

# 生物信息学试验汇报

姓名：\_\_黄栋\_\_\_\_

学号：\_\_\_\_\_

指导老师：\_\_宋晓峰\_\_\_\_

南京航空航天大学

2023 年 11 月

## 试验一 生物信息数据库的检索

### 一. 试验目的:

1. 理解生物信息学的各大门户网站以及其中的重要资源。
2. 理解重要数据库的内容及构造, 理解各数据库注释的含义。
3. 以 PubMed 为例, 学会文献数据库的基本查询检索措施。

### 二. 试验内容:

#### (1) 国际与国内的生物信息中心

国际 NCBI、EBI、ExPASy, EMBL、SIB、TIGR 以及国内 CBI、BioSino 网站的熟悉及内容的理解。

核酸序列数据库: genbank/EMBL-bank/DDBJ

NCBI

EBI

EMBL

蛋白质序列数据库:

Swiss Prot 、ExPASy

Uniprot

蛋白质构造数据库:

PDB

(2) 检索练习:

**The spike protein of SARS-Corona Virus 在 NCBI 中的核酸记录序列:**

LOCUS CS244439 3897 bp DNA linear PAT

17-JUL-2023

DEFINITION Sequence 3 from Patent WO.

ACCESSION CS244439

VERSION CS244439.1 GI:84659113

KEYWORDS .

SOURCE SARS coronavirus

ORGANISM SARS coronavirus

Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

Coronaviridae; Coronavirinae; Betacoronavirus.

REFERENCE 1

AUTHORS Altmeyer,R., Nal-Rogier,B., Chan,C., Kien,F., Kam,Y.W., Siu,Y.L.,

Tse,K.S., Staropoli,I. and Manuguerra,J.C.

TITLE Nucleic acids, polypeptides, methods of expression, and immunogenic

compositions associated with sars corona virus spike protein

JOURNAL Patent: WO -A2 3 15-DEC-2023;

INSTITUT PASTEUR (FR); Hong Kong Pasteur Research Centre Limited

(CN)

FEATURES

Location/Qualifiers

source

1..3897

/organism="SARS coronavirus"

/mol\_type="unassigned DNA"

/db\_xref="taxon:227859"

CDS

44..3847

/note="unnamed protein product"

/codon\_start=1

/protein\_id="CAJ56183.1"

/db\_xref="GI:84659114"

/translation="MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRS  
DTLYLTQDLFLPFYSNVTGFHTINHTFGNPVIPFKDGIYFAATEKSNVVRGWVFGST  
MNNKSQSVIIINNSTNVVIRACNFELCDNPFFAVSKPMGTQTHTMIFDNAFNCTFEYIS  
DAFSLDVSEKSGNFKHLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLP  
LGINITNFRAILTAFSPAQDIWGTSAAYFVGYLKPTTFMLKYDENGITITDAVDCSQN  
PLAELKCSVKSFEDKGIYQTSNFRVVPBGDVVRFPNITNLCPFGEVFNATKFPSVYA  
WERKKISNCVADYSVLYNSTFFSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVR  
QIAPGQTGVIADYNYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYRYLRHGKLRPF  
ERDISNVPFSPDGKCTPPALNCYWPLNDYGFYTTTGIGYQPYRVVLSFELLNAPAT  
VCGPKLSTDLIKNCVNFNFNGLTGTGVLTPSSKRFQPF  
FGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTNASSEVAVLYQDVNCTDVST  
AIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHVDTSYECDIPIGAGICASYHTVSLL  
RSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNFSISITTEVMPVSMAKTSVDCNMYIC  
GDSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQMYKTPTLKYFGG  
FNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICAQKFNG  
LTVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQ  
NVLYENQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNF  
GAISSVLNDILSRDKVEAEVQIDRLITGRLQSLQTYVT  
LIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLMSFPQAAPHGVVFLHVTY  
VPSQERNFTTAPAICHEGKAYFPREGVVFVNGTSWFITQRNFFSPQIITTDNTFVSGNC

DVVIGIINNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDR

LNEVAKNLNESLIDLQELGKYEQYIKWPWYVWLGFIAGLIAIVMVTILLCCMTSCCS

CLKGACSCGSCCKFDEDDSEPVLKGVKLHYTGPGGDYKDDDDK"

ORIGIN

1 ctatagggcg aattgggtac cgctagcgga tccgcgcgcc accatgttta tttcctgct

61 gtttctgact ctgaccagcg gcagtgcct ggaccggtgc accactttg atgatgtgca

121 ggctcctaatactacacgc atacttctc tatgaggggc gtgtactatc ctgatgaaat

181 ttttagatcc gacactctgt atctgactca ggatctgttt ctgccattct attctaattg

241 gacaggcttt catactatta atcatacctt tggcaaccct gtgatccctt ttaaggatgg

301 catctatttt gctgccacag agaagtccaa tgtggtgcgg ggatgggtgt tcggctctac  
361 catgaacaac aagtcaccagt ccgtgattat tattaacaat tctactaatg tggatgccg  
421 agcctgtaac ttgaactgt gtgacaaccc attctttgct gtgtctaagc ccatgggcac  
481 acagacacat actatgatct tcgataatgc ctftaattgc actttcgagt acatctctga  
541 tgccttttcc ctggatgtgt ccgaaaagtc cggcaacttt aagcacctgc gagagtttgt  
601 gtttaagaat aaggatggct ttctgtatgt gtataagggc tatcagccta tcgacgtggt  
661 gcgcgatctg ccttctggct ttaacctct gaagcctatt ttaagctgc ctctgggcat  
721 taacattaca aattttcggg ccattctgac agcctttagc cctgctcagg acatttgggg  
781 cacctctgct gccgcctatt ttgtgggcta tctgaagcca actacctta tgctgaagta  
841 tgatgaaaat ggcacaatca cagatgctgt ggattgttct cagaatccac tggctgaact  
901 gaagtgtctt gtgaagagct ttgagattga caagggaatc taccagacct ctaatttccg  
961 cgtggtgccc tctggagatg tggtagatt ccctaattatt acaaacctgt gtccttttgg  
1021 agaagtgttt aatgctacta agttcccttc tgtgtatgcc tgggagagaa agaagatttc  
1081 taattgtgtg gctgattact ctgtgctgta caactccaca ttttttagca ctttaagt  
1141 ctatggcgtg tctgccacta agctgaatga tctgtcttc tccaatgtgt atgccgattc  
1201 ttttgtgtg aaggagatg atgtgagaca gatcgcccca ggacagactg gcgtgattgc  
1261 tgattacaat tataagctgc cagatgattt catgggctgt gtgctggctt ggaatactag  
1321 gaacattgat gctactcca ctggcaatta taattacaag tatcggtatc tgagacatgg  
1381 caagctgagg ccccttgaga gagacatctc taacgtgcct ttacgccctg atggcaagcc  
1441 ttgcacccca cctgctctga atgtttattg gccactgaat gattatggct ttacaccac

1501 tactggcatt ggctaccagc cttacagagt ggtgggtgctg tcttttgaac tgctgaatgc  
1561 ccctgccaca gtgtgtggac caaagctgtc cactgacctg attaagaacc agtgtgtgaa  
1621 ctttaacttt aatggactga ctggcactgg cgtgctgact cctctagca agagatttca  
1681 gccatttcag cagtttggcc gggatgtgtc tgatttact gattccgtgc gagatcctaa  
1741 gacatctgaa atcctggaca ttcccccttg ctcttttggc ggcgtgagcg tgattacacc  
1801 tggaacaaat gcttctctg aagtggctgt gctgtatcag gatgtgaact gcactgatgt  
1861 gtctacagcc atccatgccg atcagctgac accagcttgg cgcatctatt ctactggaaa  
1921 caatgtgttc cagactcagg ccggctgtct gatcggagct gagcatgtgg acacttctta  
1981 tgagtgcgac attcctattg gagctggcat ttgtgctagt taccatacag tgtctctgct  
2041 gcggagtact agccagaagt ctattgtggc ttatactatg tctctgggcg ctgatagttc  
2101 cattgcttac tctaataaca ccattgctat cctactaac tttccatta gcattactac  
2161 agaagtgatg cctgtgtcta tggctaagac ctccgtggat tgtaatatgt acatctgcgg  
2221 agattctacc gaatgtgcta atctgctgct gcagtatggc agcttttgca cacagctgaa  
2281 tcgggctctg tctggcattg ctgctgaaca ggatcgcaac acacgggaag tgttcgctca  
2341 agtgaagcag atgtataaga cccaactct gaagtatttt ggcggcttta attttccca  
2401 gatcctgcct gacctctga agcccactaa gcggtctttt attgaggacc tgctgtttaa  
2461 caaagtgaca ctggctgatg ctggctttat gaagcagtat ggcgaatgcc tgggcgatat  
2521 taatgctaga gatctgattt gtgccagaa gtccaatggc ctgacagtgc tgcctcctct  
2581 gctgactgat gatatgattg ctgcctacac tgctgctctg gtgtctggca ctgccactgc  
2641 tggatggaca ttggcgctg gcgctgctct gcagatccct ttgctatgc agatggccta

2701 tcggttcaat ggcattggag tgaccagaa tgtgctgtat gagaaccaga agcagattgc  
2761 caaccagttt aacaaggcca ttagtcagat tcaggaatcc ctgacaacaa catccactgc  
2821 cctgggcaag ctgcaggacg tggatgaacca gaatgctcag gccctgaaca cactggtgaa  
2881 gcagctgagc agcaattttg gcgccatttc cagtgtgctg aatgatatcc tgtcccact  
2941 ggataaagtg gaggccgaag tgcagattga caggctgatt acaggcagac tgcagagcct  
3001 gcagacctat gtgacacagc agctgatcag ggctgctgaa atcagggctt ctgccaatct  
3061 ggctgctact aagatgtctg agtgtgtgct gggacagtcc aagagagtgg acttttgg  
3121 aaagggtac cacctgatgt ccttcccaca ggctgccct catggagtgg tgttctgca  
3181 tgtgacctat gtgcatccc aggagaggaa cttcaccaca gcccagcca tttgcatga  
3241 aggcaaggcc tacttcctc gggaaggcgt gttcgtgtt aatggcactt ctgggttat  
3301 tacacagcgg aacttctta gcccacagat catcactaca gacaatacat ttgtgccg  
3361 aaattgtgat gtggtgattg gcatcattaa caacacagtg tatgacctc tgcagcctga  
3421 gctggactcc ttcaaggaag agctggacaa gtacttcaag aatcatacat cccagatgt  
3481 ggatctgggc gacattccg gcattaacgc ttctgtgtg aacattcaga aggaaattga  
3541 ccgcctgaat gaagtggcta agaactgaa tgaatccctg atgacctgc aggaactggg  
3601 caagtatgag cagtatatta agtggccttg gtatgtgtgg ctgggcttca ttgctggact  
3661 gattgccatc gtgatgtga caatcctgct gtgtgcatg acctcctgtt gcagttgcct  
3721 gaagggcgct tgctcttggt gatcttgctg caagtttgat gaggatgact ctgagccagt  
3781 gctgaagggc gtgaagctgc attacacagg gcccgcggc gactacaagg acgatgacga  
3841 caagtatag atcgatgcat ggatccgtt aaaccgagct ccagctttgt tccctta

## The spike protein of SARS-Corona Virus 在 SWISS-PROT 蛋白质序列:

| Sequence                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       | Length | Mass (Da) | Tools                                                                       |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-----------|-----------------------------------------------------------------------------|
| <input type="checkbox"/> P59595 [UniParc].<br>Last modified April 23, 2003. Version 1.<br>Checksum: 43FC8750F1253034                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           | FASTA  | 422       | 46,025 <input type="text" value="Blast"/> <input type="button" value="go"/> |
| <pre> 10      20      30      40      50      60 MSDNGFQSNQ RSAPRITFGG PTDSTDNQIN GGRNGARPKQ RRPQGLENNT ASWFTALTQH  70      80      90     100     110     120 GKEELRFPRG QGVPIINTNSG PDDQIGYYRR ATRRVRGGDG KMKELSPRWY FYYLGTGPEA  130     140     150     160     170     180 SLFYGANKEG IUVVATEGAL NTPKDHIGTR NFMNNAATVL QLPQGTTLPK GFYAEGSRGG  190     200     210     220     230     240 SQASSRASSR SRGNSRNSTP GSSRGNSPAR MASGGGETAL ALLLLDRLNQ LESKVSQKQKQ  250     260     270     280     290     300 QQQGQTVTKK SAAEASKKPR QKRTATKQYN VIQAFGRRGF EQTQGNFGDQ DLIRQGTIDYK  310     320     330     340     350     360 HWPQIAQFAP SASAFFGMSR IGMEVTPSGT WLTVHGAIKL DDKDPQPKDN VILLNKHIDA  370     380     390     400     410     420 YKTFPFTEPK KDKKKKTDEA QPLPQRQKQKQ PIVTLLPAAD MDDFSRQLQN SMSGASADST  QA                     </pre> |        |           |                                                                             |

## The spike protein of SARS-Corona Virus 在 PDB 蛋白质构造序列:

|                                                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|  | <b>3C9N</b> Crystal Structure of a SARS Corona Virus Derived Peptide Bound to the Human Major Histocompatibility Complex Class I molecule HLA-B*1501<br><b>Authors:</b> Roder, G.A., Kristensen, O., Kastrup, J.S., Buus, S., Gajhede, M.<br><b>Release:</b> 2008-02-26 <b>Classification:</b> Immune System<br><b>Experiment:</b> X-RAY DIFFRACTION with resolution of 1.87 Å <b>Residue Count:</b> 384<br><b>Compound:</b> 3 Polymers [ <a href="#">Display Full Polymer Details</a>   <a href="#">Display for All Results</a> ]<br>2 Ligands [ <a href="#">Display Full Ligand Details</a>   <a href="#">Display for All Results</a> ]<br><b>Citation:</b> Structure of a SARS coronavirus-derived peptide bound to the human major histocompatibility complex class I molecule HLA-B*1501. (2008) ACTA CRYSTALLOGR., SECT. F <b>64</b> : 459-462 [ <a href="#">Display Full Abstract</a>   <a href="#">Display for All Results</a> ]<br><b>Molecule of the Month:</b> Major Histocompatibility Complex, T-Cell Receptor<br><b>Search Hit:</b> Title: Crystal Structure of a SARS Corona Virus Derived Peptide Bound to the Human Major Histocompatibility Complex Class I molecule HLA-B*1501 |
|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

### (3) 文献信息的查找与管理

有效地使用 NCBI PubMed 提供的多种重要功能，查询并下载有关课题或研究方向的论文文摘与文献全文。

查询 Influenza A Viruses 分子进化研究方向的文章。

- [Seroepidemiological Evidence of Avian Influenza A Virus Transmission to Pigs in Southern China.](#)
1. Su S, Qi W, Chen J, Zhu W, Huang Z, Xie J, Zhang G.  
J Clin Microbiol. 2012 Nov 21. [Epub ahead of print]  
PMID: 23175250 [PubMed - as supplied by publisher]  
[Related citations](#)
  2. ■ [The impact of maternally derived immunity on influenza A virus transmission in neonatal pig populations.](#)  
Allerson M, Deen J, Detmer S, Gramer M, Joo HS, Romagosa A, Torremorell M.  
Vaccine. 2012 Nov 19. doi:pii: S0264-410X(12)01616-7. 10.1016/j.vaccine.2012.11.023. [Epub ahead of print]  
PMID: 23174202 [PubMed - as supplied by publisher]  
[Related citations](#)
  3. ■ [Induction of cross-protection against influenza A virus by DNA prime-intranasal protein boost strategy based on nucleoprotein.](#)  
Luo J, Zheng D, Zhang W, Fang F, Wang H, Sun Y, Ding Y, Xu C, Chen Q, Zhang H, Huang D, Sun B, Chen Z.  
Virol J. 2012 Nov 23;9(1):286. [Epub ahead of print]  
PMID: 23173785 [PubMed - as supplied by publisher]  
[Related citations](#)

(3) NCBI 数据库简介:

## Nucleotide

该数据库由国际核苷酸序列数据库组员美国国立卫生研究院 GenBank、日本 DNA 数据库 (DDBJ) 和英国 Hinxton Hall 的欧洲分子生物学试验室数据库 (EMBL) 三部分数据构成。这三个组织联合构成国际核苷酸序列数据库协作体, 每天互换各自数据库中的新增序列记录实现数据共享。其中的序列数据也通过与基因组序列数据库(GSDB) 合作获取; 专利序列数据通过与美国专利与商标局、国际专利局合作获取。

## Genome

即基因组数据库, 提供了多种基因组、完全染色体、Contiged 序列图谱以及一体化基因物理图谱。

## Structure

即构造数据库或称分子模型数据库 (MMDB), 包括来自 X 线晶体学和三维构造的试验数据。MMDB 的数据从 PDB(Protein Data

Bank) 获得。NCBI 已经将构造数据交叉链接到书目信息、序列数据库和 NCBI 的 Taxonomy 中运用 NCBI 的 3D 构造浏览器和 Cn3D, 可以很轻易地从 Entrez 获得分子的分子构造间互相作用的图像。

### **Taxonomy**

即生物学门类数据库, 可以按生物学门类进行检索或浏览其核苷酸序列、蛋白质序列、构造等。

### **PopSet**

包括研究一种人群、一种种系发生或描述人群变化的一组组联合序列。PopSet 既包括核酸序列数据又包括蛋白质序列数据。

### **Entrez**

功能强大, 在于它的大多数记录可互相链接, 既可在同一数据库内链接, 也可在数据库之间进行链接。当运用 BLAST 软件比较某氨基酸或 DNA 序列与库中其他氨基酸或 DNA 序列差异即进行相似性检索时, 则会波及到蛋白质库或核苷酸库的库内链接。库间链接发生在核苷酸数据库内的记录与 PubMed 库中已刊登序列的引文间的链接, 或蛋白质序列记录与核苷酸序列库中编码它的核苷酸序列间的链接。

### **NCBI 数据库检索**

NCBI 数据库的检索措施很简朴, 在检索框中输入检索词, 检索词间默认逻辑关系为 AND, 检索规则基本同 PubMed。可以通过下拉菜单项选择择记录的显示格式, 一般选择 GenBank Report 格式或 FASTA Report 格式。当选择 GenBank Report 格式后, 屏幕显示较完整的基因记录, 其内容包括: 基因位点(Locus)、基因定义(Definition)、基因存取号(Accession)、

核酸编号 (NID)、关键词 (Keywords)、

来源 (Source)、组织分类 (Organism)、参考文献 (Reference)、 著者 (Author)、题目 (Title)、期刊 (Journal)、Medline 存取号 (Medline)、序列特性 (Features)、基因 (Gene)、CDS(cDNA)、等位基因 (Allele) 对等的肽 (Mat-Peptide)、计算碱基数 (Base Count)、原序列 (Origin)。而 FASTA Report 格式仅包括检出序列的简要特性描述。

## **OMIM**

孟德尔遗传学(OMIM) 数据库是人类基因和基因疾病的目录数据库。该数据库包括原文信息、图片和参照信息，同步还可以链接到 Entrez 系统 MEDLINE 数据库中有关文献和序列信息。主页如图 3 所示。

## **BLAST 相似性检索**

BLAST(Basic Local Alignment Search Tool) 是用于序列相似性检索的一种重要数据库，是辨别基因和基因特性的工具。该软件能在 15 秒内完毕整个 DNA 数据库的序列检索。BLAST 记录的有关度有明确的记录学解释，以便更轻易地将有关记录与的数据库记录相辨别。在 NCBI 主页的左工具条中，点击 BLAST 图标，即进入 BLAST 主页。BLAST 主页提供了几种 BLAST 检索软件。其中 BLAST2.0 是一种新的 BLAST 检索工具，它在原有基础上作了改善，运行速度更快，敏捷度更高，同步具有 Gapped BLAST 和 PSI-BLAST 两种软件的新功能。Gapped BLAST 容许在对准的序列中引入空位 (碱基缺失或插入)，引入空位 (Gaps) 意味着在比较两个有关序列时不会出现中断 (Break) 现象。这些空位对准的记分系统更能反应有关序列的类似程度。PSI-BLAST 的全称是 Position-Specific Iterated BLAST，即特殊位置反复 BLAST，它提供了自动、易用的概貌 (Profile) 检索，是查找序列同源的有效工具。

### 三. 试验规定:

- (1) 以其中的一种信息中心网站为例, 列举其中的重要资源 (数据库、网上分析、生物计算、数据下载等)。
- (2) 可以解释给定序列或基因组数据的含义。
- (3) 检索文献的技巧和效率。

## 试验二 序列多重比对及进化分析

### 一. 试验目的:

1. 学习序列比对工具 BLAST 以及 ClustalW 等的使用, 可以对序列数据进行初步的分析。
2. 掌握基于 DNA 序列和蛋白质序列构建系统进化树的常用措施和常用工具。

### 二. 试验内容:

1. 在 GeneBank 数据库中, 检索 10 条轮状病毒 (Homo sapiens, Rotavirus) VP7 基因的 DNA 序列, 并使用 CLUSTALW 软件对序列进行多重序列比对;

```
VP7.txt - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

>gi|815584|gb|JN706275.1| Human rotavirus A strain CU1025-KK/11 glycosylated outer capsid protein (VP7)
gene, complete cds
ATGTATGGTATTGAATATAACCAATCTAATCTTTCTGATATCAATCATTCTATTCAATTATATATTA
AATCAGTGAAGTGAATGATGACTACATTATATATAGATCTTTGTTAATTTCTGTAGCATTATTTGCTTT
GACAAGAGCTCAGAATTATGGGATTAACCTACCAATAACAGGATCAATGGACGGCCGATACGCTAACCTCT
ACTCAAGAGGGAATATTCTTAACATCCACATTATGTTTGTATTATCCAACAGAGCAAGTACTCAAAATTA
ATGATGGTGAATGGAAAGACTCATTGTCACAAATGTTTCTCATAAAAGGTTGGCCAAACAGGATCAGTCTA
TTTTAAAGAGTATTTCAAGTATTGTTGATTTTCTGTTGATCCACAATTATTTGTGATTATAAATTAGTA
TTAATGAATATGATCAAAATCTTGAATTAGATATGTCAGAGTTAGCTGATTTAATATTGAATGAATGGT
TATGTAATCCAATGGATATAACATTATACTATTATCAACAATCGGGAGAATCAAAATAAGTGGATATCAAT
GGGATCATCATGTATGTGAAAGTGTGTCCACTGAATACGCAAAACGTTAGGAATAGGTTGTCAAAACAACA
AATGATAGACTCGTTTGAATGGTTGCTGAAAATGAGAAAATAGCTATAGTGGATGTCGTTGATGGGATAA
ATCATAAAATAAAATTTGACAACACGACATGTAATTCGAAAATGTAAGAAGTTAGGTCCAAAGAGAGAA
TGTAGCTGTAATACAAAGTTGGTGGCTCTAATGTGTTAGACATAACAGCAGATCCAAACGACTAATCCACAA
ACTGAGAGAAATGATGAGAGTGAATTGGAAAAAATGGTGGCAAGTATTTTATACTATAGTAGATTATTA
ATCAGATTGTACAGGTAAATGTCCAAAAGATCAAGATCATTAAATTCGACGCTTTTATTATAGAGTATA
G
>gi|824806|gb|JX567751.1| Human rotavirus A strain RVA/Human-wt/AUS/WDP280/2011/G10P[14] VP7 gene, complete
cds
GGCTTTTAAAGAGAGAATTTCCGCTCGGCTAGCGGCTAGCTCCCTTTTAAATGTATGGTATTGAATATACCA
CAATTCCTAATCTACTTAATATCGATTATATTGTTTAAATTACATATTGAAGAGTATAACTAGAATGATGGA
CTACATAATTTATAAATTTTGTCTGATAGTCACAATTACTTCAATTGTTGTTAACGCGCAGAAATTATGGT
ATAAATTTGCCAATAAATCGGATCGATGGATGATATCATATGTAATGCTACTAAGGATGAACCAATTTCTAA
CATCAACATTTGTTTGTGTTTGTACTATCCGACAGAAAGCCAGAACAGAAATAAATGATAATGAGTGGACAAATGAC
GTTGTGCACAGCTATTTTGTACGAAGGGATGGCCAACTGGATCGGTATACCTTTTAAAGAAATATGATGATATA
GCTACTTTTTCAGTTGATCCACAACCTGTAATTTGACTATAATATAGTTTGTGATGCGATACAATTCGAGCC
TAGAAGCTTGACATGTTCGGAATTGGCGAATCTAATATTGAATGAATGGCTGTGTAACCCAAATGGACATCAC
ATTATACTACTACCAACAAACGGACGAAAGCAAAACAAAGTGGATAGCAATGGGACAACTATGCACAAATAAAA
GTGTGTCCATTGAATATCCCAACACTAGGAATAGGATGTCAACAGCAAAATACTGGAACGTTTGAAGAAAG
TTGCAACAGCTGAAAAATTTGGTGAATTACTGATGTAGTAGACGGCGTGAACCCACAATTAGATGTTACTAC
TACGACCTGTACTATCAGAAATTTGCAAGAAATTTGGGGCCAAAGAGAAAACGTAGCAGTAATACAAGTAGGT
GGTGTGTATATCTTTGATATAACATCTGTATCCAAAGCTGCTCCACAAACTGAAACGGATGATGCGAATAA
ATTGAAAAAATGGTGGCAAGTGTTTTATACTATAGTTGATTACGTGAATCAAAATTTGTACAAGCAATGTC
TAAAGAGTCCAGATCATGGATTCAGCAGCAATTTATTATAGAGTGTAGATATTATGTAGATTAGAATTG
TATGATGTGACC
>gi|824802|gb|JX567749.1| Human rotavirus A strain RVA/Human-wt/AUS/V585/2011/G10P[14] VP7 gene, complete
```

检索成果详见电子稿附件：VP7. txt 文献



```
实验2序列对比结果VP7.aln - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

gi|815584|gb|JN706275.1|TATGTCAGAGTTAGCTGATTTAATATTGAATGAATGGTTATGTAATCCAA
gi|695647|gb|JX179295.1|TATGTCAGAGTTAGCTGATTTAATATTGAATGAATGGTTATGTAATCCAA
gi|694211|gb|JX509940.1|TATGTCAGAAATTAGCTGATTTGATATTGAATGAATGGTTATGTAATCCAA
gi|824804|gb|JX567750.1|CATGTCGGAATTGGCGAATCTAATATTGAATGAATGGCTGTGTAACCCAA
gi|824800|gb|JX567748.1|CATGTCGGAATTGGCGAATCTAATATTGAATGAATGGCTGTGTAACCCAA
gi|824806|gb|JX567751.1|CATGTCGGAATTGGCGAATCTAATATTGAATGAATGGCTGTGTAACCCAA
gi|824802|gb|JX567749.1|CATGTCGGAATTGGCGAATCTAATATTGAATGAATGGCTGTGTAACCCAA
gi|824808|gb|JX567752.1|CATGTCGGAATTGGCGAATCTAATATTGAATGAATGGCTGTGTAACCCAA
gi|595614|gb|JQ037749.1|TGCATCGGAGTTAGCAGATCTTATATTGAACGAATGGCTGTGCAATCCTA
gi|595610|gb|JQ037744.1|TGCATCGGAGTTAGCAGATCTTATATTGAACGAATGGCTGTGCAATCCTA
** ** ** ** ** ** ** ** ** ***** ** ** ** ** ** **

gi|815584|gb|JN706275.1|TGGATATAACATTATACTATTATCAACAATCGGGAGAATCAAAATAAGTGG
gi|695647|gb|JX179295.1|TGGATATAACATTATACTATTATCAACAATCGGGAGAATCAAAATAAGTGG
gi|694211|gb|JX509940.1|TGGATATAACATTATACTATTATCAACAATCGGGAGAATCAAAATAAGTGG
gi|824804|gb|JX567750.1|TGGACATCACATTATACTACTACCAACAAACGGACGAAGCAAAACAAGTGG
gi|824800|gb|JX567748.1|TGGACATCACATTATACTACTACCAACAAACGGACGAAGCAAAACAAGTGG
gi|824806|gb|JX567751.1|TGGACATCACATTATACTACTACCAACAAACGGACGAAGCAAAACAAGTGG
gi|824802|gb|JX567749.1|TGGACATCACATTATACTACTACCAACAAACGGACGAAGCAAAACAAGTGG
gi|824808|gb|JX567752.1|TGGACATCACATTATACTACTACCAACAAACGGACGAAGCAAAACAAGTGG
gi|595614|gb|JQ037749.1|TGGATATATCACTTTTACTATTATCAACAAAATAGCGAATCAAAATAAATGG
gi|595610|gb|JQ037744.1|TGGATATATCACTTTTACTATTATCAACAAAATAGCGAATCAAAATAAATGG
**** ** ** * ** * ** ***** ** ** ** ** **

gi|815584|gb|JN706275.1|ATATCAATGGGATCATCATGTACTGTGAAAAGTGTGTCCACTGAATACGCA
gi|695647|gb|JX179295.1|ATATCAATGGGATCATCATGTACTGTAAAAAGTGTGTCCACTGAATACACA
gi|694211|gb|JX509940.1|ATATCAATGGGATCATCGTGTACTGTGAAAAGTGTGTCCACTGAATACACA
gi|824804|gb|JX567750.1|ATAGCAATGGGACAATCATGCACAATAAAAAGTGTGTCCATTGAATACCCA
gi|824800|gb|JX567748.1|ATAGCAATGGGACAATCATGCACAATAAAAAGTGTGTCCATTGAATACCCA
gi|824806|gb|JX567751.1|ATAGCAATGGGACAATCATGCACAATAAAAAGTGTGTCCATTGAATACCCA
gi|824802|gb|JX567749.1|ATAGCAATGGGACAATCATGCACAATAAAAAGTGTGTCCATTGAATACCCA
gi|824808|gb|JX567752.1|ATAGCAATGGGACAATCATGCACAATAAAAAGTGTGTCCATTGAATACCCA
gi|595614|gb|JQ037749.1|ATATCAATGGGAACAGACTGCACGGTAAAAAGTTTGTCCACTCAATACACA
gi|595610|gb|JQ037744.1|ATATCAATGGGAACAGACTGCACGGTAAAAAGTTTGTCCACTCAATACACA
*** ***** * ** * * ***** ***** * ***** **
```

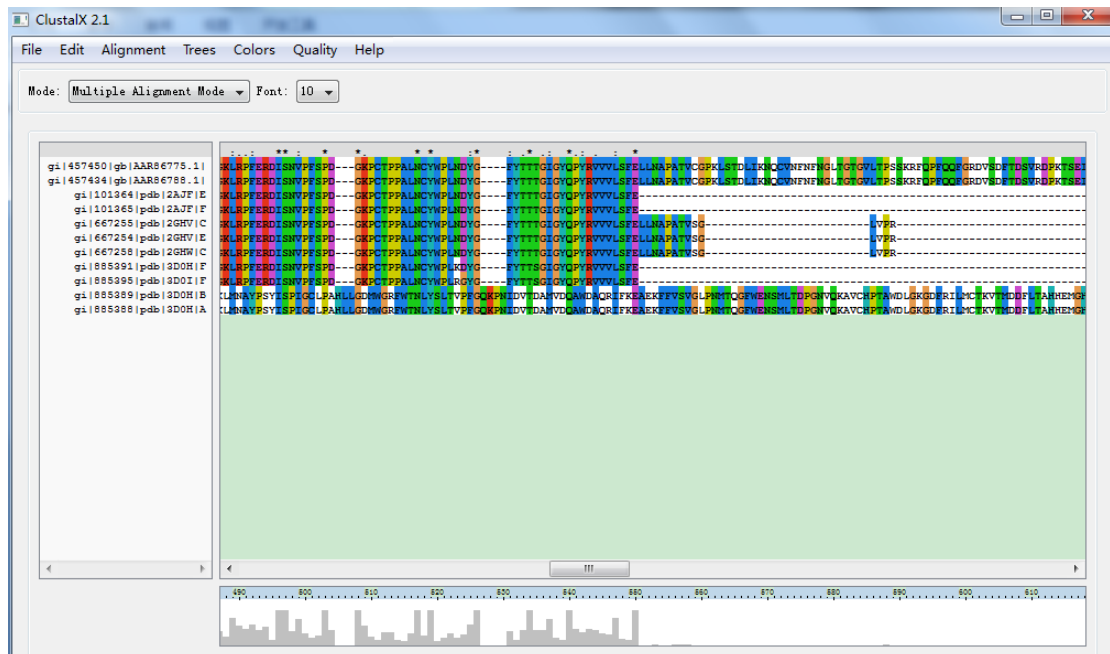
多重序列比对成果详见电子稿附件：VP7.aln 文献

2. 在 GeneBank 数据库中检索 10 条 SARS 病毒 Spike 蛋白的氨基酸序列，使用 CLUSTALX 软件对这十条序列进行多重序列比对；

```
Spike SARS.txt - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

>gi|457450|gb|AAR86775.1| spike [SARS coronavirus ShanghaiXC2]
MFIFLLFLTLTSGSDLDRCCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTDQLFLPFYSNVTGFH
TINHPTGPNVPVPPFDGIYFAATEKSNVVRGWVFGSTMMNKSQSVIIINNNTNVVIRACNFELCDNPFPAV
SKPMGTQTHTIMFDNAFNCTFEYISDAFSLDVSSEKSGNFKHLREFVFNKNDGFLYVYKGYQPIDVVRDLF
SGFNTLKPILFKLPLGINITNFRAILTAFAQDIWGTSAAYFVGYLKPTTFMLKYDENGTITDAVDCSQ
NPLAELKCSVKSEIDKGIYQTSNFRVVPSCDVVRFPNITNLCPFGGEVFNATKFPSSVYAWERKKISNCVA
DYSVLVNSTFFSTFKCYGVSATKLNLCFSDNYADSFVVKGDVVRQIAPGQTGVADIYNYKLPPDFMCCV
LAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVVFPSPDGKPCPTPALNCYWPLNDYGFYTTTGIG
YQPYRVVVLSEFLLNAPATVCGPKLSTDLIKNQCVNFMNGLTGTGVLTPSSKRFQPFQFGQGRDVSDFTD
SVRDPKTSSEILDISPCAFGGVSVITPGTNASSEVAVLYQDVNCTNVSAAIHADQLTPAWRIYSTGNVVFQ
TQAGCLIGAEHVDTSYECIPIGAGICASYHTVSLLRSTSQKSIIVAYTMSLGADSSIAYSNNTIAIPTNF
SISITTEVMPVSMAKTSVDCNMVYICGDSSTECANLLQYGSFCTQLNRALSGIAAEQDRNTREVFQVQKQ
YKTPTLKYFGGFNSQILPDLKPTKRSPFIEDLLFNKVTILADAGFMKQYGECLGINARDLICAQKFNGL
TVLPPLLTDMMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMATRFNGIGVTQNVLYENQKQIANQFN
KATISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISVSLNDILSRLDKVEAEVQIDRLIT
GRLQSLQTYVTVQQLIRAAEIRASANLAATKMSECVLQSKRVDFCGKGYHLSFPQAAPHGVVFLHVTYV
PSQERNFTTAPAIICHEGKAYFPREGVVFVNGTSWFIQNRNFFSPQIITDNTFVSGNCDVVGIIINNTVY
DPLQPELDSFKGELDKYFKNHTSPDVLGDISGINASVNNIQRIDRLNEVAKNLESILDLQELGKYEQ
YIKWPWYVWLGFIAGLIAIVMVITLLCCMTSCCSCGKACSCGSCCKFDEDDSEPVLGKVLHYY
>gi|457434|gb|AAR86788.1| spike [SARS coronavirus ShanghaiXC1]
MFIFLLFLTLTSGSDLDRCCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTDQLFLPFYSNVTGFH
TINHPTGPNVPVPPFDGIYFAATEKSNVVRGWVFGSTMMNKSQSVIIINNNTNVVIRACNFELCDNPFPAV
SKPMGTQTHTIMFDNAFNCTFEYISDAFSLDVSSEKSGNFKHLREFVFNKNDGFLYVYKGYQPIDVVRDLF
SGFNTLKPILFKLPLGINITNFRAILTAFAQDIWGTSAAYFVGYLKPTTFMLKYDENGTITDAVDCSQ
NPLAELKCSVKSEIDKGIYQTSNFRVVPSCDVVRFPNITNLCPFGGEVFNATKFPSSVYAWERKKISNCVA
DYSVLVNSTFFSTFKCYGVSATKLNLCFSDNYADSFVVKGDVVRQIAPGQTGVADIYNYKLPPDFMCCV
LAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVVFPSPDGKPCPTPALNCYWPLNDYGFYTTTGIG
YQPYRVVVLSEFLLNAPATVCGPKLSTDLIKNQCVNFMNGLTGTGVLTPSSKRFQPFQFGQGRDVSDFTD
SVRDPKTSSEILDISPCAFGGVSVITPGTNASSEVAVLYQDVNCTNVSAAIHADQLTPAWRIYSTGNVVFQ
TQAGCLIGAEHVDTSYECIPIGAGICASYHTVSLLRSTSQKSIIVAYTMSLGADSSIAYSNNTIAIPTNF
SISITTEVMPVSMAKTSVDCNMVYICGDSSTECANLLQYGSFCTQLNRALSGIAAEQDRNTREVFQVQKQ
YKTPTLKYFGGFNSQILPDLKPTKRSPFIEDLLFNKVTILADAGFMKQYGECLGINARDLICAQKFNGL
TVLPPLLTDMMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMATRFNGIGVTQNVLYENQKQIANQFN
KATISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISVSLNDILSRLDKVEAEVQIDRLIT
GRLQSLQTYVTVQQLIRAAEIRASANLAATKMSECVLQSKRVDFCGKGYHLSFPQAAPHGVVFLHVTYV
PSQERNFTTAPAIICHEGKAYFPREGVVFVNGTSWFIQNRNFFSPQIITDNTFVSGNCDVVGIIINNTVY
```

检索成果详见电子稿附件：Spike SARS.txt 文献



```
Spike SARS.aln - 记事本
文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

gi|457434|gb|AAR86788.1|QTSNFRVVPSPGDVVRFPNITNLCPFGEVFNATKFPSSVYAWERKKISNCVA
gi|101364|pdb|2AJF|E-----CPFGEVFNATKFPSSVYAWERKKISNCVA
gi|101365|pdb|2AJF|F-----CPFGEVFNATKFPSSVYAWERKKISNCVA
gi|667255|pdb|2GHV|C-----MADPNITNLCPFGEVFNATKFPSSVYAWERKKISNCVA
gi|667254|pdb|2GHV|E-----MADPNITNLCPFGEVFNATKFPSSVYAWERKKISNCVA
gi|667258|pdb|2GHV|C-----MADPNITNLCPFGEVFNATKFPSSVYAWERKKISNCVA
gi|885391|pdb|3DOH|F-----PFGEVFNATKFPSSVYAWERKKISNCVA
gi|885395|pdb|3DOI|F-----PFGEVFNATKFPSSVYAWERKKISNCVA
gi|885389|pdb|3DOH|B-----NMNNAQDKWSAFLKEQSTLAQMYPLQEIQNLTVKLQLQALQQNGSSVLSE
gi|885388|pdb|3DOH|A-----NMNNAQDKWSAFLKEQSTLAQMYPLQEIQNLTVKLQLQALQQNGSSVLSE
                        *: *: * * . : * : : *

gi|457450|gb|AAR86775.1|DYSVLN-----STFFSTFKCY-----GVSATKLNDLCFSN
gi|457434|gb|AAR86788.1|DYSVLN-----STFFSTFKCY-----GVSATKLNDLCFSN
gi|101364|pdb|2AJF|E-----DYSVLN-----STFFSTFKCY-----GVSATKLNDLCFSN
gi|101365|pdb|2AJF|F-----DYSVLN-----STFFSTFKCY-----GVSATKLNDLCFSN
gi|667255|pdb|2GHV|C-----DYSVLN-----STFFSTFKCY-----GVSATKLNDLCFSN
gi|667254|pdb|2GHV|E-----DYSVLN-----STFFSTFKCY-----GVSATKLNDLCFSN
gi|667258|pdb|2GHV|C-----DYSVLN-----STFFSTFKCY-----GVSATKLNDLCFSN
gi|885391|pdb|3DOH|F-----DYSVLN-----STFFSTFKCY-----GVSATKLNDLCFSN
gi|885395|pdb|3DOI|F-----DYSVLN-----STFFSTFKCY-----GVSATKLNDLCFSN
gi|885389|pdb|3DOH|B-----DKSKRLNTILNTMTSTIYSTGKVCNPDNPQECLLLEPGLNEIMANSLDYNE
gi|885388|pdb|3DOH|A-----DKSKRLNTILNTMTSTIYSTGKVCNPDNPQECLLLEPGLNEIMANSLDYNE
                        * * * ** : * * * * : . * * : :

gi|457450|gb|AAR86775.1|-----VYADSFVVKGDDVRQIAPGQTGVADIYNYKLP
gi|457434|gb|AAR86788.1|-----VYADSFVVKGDDVRQIAPGQTGVADIYNYKLP
gi|101364|pdb|2AJF|E-----VYADSFVVKGDDVRQIAPGQTGVADIYNYKLP
gi|101365|pdb|2AJF|F-----VYADSFVVKGDDVRQIAPGQTGVADIYNYKLP
gi|667255|pdb|2GHV|C-----VYADSFVVKGDDVRQIAPGQTGVADIYNYKLP
gi|667254|pdb|2GHV|E-----VYADSFVVKGDDVRQIAPGQTGVADIYNYKLP
gi|667258|pdb|2GHV|C-----VYADSFVVKGDDVRQIAPGQTGVADIYNYKLP
gi|885391|pdb|3DOH|F-----VYADSFVVKGDDVRQIAPGQTGVADIYNYKLP
gi|885395|pdb|3DOI|F-----VYADSFVVKGDDVRQIAPGQTGVADIYNYKLP
gi|885389|pdb|3DOH|B-----RLWAWESWRSEVGKQLRPLYEYVVLKNEMARANHVEDYGDYWRGDYEVN
gi|885388|pdb|3DOH|A-----RLWAWESWRSEVGKQLRPLYEYVVLKNEMARANHVEDYGDYWRGDYEVN
```

多重序列比对成果详见电子稿附件：Spike SARS.aln 文献

3. 使用 ClustalW 软件或其他软件包构建上述 DNA 分子系统发生树。

### 三. 试验规定：

1. 提交使用 CLUSTALX 及 PHYLIP 软件进行多重序列比对及构建系统发生树的结果；

VP7 outtree:

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。

如要下载或阅读全文，请访问：

<https://d.book118.com/316153125114010154>

