

華中科技大学

生物信息学上机实验

院 系 生命科学与技术学院

专业班级 登峰 1901 班

姓 名 张皓鸿

学 号 U201912537

2021 年 5 月 4 日

目 录

1 基因组分析	1
1.1 总结 β 属冠状病毒和 SARS-CoV-2 (2019-nCoV) 的主要特点	1
1.2 编写并运行 example4-1.pl	1
1.3 SARS-CoV-2 的基因组序列	1
1.4 SARS-CoV-2 潜在编码序列的预测	2
1.5 发现与 SARS-CoV-2 同源的冠状病毒	3
1.6 插入片段分析	3
2 序列分析	6
2.1 INS1378 与 pShuttle-SN 载体的相似性	6
2.2 SARS-CoV-2 的蛋白质序列	7
2.3 等电点与分子量分析	9
2.4 功能结构域分析	9
2.5 细胞亚定位分析	39
3 进化分析	55
3.1 获得 Beta 属冠状病毒同源基因组序	55
3.2 多序列比对	55
3.3 分子进化树构建	56
4 Spike 蛋白分析	57
4.1 SARS-CoV-2 的 Spike 的同源蛋白质分析	57
4.2 Spike 蛋白质的分子进化	57
4.3 Spike 蛋白质的 Furin 切割位点预测	60

1 基因组分析

1.1 总结 β 属冠状病毒和 SARS-CoV-2 (2019-nCoV) 的主要特点

- 1、都是 RNA 病毒
- 2、冠状病毒的 s 蛋白分为两个功能单元 S1 和 S2。S1 通过与宿主受体结合促进病毒感染。它包括两个结构域，n 端结构域和 c 端 RBD 结构域，它们通过 s 蛋白与 ACE2 受体结合而感染人类。

1.2 编写并运行 example4-1.pl

编写的 example4-1.pl 如下

```
use strict;
use warnings;

my $DNA = 'ACGGGAGGGACGGGAAAATTACTACGGCATTAGC';
print $DNA;
exit;
```

图 1-1 example4-1.pl

1.3 SARS-CoV-2 的基因组序列

得到新冠病毒基因的 fasta 文件后，通过如下 perl 程序得到其互补序列。

```

use warnings;
use FileHandle;

my $sequence = "";
open(each_line, "C:/Users/Administrator/Desktop/Bioinformatic_report1/gene.txt");
while (<each_line>){
    my $line = $_;
    chomp($line); #如果末尾有换行符，则去掉。
    if ($line!~/^>/)
    { $sequence = $sequence.$line; }
$revcom = reverse $sequence; #将字符串倒置;
$revcom =~ tr/ACGTacgt/TGCATgca/; #配对

open(ln, ">C:/Users/Administrator/Desktop/Bioinformatic_report1/matched.txt"); #输出
print ln $revcom;

close(each_line);

```

图 1-2 gene_match.pl

1.4 SARS-CoV-2 潜在编码序列的预测

预测开放阅读框如下

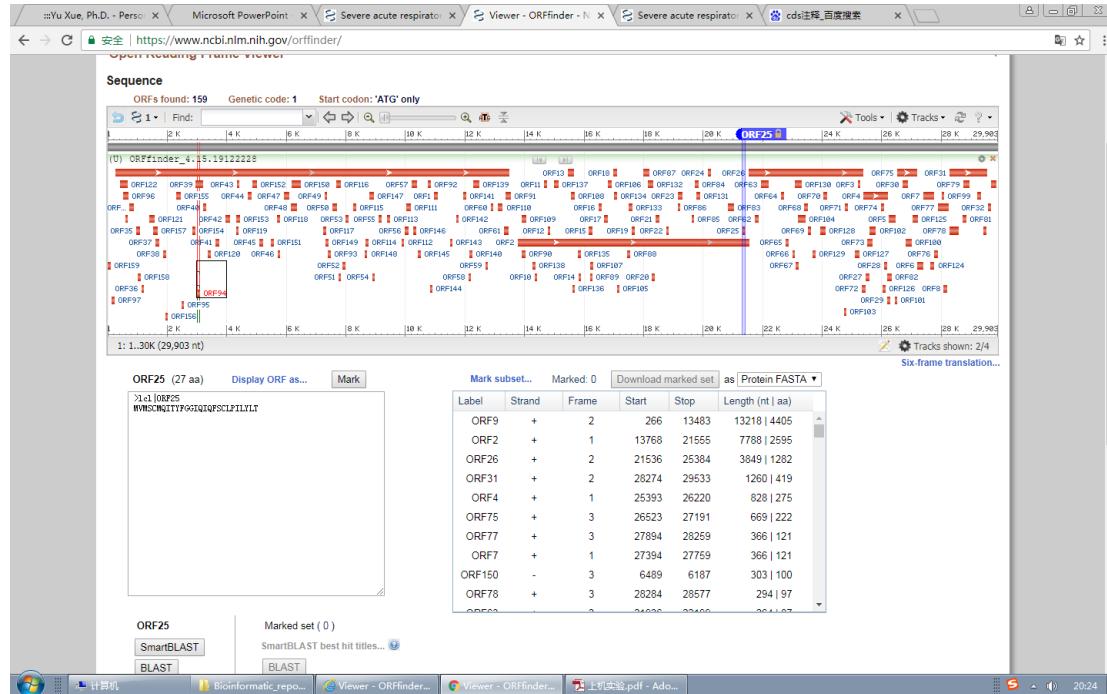


图 1-3 ORF_prediction.PNG

CDS 注释如下

```

CDS
  /locus_tag="GU280_gp01"
  /db_xref="GeneID:43740578"
  join(266..13468,13468..21555)
  /gene="ORF1ab"
  /locus_tag="GU280_gp01"
  /ribosomal_slippage
  /note="pp1ab; translated by -1 ribosomal frameshift"
  /codon_start=1
  /product="ORF1ab polyprotein"
  /protein_id="YP\_009724389.1"
  /db_xref="GeneID:43740578"
  /translation="MESLVPGFMEK THWQLSLPVLQVRDVLRGFGDSVVEVLSEARQ
HLKGDTGCLVEKEGVLPQLEQPYVFIKRSARTAPGHVMVELVAELEGIQYGRSGE
TLGVLVPHVGETPVAYRKVLLRKNGNKGAGGHSYGADLKSFDLGDELGTDPYEDFQEN
WNTKHSSGVTRELMRELNGGAYTRYVDNNFCGPDG YPLECIKDLLARAKGASCTLSEQ
LDFIDTKRGVYOCREHEHEIAWYTERSEK SYELQTPFEIKLAKKFDTFNGECPNFVFP
LNSIIKTIQPRWEKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHOGETSWQTG
DFWKATCEFOGTENLTKEGATTCGYLPQNAVVKIYCPACHNSEVGPEHSLAEHYNESG
I.KTTI.RKGGRTTARGGCWF SYVGCHNKCAYWVPRASANTGCNHTGVVGRGSREGI.NDNI."

```

图 1-4 CDS_annotation.PNG

通过比较预测的 ORF 与 CDS 注释可以发现，预测的 ORF 几乎占据了整段基因，但实际 CDS 仅为其中的一部分，说明基因并不是整段表达的。

1.5 发现与 SARS-CoV-2 同源的冠状病毒

MT461669.1	MT108784.1	HG994854.1	HG994852.1
HG994857.1	HG994855.1	MT461671.1	MT461670.1
MN996532.2	HG994858.1	HG994859.1	HG994856.1
HG994853.1	MT121216.1	MW703458.1	MT040335.1
MT040333.1	MT072864.1	MT040334.1	MT040336.1

表 1-1 20 个 SARS-CoV-2 的同源冠状病毒的序列号

1.6 插入片段分析

使用 megablast, dmegablast, blastn 搜索结果如下

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Synthetic csynthetic construct		2545	2545	100%	0	100	30857	MT461671.1
Synthetic csynthetic construct		2545	2545	100%	0	100	30347	MT461670.1
Synthetic csynthetic construct		2545	2545	100%	0	100	29903	MT461669.1
Synthetic csynthetic construct		2545	2545	100%	0	100	29891	MT108784.1
Severe acu Severe acute respiratory syndrome-related		2545	2545	100%	0	100	29903	HG994859.1
Severe acu Severe acute respiratory syndrome-related		2545	2545	100%	0	100	29903	HG994858.1
Severe acu Severe acute respiratory syndrome-related		2545	2545	100%	0	100	29903	HG994856.1
Severe acu Severe acute respiratory syndrome-related		2545	2545	100%	0	100	29903	HG994855.1
Severe acu Severe acute respiratory syndrome-related		2545	2545	100%	0	100	29903	HG994854.1
Severe acu Severe acute respiratory syndrome-related		2545	2545	100%	0	100	29901	HG994853.1
Severe acu Severe acute respiratory syndrome-related		2540	2540	100%	0	99.93	29903	HG994857.1
Severe acu Severe acute respiratory syndrome-related		2525	2525	100%	0	99.78	29900	HG994852.1
Synthetic csynthetic construct		2401	2401	100%	0	98.11	171907	MW036243.1
Synthetic csynthetic construct		2401	2401	100%	0	98.11	171918	MW030460.1
Bat corona Bat coronavirus RaTG13		1857	1857	95%	0	92.22	29855	MN996532.2
Cloning ve Cloning vector pSF_lenti_SARS-CoV-2_part		1391	1391	54%	0	99.87	13543	MT299805.1
Synthetic csynthetic construct		1391	1391	54%	0	99.87	7558	MW059035.1
Cloning ve Cloning vector pSF_lenti_SARS-CoV-2_part		1328	1328	52%	0	99.72	13507	MT299804.1
Synthetic csynthetic construct		1264	1264	50%	0	99.71	7575	MW059034.1
Severe acu Severe acute respiratory syndrome-related		1160	2144	84%	0	100	29903	HG994860.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	29521	MT121216.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	3798	MT799526.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	3798	MT799525.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	3798	MT799524.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	3798	MT799523.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	3798	MT799522.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	3798	MT799521.1
Pangolin c Pangolin coronavirus		737	737	50%	0	85.86	27213	MT084071.1
Severe acu Severe acute respiratory syndrome-related		307	307	35%	1.00E-78	78.07	29718	LC556375.1

图 1-5 megablast_result.PNG

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Synthetic csynthetic construct		2545	2545	100%	0	100	30857	MT461671.1
Synthetic csynthetic construct		2545	2545	100%	0	100	30347	MT461670.1
Synthetic csynthetic construct		2545	2545	100%	0	100	29903	MT461669.1
Synthetic csynthetic construct		2545	2545	100%	0	100	29891	MT108784.1
Severe acu Severe acute respiratory syndrome-related corona		2545	2545	100%	0	100	29903	HG994859.1
Severe acu Severe acute respiratory syndrome-related corona		2545	2545	100%	0	100	29903	HG994858.1
Severe acu Severe acute respiratory syndrome-related corona		2545	2545	100%	0	100	29903	HG994856.1
Severe acu Severe acute respiratory syndrome-related corona		2545	2545	100%	0	100	29903	HG994855.1
Severe acu Severe acute respiratory syndrome-related corona		2545	2545	100%	0	100	29903	HG994854.1
Severe acu Severe acute respiratory syndrome-related corona		2545	2545	100%	0	100	29901	HG994853.1
Severe acu Severe acute respiratory syndrome-related corona		2540	2540	100%	0	99.93	29903	HG994857.1
Severe acu Severe acute respiratory syndrome-related corona		2525	2525	100%	0	99.78	29900	HG994852.1
Synthetic csynthetic construct		2401	2401	100%	0	98.11	171907	MW036243.1
Synthetic csynthetic construct		2401	2401	100%	0	98.11	171918	MW030460.1
Bat corona Bat coronavirus RaTG13		1857	1857	95%	0	92.22	29855	MN996532.2
Cloning ve Cloning vector pSF_lenti_SARS-CoV-2_partial-S/E/T		1391	1391	54%	0	99.87	13543	MT299805.1
Synthetic csynthetic construct		1391	1391	54%	0	99.87	7558	MW059035.1
Cloning ve Cloning vector pSF_lenti_SARS-CoV-2_partial-ORF		1328	1328	52%	0	99.72	13507	MT299804.1
Synthetic csynthetic construct		1264	1264	50%	0	99.71	7575	MW059034.1
Severe acu Severe acute respiratory syndrome-related corona		1160	2144	84%	0	100	29903	HG994860.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	29521	MT121216.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	3798	MT799526.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	3798	MT799525.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	3798	MT799524.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	3798	MT799523.1
Pangolin c Pangolin coronavirus		749	749	53%	0	84.87	3798	MT799522.1
Pangolin c Pangolin coronavirus		737	737	50%	0	85.86	27213	MT084071.1
Severe acu Severe acute respiratory syndrome-related corona		307	307	35%	1.00E-78	78.07	29718	LC556375.1

图 1-6 dimegablast_result.PNG

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. ident	Acc. Len	Accession
Synthetic csynthetic construct		2486	2486	100%	0	100	30857	MT461671.1
Synthetic csynthetic construct		2486	2486	100%	0	100	30347	MT461670.1
Synthetic csynthetic construct		2486	2486	100%	0	100	29903	MT461669.1
Synthetic csynthetic construct		2486	2486	100%	0	100	29891	MT108784.1
Severe acI Severe acute respiratory syndro		2486	2486	100%	0	100	29903	HG994859.1
Severe acI Severe acute respiratory syndro		2486	2486	100%	0	100	29903	HG994858.1
Severe acI Severe acute respiratory syndro		2486	2486	100%	0	100	29903	HG994856.1
Severe acI Severe acute respiratory syndro		2486	2486	100%	0	100	29903	HG994855.1
Severe acI Severe acute respiratory syndro		2486	2486	100%	0	100	29903	HG994854.1
Severe acI Severe acute respiratory syndro		2486	2486	100%	0	100	29901	HG994853.1
Severe acI Severe acute respiratory syndro		2481	2481	100%	0	99.93	29903	HG994857.1
Severe acI Severe acute respiratory syndro		2470	2470	100%	0	99.78	29900	HG994852.1
Synthetic csynthetic construct		2369	2369	100%	0	98.11	171907	MW036243.1
Synthetic csynthetic construct		2369	2369	100%	0	98.11	171918	MW030460.1
Bat coronaviru	Bat coronavirus RaTG13	1919	1919	99%	0	90.94	29855	MN996532.2
Cloning ve	Cloning vector pSF_lenti_SARS-CoV	1360	1360	54%	0	99.87	13543	MT299805.1
Synthetic csynthetic construct		1360	1360	54%	0	99.87	7558	MW059035.1
Cloning ve	Cloning vector pSF_lenti_SARS-CoV	1299	1299	52%	0	99.72	13507	MT299804.1
Synthetic csynthetic construct		1236	1236	50%	0	99.71	7575	MW059034.1
Pangolin c	Pangolin coronavirus	1233	1233	99%	0	80.09	29805	MT040333.1
Pangolin c	Pangolin coronavirus	1229	1229	99%	0	80.01	29806	MT040335.1
Pangolin c	Pangolin coronavirus	1224	1224	99%	0	79.94	29802	MT040336.1
Pangolin c	Pangolin coronavirus	1223	1223	99%	0	79.9	29795	MT072864.1
Pangolin c	Pangolin coronavirus	1216	1216	99%	0	79.8	29801	MT040334.1
Severe acI Severe acute respiratory syndro		1133	2094	84%	0	100	29903	HG994860.1
Pangolin c	Pangolin coronavirus	967	1098	87%	0	80.35	29801	MT072865.1
Pangolin c	Pangolin coronavirus	952	952	93%	0	76.65	29521	MT121216.1
Pangolin c	Pangolin coronavirus	952	952	93%	0	76.65	3798	MT799526.1
Pangolin c	Pangolin coronavirus	952	952	93%	0	76.65	3798	MT799525.1

图 1-7 blastn_result.PNG

可以看出 megablast 的结果较少，但使用 dmegablast 与 blastn 允许错配后，可用结果变多，故该基因在冠状病毒中保守，不为人工插入序列。

2 序列分析

2.1 INS1378 与 pShuttle-SN 载体的相似性

Global Align 与 EMBOSS Water 比对结果如下

INS1378

Sequence ID: **Query_60179** Length: **1378** Number of Matches: **1**

Range 1: 1 to 1378 Graphics				▼ Next Match
NW Score	Identities	Gaps	Strand	
-7859	1106/5609(20%)	4232/5609(75%)	Plus/Plus	
Query 1	TAAC TATAAC CGGT CCTAAG GTAG CGAA AGCT CAGA TCTGG ATCT CCCG ATCCC CTAT GG	59		
Sbjct 1	CTCAG TTTAC ATTC --AA-----CTCAG GACTT GTTCT----TACCTT-----	38		
Query 60	TCG ACTCT CAGT ACAAT CTG CTGAT GCCG CATAG TTAA GGCAG TATCT GCTCC CTGCT	119		
Sbjct 39	-----TCTTTCCAAT GTTACTT GGTT---CCA----TGCTA--TACA	72		
Query 120	TGTGTGTTGGAGGTCGCTGAGTAGTGCGCGAGCAAAATTAAAGCTACAACAAGGCAAGGC	179		
Sbjct 73	TGTCTCTGGGA-----	83		
Query 180	TTGACCGACAATTGCA TGAAGAATCTGCTTAGGGTTAGGC GTTTGCGCTGCTCGCGAT	239		
Sbjct 84	---CC---AATGGTACTAAGA-----GGTTG-----	104		
Query 240	GTACGGGCCAGATATA CGC GTTGAC ATTGATT ATTGACTAGTT ATTAA TAGTA ATCA ATT	299		
Sbjct 105	-----ATA-ACCC-TGTCCT-----ACCATT TAATGATGGT-----	133		
Query 300	ACGGGGT CATTAGTT CATAG CCCATATATGGAGTTCCG GTTACATAACTTACGGTAAAT	359		
Sbjct 134	---GTTTATT--TT---GC-----TTCC-----ACTGA-----	153		
Query 360	GGCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTT	419		
Sbjct 154	-----GAAGTC--TAA-----	162		

图 2-1 global_align.PNG

Aligned_sequences: 2

1: AY862402.1

2: INS1378

Matrix: EBLOSUM62

Gap_penalty: 10.0

Extend_penalty: 0.5

Length: 1387

Identity: 916/1387 (66.0%)

Similarity: 916/1387 (66.0%)

Gaps: 71/1387 (5.1%)

Score: 4577.5

图 2-2 EMBOSS_Water.PNG

Global Align 使用的是 Needleman-Wunsh 算法，出现负分，且结果序列 gap 较多；EMBOSS Water 使用 Smith-Waterman 算法，出现负分则记为零分，结果序列 gap 较少。

2.2 SARS-CoV-2 的蛋白质序列

NC_045512 的 CDS 注释如下，通过 ncbi 下载其对应的 fasta 文件

CDS	join(266..13468,13468..21555) /gene="ORF1ab" /locus_tag="GU280_gp01" /ribosomal_slippage /note="pp1ab; translated by -1 ribosomal frameshift" /codon_start=1 /product="ORF1ab polyprotein" /protein_id="YP_009724389.1" /db_xref="GenelD:43740578"	CDS	25393..26220 /gene="ORF3a" /locus_tag="GU280_gp03" /codon_start=1 /product="ORF3a protein" /protein_id="YP_009724391.1" /db_xref="GenelD:43740569"
CDS	266..13483 /gene="ORF1ab" /locus_tag="GU280_gp01" /note="pp1a" /codon_start=1 /product="ORF1a polyprotein" /protein_id="YP_009725295.1" /db_xref="GenelD:43740578"	CDS	26245..26472 /gene="E" /locus_tag="GU280_gp04" /note="ORF4; structural protein; E protein" /codon_start=1 /product="envelope protein" /protein_id="YP_009724392.1" /db_xref="GenelD:43740570"
CDS	21563..25384 /gene="S" /locus_tag="GU280_gp02" /gene_synonym="spike glycoprotein" /note="structural protein; spike protein" /codon_start=1 /product="surface glycoprotein" /protein_id="YP_009724390.1" /db_xref="GenelD:43740568"	CDS	26523..27191 /gene="M" /locus_tag="GU280_gp05" /note="ORF5; structural protein" /codon_start=1 /product="membrane glycoprotein" /protein_id="YP_009724393.1" /db_xref="GenelD:43740571"

图 2-3 cds_list_1.PNG

CDS	27202..27387 /gene="ORF6" /locus_tag="GU280_gp06" /codon_start=1 /product="ORF6 protein" /protein_id="YP_009724394.1" /db_xref="GenelD:43740572"	CDS	27894..28259 /gene="ORF8" /locus_tag="GU280_gp09" /codon_start=1 /product="ORF8 protein" /protein_id="YP_009724396.1" /db_xref="GenelD:43740577"
CDS	27394..27759 /gene="ORF7a" /locus_tag="GU280_gp07" /codon_start=1 /product="ORF7a protein" /protein_id="YP_009724395.1" /db_xref="GenelD:43740573"	CDS	28274..29533 /gene="N" /locus_tag="GU280_gp10" /note="ORF9; structural protein" /codon_start=1 /product="nucleocapsid phosphoprotein" /protein_id="YP_009724397.2" /db_xref="GenelD:43740575"
CDS	27756..27887 /gene="ORF7b" /locus_tag="GU280_gp08" /codon_start=1 /product="ORF7b" /protein_id="YP_009725318.1" /db_xref="GenelD:43740574"	CDS	29558..29674 /gene="ORF10" /locus_tag="GU280_gp11" /codon_start=1 /product="ORF10 protein" /protein_id="YP_009725255.1" /db_xref="GenelD:43740576"

图 2-4 cds_list_2.PNG

2.3 等电点与分子量分析

序号	等电点 pI	分子量 Mw
CDS_1	6.32	794057.79
CDS_2	6.04	489988.91
CDS_3	6.24	141178.47
CDS_4	5.55	31122.94
CDS_5	8.57	8365.04
CDS_6	9.51	25146.62
CDS_7	4.60	7272.54
CDS_8	8.23	13744.17
CDS_9	4.17	5180.27
CDS_10	5.42	13831.01
CDS_11	10.07	45625.70
CDS_12	7.93	4449.23

表 2-1 编码蛋白等电点与分子量

2.4 功能结构域分析

CDS_1 功能结构域如下：

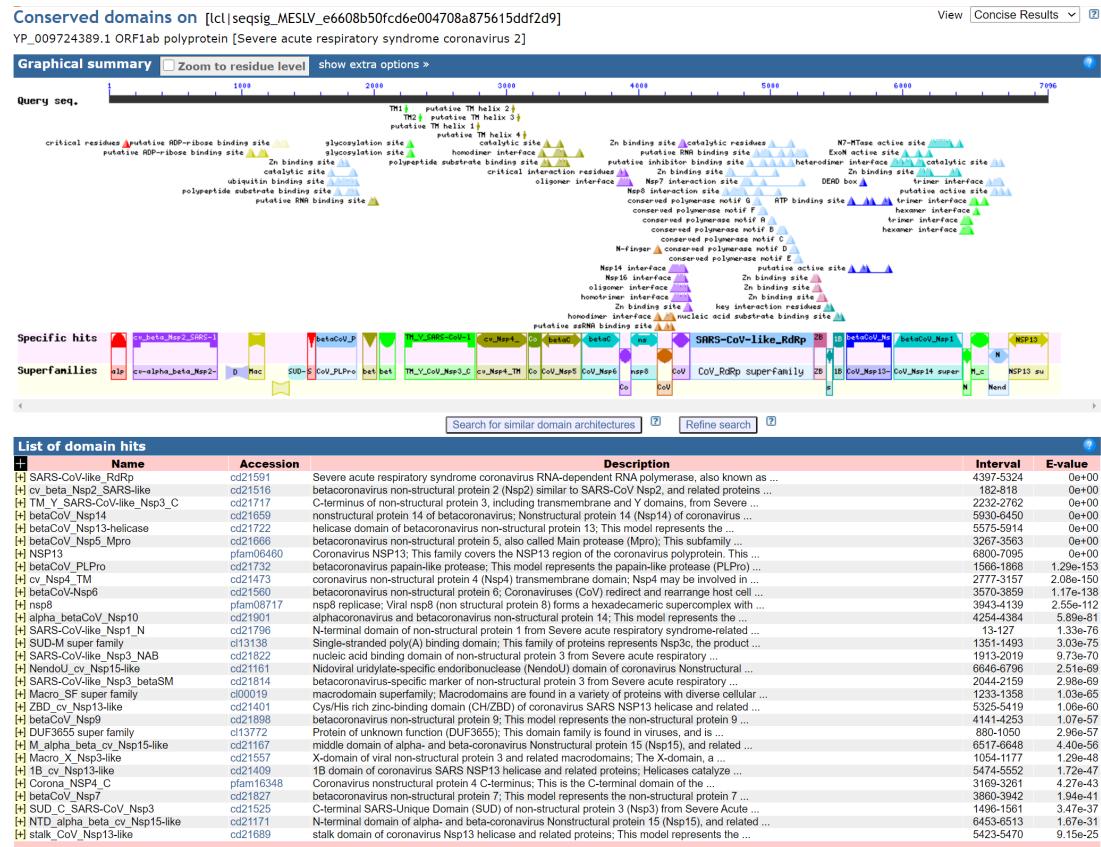


图 2-5 cds_1_cd_search.PNG

EMBL-EBI 

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Pfam
keyword search 

Protein: R1AB_SARS2 (PODTD1)

Summary

Sequence
Structures
TreeFam

Jump to... 
enter ID/acc 

This is the summary of UniProt entry **R1AB_SARS2# (PODTD1)**.

Descriptions: Replicase polyprotein 1ab EC=3.4.22.- EC=2.7.7.48 EC=3.6.4.12 EC=3.6.4.13 EC=3.1.13.- EC=3.1.-, EC=2.1.1.-
Source organism: Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)
Length: 7096 amino acids

Reference Proteomes: ✓

Please note: when we start each new Pfam data release, we take a copy of the UniProt sequence database. This snapshot of UniProt forms the basis of the overview that you see here. It is important to note that, although some UniProt entries may be removed after a Pfam release, these entries will not be removed from Pfam until the next Pfam data release.

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will take you to the page describing that Pfam entry. The table below gives the domain boundaries for each of the domains. [More...](#)



Download the data used to generate the domain graphic in JSON format.

Source	Domain	Start	End
Pfam	bCoV_NSP1	7	144
disorder	n/a	156	157
disorder	n/a	160	161
disorder	n/a	168	169
Pfam	CoV_NSP2_N	182	423
Pfam	CoV_NSP2_C	652	818
low_complexity	n/a	742	760
Pfam	bCoV_NSP2_N	880	1051
disorder	n/a	926	956
low_complexity	n/a	926	943
disorder	n/a	960	1008
low_complexity	n/a	970	986
low_complexity	n/a	998	1005
Pfam	Macro	1058	1165
disorder	n/a	1077	1078
disorder	n/a	1217	1220
disorder	n/a	1228	1227
low_complexity	n/a	1237	1255
Pfam	bCoV_SUD_M	1351	1493
Pfam	bCoV_SUD_C	1494	1561
Pfam	CoV_endonuclease	1564	1882
Pfam	bCoV_NAK	1899	2019
low_complexity	n/a	2172	2183
transmembrane	n/a	2229	2253
Pfam	CoV_NSP2_C	2260	2749
transmembrane	n/a	2328	2351
transmembrane	n/a	2363	2387
low_complexity	n/a	2546	2558
transmembrane	n/a	2773	2790
Pfam	CoV_NSP4_N	2788	3142
transmembrane	n/a	3043	3065
low_complexity	n/a	3043	3057
transmembrane	n/a	3077	3098
transmembrane	n/a	3104	3124
transmembrane	n/a	3136	3154
Pfam	CoV_NSP4_C	3166	3262
transmembrane	n/a	3166	3185
Pfam	Peptidase_C3D	3292	3582
transmembrane	n/a	3582	3605
low_complexity	n/a	3583	3599
Pfam	CoV_NSP6	3597	3859
transmembrane	n/a	3636	3656
transmembrane	n/a	3683	3697
transmembrane	n/a	3708	3725
transmembrane	n/a	3731	3755
transmembrane	n/a	3767	3789
Pfam	CoV_NSP7	3860	3942
low_complexity	n/a	3869	3885
low_complexity	n/a	3912	3920
Pfam	CoV_NSP8	3943	4139
low_complexity	n/a	3973	3989
low_complexity	n/a	4040	4052
Pfam	CoV_NSP9	4141	4253
Pfam	CoV_NSP10	4262	4384
Pfam	CoV_RP01_N	4406	4758
Pfam	RdBP_1	4858	5195
Pfam	Viral_helicase1	5534	5900
Pfam	CoV_ExoN	5928	6450
disorder	n/a	6056	6059
disorder	n/a	6061	6062
Pfam	CoV_NSP15_N	6453	6513
Pfam	CoV_NSP15_M	6514	6637
Pfam	CoV_NSP15_C	6643	6796
Pfam	CoV_Methyltr_2	6800	7096

Show or hide domain scores.

 Pfam is part of the **ELIXIR** infrastructure
Pfam is an Elixir service [Read more](#)

图 2-6 cds_1_Pfam.PNG

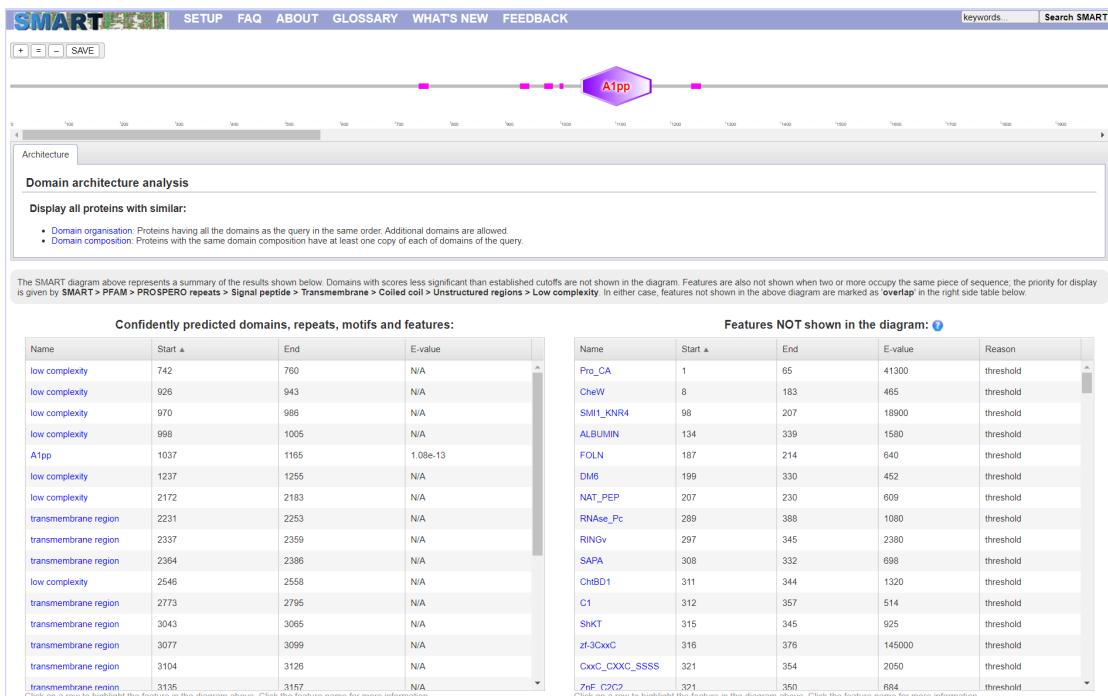
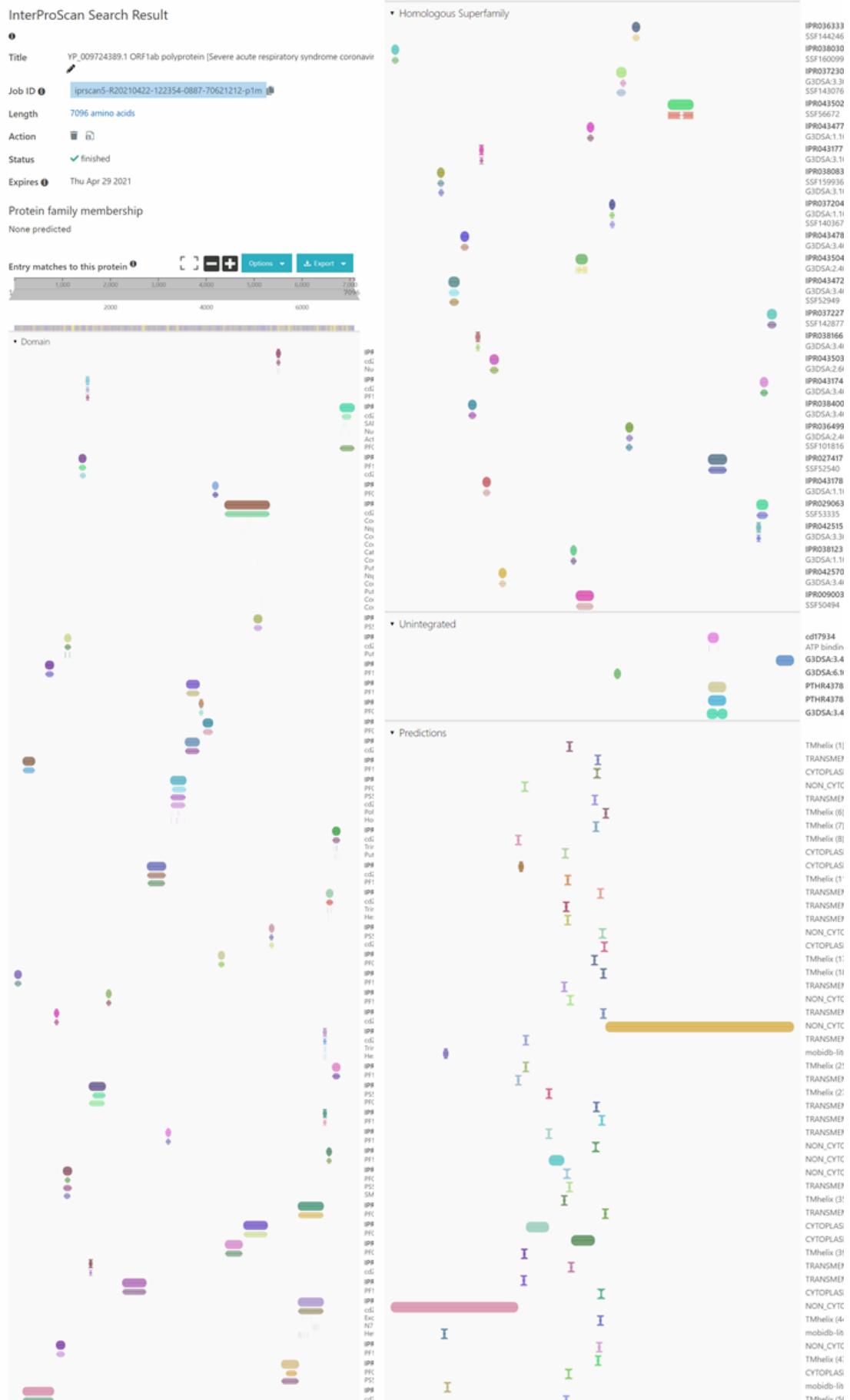


图 2-7 cds_1_SMART.PNG



CDS 2 功能结构域如下：

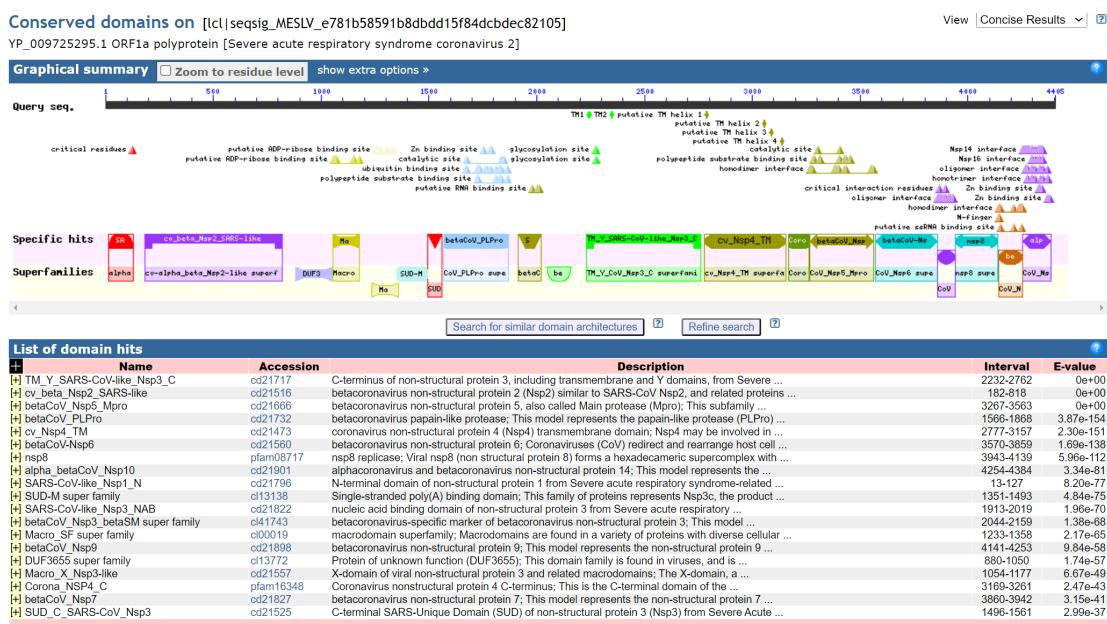


图 2-9 cds 2 cd search.PNG

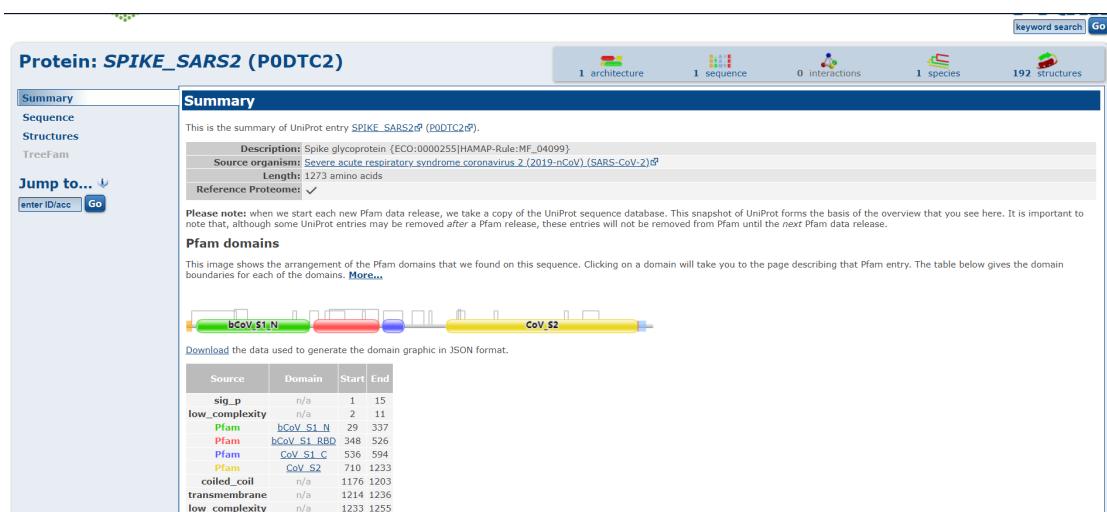


图 2-10 cds 2 Pfam.PNG

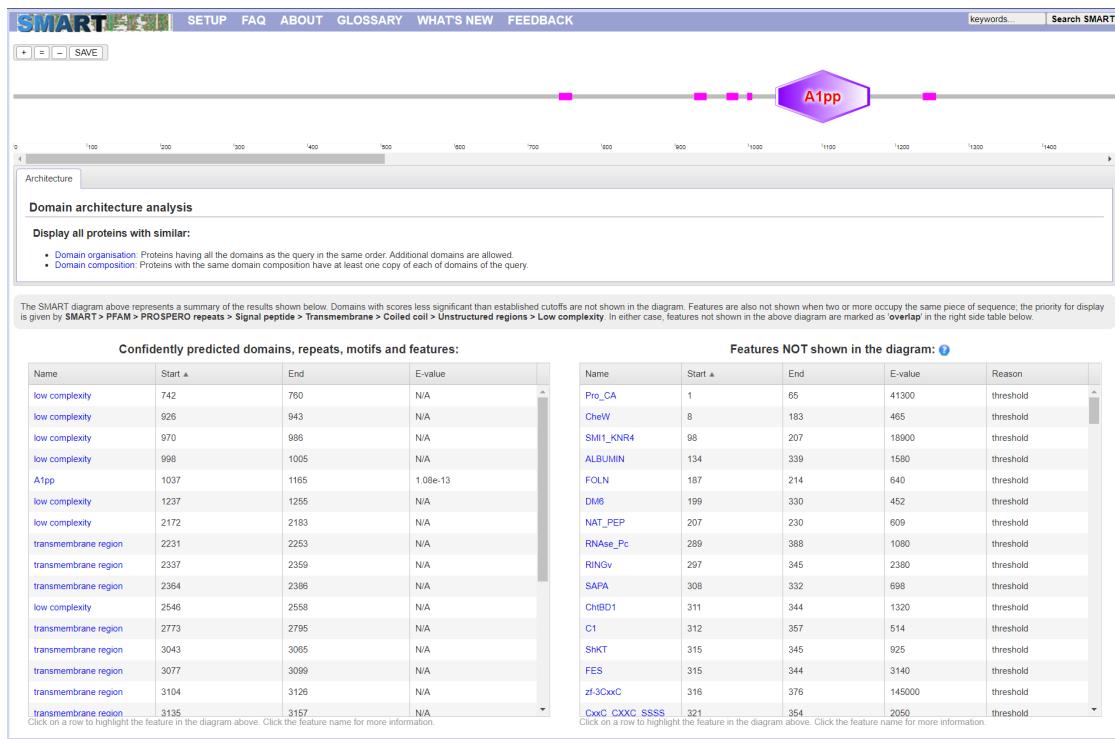


图 2-11 cds 2 SMART.PNG

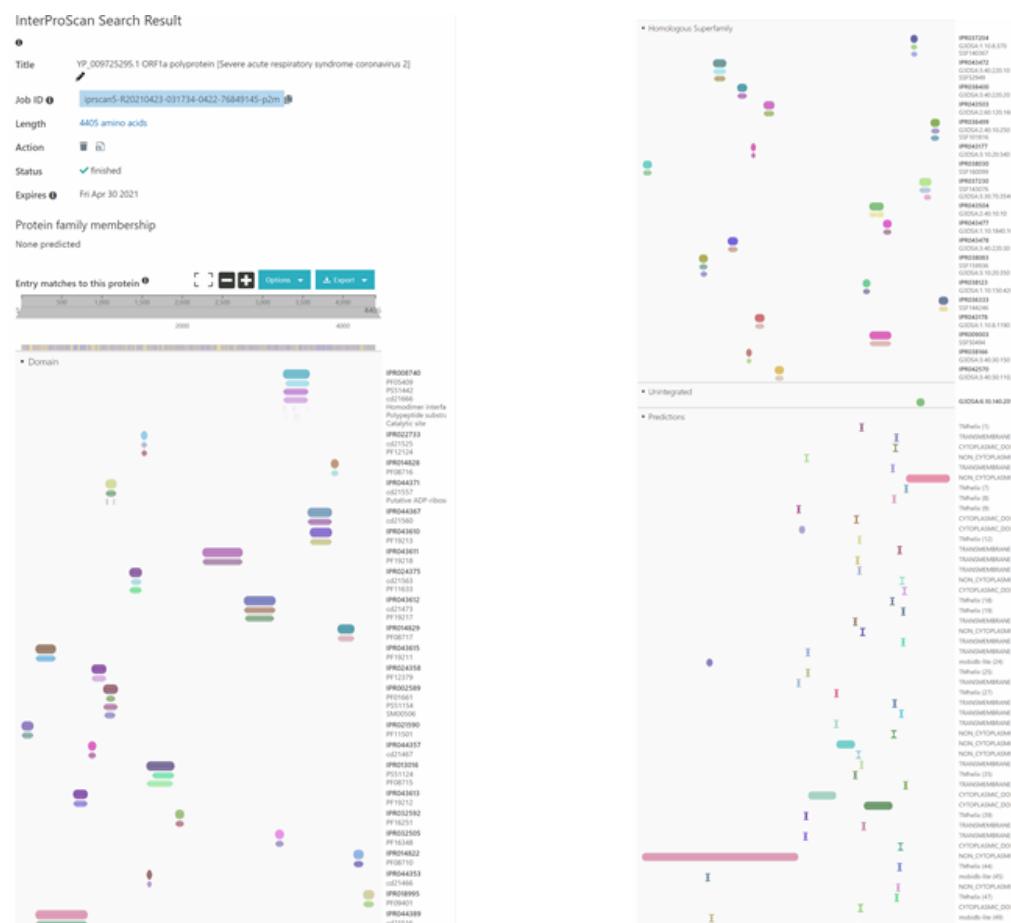


图 2-12 cds_2_InterPro.PNG

CDS_3 功能结构域如下：

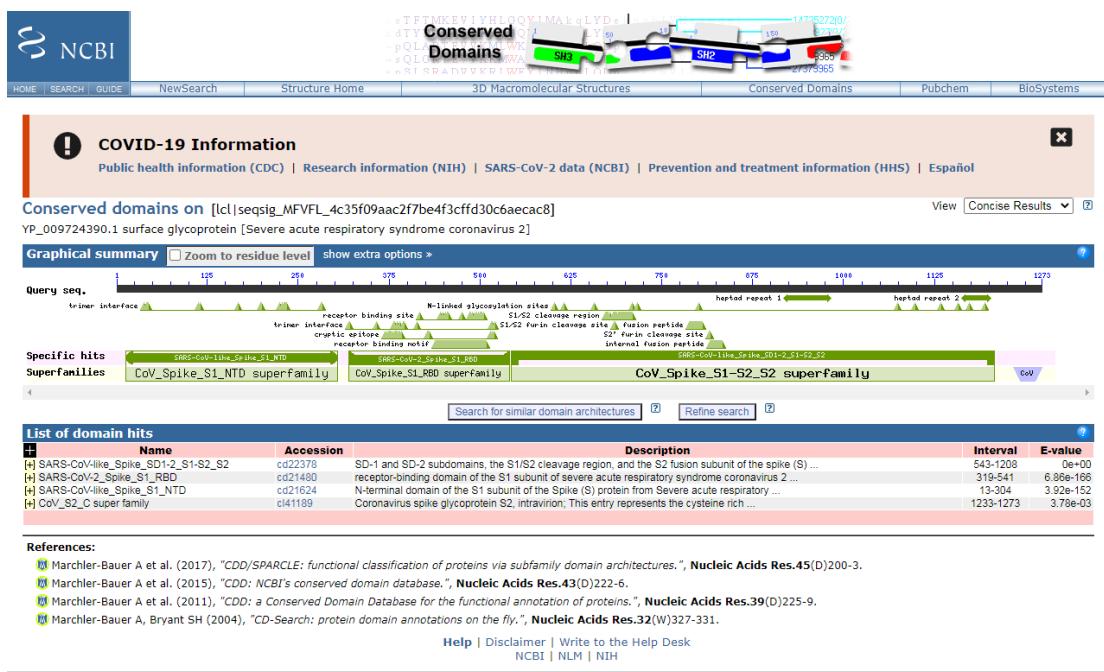


图 2-13 cds_3_cd_search.PNG

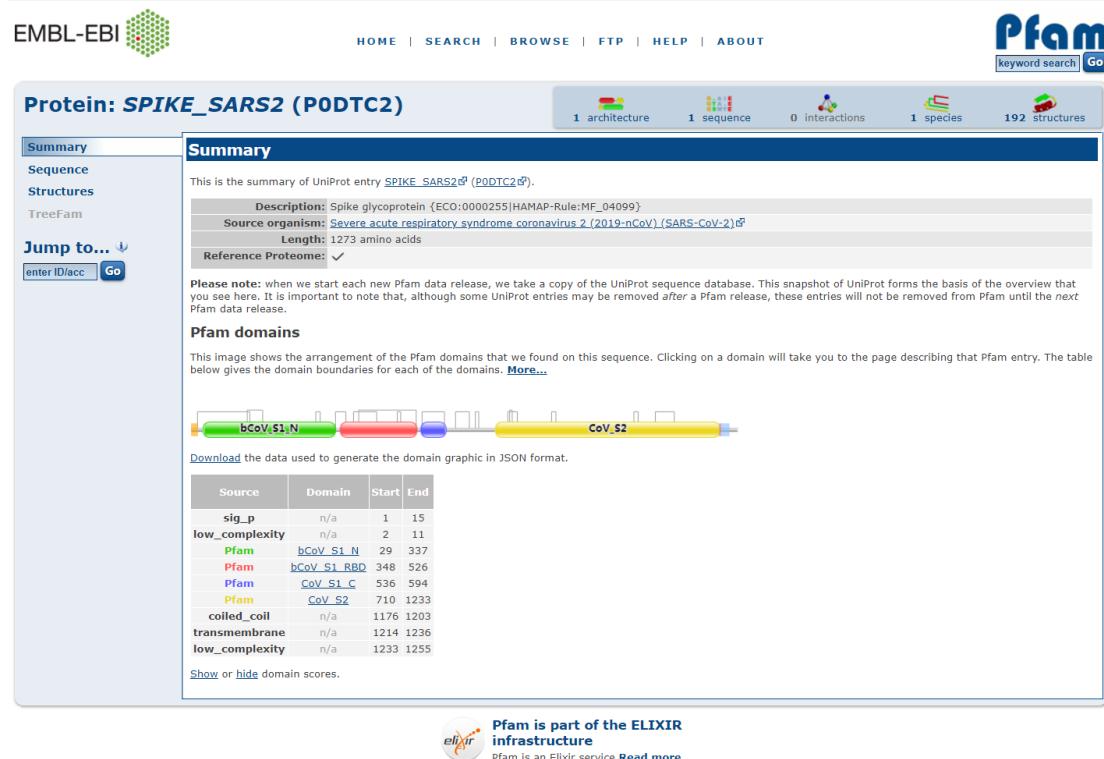


图 2-14 cds_3_Pfam.PNG

Confidently predicted domains, repeats, motifs and features:				Features NOT shown in the diagram: ⓘ				
Name	Start ▲	End	E-value	Name	Start ▲	End	E-value	Reason
low complexity	2	11	N/A	DCX	184	255	1240	threshold
coiled coil	1168	1205	N/A	THN	258	381	814	threshold
transmembrane region	1214	1236	N/A	DBC1	266	364	109000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	PSI	290	337	1980	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	ChtBD2	299	363	1510	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	s48_45	332	440	118000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	Elicitin	334	415	174000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	ChtBD3	350	395	1440	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	DM13	519	605	2500	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	Malic_M	522	692	103000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	Int_alpha	561	624	1230	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	HTH_ARSR	571	655	935	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	GHB	709	761	844	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	WR1	725	762	424	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	X8	733	782	4620	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	MUTsd	754	1054	2380	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	HSF	755	847	1500	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	Spc7	773	1043	36300	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	GLUCA	828	854	224	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	AICART_IIMPCHas	835	1029	37200	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	ILWEQ	838	1010	1050	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	PKS_MT	861	1029	57300	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	M16C_associated	864	1040	93100	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	MA	886	1048	711	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	Fib_alpha	882	1012	99600	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	IL6	886	1028	323	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	OmpH	890	1041	101000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	WIF	893	1021	1410	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	Drf_GBD	894	1029	78700	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	MHC_II_beta	901	971	119000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	HBM	903	1209	22600	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	SPEC	904	1005	452	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	BURP	908	1106	109000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	SATase_N	911	1006	91300	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	AD	914	987	112000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	LAMTOR	916	992	9010	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	SynN	919	1018	1320	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	ZM	921	941	417	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	BBC	923	1027	91.5	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	MIF4G	924	1204	2030	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	t_SNARE	926	994	542	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	Ifabd	926	1022	587	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	SMC_hinge	928	1019	111000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	HAMP	929	989	557	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	L27	937	998	1010	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	Dak2	938	1044	20600	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	BAG	940	1012	1230	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	SANT	947	990	1140	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	PBPb	958	1192	870	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	IBN_N	958	1014	114000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	ING	962	1047	11400	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	A2ML_recep	968	1030	55900	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	G_gamma	975	1032	199000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	Sm	978	1070	1670	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	Matrillin_ccoil	979	1021	191000	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	KNOX2	1135	1171	39200	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	BRIGHT	1143	1212	749	threshold
Click on a row to highlight the feature in the diagram above. Click the feature name for more information.								
low complexity	1233	1255	N/A	low complexity	1233	1255	N/A	overlap

图 2-15 cds_3_SMART.PNG

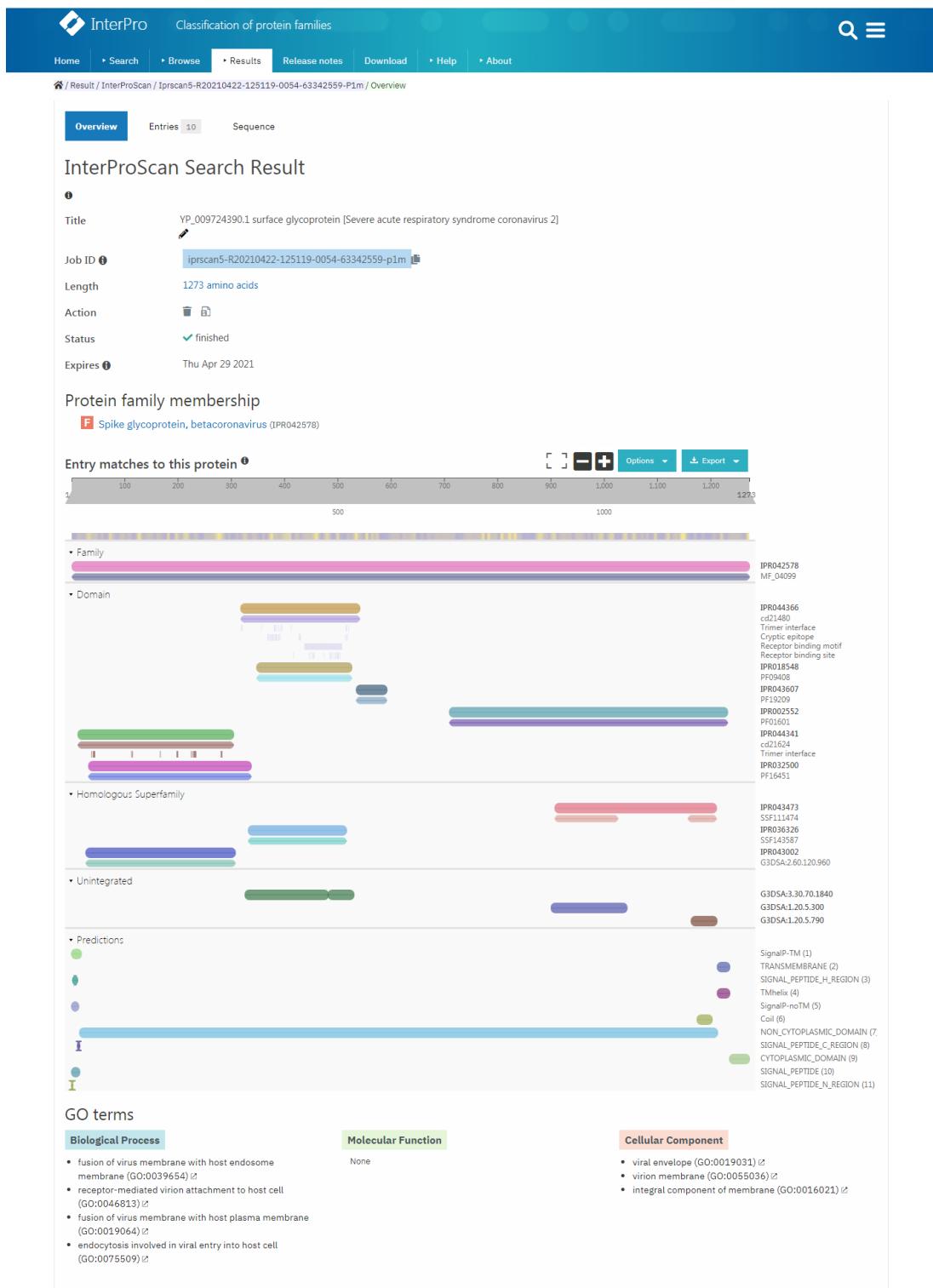


图 2-16 cds_3_InterPro.PNG

CDS_4 功能结构域如下：

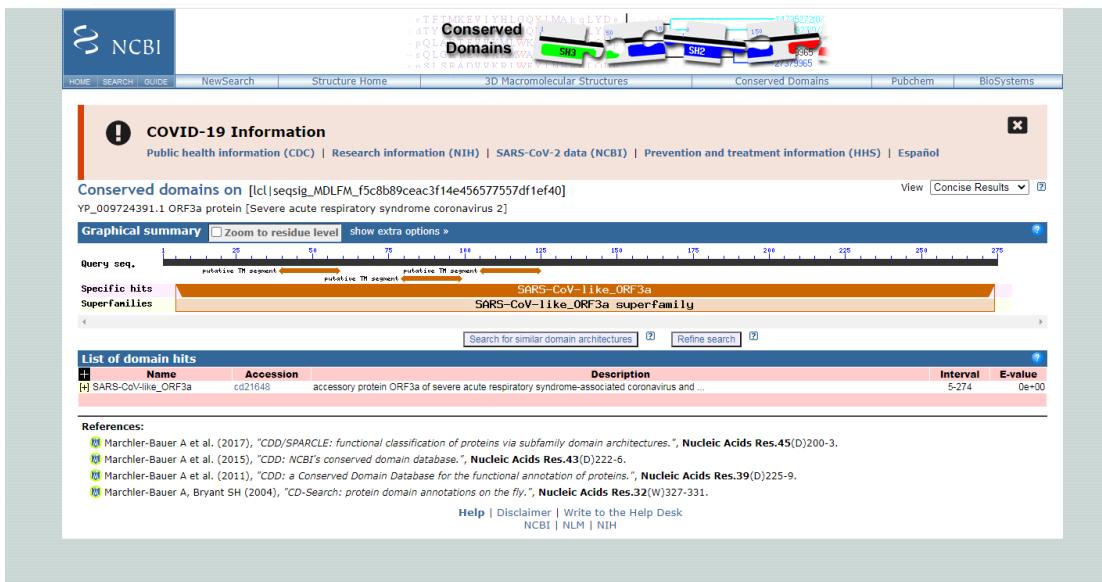


图 2-17 cds_4_cd_search.PNG

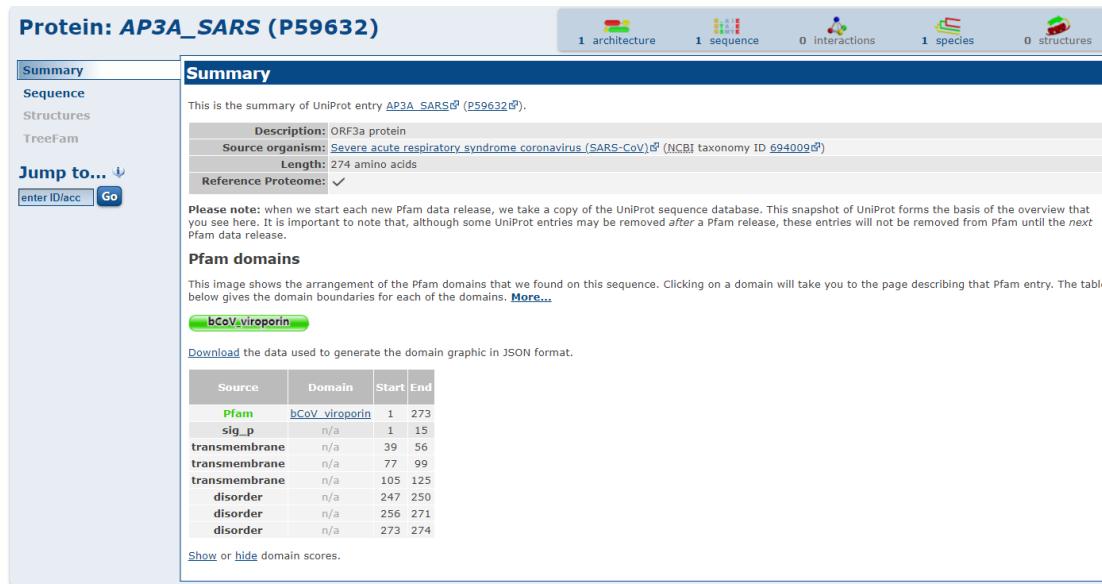


图 2-18 cds_4_Pfam.PNG

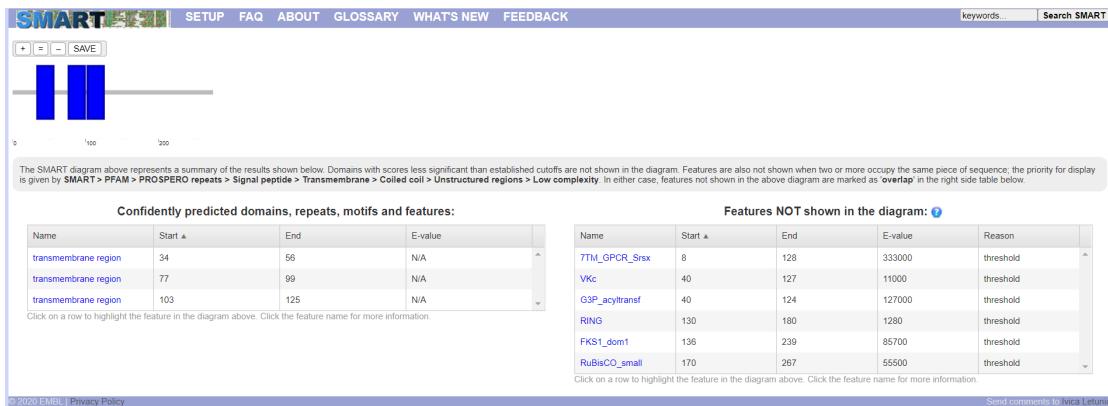


图 2-19 cds_4 SMART.PNG

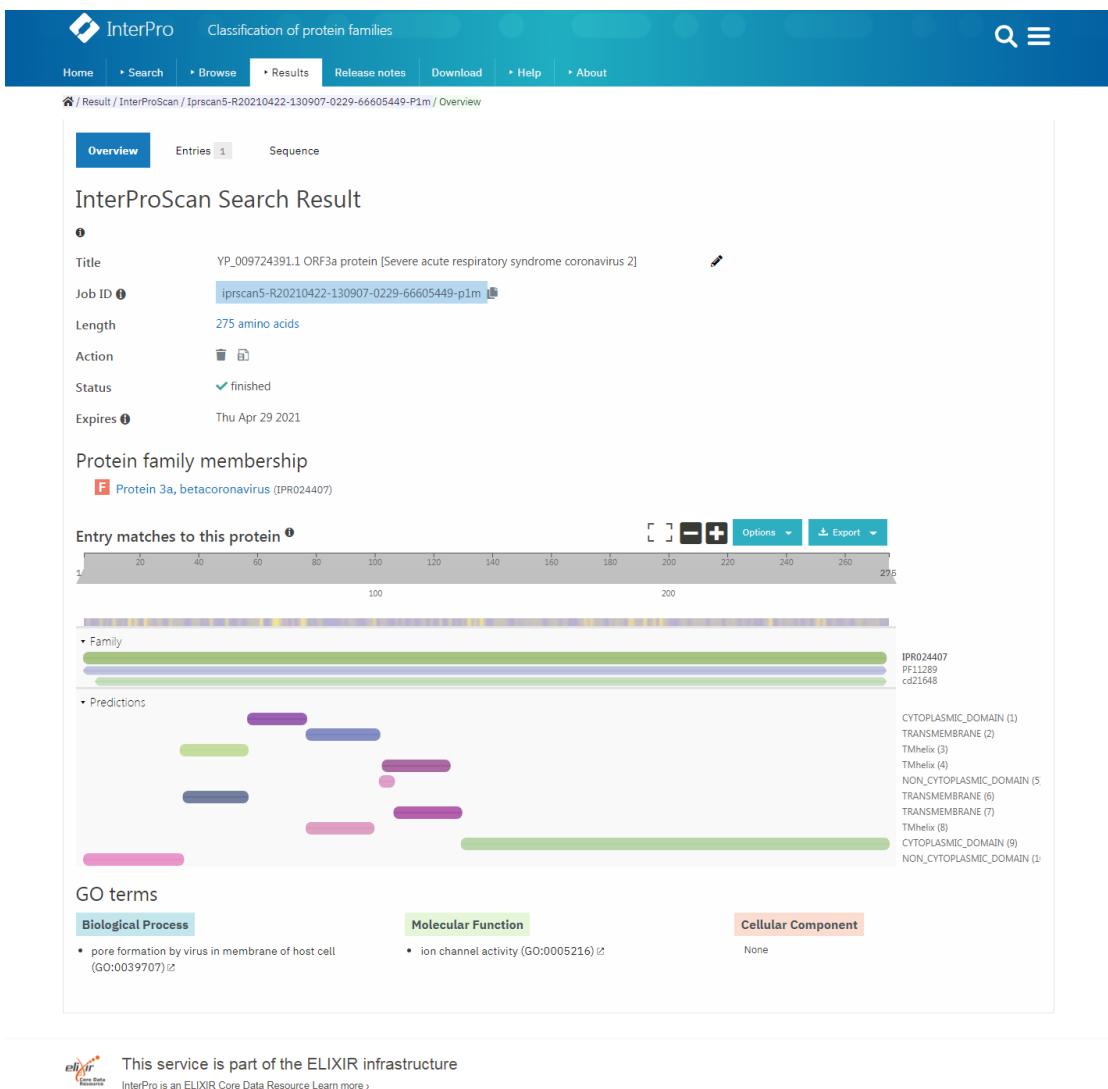


图 2-20 cds_4_InterPro.PNG

CDS_5 功能结构域如下：

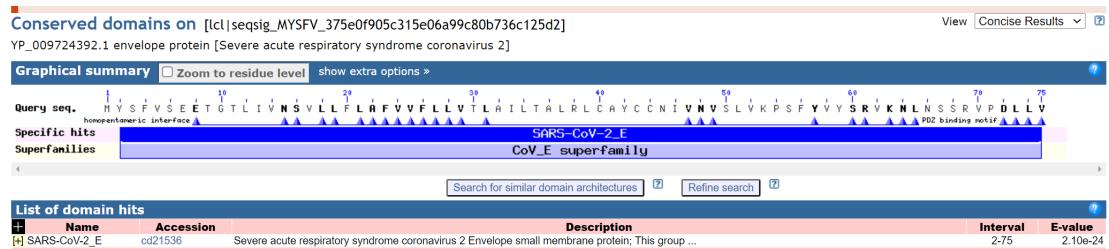


图 2-21 cds_5_cd_search.PNG

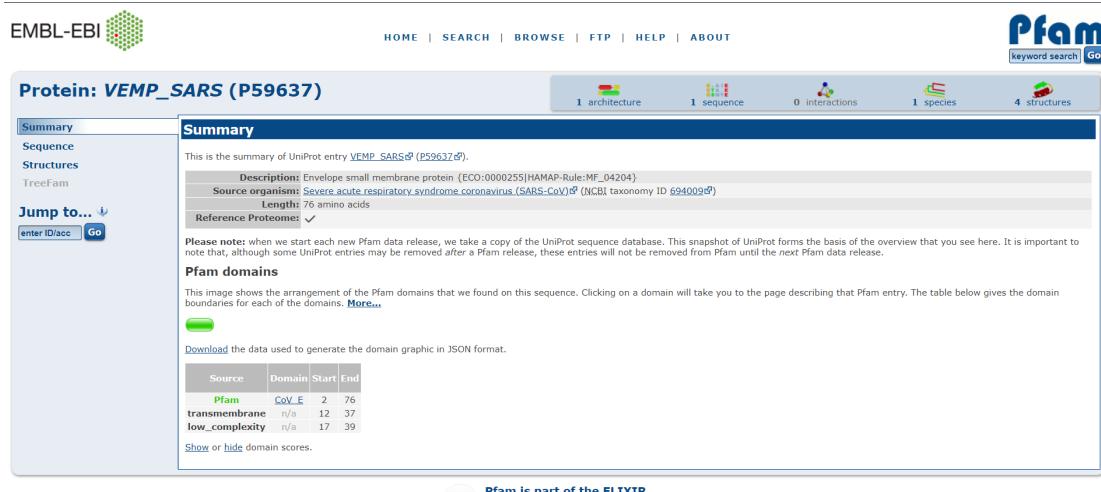


图 2-22 cds_5_Pfam.PNG

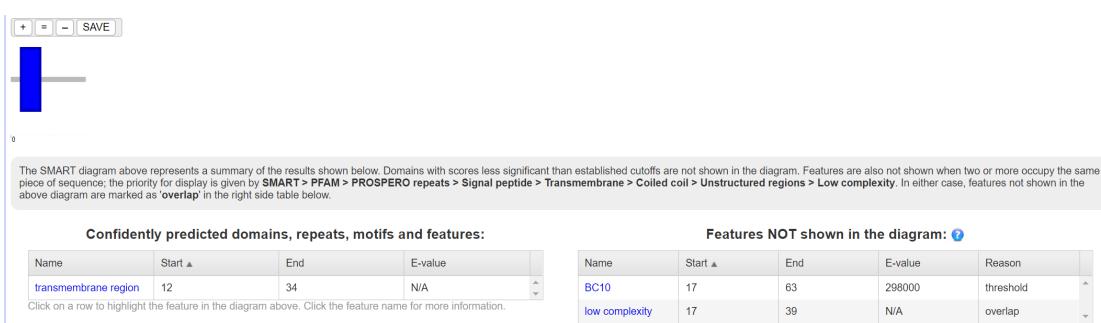


图 2-23 cds_5_SMART.PNG

InterProScan Search Result



Title YP_009724392.1 envelope protein [Severe acute respiratory syndrome coronavirus 2]

Job ID [iprscan5-R20210422-135236-0253-92852798-p1m](#)

Length 75 amino acids

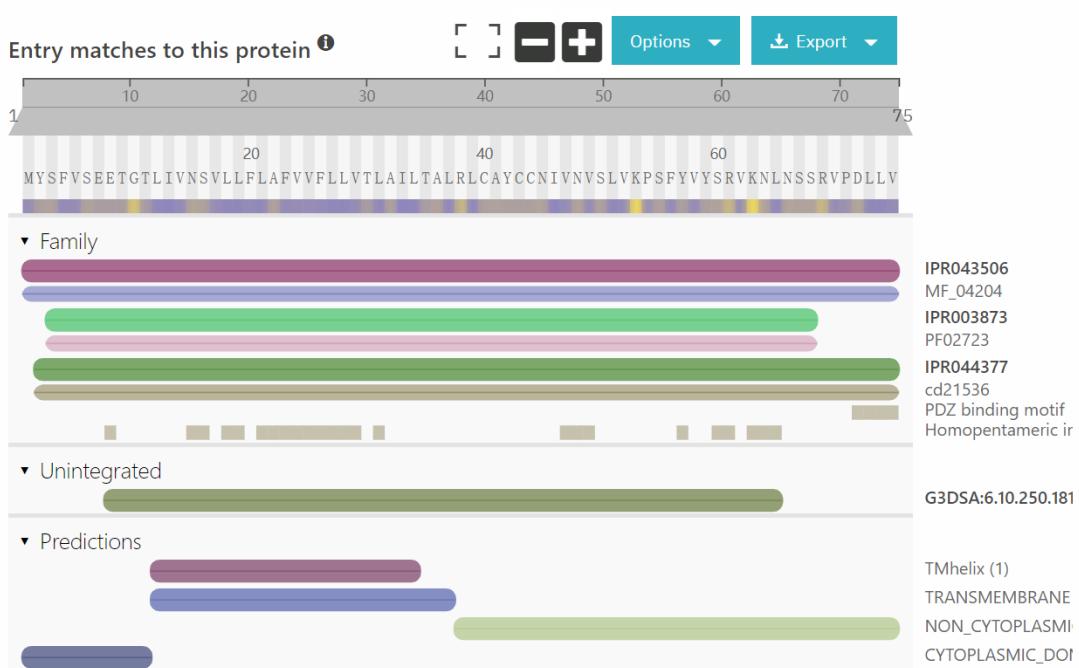


Status ✓ finished

Expires ⓘ Thu Apr 29 2021

Protein family membership

- ▼ Envelope small membrane protein, coronavirus (IPR003873)
- ▼ Envelope small membrane protein, betacoronavirus (IPR043506)
- Envelope small membrane protein, SARS-CoV-2-like (IPR044377)



GO terms

Biological Process

- virion assembly (GO:0019068)
- pore formation by virus in membrane of host cell (GO:0039707)
- viral budding from Golgi membrane (GO:0046760)

Molecular Function

None

Cellular Component

- integral component of membrane (GO:0016021)



This service is part of the ELIXIR infrastructure

InterPro is an ELIXIR Core Data Resource Learn more >

CDS_6 功能结构域如下：

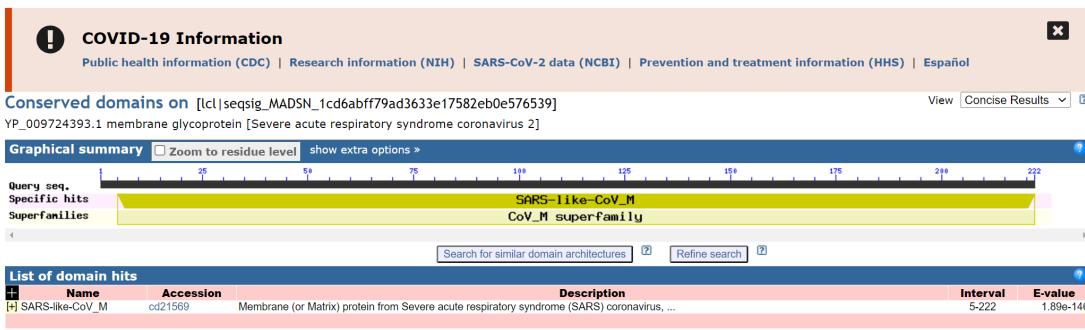


图 2-25 cds_6_cd_search.PNG

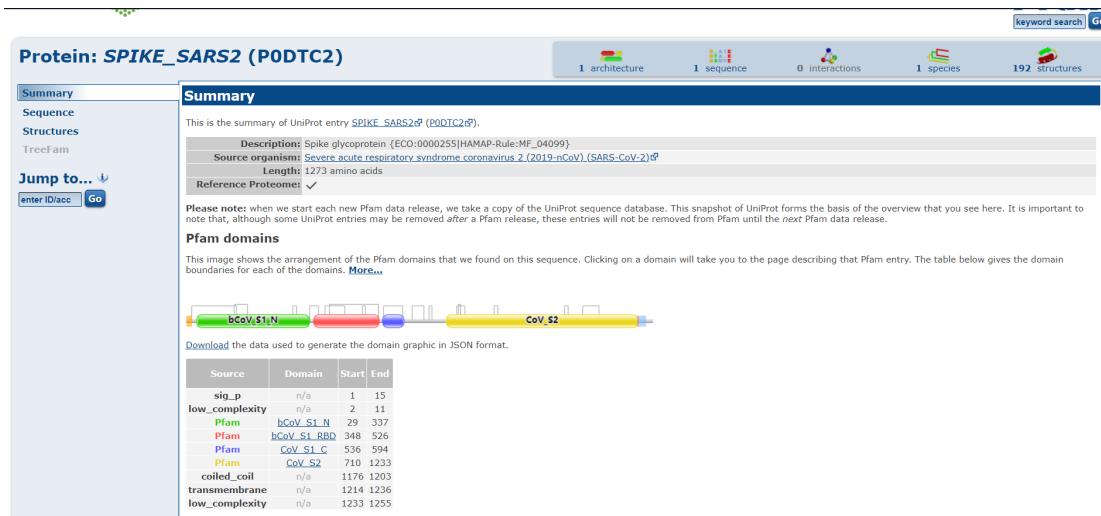


图 2-26 cds_6_Pfam.PNG

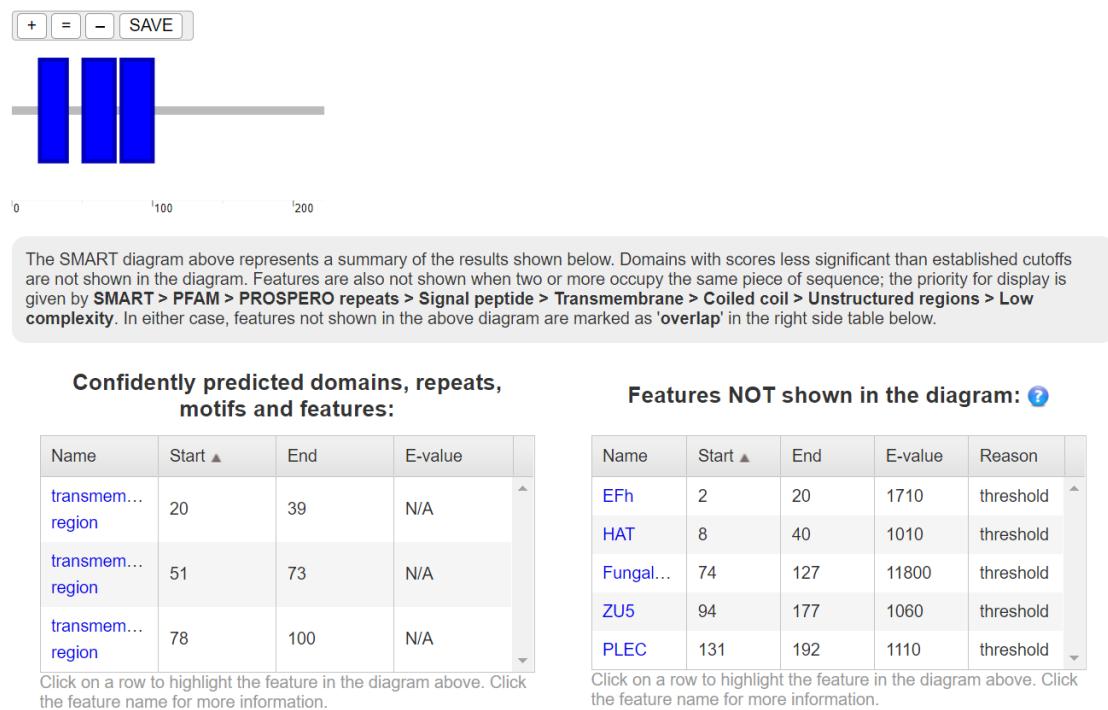


图 2-27 cds_6_SMART.PNG

InterProScan Search Result

i	
Title	YP_009724393.1 membrane glycoprotein [Severe acute respiratory syndrome coronavirus 2]
Job ID i	iprscan5-R20210422-131819-0714-9828333-p1m
Length	222 amino acids
Action	
Status	finished
Expires i	Thu Apr 29 2021

Protein family membership

- ▼ M matrix/glycoprotein, coronavirus (IPR002574)
- ▼ M matrix/glycoprotein, SARS-CoV-like (IPR044361)



图 2-28 cds_6_InterPro.PNG

CDS_7 功能结构域如下：

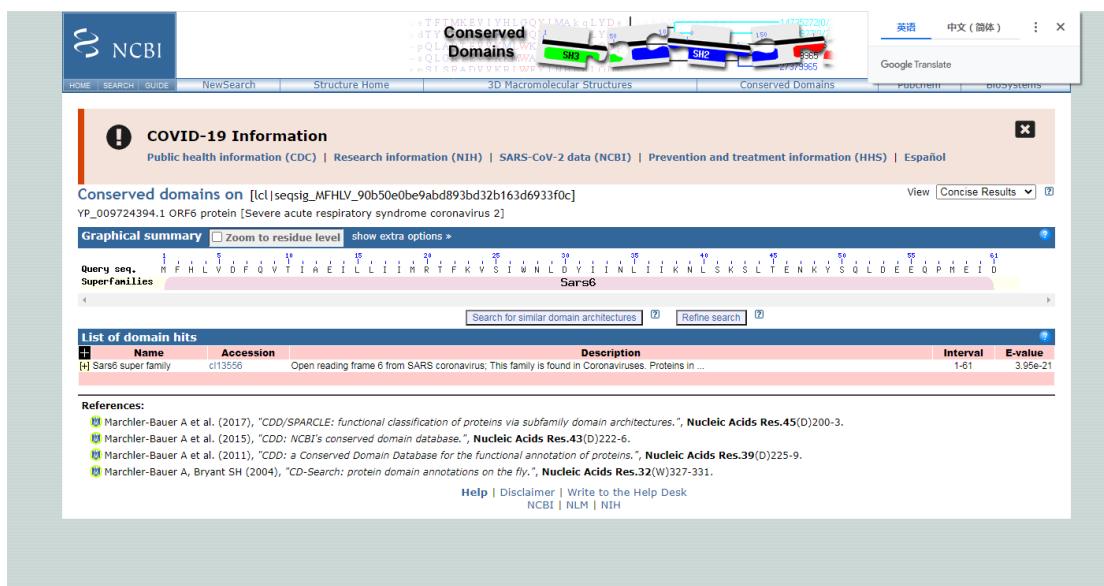


图 2-29 cds_7_cd_search.PNG

EMBL-EBI 

HOME | SEARCH | BROWSE | FTP | HELP | ABOUT

Pfam 
keyword search Go

Protein: R1AB_SARS (P0C6X7)

Summary

Sequence
Structures
TreeFam
Jump to... ↻
enter ID/acc

Summary

This is the summary of UniProt entry [R1AB_SARS](#) (P0C6X7).

Description: Replicase polyprotein 1ab EC=3.4.19.12 EC=3.4.22.69 EC=3.4.22.- EC=2.7.7.48 EC=3.6.4.12 EC=3.6.4.13 EC=2.1.1.- EC=3.1.13.- EC=3.1.- EC=2.1.1.-

Source organism: Severe acute respiratory syndrome coronavirus (SARS-CoV) (NCBI taxonomy ID 694009)

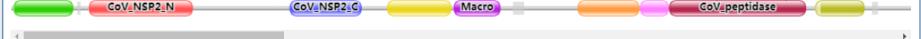
Length: 7073 amino acids

Reference Proteome: ✓

Please note: when we start each new Pfam data release, we take a copy of the UniProt sequence database. This snapshot of UniProt forms the basis of the overview that you see here. It is important to note that, although some UniProt entries may be removed after a Pfam release, these entries will not be removed from Pfam until the next Pfam data release.

Pfam domains

This image shows the arrangement of the Pfam domains that we found on this sequence. Clicking on a domain will take you to the page describing that Pfam entry. The table below gives the domain boundaries for each of the domains. [More...](#)



[Download](#) the data used to generate the domain graphic in JSON format.

Source	Domain	Start	End
Pfam	bCoV_NSP1	7	144
disorder	n/a	156	158
disorder	n/a	160	161
Pfam	CoV_NSP2_N	182	423
disorder	n/a	384	385
Pfam	CoV_NSP2_C	652	818
Pfam	bCoV_NSP3_N	881	1029
disorder	n/a	920	926
disorder	n/a	928	956
low_complexity	n/a	928	950
disorder	n/a	958	960
disorder	n/a	962	1007
low_complexity	n/a	968	987
low_complexity	n/a	985	1002
Pfam	Macro	1036	1143
disorder	n/a	1054	1057
disorder	n/a	1175	1198
Pfam	bCoV_SUD_M	1327	1469
low_complexity	n/a	1447	1459
Pfam	bCoV_SUD_C	1473	1538
low_complexity	n/a	1528	1539
Pfam	CoV_peptidase	1541	1859
Pfam	bCoV_NAR	1883	1996
disorder	n/a	2015	2027
transmembrane	n/a	2145	2162
transmembrane	n/a	2209	2231
low_complexity	n/a	2210	2222
Pfam	CoV_NSP3_C	2237	2726
transmembrane	n/a	2314	2333
transmembrane	n/a	2340	2371
low_complexity	n/a	2618	2628
Pfam	CoV_NSP4_N	2720	3119
transmembrane	n/a	2749	2768
low_complexity	n/a	2755	2764
transmembrane	n/a	2993	3008
transmembrane	n/a	3020	3042
transmembrane	n/a	3054	3078
transmembrane	n/a	3084	3103
transmembrane	n/a	3115	3134
Pfam	CoV_NSP4_C	3143	3239
transmembrane	n/a	3146	3163
Pfam	Peptidase_C30	3269	3559
transmembrane	n/a	3559	3582
low_complexity	n/a	3561	3571
Pfam	CoV_NSP6	3574	3836
transmembrane	n/a	3588	3606
transmembrane	n/a	3613	3634
transmembrane	n/a	3660	3678
transmembrane	n/a	3685	3701
transmembrane	n/a	3721	3743
transmembrane	n/a	3755	3774
low_complexity	n/a	3768	3777
Pfam	CoV_NSP7	3837	3919
low_complexity	n/a	3846	3862
low_complexity	n/a	3889	3897
Pfam	CoV_NSP8	3920	4116
low_complexity	n/a	3931	3946
low_complexity	n/a	3950	3966
low_complexity	n/a	4017	4029
Pfam	CoV_NSP9	4118	4230
Pfam	CoV_NSP10	4239	4361
Pfam	CoV_Rpol_N	4383	4735
Pfam	RdRP_1	4835	5172
Pfam	Viral_helicase1	5611	5877
Pfam	CoV_ExonL	5905	6427
disorder	n/a	6033	6039
Pfam	CoV_NSP15_N	6430	6490
Pfam	CoV_NSP15_M	6491	6614
disorder	n/a	6583	6584
Pfam	CoV_NSP15_C	6620	6773
Pfam	CoV_Methyltr_2	6777	7073

Show or [hide](#) domain scores.

Pfam is part of the ELIXIR infrastructure
Pfam is an Elixir service [Read more](#)

Comments or questions on the site? Send a mail to pftam-help@ebi.ac.uk.

图 2-30 cds_7_Pfam.PNG

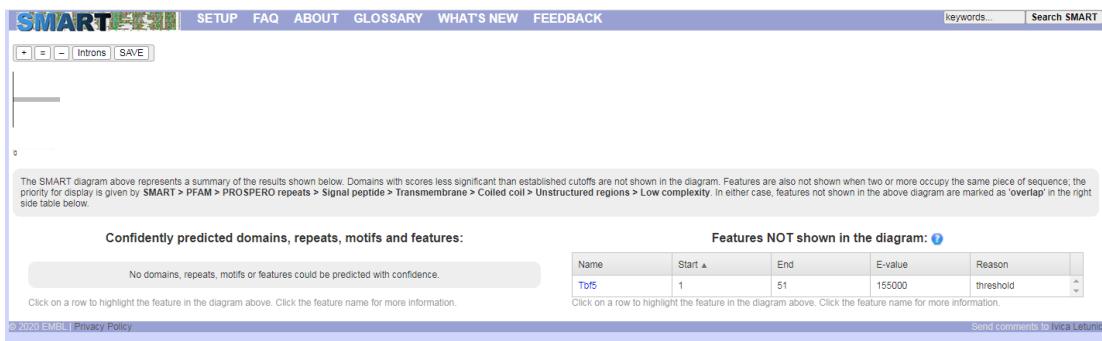
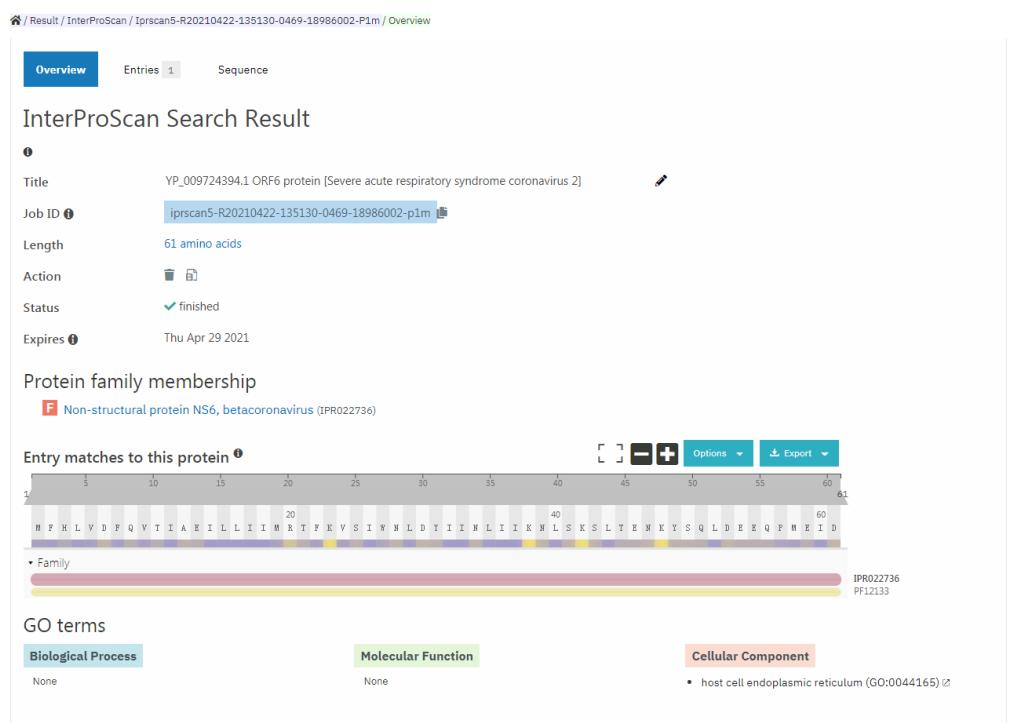


图 2-31 cds_7 SMART.PNG



This service is part of the ELIXIR infrastructure
InterPro is an ELIXIR Core Data Resource Learn more ↗

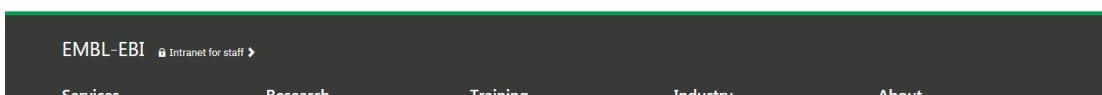


图 2-32 cds_7 InterPro.PNG

CDS_8 功能结构域如下：

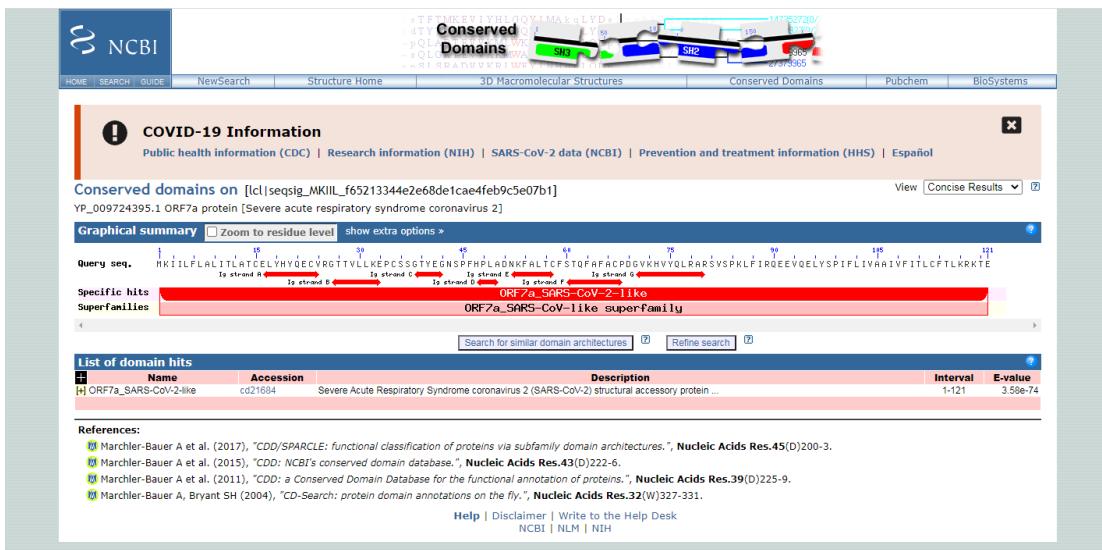


图 2-33 cds_8_cd_search.PNG

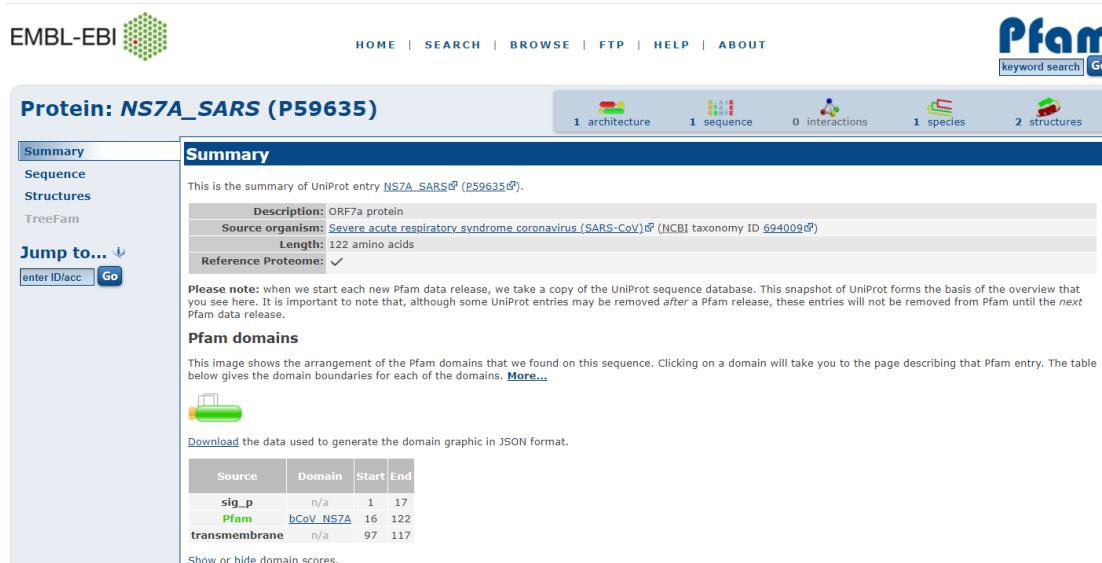


图 2-34 cds_8_Pfam.PNG

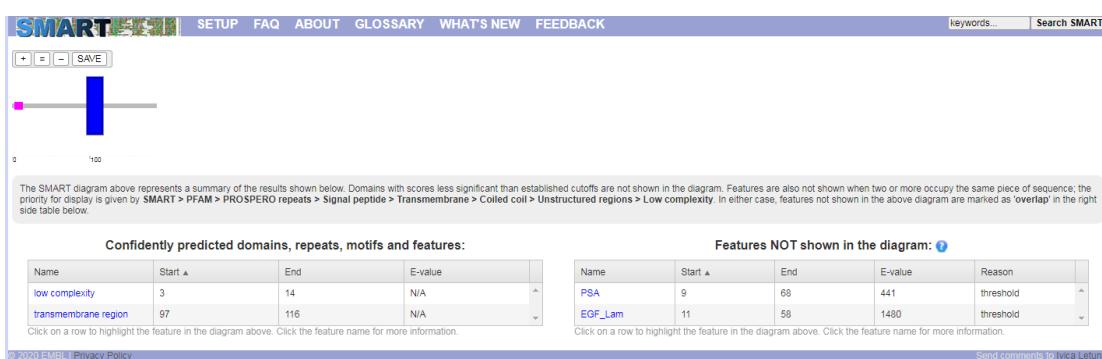


图 2-35 cds_8_SMART.PNG

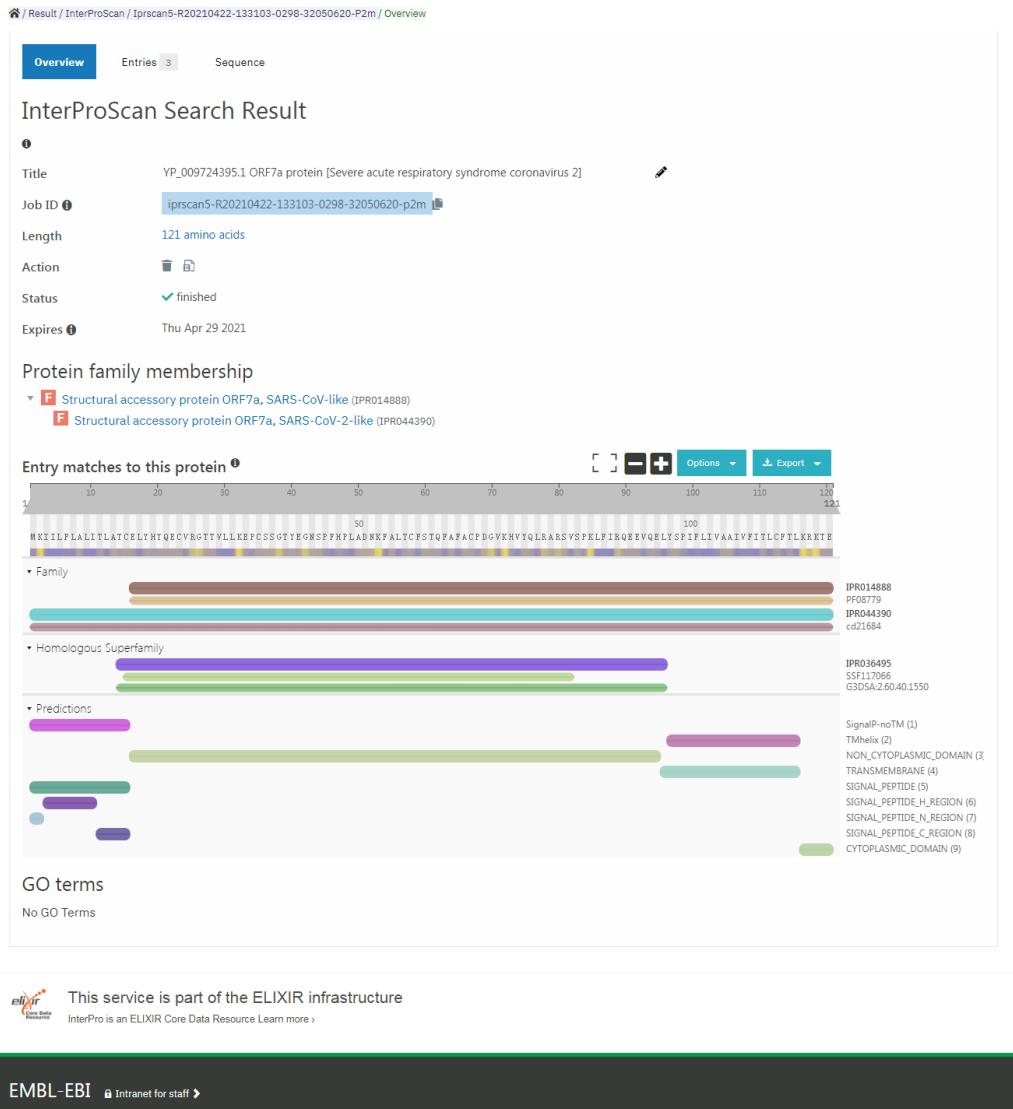


图 2-36 cds_8_InterPro.PNG

CDS_9 功能结构域如下：

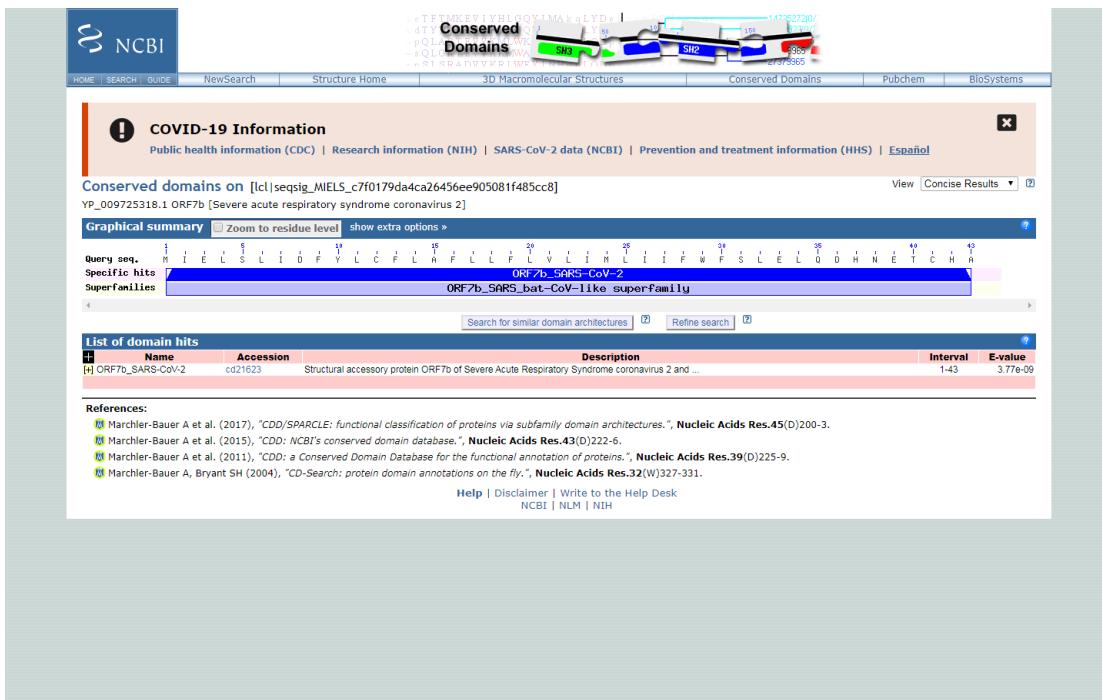


图 2-37 cds_9_cd_search.PNG

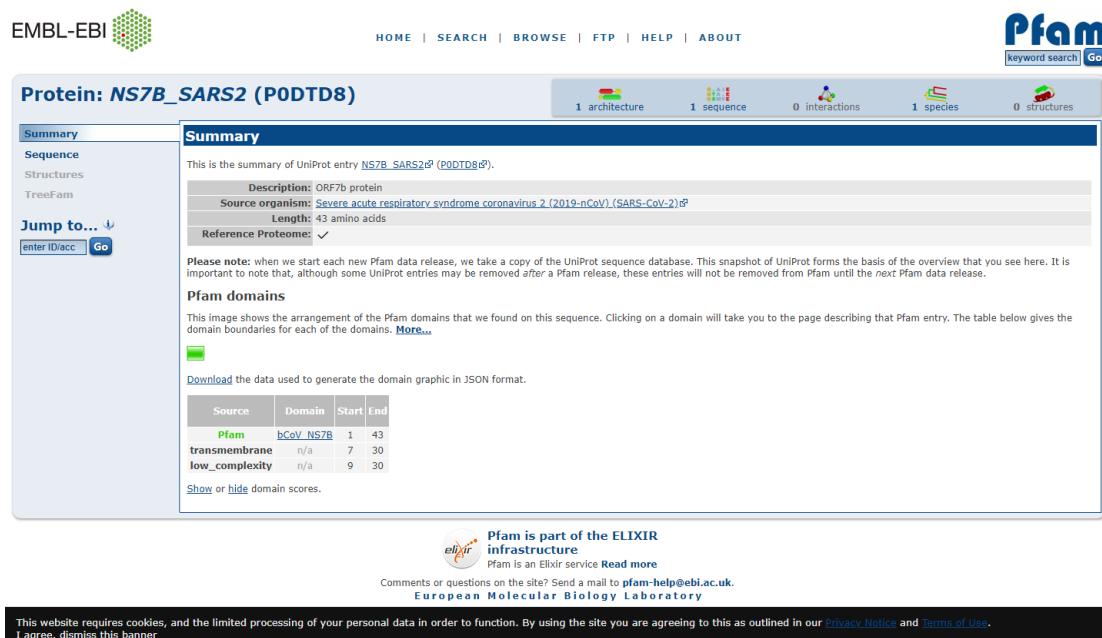


图 2-38 cds_9_Pfam.PNG

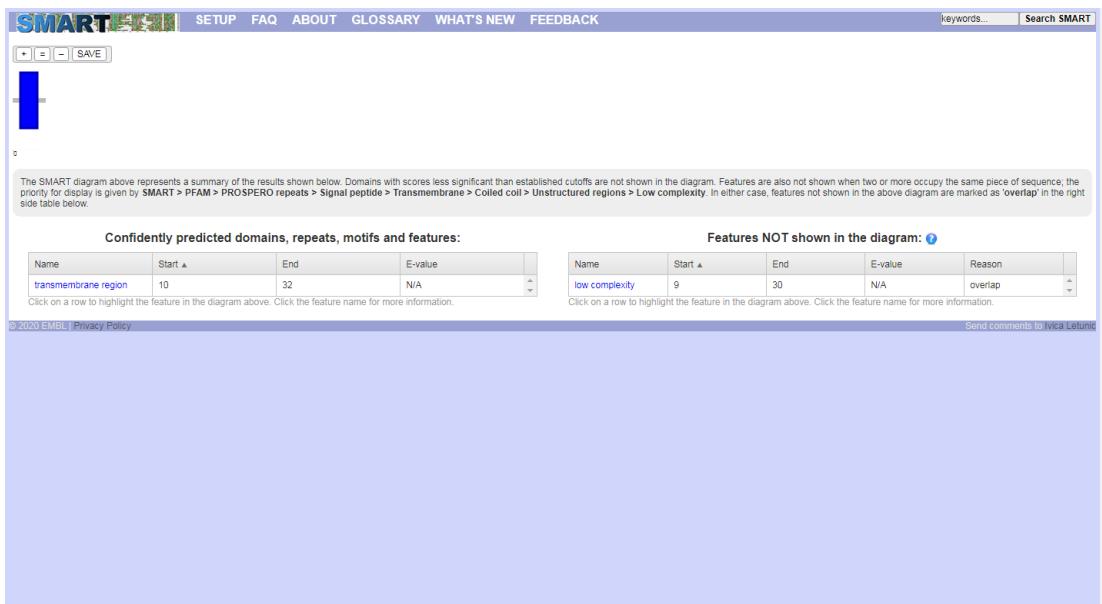
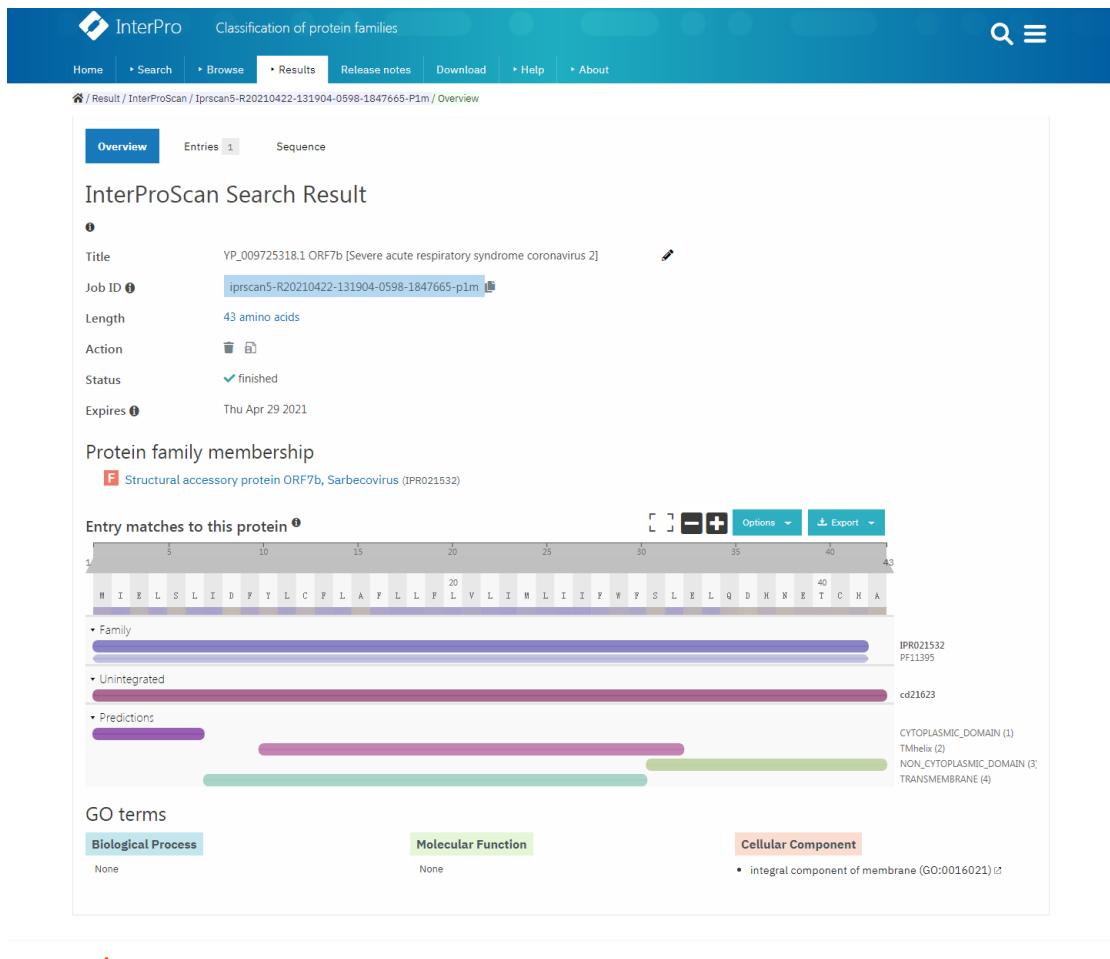


图 2-39 cds_9_SMART.PNG



This service is part of the ELIXIR infrastructure
InterPro is an ELIXIR Core Data Resource Learn more >

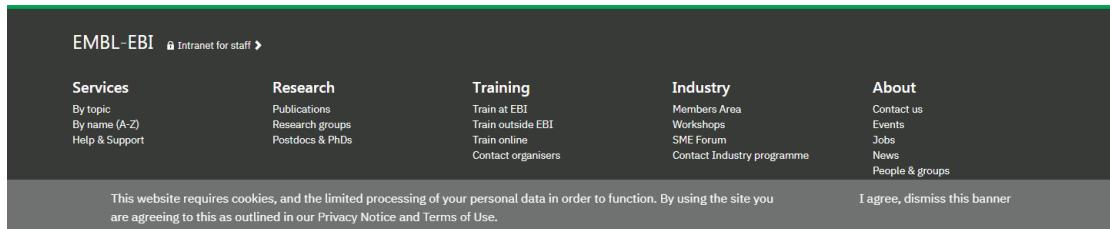


图 2-40 cds_9_InterPro.PNG

CDS_10 功能结构域如下：

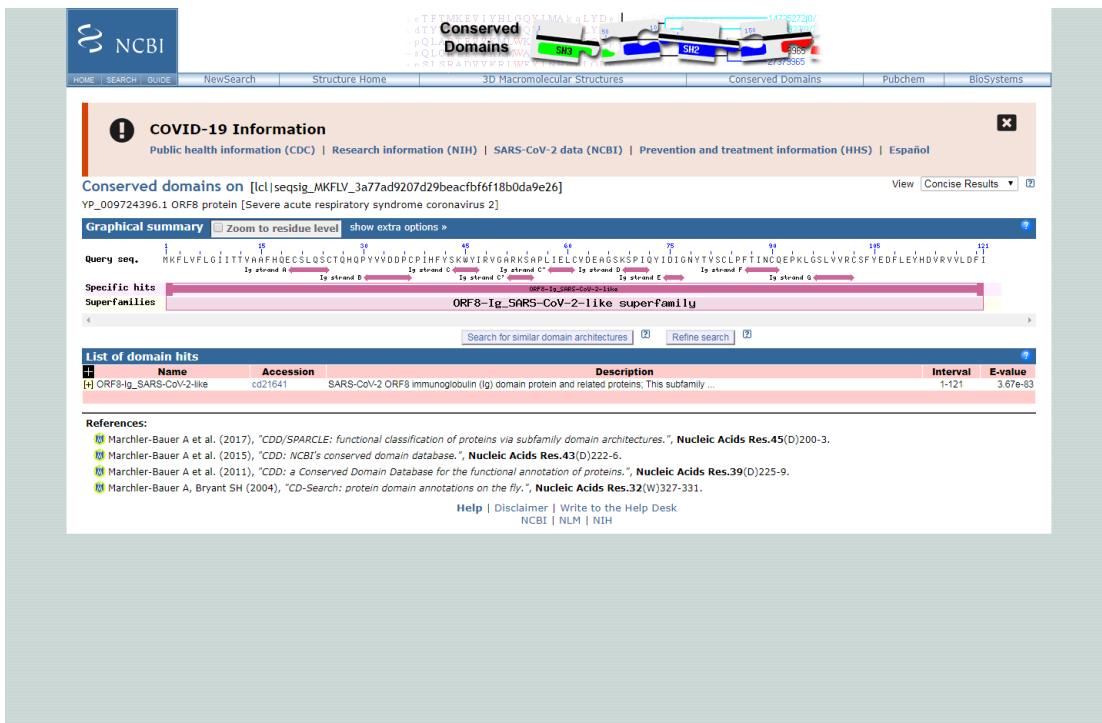


图 2-41 cds_10_cd_search.PNG

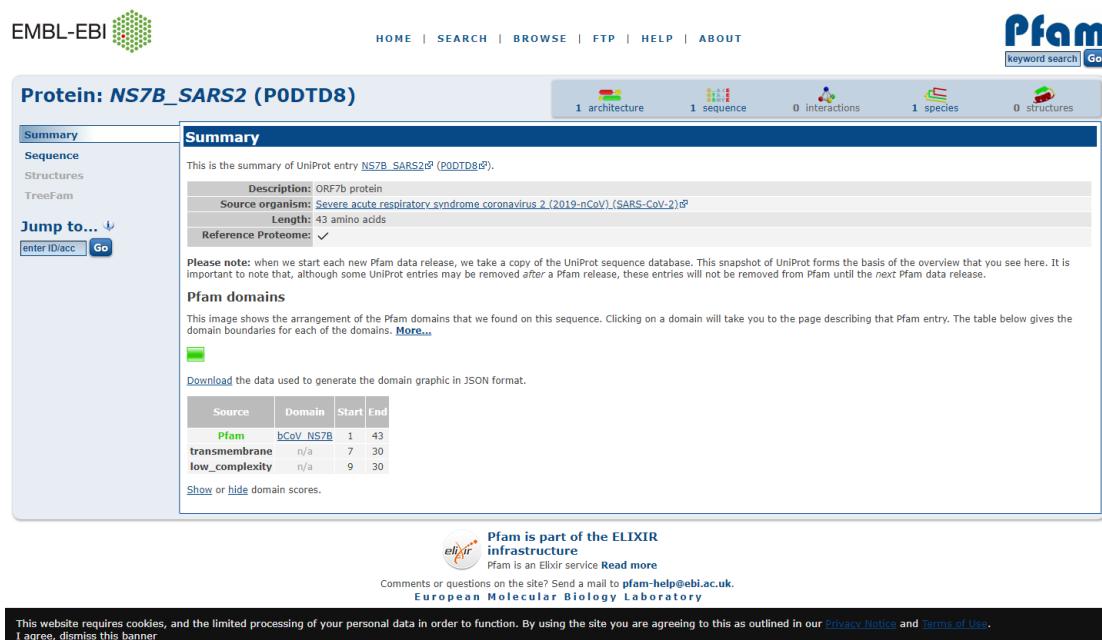


图 2-42 cds_10_Pfam.PNG

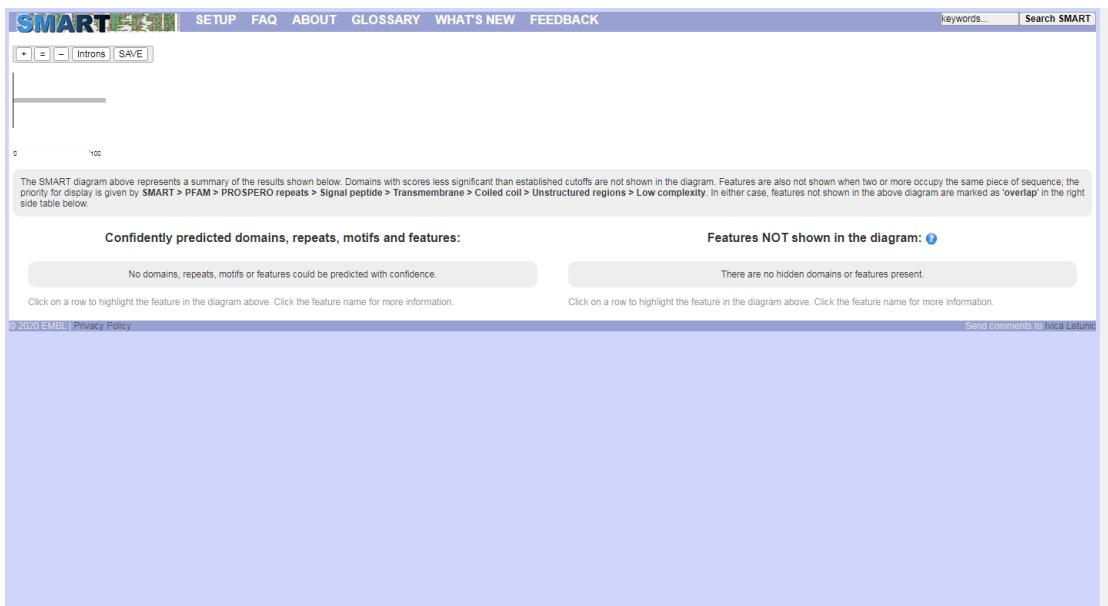


图 2-43 cds_10_SMART.PNG

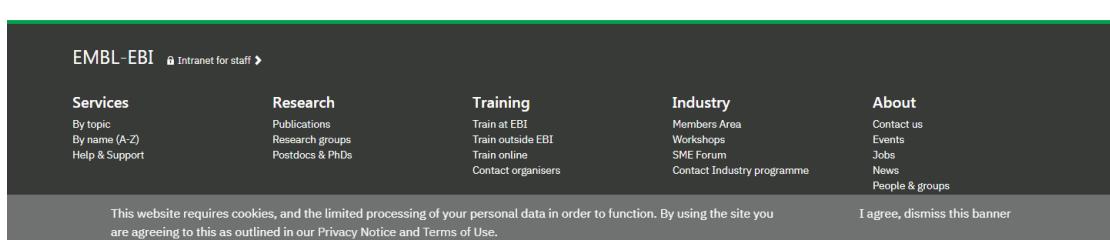
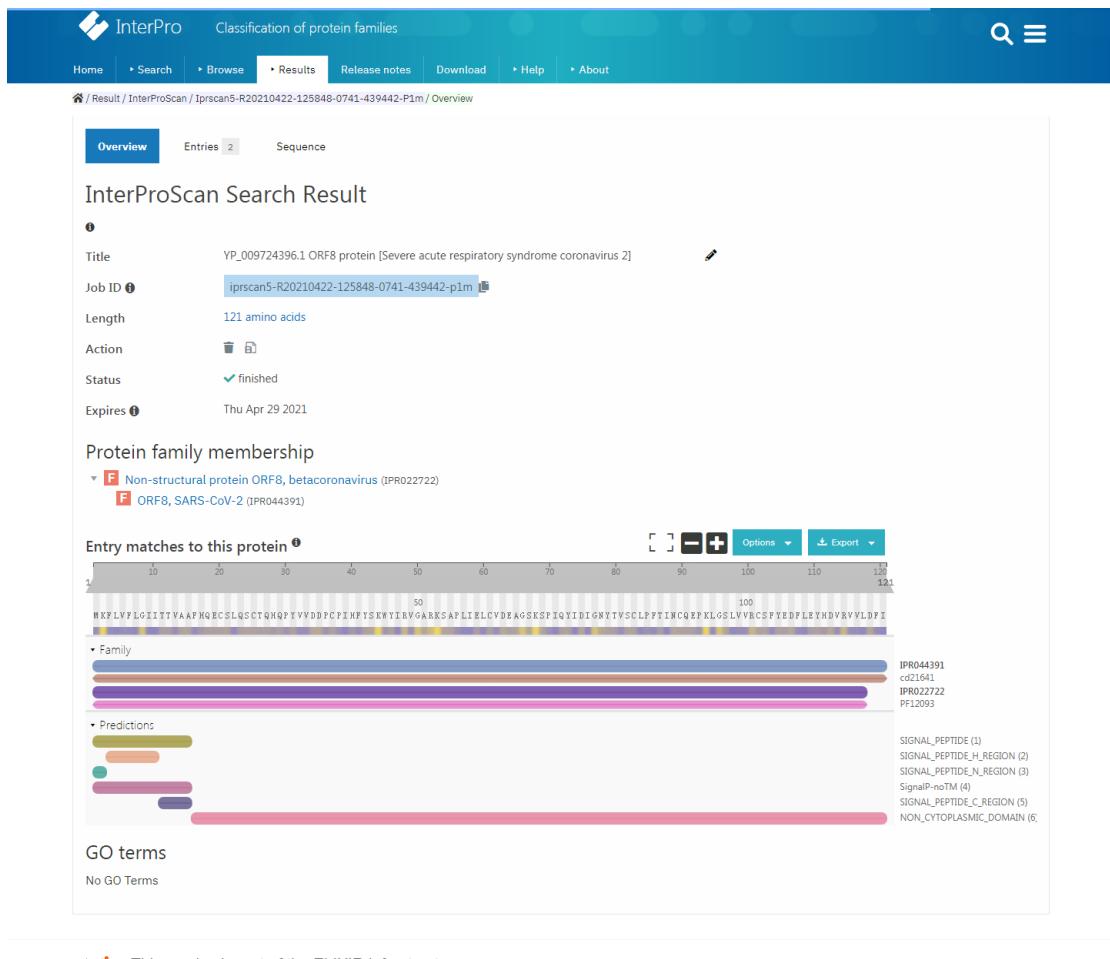


图 2-44 cds_10_InterPro.PNG

CDS_11 功能结构域如下：

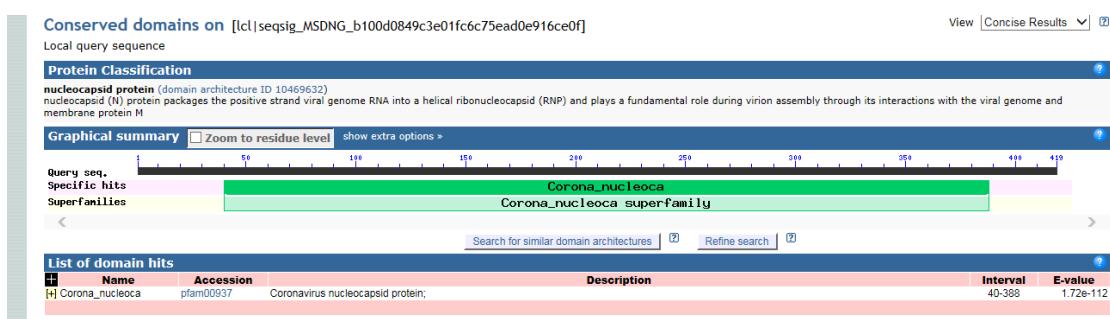


图 2-45 cds_11_cd_search.PNG

Sequence search results

Show the detailed description of this results page.
We found 1 Pfam-A match to your search sequence (all significant)

Corona_nucleo

Show the search options and sequence that you submitted.
Return to the search form to look for Pfam domains on a new sequence.

Significant Pfam-A Matches

Show or hide all alignments.

Family	Description	Entry type	Clan	Envelope		Alignment		HMM		HMM length	Bit score	E-value	Predicted active sites	Show/hide alignment
				Start	End	Start	End	From	To					
Corona_nucleo	Coronavirus nucleocapsid protein	Domain	n/a	14	377	14	377	1	364	364	547.6	1.2e-164	n/a	Show

图 2-46 cds_11_Pfam.PNG

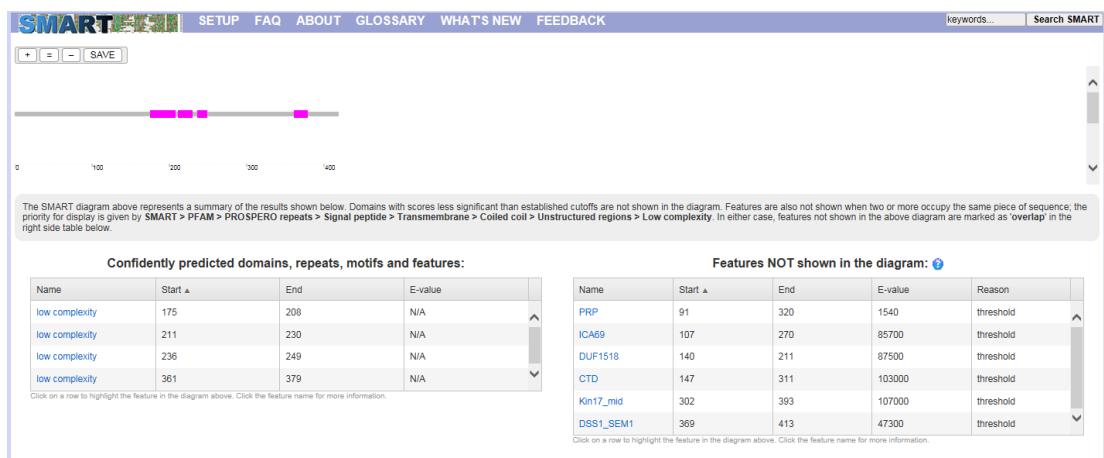


图 2-47 cds_11_SMART.PNG

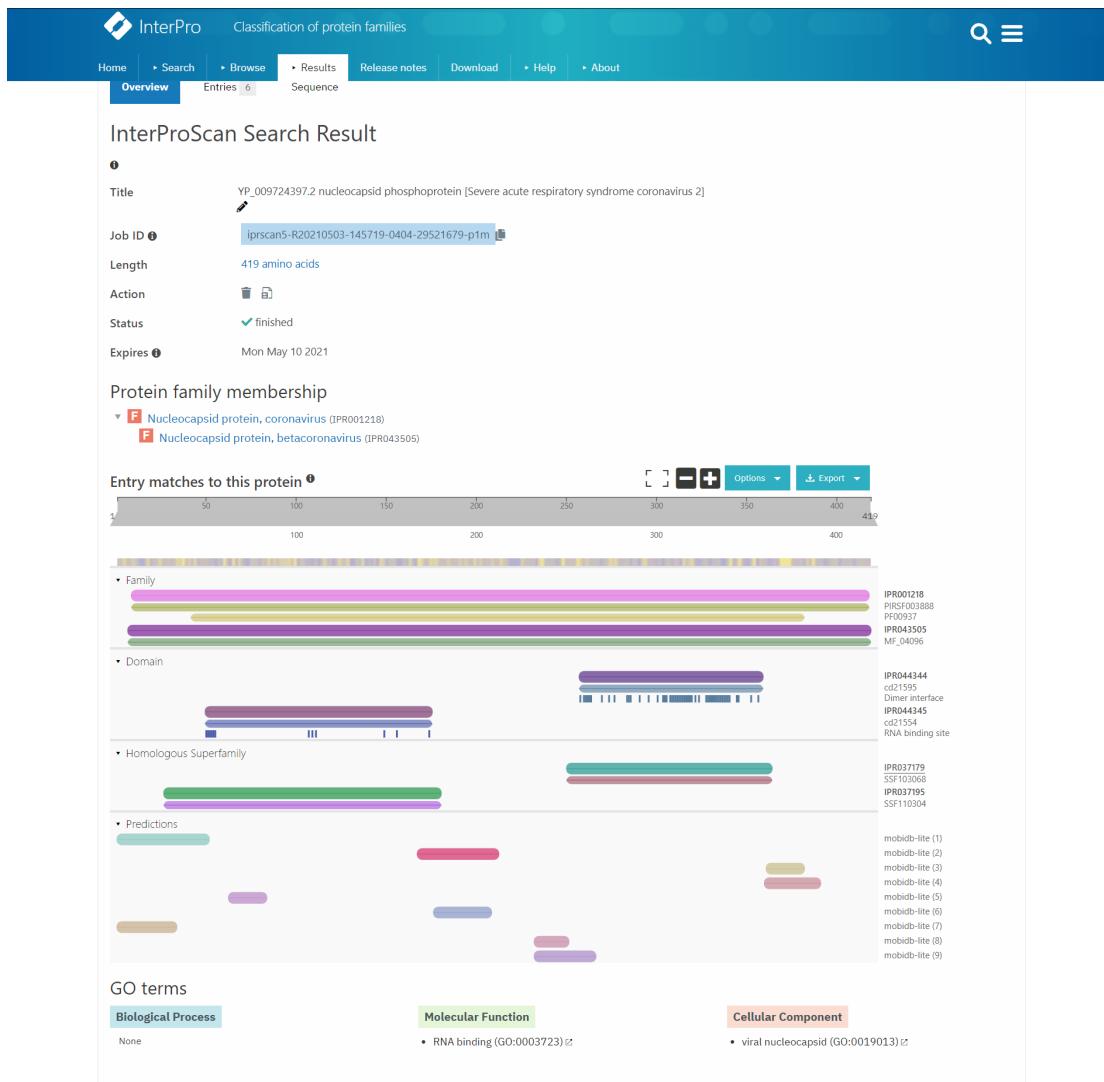


图 2-48 cds_11_InterPro.PNG

CDS_12 功能结构域如下：

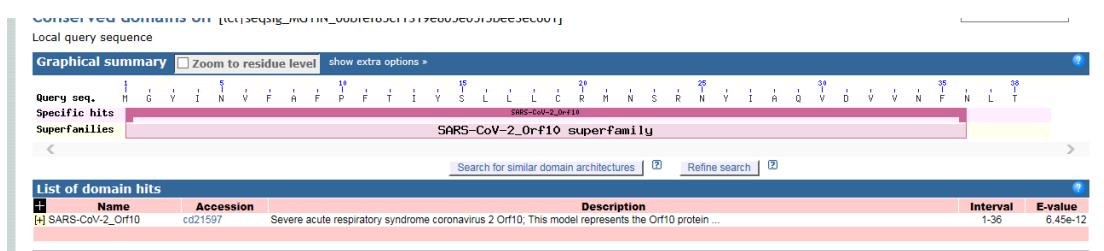


图 2-49 cds_12_cd_search.PNG

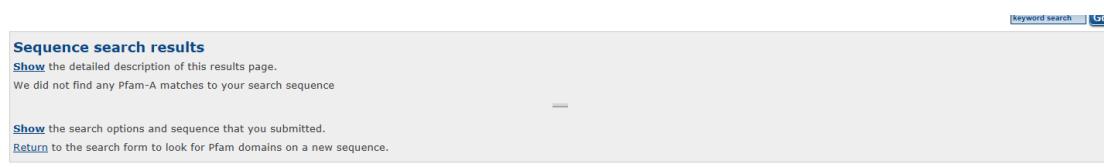


图 2-50 cds_12_Pfam.PNG

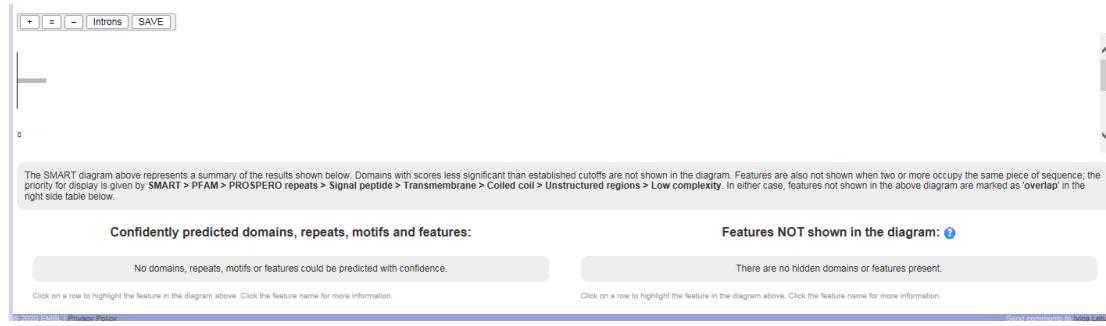


图 2-51 cds_12_SMART.PNG

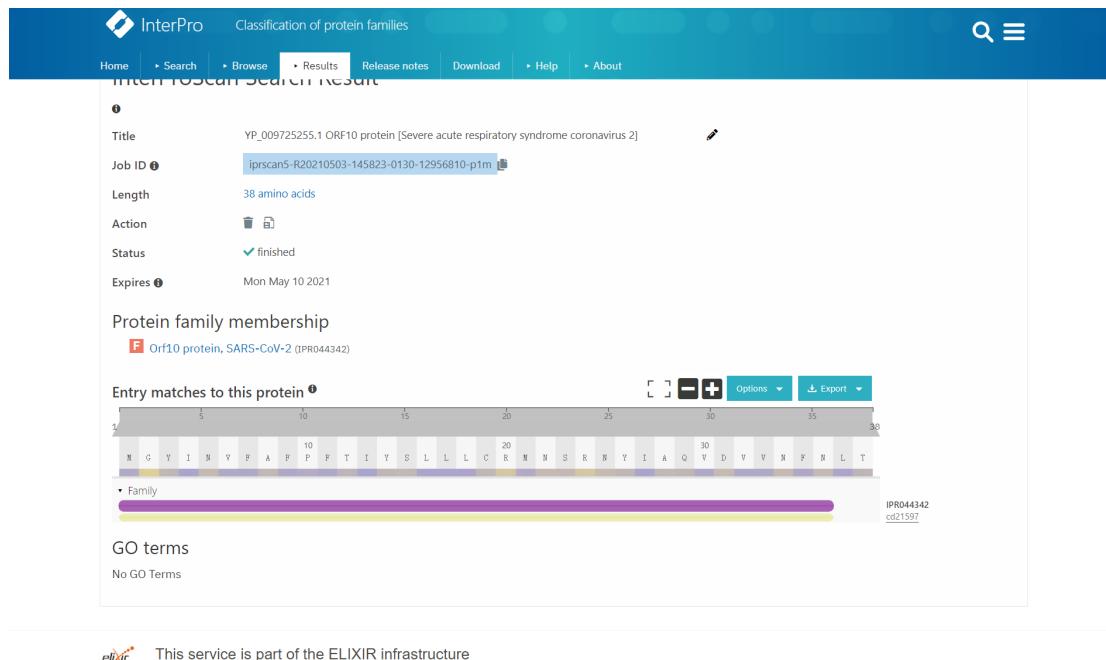


图 2-52 cds_12_InterPro.PNG

2.5 细胞亚定位分析

CDS_1 亚细胞定位如下：

TargetP-2.0

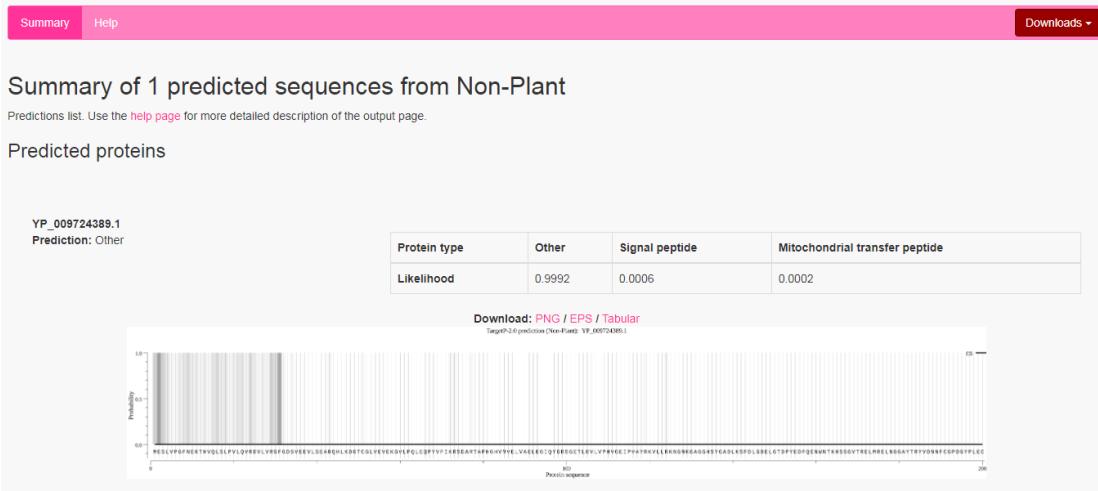


图 2-53 cds_1_TargetP.PNG

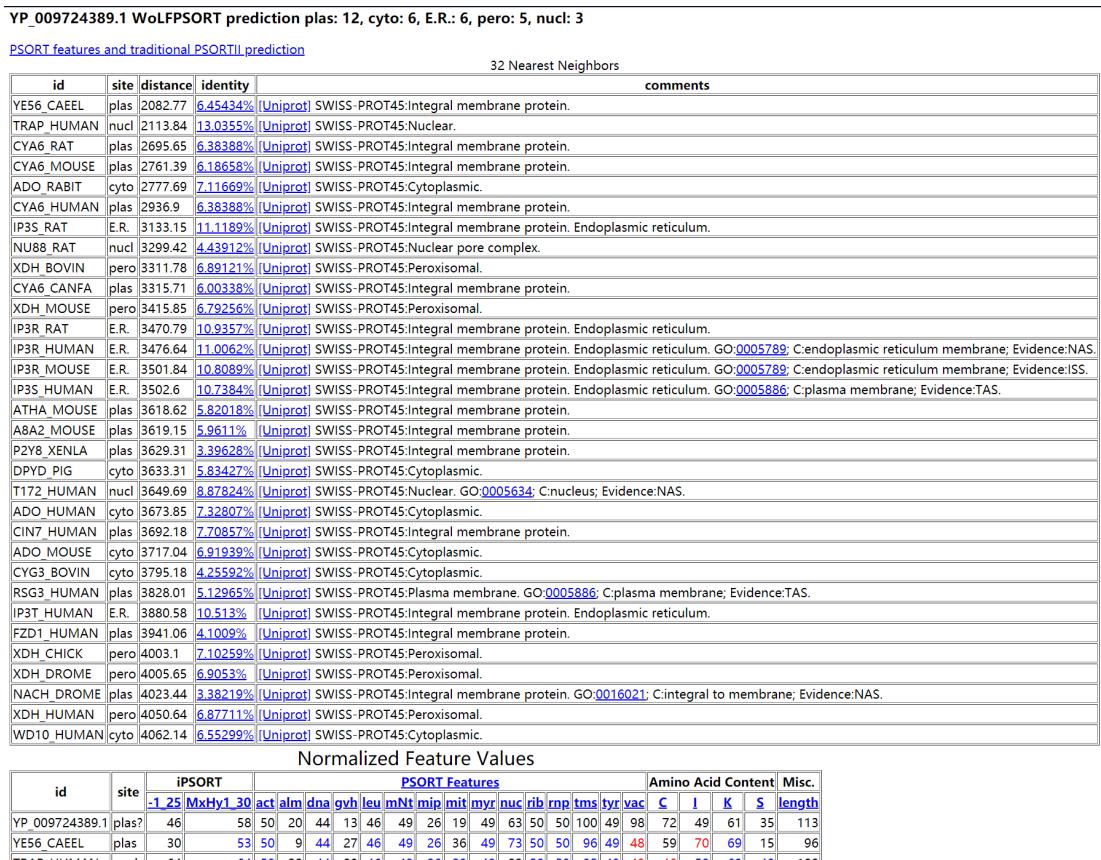


图 2-54 cds_1_Wolf.PNG

CDS_2 亚细胞定位如下：

Summary of 1 predicted sequences from Non-Plant

Predictions list. Use the [help page](#) for more detailed description of the output page.

Predicted proteins

YP_009725295.1

Prediction: Other

Protein type	Other	Signal peptide	Mitochondrial transfer peptide
Likelihood	0.9992	0.0006	0.0002

[Download: PNG / EPS / Tabular](#)

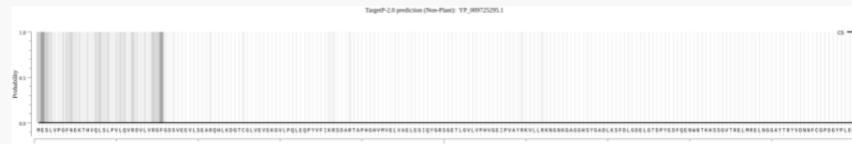


图 2-55 cds_2_TargetP.PNG

id	site	distance	identity	comments
TRAP_HUMAN	nuc1	1812.9	13.9197% [Uniprot] SWISS-PROT45:Nuclear.	
YE56_CAEEL	plas	1900.84	8.89898% [Uniprot] SWISS-PROT45:Integral membrane protein.	
ADO_RABIT	cyto	2599.13	9.67083% [Uniprot] SWISS-PROT45:Cytoplasmic.	
NU88_RAT	nuc1	2702.58	5.94779% [Uniprot] SWISS-PROT45:Nuclear pore complex.	
CYA6_RAT	plas	2703.5	8.85358% [Uniprot] SWISS-PROT45:Integral membrane protein.	
CYA6_MOUSE	plas	2796.8	8.74007% [Uniprot] SWISS-PROT45:Integral membrane protein.	
IP3S_RAT	E.R.	2874.07	14.37% [Uniprot] SWISS-PROT45:Integral membrane protein. Endoplasmic reticulum.	
P2Y8_XENLA	plas	2897.97	4.99432% [Uniprot] SWISS-PROT45:Integral membrane protein.	
RSG3_HUMAN	plas	2969.08	7.10556% [Uniprot] SWISS-PROT45:Plasma membrane. GO: 0005886 : C:plasma membrane; Evidence:TAS.	
CYA6_HUMAN	plas	2971.17	8.92168% [Uniprot] SWISS-PROT45:Integral membrane protein.	
XDH_BOVIN	pero	3001.77	9.82974% [Uniprot] SWISS-PROT45:Peroxisomal.	
CYG3_BOVIN	cyto	3022.28	5.85698% [Uniprot] SWISS-PROT45:Cytoplasmic.	
XDH_MOUSE	pero	3064.79	9.2168% [Uniprot] SWISS-PROT45:Peroxisomal.	
NACH_DROME	plas	3150.38	4.85812% [Uniprot] SWISS-PROT45:Integral membrane protein. GO: 0016021 : C:integral to membrane; Evidence:NAS.	
IP3R_HUMAN	E.R.	3166.25	13.6209% [Uniprot] SWISS-PROT45:Integral membrane protein. Endoplasmic reticulum. GO: 0005789 : C:endoplasmic reticulum membrar	
IP3R_RAT	E.R.	3214.43	14.0295% [Uniprot] SWISS-PROT45:Integral membrane protein. Endoplasmic reticulum.	
IP3R_MOUSE	E.R.	3218.53	13.6663% [Uniprot] SWISS-PROT45:Integral membrane protein. Endoplasmic reticulum. GO: 0005789 : C:endoplasmic reticulum membrar	
IP3S_HUMAN	E.R.	3232.4	14.37% [Uniprot] SWISS-PROT45:Integral membrane protein. Endoplasmic reticulum. GO: 0005886 : C:plasma membrane; Evidence:TAS	
DYPD_PIG	cyto	3269.67	8.71737% [Uniprot] SWISS-PROT45:Cytoplasmic.	
CIN7_HUMAN	plas	3295.85	11.0329% [Uniprot] SWISS-PROT45:Integral membrane protein.	
T100_HUMAN	nuc1	3319.82	8.05902% [Uniprot] SWISS-PROT45:Nuclear. GO: 0005634 : C:nucleus; Evidence:IDA.	
A8A2_MOUSE	plas	3337.87	9.01249% [Uniprot] SWISS-PROT45:Integral membrane protein.	
T172_HUMAN	nuc1	3356.97	12.2134% [Uniprot] SWISS-PROT45:Nuclear. GO: 0005634 : C:nucleus; Evidence:NAS.	
CYA6_CANFA	plas	3371.69	8.74007% [Uniprot] SWISS-PROT45:Integral membrane protein.	
ADO_HUMAN	cyto	3426.75	9.94325% [Uniprot] SWISS-PROT45:Cytoplasmic.	
ATHA_MOUSE	plas	3426.98	7.80931% [Uniprot] SWISS-PROT45:Integral membrane protein.	
NU88_HUMAN	nuc1	3440.32	6.1975% [Uniprot] SWISS-PROT45:Nuclear pore complex. GO: 0005643 : C:nuclear pore; Evidence:NAS.	
ADO_MOUSE	cyto	3449.07	9.71623% [Uniprot] SWISS-PROT45:Cytoplasmic.	
DYPD_HUMAN	cyto	3509.72	8.26334% [Uniprot] SWISS-PROT45:Cytoplasmic. GO: 0005737 : C:cytoplasm; Evidence:NAS.	
WD10_HUMAN	cyto	3525.38	8.98978% [Uniprot] SWISS-PROT45:Cytoplasmic.	
IP3T_HUMAN	E.R.	3539.17	13.7571% [Uniprot] SWISS-PROT45:Integral membrane protein. Endoplasmic reticulum.	
MRP4_HUMAN	plas	3572.65	9.62543% [Uniprot] SWISS-PROT45:Integral membrane protein.	

Normalized Feature Values

id	site	iPSORT		PSORT Features												Amino Acid Content				Misc.				
		-1	25	MxHy1	30	act	alm	dna	gvh	leu	mNt	mip	mit	myr	nuc	rib	rnp	tms	tyr	vac	C	I	K	S
YP_009725295.1	plas?	46	58	50	20	44	13	46	49	26	19	49	63	50	50	100	49	48	70	50	63	40	100	
TRAP_HUMAN	nuc1	64	64	50	32	44	28	46	49	26	20	49	83	50	50	95	49	48	46	59	60	40	100	
YE56_CAEEL	plas	30	53	50	9	44	27	46	49	26	36	49	73	50	50	96	49	48	59	70	69	15	96	
ADO_RABIT	cyto	18	33	50	43	44	11	93	49	26	31	49	67	50	50	63	49	48	65	65	58	25	97	
NU88_RAT	nuc1	18	40	50	30	44	12	96	49	26	0	49	29	50	50	63	49	48	65	58	62	43	89	

图 2-56 cds_2_Wolf.PNG

CDS_3 亚细胞定位如下：

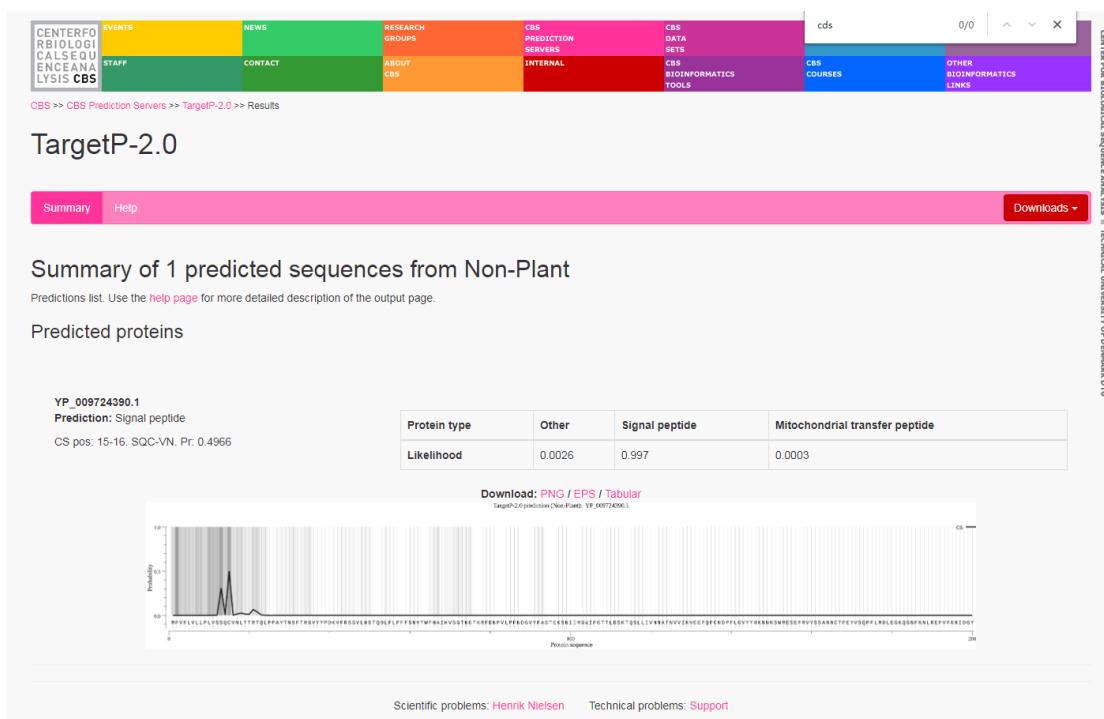


图 2-57 cds_3_TargetP.PNG

id	site	distance	identity	comments
LHN2_BOVIN	plas	860.47	13.8024% [Uniprot]	SWISS-PROT45:Integral membrane protein.
LHN2_HUMAN	plas	877.989	13.2282% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0016021 ; C:integral to membrane; Evidence:NAS.
MGR1_HUMAN	plas	1038.61	12.4122% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0005887 ; C:integral to plasma membrane; Evidence:TAS.
LHN2_RAT	plas	1428.48	13.7189% [Uniprot]	SWISS-PROT45:Integral membrane protein.
MGR4_RAT	plas	1473.09	13.1186% [Uniprot]	SWISS-PROT45:Integral membrane protein.
FZD6_MOUSE	plas	1537.92	12.4116% [Uniprot]	SWISS-PROT45:Integral membrane protein.
GLK4_RAT	plas	1583.59	15.5538% [Uniprot]	SWISS-PROT45:Integral membrane protein.
CD97_MOUSE	plas	1604.88	14.454% [Uniprot]	SWISS-PROT45:Integral membrane protein.
MGR6_RAT	plas	1651.75	13.6685% [Uniprot]	SWISS-PROT45:Integral membrane protein.
FZD6_HUMAN	plas	1774.51	12.8044% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0005887 ; C:integral to plasma membrane; Evidence:TAS.
GLK4_HUMAN	plas	1844.48	15.6324% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0005887 ; C:integral to plasma membrane; Evidence:TAS.
NPC1_CAEEL	plas	1880.37	11.1913% [Uniprot]	SWISS-PROT45:Integral membrane protein.
GAC3_MOUSE	plas	2181.73	10.6834% [Uniprot]	SWISS-PROT45:Integral membrane protein.
S291_MOUSE	plas	2203.01	10.1335% [Uniprot]	SWISS-PROT45:Integral membrane protein.
GAC3_RAT	plas	2317.51	10.6049% [Uniprot]	SWISS-PROT45:Integral membrane protein.
MGR7_RAT	plas	2343.76	14.9254% [Uniprot]	SWISS-PROT45:Integral membrane protein.
GAE_RAT	plas	2370.41	10.9191% [Uniprot]	SWISS-PROT45:Integral membrane protein.
SMO_HUMAN	plas	2433.55	13.5899% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0016021 ; C:integral to membrane; Evidence:TAS.
NAC3_RAT	plas	2440.35	13.2757% [Uniprot]	SWISS-PROT45:Integral membrane protein.
KCD2_RABBIT	plas	2450.59	12.7258% [Uniprot]	SWISS-PROT45:Integral membrane protein.
PTRR_DIDMA	plas	2454.38	10.8405% [Uniprot]	SWISS-PROT45:Integral membrane protein.
MGR6_HUMAN	plas	2480.42	13.4741% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0005887 ; C:integral to plasma membrane; Evidence:TAS.
HRA1_HUMAN	extr	2491.11	10.1335% [Uniprot]	SWISS-PROT45:Secreted. GO: 0005615 ; C:extracellular space; Evidence:TAS.
KCD2_MOUSE	plas	2513.9	12.7258% [Uniprot]	SWISS-PROT45:Integral membrane protein.
S291_RAT	plas	2519.63	10.8405% [Uniprot]	SWISS-PROT45:Integral membrane protein.
NAC3_HUMAN	plas	2529.84	13.4328% [Uniprot]	SWISS-PROT45:Integral membrane protein.
SMO_DROME	plas	2550.7	13.6685% [Uniprot]	SWISS-PROT45:Integral membrane protein.
ACH4_RAT	plas	2572.85	12.176% [Uniprot]	SWISS-PROT45:Integral membrane protein.
CP11_MACFA	E.R.	2584.8	11.3904% [Uniprot]	SWISS-PROT45:Membrane-bound. Endoplasmic reticulum.
ACHP_HUMAN	plas	2591.68	10.8405% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0005892 ; C:nicotinic acetylcholine-gated receptor-chan...; Evidence:TAS.
GP83_MOUSE	plas	2597.29	10.2907% [Uniprot]	SWISS-PROT45:Integral membrane protein.
FZD5_HUMAN	plas	2619.1	12.0974% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0005887 ; C:integral to plasma membrane; Evidence:TAS.

Normalized Feature Values																								
id	site	iPSORT			PSORT Features					Amino Acid Content					Misc.									
		-1	-2	-3	act	alm	dna	gvh	leu	mNt	mip	mit	myr	nuc	rib	rnp	tms	tyr	vac	C	I	K	S	length

图 2-58 cds_3_Wolf.PNG

CDS_4 亚细胞定位如下：

TargetP-2.0

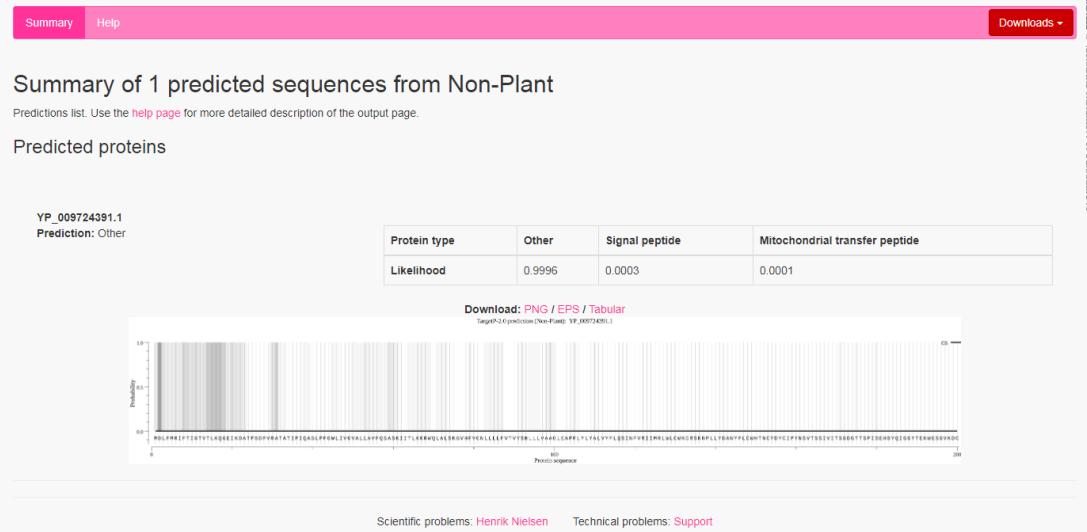


图 2-59 cds_4_TargetP.PNG

O5DE_HUMAN	plas	482.012	15.2381%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
O8B2_HUMAN	plas	851.891	15.142%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
O4A5_HUMAN	plas	894.264	15.873%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
O8B3_HUMAN	plas	895.049	15.1899%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
O9I1_HUMAN	plas	950.251	11.465%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OAV1_HUMAN	plas	956.406	13.9159%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
O1L1_HUMAN	plas	998.877	12.2186%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
O6C4_HUMAN	plas	1091.84	12.6214%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OR07_MOUSE	plas	1144.12	11.4198%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
O4AG_HUMAN	plas	1163.19	15.2439%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OLF5_CHICK	plas	1179.47	14.0575%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OXB4_HUMAN	plas	1194.91	13.5484%	[Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0016021 ; C:integral to membrane; Evidence:NAS.
OXG1_HUMAN	plas	1267.9	13.3956%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OAA3_HUMAN	plas	1318.3	13.0573%	[Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0016021 ; C:integral to membrane; Evidence:NAS.
O7G1_HUMAN	plas	1331.54	14.4144%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OYL1_HUMAN	plas	1363.24	13.0573%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
O1L3_HUMAN	plas	1501.82	13.8889%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
O2T4_HUMAN	plas	1520.5	14.5062%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
CLDZ_BRARE	plas	1532.03	13.8182%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OPSD_ANGAN	plas	1575.34	14.4886%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OLF9_RAT	plas	1576.63	14.0127%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OPSB_CONCO	plas	1589.22	14.8571%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
O5M8_HUMAN	plas	1663.68	13.141%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OR13_MOUSE	plas	1668.98	16.0883%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OR08_MOUSE	plas	1680.63	13.0303%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OPS1_CALVI	plas	1696.14	13.2075%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
O5F1_HUMAN	plas	1733.13	12.4204%	[Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0016021 ; C:integral to membrane; Evidence:NAS.
OK3_KUMAN	plas	1734.78	11.745%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OAJ1_HUMAN	plas	1745.88	14.3302%	[Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0005887 ; C:integral to plasma membrane; Evidence:TAS.
O1N2_HUMAN	plas	1747.72	13.6076%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
OXG2_HUMAN	plas	1787.13	14.557%	[Uniprot]	SWISS-PROT45:Integral membrane protein.
CLDD_MOUSE	plas	1823.44	12.3636%	[Uniprot]	SWISS-PROT45:Integral membrane protein.

Normalized Feature Values

id	site	IP SORT		PSORT Features												Amino Acid Content				Misc.				
		-1_25	MxHy_1_30	act	alm	dna	gvh	leu	mNt	mip	mit	myr	nuc	rib	rnp	tms	tyr	vac	C	I	K	S		
YP_009724391.1	plas?	46		64	50	30	44	65	46	49	61	67	49	29	50	50	76	49	48	59	88	26	63	36
O5DE_HUMAN	plas	46	57	50	18	44	76	46	49	61	55	49	29	50	50	87	49	48	66	85	24	68	42	
O8B2_HUMAN	plas	46	65	50	26	44	56	46	49	57	47	49	29	50	50	87	49	48	76	93	24	82	42	
O4A5_HUMAN	plas	46	61	50	22	44	69	46	49	54	28	49	29	50	50	92	49	48	83	94	23	57	42	
O8B3_HUMAN	plas	46	65	50	25	44	56	46	49	57	47	49	29	50	50	82	49	48	76	96	24	82	42	
O9I1_HUMAN	plas	46	57	50	21	44	66	46	49	63	65	49	29	50	50	87	49	48	76	91	39	46	42	

图 2-60 cds_4_Wolf.PNG

CDS_5 亚细胞定位如下：

Summary of 1 predicted sequences from Non-Plant

Predictions list. Use the [help page](#) for more detailed description of the output page.

Predicted proteins

YP_009724393.1

Prediction: Other

Protein type	Other	Signal peptide	Mitochondrial transfer peptide
Likelihood	0.9998	0.0001	0.0001

Download: [PNG](#) / [EPS](#) / [Tabular](#)

TargetP-2.0 prediction (Non-Plant) YP_009724393.1

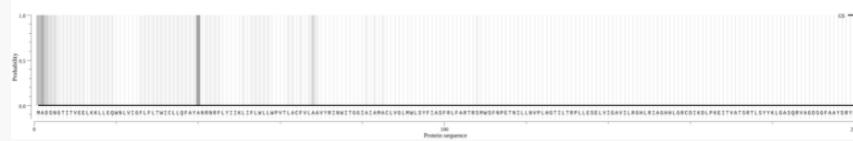


图 2-61 cds_5_TargetP.PNG

id	site	distance	identity	comments
LAPP_HAEOF	extr	1734.92	10.8844% [Uniprot]	SWISS-PROT45:Secreted.
LPP2_RAT	plas	2375.83	10.8696% [Uniprot]	SWISS-PROT45:Integral membrane protein.
CLD6_HUMAN	plas	2394.32	11.1363% [Uniprot]	SWISS-PROT45:Integral membrane protein.
EMP3_MOUSE	plas	2625.75	15.3374% [Uniprot]	SWISS-PROT45:Integral membrane protein.
CXO1_CONST	extr	2643.12	12% [Uniprot]	SWISS-PROT45:Secreted.
CLD6_MOUSE	plas	2645.37	9.58904% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: Q0005923 ; C:tight junction; Evidence:IDA.
ANFB_FUGRU	extr	2803.78	15.0376% [Uniprot]	SWISS-PROT45:Secreted.
CXO2_CONST	extr	2823.64	12% [Uniprot]	SWISS-PROT45:Secreted.
CXOA_CONST	extr	2859.74	10.6667% [Uniprot]	SWISS-PROT45:Secreted.
CXO6_CONGE	extr	2869.23	17.9487% [Uniprot]	SWISS-PROT45:Secreted.
DEF5_HUMAN	extr	2915.23	8.51064% [Uniprot]	SWISS-PROT45:Secreted.
BXD1_BOMMO	extr	2989.54	11.1111% [Uniprot]	SWISS-PROT45:Secreted.
MMHB_AGKHA	extr	3062.77	10.9589% [Uniprot]	SWISS-PROT45:Secreted. GO: Q0005576 ; C:extracellular; IC.
CLDJ_MOUSE	plas	3107.87	13.8393% [Uniprot]	SWISS-PROT45:Integral membrane protein.
O2Z1_HUMAN	plas	3131.71	9.87261% [Uniprot]	SWISS-PROT45:Integral membrane protein.
LPP2_MOUSE	plas	3143.06	10.5072% [Uniprot]	SWISS-PROT45:Integral membrane protein.
IM18_DROME	extr	3229.38	11.8421% [Uniprot]	SWISS-PROT45:Secreted.
CLDJ_HUMAN	plas	3252.45	12.9464% [Uniprot]	SWISS-PROT45:Integral membrane protein.
HEPC_MORCS	extr	3262.69	12.6437% [Uniprot]	SWISS-PROT45:Secreted.
O3A3_PANTR	plas	3315.97	9.52381% [Uniprot]	SWISS-PROT45:Integral membrane protein.
ITH2_HIRMA	extr	3421.74	13.0952% [Uniprot]	SWISS-PROT45:Secreted.
CHH2_PENJP	extr	3448.49	18.3333% [Uniprot]	SWISS-PROT45:Secreted.
CXA4_HUMAN	plas	3494.2	8.73494% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: Q0005887 ; C:integral to plasma membrane; Evidence:TAS.
HPC1_SALSA	extr	3533.45	10.4651% [Uniprot]	SWISS-PROT45:Secreted.
CLDI_HUMAN	plas	3567.43	9.1954% [Uniprot]	SWISS-PROT45:Integral membrane protein.
GUAU_DIDMA	extr	3577.1	11.9266% [Uniprot]	SWISS-PROT45:Secreted.
O3A3_HUMAN	plas	3580.91	8.57143% [Uniprot]	SWISS-PROT45:Integral membrane protein.
GP41_HUMAN	plas	3713.54	10.1156% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: Q0005887 ; C:integral to plasma membrane; Evidence:TAS.
CRVP_LATSE	extr	3717.64	9.66387% [Uniprot]	SWISS-PROT45:Secreted.
BXB2_BOMMO	extr	3742.06	14.6067% [Uniprot]	SWISS-PROT45:Secreted.
PIGL_HUMAN	E.R.	3751.44	10.3175% [Uniprot]	SWISS-PROT45:Endoplasmic reticulum membrane. GO: Q0005783 ; C:Endoplasmic reticulum; Evidence:NAS.
SSRD_RAT	E.R.	3781.62	10.9827% [Uniprot]	SWISS-PROT45>Type I membrane protein. Endoplasmic reticulum.

Normalized Feature Values

图 2-62 cds_5_Wolf.PNG

CDS_6 亚细胞定位如下：

Summary of 1 predicted sequences from Non-Plant

Predictions list. Use the [help page](#) for more detailed description of the output page.

Predicted proteins

YP_009724393.1

Prediction: Other

Protein type	Other	Signal peptide	Mitochondrial transfer peptide
Likelihood	0.9998	0.0001	0.0001

[Download: PNG / EPS / Tabular](#)

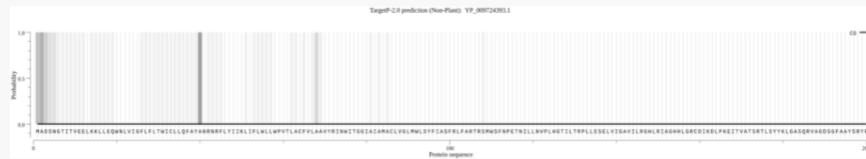


图 2-63 cds_6_TargetP.PNG

id	site	distance	identity	comments
CLF5_HUMAN	plas	892.055	15.2838% (Uniprot) SWISS-PROT45:Integral membrane protein.	
OPSP_ICTPU	plas	1012.78	16.185% (Uniprot) SWISS-PROT45:Integral membrane protein.	
OXB2_HUMAN	plas	1099.71	14.7436% (Uniprot) SWISS-PROT45:Integral membrane protein. GO: 0016021 ; C:integral to membrane; Evidence:NAS.	
FML1_HUMAN	plas	1170.1	14.245% (Uniprot) SWISS-PROT45:Integral membrane protein. GO: 0005886 ; C:plasma membrane; Evidence:TAS.	
O5H2_HUMAN	plas	1347.92	18.7702% (Uniprot) SWISS-PROT45:Integral membrane protein.	
FML2_HUMAN	plas	1358.06	14.6893% (Uniprot) SWISS-PROT45:Integral membrane protein. GO: 0005886 ; C:plasma membrane; Evidence:TAS.	
O9Q1_HUMAN	plas	1580.07	16.4516% (Uniprot) SWISS-PROT45:Integral membrane protein.	
OZB4_HUMAN	plas	1743.61	15.047% (Uniprot) SWISS-PROT45:Integral membrane protein.	
CLF5_MOUSE	plas	1786.01	14.8649% (Uniprot) SWISS-PROT45:Integral membrane protein.	
O4X1_HUMAN	plas	1814.96	18.0328% (Uniprot) SWISS-PROT45:Integral membrane protein.	
O4CF_HUMAN	plas	1818.37	16.1392% (Uniprot) SWISS-PROT45:Integral membrane protein.	
OXE1_HUMAN	plas	1819.55	17.6656% (Uniprot) SWISS-PROT45:Integral membrane protein.	
OYE6_HUMAN	plas	1865.24	16.2939% (Uniprot) SWISS-PROT45:Integral membrane protein.	
MAL_HUMAN	plas	1898.09	18.0118% (Uniprot) SWISS-PROT45:Integral membrane protein. GO: 0005624 ; C:membrane fraction; Evidence:TAS.	
OR04_MOUSE	plas	1940.16	16.879% (Uniprot) SWISS-PROT45:Integral membrane protein.	
O4D6_HUMAN	plas	2013.69	15.2866% (Uniprot) SWISS-PROT45:Integral membrane protein.	
FML1_MOUSE	plas	2026.62	14.5299% (Uniprot) SWISS-PROT45:Integral membrane protein.	
OYA1_HUMAN	plas	2028.8	18.9103% (Uniprot) SWISS-PROT45:Integral membrane protein. GO: 0005887 ; C:integral to plasma membrane; Evidence:TAS.	
OCD3_HUMAN	plas	2032.6	16.4557% (Uniprot) SWISS-PROT45:Integral membrane protein.	
OYD1_HUMAN	plas	2072.4	16.3522% (Uniprot) SWISS-PROT45:Integral membrane protein. GO: 0016021 ; C:integral to membrane; Evidence:NAS.	
OAC1_HUMAN	plas	2079.62	16.3462% (Uniprot) SWISS-PROT45:Integral membrane protein.	
OXE2_RAT	plas	2089.18	15.3125% (Uniprot) SWISS-PROT45:Integral membrane protein.	
INX8_CAEEL	plas	2095.23	15.1436% (Uniprot) SWISS-PROT45:Integral membrane protein.	
VATO_HUMAN	lyso	2119.66	12.1076% (Uniprot) SWISS-PROT45:Multipass integral membrane protein. Vacuolar. GO: 0016021 ; C:integral to membrane; Evidence:TAS.	
OYE1_HUMAN	plas	2135.83	14.9351% (Uniprot) SWISS-PROT45:Integral membrane protein.	
VATL_NEPMO	lyso	2157.2	16.2162% (Uniprot) SWISS-PROT45:Multipass integral membrane protein. Vacuolar.	
T2RG_HUMAN	plas	2169.27	12.7148% (Uniprot) SWISS-PROT45:Integral membrane protein. GO: 0016021 ; C:integral to membrane; Evidence:NAS.	
OCD2_HUMAN	plas	2197.95	14.9837% (Uniprot) SWISS-PROT45:Integral membrane protein.	
VAL2_CAEEL	lyso	2207.63	13.964% (Uniprot) SWISS-PROT45:Integral membrane protein. Vacuolar. GO: 0016021 ; C:integral to membrane; Evidence:NAS.	
O8K3_HUMAN	plas	2223.33	14.4231% (Uniprot) SWISS-PROT45:Integral membrane protein.	
NTCI_RABIT	plas	2224.12	15.8501% (Uniprot) SWISS-PROT45:Integral membrane protein.	
O4C3_HUMAN	plas	2224.7	15.894% (Uniprot) SWISS-PROT45:Integral membrane protein.	

Normalized Feature Values

id	site	iPSORT										PSORT Features										Amino Acid Content		Misc.	
		-1	25	MxHy1	30	act	alm	dna	gvh	leu	mNt	mip	mit	myr	nuc	rib	rnp	tms	tyr	vac	C	I	K	S	length
YP_009724393.1	plas?	18	63	50	25	44	86	46	49	26	17	49	29	50	50	76	49	48	40	96	15	42	29		
CLF5_HUMAN	plas	30	51	50	14	44	77	46	49	26	59	49	29	50	50	78	49	48	39	83	4	49	30		
OPSP_ICTPU	plas	11	50	50	21	44	84	46	49	26	25	49	29	50	50	82	49	48	60	92	8	39	49		

图 2-64 cds_6_Wolf.PNG

CDS_7 亚细胞定位如下：

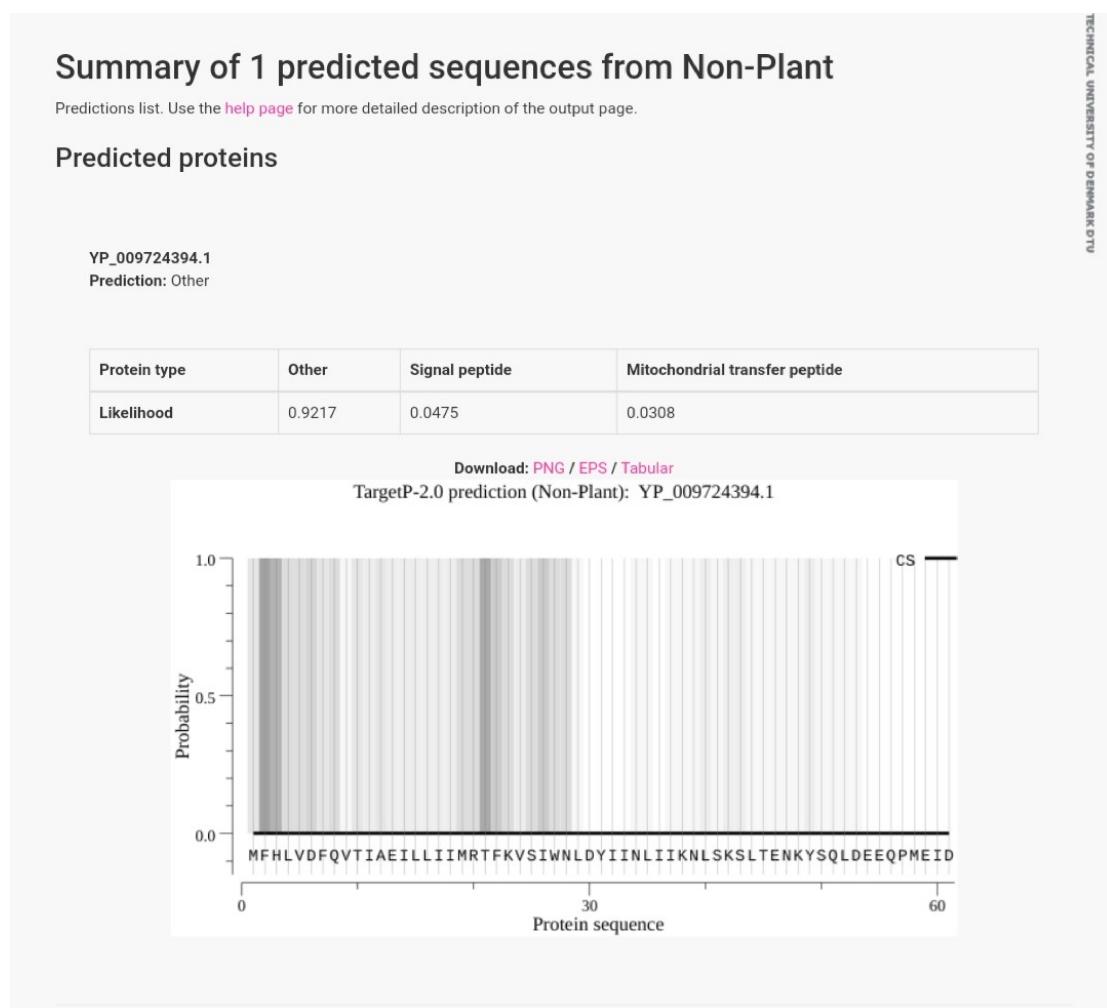


图 2-65 cds_7_TargetP.PNG

id	site	distance	identity	comments	
DL2A_HUMAN	cyto	2736.38	9.375%	[Uniprot] SWISS-PROT45:Cytoplasmic. GO: 0005868 ; C:cytoplasmic dynein complex; Evidence:NAS.	
MD21_HUMAN	nucl	2901.51	10.731%	[Uniprot] SWISS-PROT45:Nuclear. GO: 0000776 ; C:kinetochore; Evidence:TAS.	
DL2B_HUMAN	cyto	3143.3	13.5417%	[Uniprot] SWISS-PROT45:Cytoplasmic. GO: 0005868 ; C:cytoplasmic dynein complex; Evidence:NAS.	
DL2A_MOUSE	cyto	3174.71	9.375%	[Uniprot] SWISS-PROT45:Cytoplasmic. GO: 0005868 ; C:cytoplasmic dynein complex; Evidence:ISS.	
ODC_HUMAN	mito	3195.63	9.36455%	[Uniprot] SWISS-PROT45:Integral membrane protein. Mitochondrial inner membrane.	
DH1_MOUSE	pero	3205.06	7.87671%	[Uniprot] SWISS-PROT45:Microsomal.	
DL2B_MOUSE	cyto	3227.51	13.5417%	[Uniprot] SWISS-PROT45:Cytoplasmic. GO: 0005868 ; C:cytoplasmic dynein complex; Evidence:ISS.	
RER1_HUMAN	golg	3276.57	9.69388%	[Uniprot] SWISS-PROT45:Integral membrane protein. Golgi. GO: 0030173 ; C:integral to Golgi membrane; Evidence:IDA.	
RFA3_HUMAN	nucl	3356.86	14.876%	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005662 ; C:DNA replication factor A complex; Evidence:TAS.	
FAB2_AMBBE	cyto	3452.99	8.8%	[Uniprot] SWISS-PROT45:Cytoplasmic.	
ADT3_BOVIN	mito	3474.32	5.7047%	[Uniprot] SWISS-PROT45:Integral membrane protein. Mitochondrial inner membrane.	
ATPF_BOVIN	mito	3494.91	10.2804%	[Uniprot] SWISS-PROT45:Mitochondrial.	
ADT3_HUMAN	mito	3525.26	6.04027%	[Uniprot] SWISS-PROT45:Integral membrane protein. Mitochondrial inner membrane. GO: 0005744 ; C:mitochondrial inner membrane pres	
DH1_SAISC	pero	3735.67	6.52921%	[Uniprot] SWISS-PROT45:Microsomal.	
ADT2_MOUSE	mito	3871.62	7.38255%	[Uniprot] SWISS-PROT45:Integral membrane protein. Mitochondrial inner membrane.	
FABL_PIG	cyto	3880.4	10.2362%	[Uniprot] SWISS-PROT45:Cytoplasmic.	
ADT2_RAT	mito	4008.25	7.38255%	[Uniprot] SWISS-PROT45:Integral membrane protein. Mitochondrial inner membrane.	
ADT2_HUMAN	mito	4034.54	7.38255%	[Uniprot] SWISS-PROT45:Integral membrane protein. Mitochondrial inner membrane. GO: 0005887 ; C:integral to plasma membrane; Evid	
PHB_HUMAN	cyto	4037.46	8.82353%	[Uniprot] SWISS-PROT45:Cytoplasmic.	
PHB_MOUSE	cyto	4037.46	8.82353%	[Uniprot] SWISS-PROT45:Cytoplasmic.	
CECB_DROME	extr	4084.11	15.625%	[Uniprot] SWISS-PROT45:Secreted. GO: 0005576 ; C:extracellular; Evidence:IDA.	
CECD_ANTP	extr	4091.36	9.83607%	[Uniprot] SWISS-PROT45:Secreted.	
FABL_HUMAN	cyto	4139.43	10.2362%	[Uniprot] SWISS-PROT45:Cytoplasmic.	
CECB_DROMA	extr	4235.52	15.625%	[Uniprot] SWISS-PROT45:Secreted. GO: 0005576 ; C:extracellular; Evidence:ISS.	
KAD_CYPCA	cyto	4361.15	7.77202%	[Uniprot] SWISS-PROT45:Cytoplasmic.	
ADT1_MOUSE	mito	4375.89	5.36913%	[Uniprot] SWISS-PROT45:Integral membrane protein. Mitochondrial inner membrane.	
DH1_SHEEP	pero	4391.42	6.50685%	[Uniprot] SWISS-PROT45:Microsomal.	
RLA0_DROME	cyto_nucl	4423.9	8.20189%	[Uniprot] SWISS-PROT45:Nuclear and cytoplasmic. GO: 0005830 ; C:cytosolic ribosome (sensu Eukarya); Evidence:NAS.	
NDKA_RAT	cyto_nucl	4439.17	11.8421%	[Uniprot] SWISS-PROT45:Nuclear and cytoplasmic.	
ADT1_RAT	mito	4447.1	5.36913%	[Uniprot] SWISS-PROT45:Integral membrane protein. Mitochondrial inner membrane.	
CECB_DROSE	extr	4448.72	15.625%	[Uniprot] SWISS-PROT45:Secreted. GO: 0005576 ; C:extracellular; Evidence:ISS.	
NUCG_CAEEL	mito	4546.97	7.79221%	[Uniprot] SWISS-PROT45:Mitochondrial.	

Normalized Feature Values

id	site	iPSORT		PSORT Features												Amino Acid Content				Misc.			
		-1_25	MxHy_1_30	act	alm	dna	gvh	leu	mNt	mp	mit	myr	nuc	rib	rnp	tms	tyr	vac	C	I	K	S	length
YP_009724394.1	cyto?	46	76	50	46	44	24	98	49	26	10	49	29	50	50	63	49	48	3	101	67	38	4
DL2A_HUMAN	cyto	46	62	50	89	44	44	46	49	26	6	49	29	50	50	27	49	48	3	99	76	32	10

图 2-66 cds_7_Wolf.PNG

CDS_8 亚细胞定位如下：

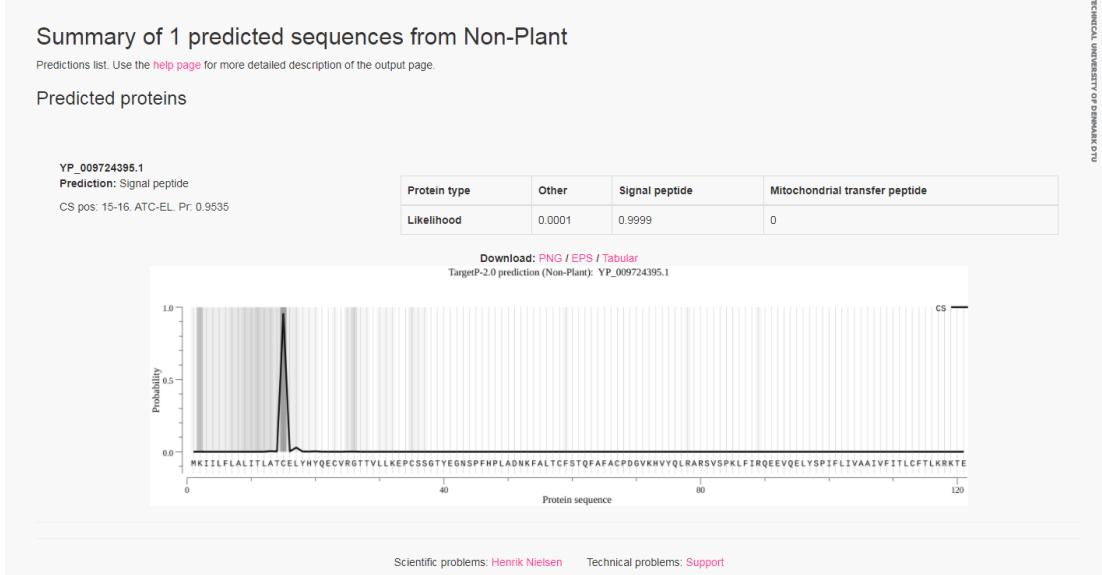


图 2-67 cds_8_TargetP.PNG

YP_009724395.1 WoLFPSORT prediction extr: 27, plas: 5

PSORT features and traditional PSORTII prediction

32 Nearest Neighbors

id	site	distance	identity	comments
ANGI_BOVIN	extr	679.636	13.5135%	[Uniprot] SWISS-PROT45:Secreted.
ALAL_LATMA	extr	708.488	11.5702%	[Uniprot] SWISS-PROT45:Secreted.
DEFI_TENMO	extr	1095.29	14.0496%	[Uniprot] SWISS-PROT45:Secreted.
CLDY_BRARE	plas	1191.64	12.55%	[Uniprot] SWISS-PROT45:Integral membrane protein.
ANGI_AOTTR	extr	1249.18	13.6054%	[Uniprot] SWISS-PROT45:Secreted.
LP2_MESMA	extr	1433.27	16.3934%	[Uniprot] SWISS-PROT45:Secreted.
MUP3_MOUSE	extr	1557.01	15.7609%	[Uniprot] SWISS-PROT45:Secreted.
ANGI_PONPY	extr	1573.38	17.0068%	[Uniprot] SWISS-PROT45:Secreted.
CXK1_CONTE	extr	1633.58	15.7025%	[Uniprot] SWISS-PROT45:Secreted.
ISK3_MOUSE	extr	1671.66	15.7025%	[Uniprot] SWISS-PROT45:Secreted.
APOD_HUMAN	extr	1771.6	14.2857%	[Uniprot] SWISS-PROT45:Secreted. GO: 0005615 ; C:extracellular space; Evidence:TAS.
CXMB_CONNMR	extr	1772.91	13.2231%	[Uniprot] SWISS-PROT45:Secreted.
MUP6_MOUSE	extr	1776.4	18.3333%	[Uniprot] SWISS-PROT45:Secreted.
MUP1_MOUSE	extr	1788	18.3333%	[Uniprot] SWISS-PROT45:Secreted.
ATT_APPLCA	extr	1863.95	12.3967%	[Uniprot] SWISS-PROT45:Secreted.
MUP2_MOUSE	extr	1926.77	18.3333%	[Uniprot] SWISS-PROT45:Secreted.
COL_MYOCO	extr	1937.44	14.0496%	[Uniprot] SWISS-PROT45:Secreted.
COAG_LIMPO	extr	2041.02	13.3333%	[Uniprot] SWISS-PROT45:Secreted.
G171_HUMAN	plas	2170.27	10.6583%	[Uniprot] SWISS-PROT45:Integral membrane protein. GO: 0016021 ; C:integral to membrane; Evidence:NAS.
IL8_PIG	extr	2210.12	17.3554%	[Uniprot] SWISS-PROT45:Secreted.
OAG1_HUMAN	plas	2213.31	10.6312%	[Uniprot] SWISS-PROT45:Integral membrane protein.
NTX2_NAJSP	extr	2246.36	14.7541%	[Uniprot] SWISS-PROT45:Secreted.
NTX1_NAJSP	extr	2298.64	14.0496%	[Uniprot] SWISS-PROT45:Secreted.
NTX4_NAJSP	extr	2362.48	14.7541%	[Uniprot] SWISS-PROT45:Secreted.
NTX3_NAJSP	extr	2414.76	14.876%	[Uniprot] SWISS-PROT45:Secreted.
IL8_RABIT	extr	2442.46	19.8347%	[Uniprot] SWISS-PROT45:Secreted.
EMP3_HUMAN	plas	2507.62	15.9509%	[Uniprot] SWISS-PROT45:Integral membrane protein.
LCA_SHEEP	extr	2510.24	13.3803%	[Uniprot] SWISS-PROT45:Secreted.
LCA_TRIVU	extr	2539.64	15.7143%	[Uniprot] SWISS-PROT45:Secreted.
P2YS_CHICK	plas	2560.76	13.3117%	[Uniprot] SWISS-PROT45:Integral membrane protein.
LCA_BOSMU	extr	2591.97	14.7887%	[Uniprot] SWISS-PROT45:Secreted.
LCA_BOVIN	extr	2591.97	14.7887%	[Uniprot] SWISS-PROT45:Secreted.

Normalized Feature Values

id	site	iPSORT		PSORT Features												Amino Acid Content				Misc.				
		-1	25	MxHy1	30	act	alm	dna	gyh	leu	mNt	nip	mit	myr	nuc	rib	rnp	tms	tyr	vac	C	I	K	S
YP_009724395.1	extr?	30		91	50	5	44	89	46	49	26	61	49	29	50	50	74	49	48	88	76	56	24	14
ANGI_BOVIN	extr	46		92	50	16	44	88	46	49	26	30	49	29	50	50	63	49	48	82	78	60	12	19
ALAL_LATMA	extr	46		aa1	cc1	aa1	aa1	cc1	cc1	cc1	cc1	cc1	cc1	cc1	cc1	cc1								

图 2-68 cds_8_Wolf.PNG

CDS_9 亚细胞定位如下：

CENTER FOR BIOLOGICALS SIGNAL PROCESSING AND ANALYSIS ■ THE TECHNICAL UNIVERSITY OF BERGAKADEMIE DÜSSELDORF

TargetP-2.0

Summary Help Downloads ▾

Summary of {{ JSON.INFO.size }} predicted sequences from {{ JSON.ORG }}

Predictions list. Use the [help page](#) for more detailed description of the output page.

Failed jobs: {{ JSON.INFO.failedjobs }} ⓘ

Predicted proteins

Protein type	{{ loc }}
Likelihood	{{ prob }}

[Download: PNG / EPS / Tabular](#)

TargetP-2.0. Starting fasta file reading... Total proteins read: 1. Organism: Non-Plant. Model parameters path: /usr/cbs/bio/src/targetp-2.0/parameters Starting protein prediction... Completed.

Instructions

1. Specify the input sequences

All the input sequences must be in one-letter amino acid code. The allowed alphabet (not case sensitive) is as follows:

A C D E F G H I K L M N P Q R S T V W Y and X (unknown)

All the alphabetic symbols not in the allowed alphabet will be converted to X before processing. All the non-alphabetic symbols, including white space and digits, will be ignored.

The sequences can be input in the following two ways:

- Paste a single sequence (just the amino acids) or a number of sequences in **FASTA** format into the upper window of the main server page.
- Select a **FASTA** file on your local disk, either by typing the file name into the lower window or by browsing the disk.

Both ways can be employed at the same time: all the specified sequences will be processed. However, there may be not more than **5,000 sequences** in one submission. The sequences may not be longer than **10,000** amino acids.

2. Customize your run

- Organism group:**
It is important for performance that you choose the correct organism group — **Archaea**, **Eukaryotes**, **Gram-negative bacteria** or **Gram-positive bacteria** — since the signal peptides of these three groups are known to differ from each other.
Gram-positive bacteria correspond to **Actinobacteria** and **Firmicutes** in the **NCBI Taxonomy**.
Gram-negative bacteria are all other **eubacteria**, except **Tenericutes** (including **Mycoplasma**), which seem to lack a type I signal peptidase and therefore do not have standard signal peptides.
- Output format:**
You can choose between four output formats:
Standard
Appropriate for most users. Shows one plot and one summary per sequence.
Short
Convenient if you submit lots of sequences. Shows only one line of output per sequence and no graphics.

3. Submit the job

Click on the "Submit" button. The status of your job (either 'queued' or 'running') will be displayed and constantly updated until it terminates and the server output appears in the browser window.
At any time during the wait you may enter your e-mail address and simply leave the window. Your job will continue; you will be notified by e-mail when it has terminated. The e-mail message will contain the URL under which the results are stored; they will remain on the server for 24 hours for you to collect them.

Example Outputs

By default the server produces the following output for each input sequence. The example below shows the output for thioredoxin domain containing protein 4 precursor (endoplasmic reticulum protein ERP44), taken from the [Uniprot](#) entry **ERP44_HUMAN**. The signal peptide prediction is consistent with the database annotation.

One annotation is attributed to each protein, the one that has the highest probability. The protein can have a Sec signal peptide (Sec/SPI), a Lipoprotein signal peptide (Sec/SPII), a Tat signal peptide (Tat/SPI) or No signal peptide at all (Other).

If a signal peptide is predicted, the cleavage site position is reported as well.

On the plot, three likelihood probabilities are reported on the plot, i.e. **SP(Sec/SPI)** / **LIGO(Sec/SPII)** / **TAT(Tat/SPI)** (depending on what type of signal peptide is predicted), CS (the cleavage site) and OTHER (the probability that the sequence does not have any kind of signal peptide).

Example: secretory protein - standard output format

Protein type	Signal Peptide (Sec/SPI)	Other
Likelihood	0.8645	0.1355

[Download: PNG / EPS / Tabular](#)

SignalP-5.0 prediction (Eukaryote): sp_Q9BS26_ERP44_HUMAN

Probability

SP (Sec/SPI) OTHER

49

Protein sequence

CS

SignalP-5.0 prediction (Eukaryote): sp_Q9BS26_ERP44_HUMAN

Example: secretory protein - short output format

Protein type	Signal Peptide (Sec/SPI)	Other
Likelihood	0.8645	0.1355

YP_009725318.1 WoLPPSORT prediction plas: 22, extr: 10

PSORT features and traditional PSORTII prediction

32 Nearest Neighbors

id	site	distance	identity	comments	
DIAP_GASAT	extr	4336.6	12.3077% [Uniprot]	SWISS-PROT45:Secreted.	
CLDY_BRARE	plas	5128.42	6.73077% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
ATT_APLCA	extr	5224.36	7.89474% [Uniprot]	SWISS-PROT45:Secreted.	
MIH_PENJP	extr	5286.84	10.4762% [Uniprot]	SWISS-PROT45:Secreted.	
CX51_CONGL	extr	5441.78	12.6984% [Uniprot]	SWISS-PROT45:Secreted.	
EMP3_HUMAN	plas	5521.32	9.81595% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
EMP2_HUMAN	plas	5546.46	9.58084% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0016021 ; C:integral to membrane; Evidence:TAS.	
UPKA_HUMAN	plas	5573.41	6.58915% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
OAG1_HUMAN	plas	5620.72	6.64452% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
O4N4_HUMAN	plas	5632.85	6.64557% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
OYK2_HUMAN	plas	5786.63	6.36943% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
OYE5_HUMAN	plas	6001.52	5.39683% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
UPKA_BOVIN	plas	6037.75	6.58915% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
O4N5_HUMAN	plas	6089.24	7.46753% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
O4N2_HUMAN	plas	6162.96	7.49186% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
OYE4_HUMAN	plas	6266.76	6.73077% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
PAFR_MOUSE	plas	6337.27	6.15836% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
DEF1_HUMAN	extr	6420.17	8.51064% [Uniprot]	SWISS-PROT45:Secreted.	
DEF3_HUMAN	extr	6420.17	8.51064% [Uniprot]	SWISS-PROT45:Secreted.	
OYJ3_HUMAN	plas	6423.34	7.3955% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
OYA1_HUMAN	plas	6453.81	5.76923% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0005887 ; C:integral to plasma membrane; Evidence:TAS.	
PAFR_HUMAN	plas	6460.82	5.84795% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0005887 ; C:integral to plasma membrane; Evidence:TAS.	
O2A7_HUMAN	plas	6479.89	7.09677% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
CLD9_HUMAN	plas	6483.85	7.37327% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
ODG1_HUMAN	plas	6539.55	5.53746% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
CXET_CONTE	extr	6597.43	20.8955% [Uniprot]	SWISS-PROT45:Secreted.	
APAM_APIME	extr	6633.5	19.5652% [Uniprot]	SWISS-PROT45:Secreted.	
OYE1_HUMAN	plas	6671.97	5.84416% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
M6A_HUMAN	plas	6672.38	5.39568% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
M6A_MOUSE	plas	6672.38	5.39568% [Uniprot]	SWISS-PROT45:Integral membrane protein.	
ISK4_PIG	extr	6681.07	10.4651% [Uniprot]	SWISS-PROT45:Secreted.	
CXK_CONVR	extr	6684.68	16.4179% [Uniprot]	SWISS-PROT45:Secreted.	

Normalized Feature Values

id	site	iPSORT										PSORT Features										Amino Acid Content				Misc.
		-1	25	MxHy1	30	act	alm	dna	gyh	leu	mNt	mip	mit	myr	nuc	rib	rnp	tms	tyr	vac	C	I	K	S		
YP_009725318.1	plas?	18	100	50	0	44	67	46	49	26	39	49	29	50	50	63	99	48	86	100	1	11	2			
DIAP_GASAT	extr	46	97	50	27	44	75	46	49	26	59	49	29	50	50	63	49	48	95	88	3	4	5			
CLDY_BRARE	plas	30	82	50	10	44	77	46	49	26	28	49	29	50	50	78	49	48	89	88	32	24	27			

图 2-70 cds_9_Wolf.PNG

CDS_10 亚细胞定位如下：

CENTER FOR BIOLOGICALS SIGNAL PROCESSING AND ANALYSIS ■ THE TECHNICAL UNIVERSITY OF BERGAKADEMIE DORTMUND

CENTROFORBIOLOGICALSIGNALPROCESSINGANDANALYSIS	EVENTS	NEWS	RESEARCH GROUPS	CBS PREDICTION SERVERS	CBS DATA SETS	PUBLICATIONS	EDUCATION
STAFF	CONTACT	ABOUT CBS	INTERNAL	CBS BIOINFORMATICS TOOLS	CBS COURSES	OTHER BIOINFORMATICS LINKS	

CBS >> CBS Prediction Servers >> TargetP-2.0 >> Results

TargetP-2.0

Summary
Help
Downloads ▾

Summary of {{ JSON.INFO.size }} predicted sequences from {{ JSON.ORG }}

Predictions list. Use the [help page](#) for more detailed description of the output page.

Failed jobs: {{ JSON.INFO.failedjobs }} ⓘ

Predicted proteins

{{ sequence.Name }}
Prediction: {{ sequence.Prediction }}
{{ sequence.CS_pos }}

Protein type	{{ loc }}
Likelihood	{{ prob }}

[Download: PNG / EPS / Tabular](#)

TargetP-2.0. Starting fasta file reading... Total proteins read: 1. Organism: Non-Plant. Model parameters path: /usr/cbs/bio/src/targetp-2.0/parameters Starting protein prediction... Completed.

Instructions

1. Specify the input sequences

All the input sequences must be in one-letter amino acid code. The allowed alphabet (not case sensitive) is as follows:

A C D E F G H I K L M N P Q R S T V W Y and X (unknown)

All the alphabetic symbols not in the allowed alphabet will be converted to X before processing. All the non-alphabetic symbols, including white space and digits, will be ignored.

The sequences can be input in the following two ways:

- Paste a single sequence (just the amino acids) or a number of sequences in **FASTA** format into the upper window of the main server page.
- Select a **FASTA** file on your local disk, either by typing the file name into the lower window or by browsing the disk.

Both ways can be employed at the same time: all the specified sequences will be processed. However, there may be not more than **5,000 sequences** in one submission. The sequences may not be longer than **10,000** amino acids.

2. Customize your run

- Organism group:**
It is important for performance that you choose the correct organism group — **Archaea**, **Eukaryotes**, **Gram-negative bacteria** or **Gram-positive bacteria** — since the signal peptides of these three groups are known to differ from each other.
Gram-positive bacteria correspond to **Actinobacteria** and **Firmicutes** in the **NCBI Taxonomy**.
Gram-negative bacteria are all other **eubacteria**, except **Tenericutes** (including **Mycoplasma**), which seem to lack a type I signal peptidase and therefore do not have standard signal peptides.
- Output format:**
You can choose between four output formats:
Standard
Appropriate for most users. Shows one plot and one summary per sequence.
Short
Convenient if you submit lots of sequences. Shows only one line of output per sequence and no graphics.

3. Submit the job

Click on the "**Submit**" button. The status of your job (either 'queued' or 'running') will be displayed and constantly updated until it terminates and the server output appears in the browser window.
At any time during the wait you may enter your e-mail address and simply leave the window. Your job will continue; you will be notified by e-mail when it has terminated. The e-mail message will contain the URL under which the results are stored; they will remain on the server for 24 hours for you to collect them.

Example Outputs

By default the server produces the following output for each input sequence. The example below shows the output for thioredoxin domain containing protein 4 precursor (endoplasmic reticulum protein ERP44), taken from the [Uniprot](#) entry [ERP44_HUMAN](#). The signal peptide prediction is consistent with the database annotation.

One annotation is attributed to each protein, the one that has the highest probability. The protein can have a Sec signal peptide (Sec/SP), a Lipoprotein signal peptide (Sec/SPII), a Tat signal peptide (Tat/SP) or No signal peptide at all (Other).

If a signal peptide is predicted, the cleavage site position is reported as well.

On the plot, three likelihood probabilities are reported on the plot, i.e. **SP(Sec/SP)** / **LIP(Sec/SPII)** / **TAT(Tat/SP)** (depending on what type of signal peptide is predicted), CS (the cleavage site) and OTHER (the probability that the sequence does not have any kind of signal peptide).

Example: secretory protein - standard output format

sp_098326_ERP44_HUMAN
Predictions: Signal peptide (Sec/SP)
Cleavage site between pos. 29 and 30: VTT-E; Probability: 0.6938

Protein type	Signal Peptide (Sec/SP)	Other
Likelihood	0.8645	0.1355

[Download: PNG / EPS / Tabular](#)

SignalP-5.0 prediction (Eukaryote): sp_098326_ERP44_HUMAN

Example: secretory protein - short output format

sp_098326_ERP44_HUMAN
Prediction: Signal peptide (Sec/SP)
Cleavage site between pos. 29 and 30: VTT-E; Probability: 0.6938

Protein type	Signal Peptide (Sec/SP)	Other
Likelihood	0.8645	0.1355

queryProtein WoLFPSORT prediction plas: 16, extr: 14, lyso: 2

PSORT features and traditional PSORTII prediction

32 Nearest Neighbors				
id	site	distance	identity	comments
MAK2_MESMA	extr	1618.27	14.2857% [Uniprot]	SWISS-PROT45:Secreted.
CRVP_TRIJE	extr	1822.11	13.3333% [Uniprot]	SWISS-PROT45:Secreted.
ASFP_BOVIN	extr	1992.83	11.9403% [Uniprot]	SWISS-PROT45:Secreted.
CRVP_AKGHA	extr	2034	10.4167% [Uniprot]	SWISS-PROT45:Secreted.
PA21_OPHHA	extr	2149.57	12.5828% [Uniprot]	SWISS-PROT45:Secreted.
O4X2_HUMAN	plas	2253.8	10.231% [Uniprot]	SWISS-PROT45:Integral membrane protein.
CRVP_TRIMU	extr	2653.75	10.8333% [Uniprot]	SWISS-PROT45:Secreted.
UPKA_BOVIN	plas	2694.82	12.4031% [Uniprot]	SWISS-PROT45:Integral membrane protein.
OCD2_HUMAN	plas	2718.62	12.3779% [Uniprot]	SWISS-PROT45:Integral membrane protein.
CATB_RAT	lyso	2816.04	10.3245% [Uniprot]	SWISS-PROT45:Lysosomal.
O2A1_HUMAN	plas	2858.7	10.3226% [Uniprot]	SWISS-PROT45:Integral membrane protein.
CLDR_HUMAN	plas	2915.4	12.3288% [Uniprot]	SWISS-PROT45:Integral membrane protein.
P2Y5_CHICK	plas	2936.52	11.039% [Uniprot]	SWISS-PROT45:Integral membrane protein.
PIF_BOMMO	extr	2967.87	13.4454% [Uniprot]	SWISS-PROT45:Secreted.
OXA17_HUMAN	plas	2999.94	10.2564% [Uniprot]	SWISS-PROT45:Integral membrane protein.
O2B6_HUMAN	plas	3032.54	9.90415% [Uniprot]	SWISS-PROT45:Integral membrane protein.
O7AA_HUMAN	plas	3037.12	10.356% [Uniprot]	SWISS-PROT45:Integral membrane protein.
APAM_APIME	extr	3063.28	12.605% [Uniprot]	SWISS-PROT45:Secreted.
PA23_BUNMU	extr	3093.65	15.6463% [Uniprot]	SWISS-PROT45:Secreted.
COGS_HYPLI	extr	3097.76	12.3077% [Uniprot]	SWISS-PROT45:Secreted.
ATT_APLCA	extr	3105.57	16.8067% [Uniprot]	SWISS-PROT45:Secreted.
OCD3_HUMAN	plas	3115.81	13.2911% [Uniprot]	SWISS-PROT45:Integral membrane protein.
OZB4_HUMAN	plas	3130.95	10.3448% [Uniprot]	SWISS-PROT45:Integral membrane protein.
CRVP_AKGPI	extr	3134.61	12.9167% [Uniprot]	SWISS-PROT45:Secreted.
IM23_DROME	extr	3139.21	10.8333% [Uniprot]	SWISS-PROT45:Secreted. GO: 0005576 : C:extracellular; Evidence:IDA.
COL_MYOCO	extr	3153.22	12.5% [Uniprot]	SWISS-PROT45:Secreted.
VATO_HUMAN	lyso	3161.95	14.1463% [Uniprot]	SWISS-PROT45:Multipass integral membrane protein. Vacuolar. GO: 0016021 : C:integral to membrane; Evidence:TAS.
O4X1_HUMAN	plas	3166.1	10.8197% [Uniprot]	SWISS-PROT45:Integral membrane protein.
LMIP_HUMAN	plas	3166.31	13.2184% [Uniprot]	SWISS-PROT45:Integral membrane protein. GO: 0030054 : C:cell junction; Evidence:NAS.
UPKA_HUMAN	plas	3184.79	12.0155% [Uniprot]	SWISS-PROT45:Integral membrane protein.
O2AC_HUMAN	plas	3190.45	10.9677% [Uniprot]	SWISS-PROT45:Integral membrane protein.
O4CF_HUMAN	plas	3229.97	12.6582% [Uniprot]	SWISS-PROT45:Integral membrane protein.

Figure 2.72 shows a sequence logo for PSORT Features. The x-axis lists amino acids: gvh, leu, mNt, mip, mit, myr, nuc, rib, rnp, tms, tyr, vac. The y-axis represents normalized feature values. The logo shows high conservation for 'g' at position 1, 'v' at position 2, 'h' at position 3, 'l' at position 4, 'e' at position 5, 'u' at position 6, 'm' at position 7, 'N' at position 8, 't' at position 9, 'i' at position 10, 'p' at position 11, 'i' at position 12, 't' at position 13, 'y' at position 14, 'r' at position 15, 'b' at position 16, 'r' at position 17, 'n' at position 18, 'p' at position 19, 't' at position 20, 'm' at position 21, 's' at position 22, 'y' at position 23, 't' at position 24, 'r' at position 25, 'a' at position 26, and 'c' at position 27.

CDS 11 亚细胞定位如下：



图 2-73 cds 11 TargetP.PNG

queryProtein WolFPsort prediction nucl: 29, cysk: 2, cyto: 1

PSORT features and traditional PSORTIII prediction

32 Nearest Neighbors				
id	site	distance	identity	comments
CYL2_HUMAN	cysk	1292.87	15_27459	[Uniprot] SWISS-PROT45:Calyx; sperm head cytoskeletal structure.
RGE5_HUMAN	nucl	1650.59	14_65526	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005634 : C:nucleus; Evidence:IDA.
LAM3_XENLA	nucl	1940.08	14_87989	[Uniprot] SWISS-PROT45:Nucleoplasmic side of the inner nuclear membrane.
LAM1_HUMAN	nucl	1991.08	13_50393	[Uniprot] SWISS-PROT45:Nucleoplasmic side of the inner nuclear membrane. GO: 0005633 : C:lamin filament; Evidence:TAS.
TTL1_MOUSE	nucl	1995.29	13_01678	[Uniprot] SWISS-PROT45:Nuclear.
LAMC_MOUSE	nucl	2080.8	14_82719	[Uniprot] SWISS-PROT45:Nucleoplasmic side of the inner nuclear membrane. GO: 0005638 : C:lamin filament; Evidence:IDA.
LAM1_MOUSE	nucl	2083.03	14_13975	[Uniprot] SWISS-PROT45:Nucleoplasmic side of the inner nuclear membrane. GO: 0005634 : C:nucleus; Evidence:IDA.
CYL2_BOVIN	cysk	2203.53	14_34458	[Uniprot] SWISS-PROT45:Calyx; sperm head cytoskeletal structure.
LAMC_MOUSE	nucl	2389.39	13_24049	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005634 : C:nucleus; Evidence:IDA.
ATRX_DROME	nucl	2465	8_314265	[Uniprot] SWISS-PROT45:Nuclear.
ZHX1_RAT	nucl	2548.39	11_56938	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005634 : C:nucleus; Evidence:IDA.
NCR2_MOUSE	nucl	2589.09	6_027516	[Uniprot] SWISS-PROT45:Nuclear.
MTA1_MOUSE	nucl	2619.88	12_16783	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005634 : C:nucleus, Evidence:IDA.
LAM2_CHICK	nucl	2630.75	12_16679	[Uniprot] SWISS-PROT45:Nucleoplasmic side of the inner nuclear membrane.
T2FA_MOUSE	nucl	2649.28	14_12219	[Uniprot] SWISS-PROT45:Nuclear.
LAM1_HUMAN	nucl	2804.68	13_8	[Uniprot] SWISS-PROT45:Nucleoplasmic side of the inner nuclear membrane. GO: 0005652 : C:nuclear lamina; Evidence:NAS.
ILF3_HUMAN	nucl	2805.26	11_96876	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005634 : C:nucleus; Evidence:TAS.
LAM2_XENLA	nucl	2855.38	14_44626	[Uniprot] SWISS-PROT45:Nucleoplasmic side of the inner nuclear membrane.
NCR2_HUMAN	nucl	2919.6	6_760839	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005634 : C:nucleus; Evidence:TAS.
LAM1_HUMAN	nucl	2995.72	13_85549	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005633 : C:lamin filament; Evidence:TAS.
DSRA_HUMAN	nucl	3034.6	11_50088	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005634 : C:nucleus; Evidence:TAS.
T2FA_HUMAN	nucl	3079.92	14_31539	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005674 : C:transcription factor TFIIF complex; Evidence:TAS.
Y11_MOUSE	nucl	3141.16	11_69458	[Uniprot] SWISS-PROT45:Nuclear.
DD21_MOUSE	nucl	3154.73	11_28089	[Uniprot] SWISS-PROT45:Nuclear; nucleolar.
SFR6_HUMAN	nucl	3157.07	12_17189	[Uniprot] SWISS-PROT45:Nuclear.
LAMA_RAT	nucl	3158.1	14_15838	[Uniprot] SWISS-PROT45:Nuclear.
LAMA_MOUSE	nucl	3241.05	14_15839	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005634 : C:nucleus; Evidence:IDA.
WAS2_HUMAN	cyto	3247.97	16_29198	[Uniprot] SWISS-PROT45:Cyttoplasmic. GO: 0013629 : C:actin cytoskeleton; Evidence:TAS.
NLPE_MOUSE	nucl	3298.26	10_02399	[Uniprot] SWISS-PROT45:Nuclear.
DSRA_RAT	nucl	3305.65	10_80859	[Uniprot] SWISS-PROT45:Nuclear.
ZHX1_HUMAN	nucl	3320.37	11_79849	[Uniprot] SWISS-PROT45:Nuclear. GO: 0005634 : C:nucleus; Evidence:IDA.
SRY_CALJA	nucl	3331.87	11_45589	[Uniprot] SWISS-PROT45:Nuclear.

图 2-74 cds_11_Wolf.PNG

CDS_12 亚细胞定位如下：

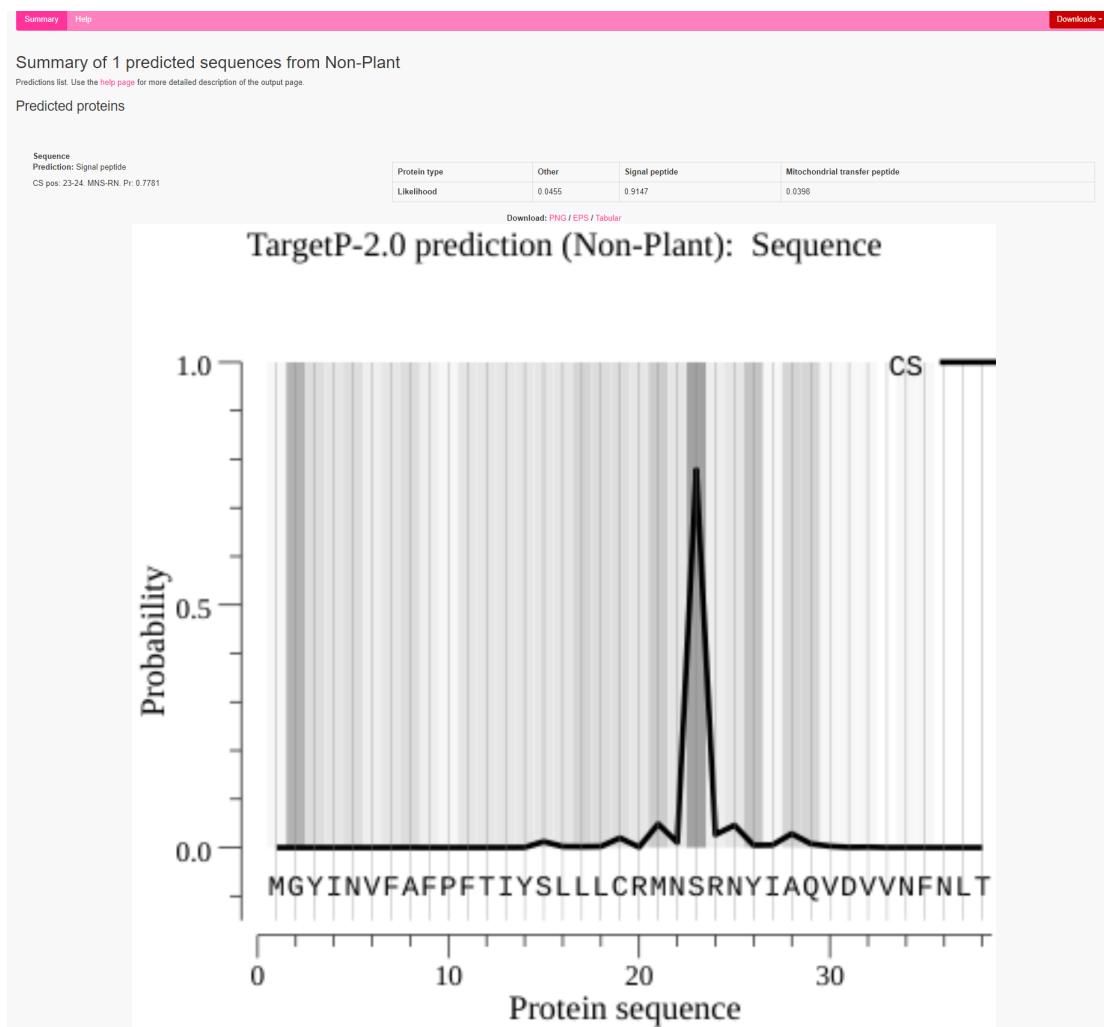


图 2-75 cds_12_TargetP.PNG

queryProtein WoLMP50K1 prediction extr: 14, plas: 12, cyto: 3.5, cyto_nucl: 3.5, nucl: 2.5					
PSORT features and traditional PSORTII prediction					
82 Nearest Neighbors					
id	site	distance	identity	comments	
P131_HUMAN	nucl	2950.0	10.79149	[Uniprot] SWISS-PROT45:Nuclear.	
APAM_APIME	extr	5311.79	15.21749	[Uniprot] SWISS-PROT45:Secreted.	
BD12_BOVIN	extr	5320.48	7.894749	[Uniprot] SWISS-PROT45:Secreted.	
CEC3_MANSE	extr	4028.2	0%	[Uniprot] SWISS-PROT45:Secreted.	
CEC1_MANSE	extr	4137.87	0%	[Uniprot] SWISS-PROT45:Secreted.	
BD01_BOVIN	extr	4147.68	5.285169	[Uniprot] SWISS-PROT45:Secreted.	
PIBR_TPLE	extr	4218.9	7.42749	[Uniprot] SWISS-PROT45:Secreted.	
BD11_BOVIN	extr	4306.8	7.894749	[Uniprot] SWISS-PROT45:Secreted.	
SCA4_HUMAN	plas	4828.04	6.98699	[Uniprot] SWISS-PROT45:Integral membrane protein.	
KAD5_HUMAN	cyto	4834.02	7.070719	[Uniprot] SWISS-PROT45:Cyttoplasmic. GO:0005829; Cytosol; Evidence:TAS.	
T145_HUMAN	plas	4838.18	6.93669	[Uniprot] SWISS-PROT45:Integral membrane protein. GO:0005624; C:membrane fraction; Evidence:TAS.	
AQP_AEDAE	plas	4835.16	5.220889	[Uniprot] SWISS-PROT45:Integral membrane protein.	
LMIP_BOVIN	plas	4924.94	9.248589	[Uniprot] SWISS-PROT45:Integral membrane protein.	
CXO1_CONVE	extr	5045.8	13.15799	[Uniprot] SWISS-PROT45:Secreted. GO:0005576; C:extracellular; Evidence:NAS.	
CEC2_MANSE	extr	5131.47	0%	[Uniprot] SWISS-PROT45:Secreted.	
CKK_CONVE	extr	5190.5	8.955229	[Uniprot] SWISS-PROT45:Secreted.	
UPKB_BOVIN	plas	5212.8	5.019319	[Uniprot] SWISS-PROT45:Integral membrane protein.	
ID5_HUMAN	nucl	5226.72	11.76479	[Uniprot] SWISS-PROT45:Nuclear.	
UPKB_MUSVI	plas	5380	5.019319	[Uniprot] SWISS-PROT45:Integral membrane protein.	
BD10_BOVIN	extr	5461.06	12.88	[Uniprot] SWISS-PROT45:Secreted.	
ODG1_HUMAN	plas	5469.12	4.885699	[Uniprot] SWISS-PROT45:Integral membrane protein.	
OAGC_HUMAN	plas	5499	5.806489	[Uniprot] SWISS-PROT45:Integral membrane protein.	
OYI3_HUMAN	plas	5502.49	5.466249	[Uniprot] SWISS-PROT45:Integral membrane protein.	
CXPH_CONVE	extr	5519.05	5.953389	[Uniprot] SWISS-PROT45:Secreted.	
PPA2_DROME	cyto	5525.83	7.926639	[Uniprot] SWISS-PROT45:Cyttoplasmic.	
SPL_CYPCA	extr	5628.96	12.88	[Uniprot] SWISS-PROT45:Secreted.	
CLDG_MOUSE	plas	5607.84	5.957499	[Uniprot] SWISS-PROT45:Integral membrane protein.	
PR40_HUMAN	cyto_nucl	5634.6	4.363649	[Uniprot] SWISS-PROT45:Cytoplasmic and nuclear; nucleolar. GO:0000176; C:nuclear exosome (RNase complex); Evidence:IDA.	
IM23_DROME	extr	5647.04	7.692319	[Uniprot] SWISS-PROT45:Secreted. GO:0005576; C:extracellular; Evidence:IDA.	
LMIP_MOUSE	plas	5650.42	9.826599	[Uniprot] SWISS-PROT45:Integral membrane protein.	
O244_HUMAN	plas	5677.93	8.161299	[Uniprot] SWISS-PROT45:Integral membrane protein.	
SDOC_CERCA	cyto	5724.23	6.578999	[Uniprot] SWISS-PROT45:Cytoplasmic.	

图 2-76 cds_12_Wolf.PNG

3 进化分析

3.1 获得 Beta 属冠状病毒同源基因组序

获得的同源基因组序所用 accession number 如下

MG596803.1	LC556375.1	MT040335.1	MT040336.1
MG772934.1	MW251310.1	MN996532.2	HG994853.1
KF294457.1	MT782115.1	FJ211859.1	KY417145.1
KF294455.1	KY417146.1	MK211376.1	AY304486.1
EU371562.1	AY278554.2	AY394987.1	AY559093.1
JX163925.1	FJ882963.1		

表 3-1 用于多序列对比的同源冠状病毒的序列号

3.2 多序列比对

由于比对结果过长，仅展示两种方法的系统发生树

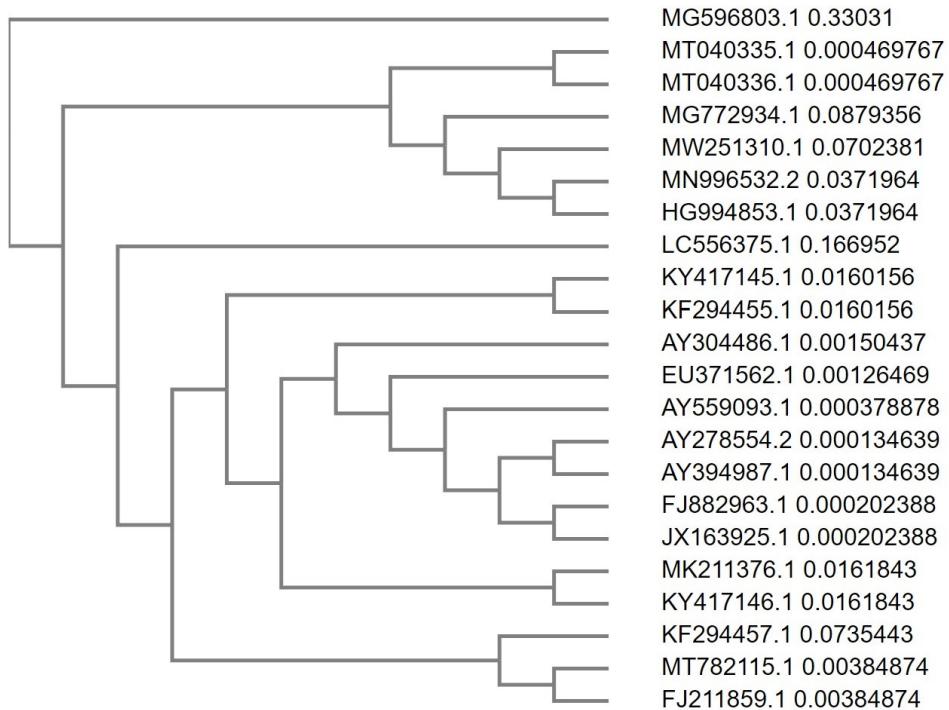


图 3-1 Clustal Omega 结果

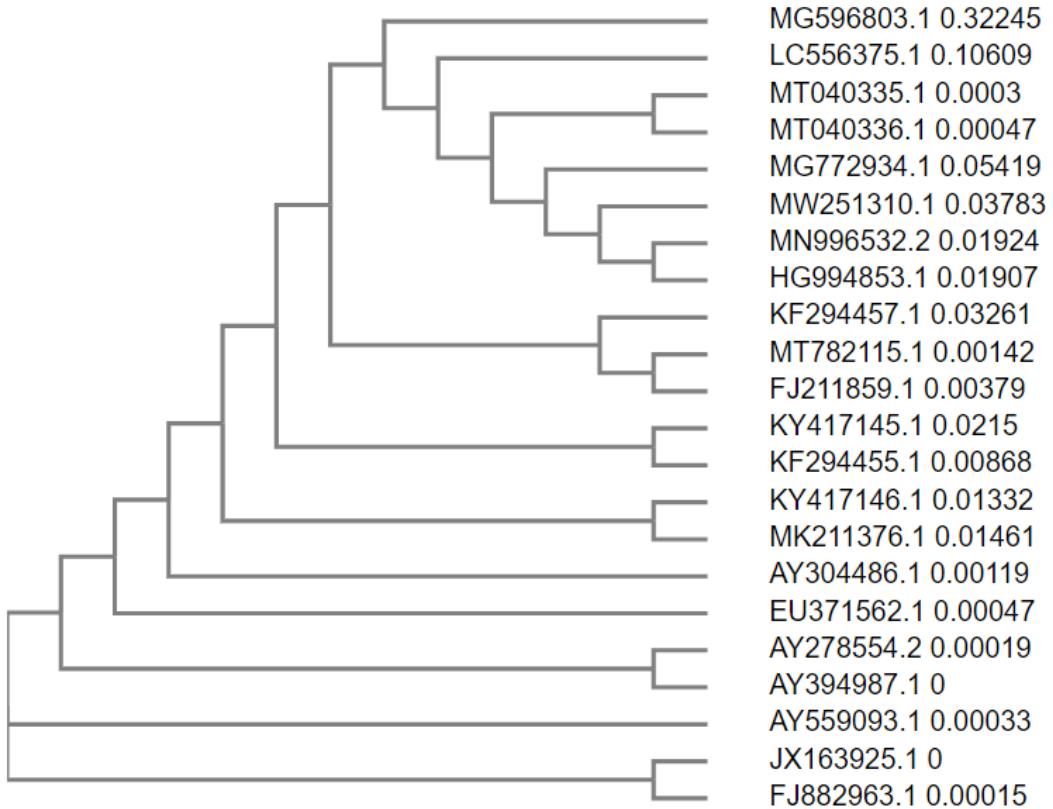


图 3-2 MUSCLE 结果

3.3 分子进化树构建

使用 NJ 法进行分子进化树构建，并开展 Bootstrap 检验如下

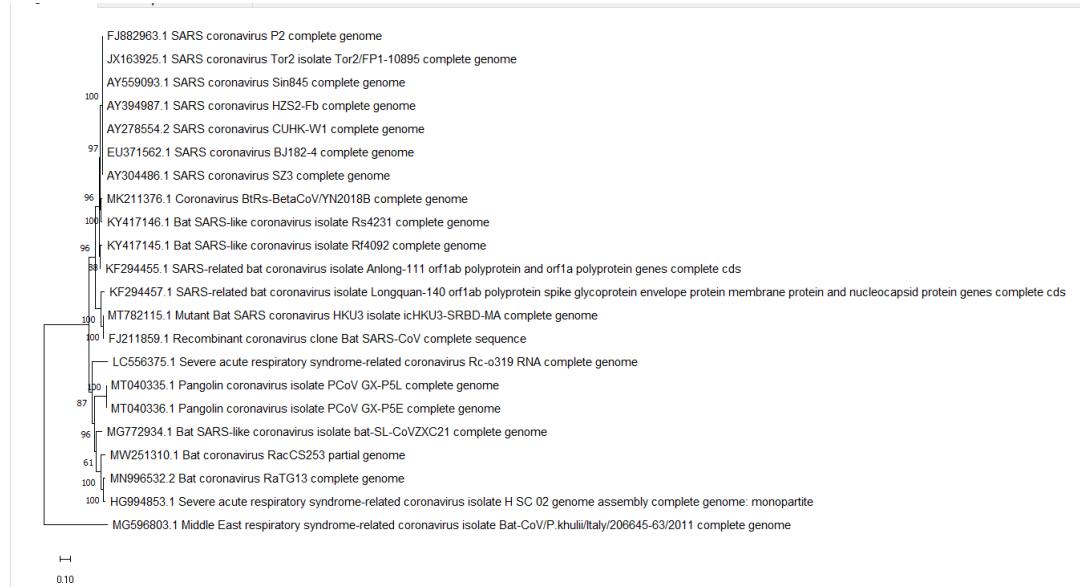


图 3-3 NJ 法构建分子进化树

4 Spike 蛋白分析

4.1 SARS-CoV-2 的 Spike 的同源蛋白质分析

通过 blast 得到 Spike 在其他冠状病毒中的同源蛋白质，通过 muscle 进行多序列比对

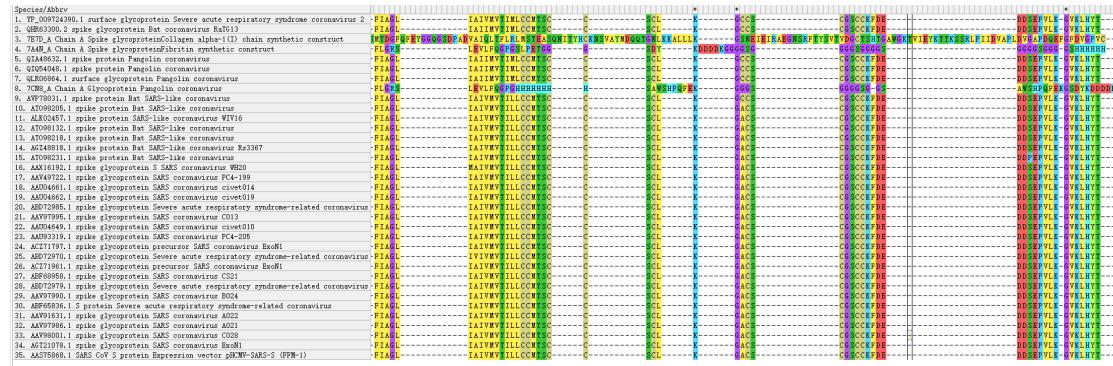


图 4-1 MUSCLE 结果

4.2 Spike 蛋白质的分子进化

分别使用 NJ, MP 和 ML 法构建分子进化树，并开展 Bootstr 检验如下

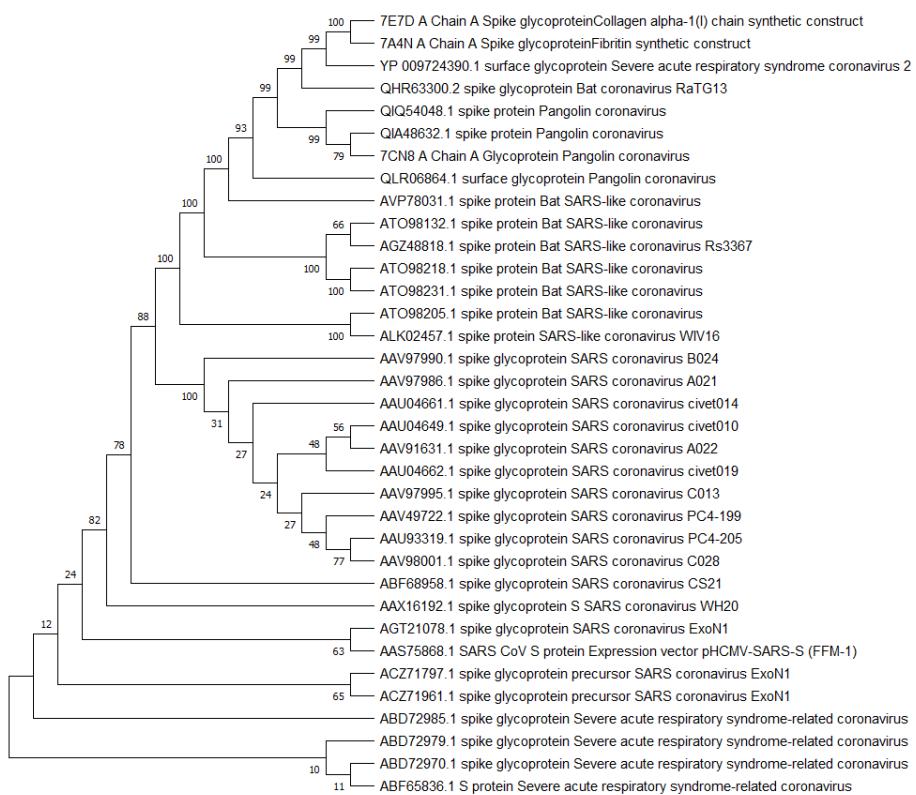


图 4-2 NJ 法.PNG

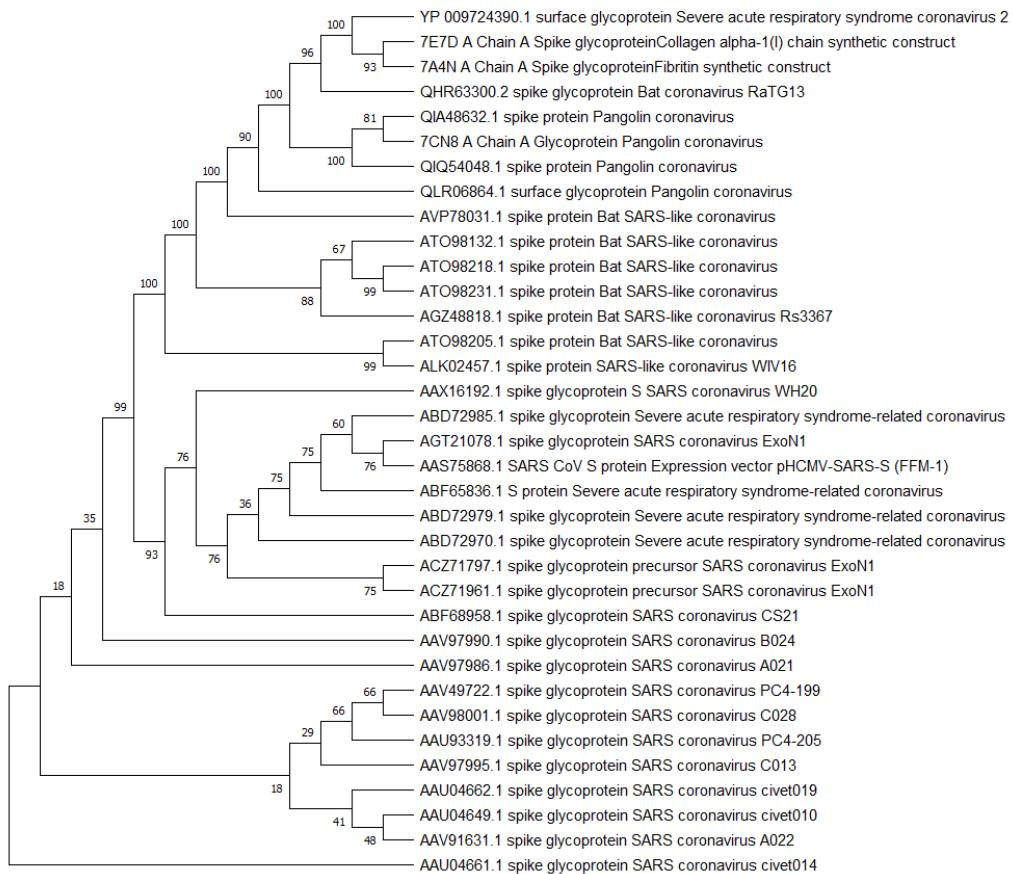


图 4-3 MP 法.PNG

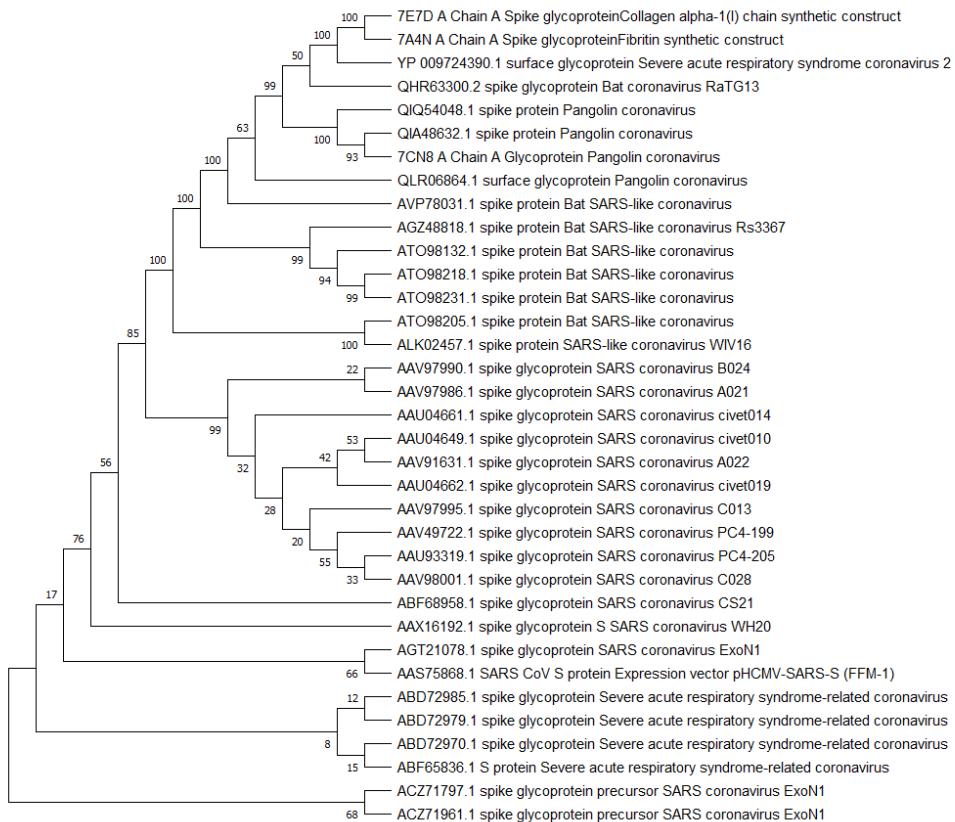


图 4-4 ML 法.PNG

4.3 Spike 蛋白质的 Furin 切割位点预测

使用 ProP 对 Spike 蛋白质的 Furin 切割位点预测结果如下

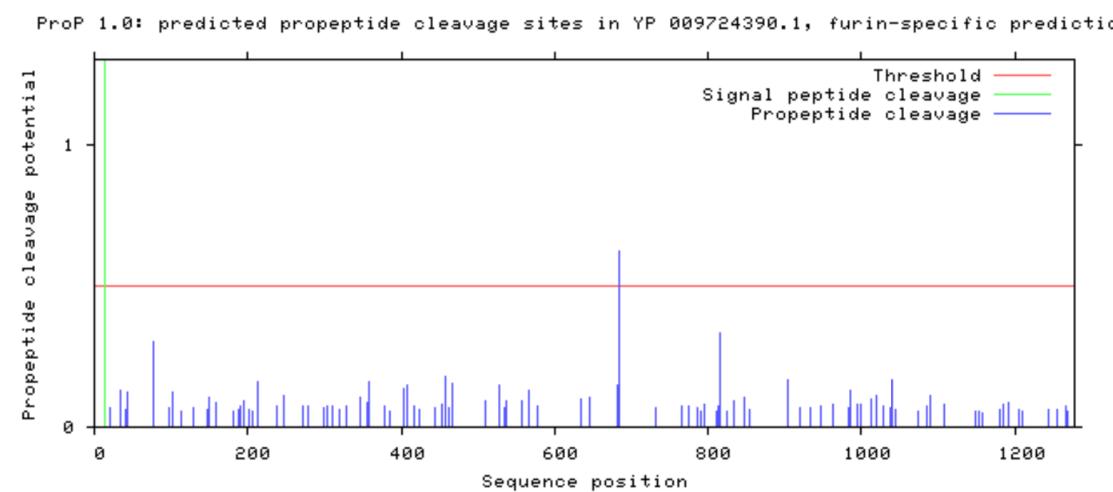


图 4-5 ProP.PNG