Architectures

« First « Previous Page 1 of 2 Next » Last »

7041 SEQUENCES	Exact match with query architecture: <a href="#">Serp, example: OVALX_CHICK</a>	<a href="#">View Scores</a>
Show All	Sequence Features:  232	
567 SEQUENCES	with domain architecture: <a href="#">Serp, Serpin, example: A0A226NI85_CALSU</a>	<a href="#">View Scores</a>
Show All	Sequence Features:  861	
40 SEQUENCES	with domain architecture: <a href="#">Serp, Serpin, Serpin, example: A0A218UPK0_9PASE</a>	<a href="#">View Scores</a>
Show All	Sequence Features:  1175	
17 SEQUENCES	with domain architecture: <a href="#">Serp, P34-Arc, example: A0A0V0W9A4_9BILA</a>	<a href="#">View Scores</a>
Show All	Sequence Features:  726	
15 SEQUENCES	with domain architecture: <a href="#">Serp, Serpin, Serpin, Serpin, example: A0A0P7XSN6_SCLFO</a>	<a href="#">View Scores</a>
Show All	Sequence Features:  581	

7041 SEQUENCES	Exact match with query architecture: <a href="#">Serp, example: OVALX_CHICK</a>	<a href="#">View Scores</a>
Hide All	Sequence Features:  232	
<a href="#">OVALX_CHICK</a> <a href="#">Gallus gallus</a>	Sequence Features:  232	
<a href="#">A0A1D5P158_CHICK</a> <a href="#">Gallus gallus</a>	Sequence Features:  388	
<a href="#">G1MZH2_MELGA</a> <a href="#">Meleagris gallopavo</a>	Sequence Features:  232	
<a href="#">A0A226PWY1_COLVI</a> <a href="#">Colinus virginianus</a>	Sequence Features:  388	
<a href="#">OVALX_CHICK</a> <a href="#">Gallus gallus</a>	Sequence Features:  388	
<a href="#">G1MZF4_MELGA</a> <a href="#">Meleagris gallopavo</a>	Sequence Features:  388	
<a href="#">E1BTF4_CHICK</a> <a href="#">Gallus gallus</a>	Sequence Features:  388	

## HMMscan

Hier habe ich von [https://en.wikipedia.org/wiki/FASTA\\_format](https://en.wikipedia.org/wiki/FASTA_format) eine Proteinsequenz von Calmodulin im HMMscan gesucht, Dort wird mit einem Hidden Markov Model nach Übereinstimmungen mit Proteinsequenzen gesucht. Im Ergebnis wird dementsprechend angegeben, wo Deletionen (Punkte in der gefundenen Sequenz)/Insertionen (Minusse in der Ausgangssequenz) stattgefunden haben müssten.

v	EF-hand_9	PF14658.6	CL0220	EF-hand domain		13	76	2.3e-09	1.3e-12
<p>Model 3 vEvcltqtGrVpvsrIldyLlvtacqDeesseLeelakSLlPdgedasvldtflrvmr 64 F+++D + G+++ ++l +++r ++ +p+e+eL++ +e+D dg+ +++d+ fl++m + Query 15 AFSLFDKGDGDTITTRKELGTVMRSLGQ-NPTEAELQDMINEVDADGN-STIDFPEFLTMMAR 74 PP 6*****.*****.*****76</p>									
>	EF-hand_4	PF12763.7	CL0220	Cytoskeletal-regulatory complex EF hand		75	147	5.2e-05	2.9e-08
v	EFhand_Ca_insen	PF08726.10	CL0220	Ca2+ insensitive EF hand		78	147	8.2e-05	4.6e-08
<p>Model 1 vvdseaeqveeaFfal.aggKpYvTeedLrrellp...eqaeYlerMpeysgpe 50 d+ds e++++eFr ++g Y+ + +Lr+ ++ + ++ +++M + + + Query 78 DTDSEEEIREAFRVFdKDGNGYISAAELRHVMTNlgKLTDEEVDEMIREADID 131 PP 89*****g*****966541103335555555555544</p>									

## HMMsearch

In diesem Tool wurden drei Sequenzen sehr unterschiedlicher Herkunft (Cytochrom B, Calmodulin und Prolactin) verwendet und mit T-Coffee aneinanderkombiniert.

T

COFFEE

HomeHistoryTutorialReferencesContactsProjectsDownload

Expresso

Aligns protein sequences using structural information

Sequences input

Paste or upload your set of sequences in FASTA format

Sequences to align

Click here to use the sample file

GLMPFLHTSKHRSMMRLPLSQALFWTLTMDLLTLTWIGSQPVEYPYTIIGQMASILYFSIILAFLPIAGX  
IENY  
>LCBO - Prolactin precursor - Bovine  
MSKGSQKGSRLLLLVVSNLLLCQGVVSTPVCNPGNQCQVSLRDLFDRAVMVSHYIHDLS  
EMFNEFDKRYAQKGFITMALNSCHTSSLPTPEDKEQAQQTHEVLMSLILGLRSWNPPLYHL  
VTEVRGMKGAPDAILSRATEIEENKRLLEGMEMIFGQVIPGAKETEPYFVWSGLPSLQTKDED  
ARYSAFYNLLHCLRRDSSRIDTVLKLINCRIIYNNNC\*

- OR - Click here to upload a file

Show more options

Your email address

SubmitReset

## Espresso alignment result

MSPA

The multiple sequence alignment result as produced by T-coffee.

T-COFFEE, Version\_11.00.d625267 (2016-01-11 15:25:41 - Revision d625267 - Build 507)  
Cedric Notredame  
SCORE=26

\* **BAD AVG GOOD**

MCHU : 21  
gi|5524211|ab|A : 29  
LCB0 : 28  
cons : 2

MCHU ADLTE-----EQIAEFKEAFSLFDKGDGDTITTKELGTVMRSLGO  
gi|5524211|gb|A LCLYTHIGRNIYYGSYLSETWNTGIMLLITMATAFMGYVLPWGOMSFWATVITNLSFATP  
LCB0 MSDKGS-----SQKGSRLLLLVSNLLLCQGVVSTPVCNPGPGNC

cons : . . . .

MCHU NPTEAELDMINEVDADNGTIDFPEFLTMARKMKDTSSEEEIREFRVFDKDGNGYI----  
gi|5524211|gb|A YGTNLVEWVWGGFSVDKATLNRFFAFHFIPLPTMVALAGVHLTFLHETGSNNPLGLTS  
LCB0 QVSLRDLFDRAVMVSHYIHLDSSEMFNEFDKRYAQKGKITMALNSCHTSSLPTPEDKEAQAO

cons : . . . .

MCHU -----  
gi|5524211|gb|A DSDKIPFHPYYTIKDFLGLLLTLTLTLALLSPDMLGPDPNHMP  
LCB0 THHEVLSLTLGLLRSWNDPLYHLVTEVRGMKGAPDILSRATEIEENKRLLEGMEIMIFGV

cons : . . . .

MCHU -----S-AAELRHVMTNL-----GEKLT  
gi|5524211|gb|A ADPLNPLHLKPEWYFLFAYALLRSPVKNLGGVLAFLSIVILGLMPFLHTSKHRSMMPLRLS  
LCB0 IPGAKETEPYPVWSGLPSLTK-----DEDAR

cons : . . . .

MCHU DEEVDEMIREADIDGGGVNYEEFVOMMTAK-----  
gi|5524211|gb|A QALFWTLTMDLLTLTWIGSQPVEYPTIIGOMASILYFSIILAFPLPIAGXIENY  
LCB0 YSAFYNNLHCLRDSSKIDTYLKLNCRIIYNNC-----

cons : . . . .

Diese Daten wurden nun mit hmmSearch analysiert bzw. es wurde ein HMM-Consensusmotiv erstellt und dann wurden wieder wie bei phmmer verschiedene Proteinsequenzen/Domänen etc. aufgelistet.

**HMMSEARCH Results**

Search Again

Score Taxonomy Domain Download

Distribution of Significant Hits

Taxonomic Group	Percentage of Hits
Bacteria	~15%
Eukaryota	~45%
Archaea	~10%
Viruses	~5%
Unclassified Sequences	~20%
Other Sequences	~5%

« First « Previous Page 1 of 118 Next » Last »

Significant Query Matches (9027) in uniprotrefprot (v.2019\_05)

Customise

Target	Species	E-value
K1PMY9_Cragi	Crassostrea gigas	9.1e-119

Query	Target Envelope	Target Alignment	Bias	Accuracy	% Identity (count)	% Similarity (count)	Bit Score	E-value				
start	end	start	end	start	end			Ind.	Cond.			
1	99	215	316	215	313	0.06	0.98	19.2 (19)	97.0 (96)	110.9	1.3e-28	3.6e-32

```

.....*.....*.....*.....*.....*.....*.....*.....*.....*
Query   1  |dkktelqitsatallgvvlkigmsfvgatviunlf@aipqigtndleciaggwsvdkatlnrfafefilrftmvala 80
      +d+tte+q+++++++k+g++++++t++a+s++++++l+d++++v+d+++++f+f+++++tm++ +
Target 215 |ADQITEEQIAEFPEAFSLFDKDGSGTITTRKLGTVMRSLGQNPTAEALQDMINEVDADNGTIDFPEFLTMARKMKDSE 294
PP      69*****
    
```

```

.....*.....
Query   81 |gvellflaetgsnnplelk 99
      ++e+l+a+++++++
Target 295 |SEELREAPRVFDKGNGF 313
PP      *****985
    
```

Natürlich wurde diesmal keine hundertprozentig Übereinstimmung gefunden, da es sich bei der Eingabe um ein Consensusmotiv aus drei völlig verschiedenen Proteinen handelte.

# jackHMMer

Es wurde wieder die Proteinsequenz von HMMScan verwendet. Die Ergebnisse sind ähnlich denen von phmmer, so werden wieder Listen von unterschiedlich gut übereinstimmenden Proteinen geordnet nach Ähnlichkeit (Score), Verwandtschaft (Taxonomy) und Proteindomäne (domain) angezeigt.

## JACKHMMER Results

Search Again

Score

Taxonomy

Domain

Download

Iteration 1

Return to the [Results Summary](#) | [Jump to threshold](#)

Sequence selection

- ☒ above threshold
- ☐ unselect all

Continue the search

Start Iteration 2

## Sequence Matches and Features

Pfam 149

☐ disorder ☒ coiled-coil ☒ tm & signal peptide

Continue the search

Start Iteration 2

## Distribution of Significant Hits



« First « Previous Page 1 of 494 Next » Last »



Significant Query Matches (39867) in uniprotrefprot (v.2019\_05)

Customise

Target		Species		E-value	
K1PMY9_CRAGI		Crassostrea gigas		7.7e-288	
Query		Target Envelope		Target Alignment	
start	end	start	end	start	end
11	148	64	214	73	214
Query		Target		Bias	
11 efkeas...lfdkdgdtittkeltgtv...lqgnp...elqdmnevdaadngtldfp...ltnma...m...tdtdee...i		73 DYKLI...FDK...DGGY...SSDDVRN...VLSYDMLSTEGELQDVVAELDKKGDGLITLLEFVSVMN...HK...ifSKKDEKDLEPR		2.19	
148		214		0.91	
Query		Target		% Identity (count)	
11 efkeas...lfdkdgdtittkeltgtv...lqgnp...elqdmnevdaadngtldfp...ltnma...m...tdtdee...i		73 DYKLI...FDK...DGGY...SSDDVRN...VLSYDMLSTEGELQDVVAELDKKGDGLITLLEFVSVMN...HK...ifSKKDEKDLEPR		36.2 (50)	
Query		Target		% Similarity (count)	
11 efkeas...lfdkdgdtittkeltgtv...lqgnp...elqdmnevdaadngtldfp...ltnma...m...tdtdee...i		73 DYKLI...FDK...DGGY...SSDDVRN...VLSYDMLSTEGELQDVVAELDKKGDGLITLLEFVSVMN...HK...ifSKKDEKDLEPR		65.2 (90)	
Query		Target		Bit Score	
11 efkeas...lfdkdgdtittkeltgtv...lqgnp...elqdmnevdaadngtldfp...ltnma...m...tdtdee...i		73 DYKLI...FDK...DGGY...SSDDVRN...VLSYDMLSTEGELQDVVAELDKKGDGLITLLEFVSVMN...HK...ifSKKDEKDLEPR		81.0	
Query		Target		E-value	
11 efkeas...lfdkdgdtittkeltgtv...lqgnp...elqdmnevdaadngtldfp...ltnma...m...tdtdee...i		73 DYKLI...FDK...DGGY...SSDDVRN...VLSYDMLSTEGELQDVVAELDKKGDGLITLLEFVSVMN...HK...ifSKKDEKDLEPR		2.3e-19	
Query		Target		Cond.	
11 efkeas...lfdkdgdtittkeltgtv...lqgnp...elqdmnevdaadngtldfp...ltnma...m...tdtdee...i		73 DYKLI...FDK...DGGY...SSDDVRN...VLSYDMLSTEGELQDVVAELDKKGDGLITLLEFVSVMN...HK...ifSKKDEKDLEPR		2.7e-22	

Jedoch ist die Suche hier wiederholbar möglich, wobei mehr Sequenzen gefunden werden (New) und die Gesamtzahl angezeigt wird (Total).

Search Again

Jackhmmmer Summary					
Iteration	Results	Hits			
		New	Lost 	Dropped 	Total
1	6BCA69DC-90FD-11E9-A25C-D898DBC3747A.1	+39867	-	-	39867
2	6BCA69DC-90FD-11E9-A25C-D898DBC3747A.2	+35003	-2	6	74862
3	6BCA69DC-90FD-11E9-A25C-D898DBC3747A.3	+6935	-21	299	81477
4	6BCA69DC-90FD-11E9-A25C-D898DBC3747A.4	+3046	-30	840	83653
5	6BCA69DC-90FD-11E9-A25C-D898DBC3747A.5	+2388	-55	980	85006

Start Next Iteration