

华中科技大学

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 4 月 29 日

目 录

1	基因组分析	1
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	2. 编写并运行 example4-1.pl	1
1.3	3. SARS-CoV-2 的基因组序列	1
1.4	4. SARS-CoV-2 潜在编码序列的预测	2
1.5	5. 发现与 SARS-CoV-2 同源的冠状病毒	3
1.6	6. 插入片段分析	4
2	序列分析	4
2.1	1. INS1378 与 pShuttle-SN 载体的相似性	4
2.2	2. SARS-CoV-2 的蛋白质序列	4
2.3	3. 等电点与分子量分析	4
2.4	4. 功能结构域分析	4
2.5	5. 细胞亚定位分析	4

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

1.2 2. 编写并运行 example4-1.pl

编写的 example4-1.pl 如下

```
use strict;
use warnings;

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

图 1-1 example4-1.pl

1.3 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 4.SARS-CoV-2 潜在编码序列的预测

预测开放阅读框如下

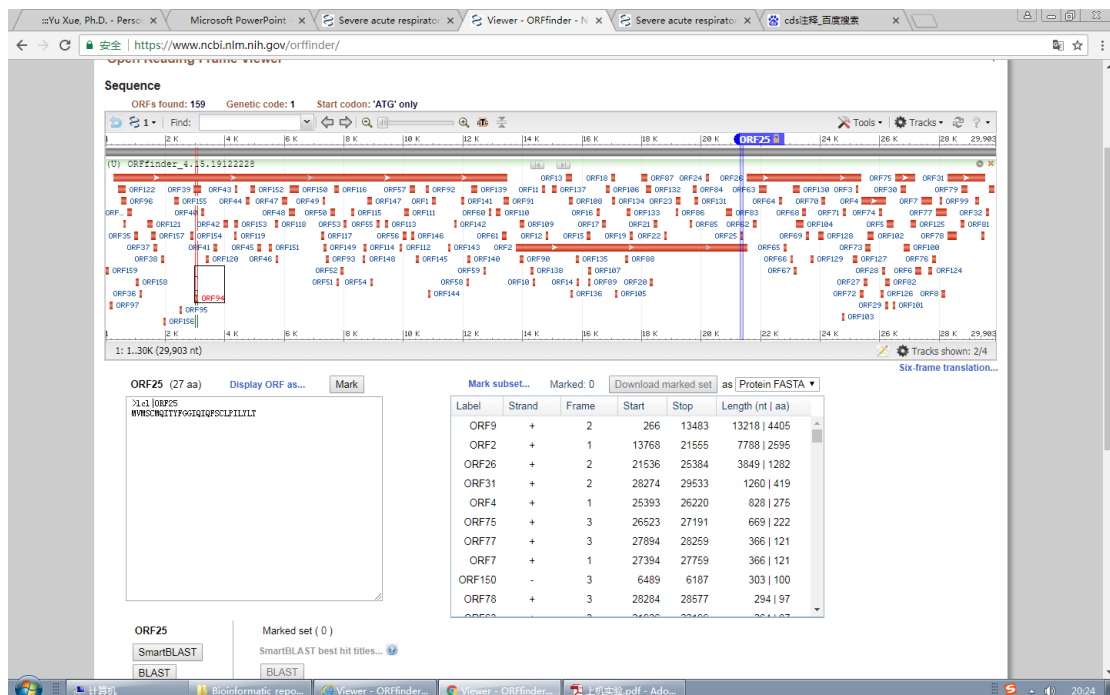


图 1-3 ORF_prediction.PNG

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="pplab; translated by -1 ribosomal frameshift"
/codon_start=1
/product="ORF1ab polyprotein"
/protein_id="YP_009724389.1"
/db_xref="GeneID:43740578"
/translation="MESLVPGFNEK THWQLSLPVLQVRDVLVRGFGDSVEEVLSEARQ
HLKDG TCGLVEVEKG VLPQLEQPYVFIKRS DARTAPHGHVMVELVAELEGIQYGRSGE
TLGVL VPHVGEIPVAYRKVLLRKNGNKGAGGHSYGADLK SFDLGDELGTDPYEDFQEN
WNTKHSSGVITRELMRELNGGAYTRYVDNNFCGPDGYPLECIKDLLARAGKASCTLSEQ
LDFIDTKRGVYOCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDTFNGECPNFVFP
LNSIIKT IQPRVEKKLDGFMGRIRSVYPVASPNECNQMCLSTLMKCDHOGETSWQTG
DFVKATCEFOG TENLTKEGATTCGYLPQNAVVKIYCPACHNSEVWPEHSLAEYHNESG
I.KTTLRKGGRTIARGGCVF SYWGCHNKCA YWVPRASANTGCNHTGVVGE GSEGI.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 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 6. 插入片段分析

2 序列分析

2.1 1.INS1378 与 pShuttle-SN 载体的相似性

2.2 2.SARS-CoV-2 的蛋白质序列

2.3 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 4. 功能结构域分析

2.5 5. 细胞亚定位分析