生物信息学上机实验

**实验一 生物信息数据库的检索**

1. **实验目的：**

1.了解生物信息学的各大门户网站以及其中的主要资源。

2.了解主要数据库的内容及结构，理解各数据库注释的含义。

3.了解NCBI、EBI数据库的检索系统ENTREZ，并掌握文献、序列的快速高效检索方法。

1. **实验内容：**
2. 访问NCBI、EMBL-EBI和DDBJ主页，打开数据库的SITE MAP页面，了解各数据库的结构和主要内容。整理记录以下信息：  
   1）NCBI、EMBL-EBI和DDBJ的英文及中文全名；

**答：NCBI: National Center Bioinformatics Institute.美国国立生物技术信息中心；**

**EMBL-EBI：European Molecular Biology Laboratory-European Bioinformatics Institute.欧洲分钟生物学实验室-欧洲生物信息学研究所。**

**DDBJ：DNA Data Bank of Japan.日本DNA数据库。**  
2）在NCBI和EBI所管理的生物学数据库或所提供的服务（工具）里，各选出5个（5个NCBI，5个EBI），列表整理出这些数据库或服务的简称和全称（英文名称即可）。  
答：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| NCBI | GSS  (Genome Survey Sequence) | EST (Expressed Sequence Tag) | RefSeq  (reference  Sequence) | GEO  (Gene Expression Omnibus) | PDB  (the Protein Date Bank) |
| EBI | ENA  (European Nucleotide Archive) | Ensembl | Ensembl Genomes |  |  |

https://www.ebi.ac.uk/submission/

1. 分别进入三大核酸序列数据库，Genbank、ENA（EMBL）和DDBJ，了解其结构，记录以下内容：  
   1）访问地址

答：Genbank:https://www.ncbi.nlm.nih.gov/genbank/;

ENA(EMBL):http://www.ebi.ac.uk/ena;

DDBJ:http://www.ddbj.nig.ac.jp/  
2）数据库全称

答：Genbank:Genbank;

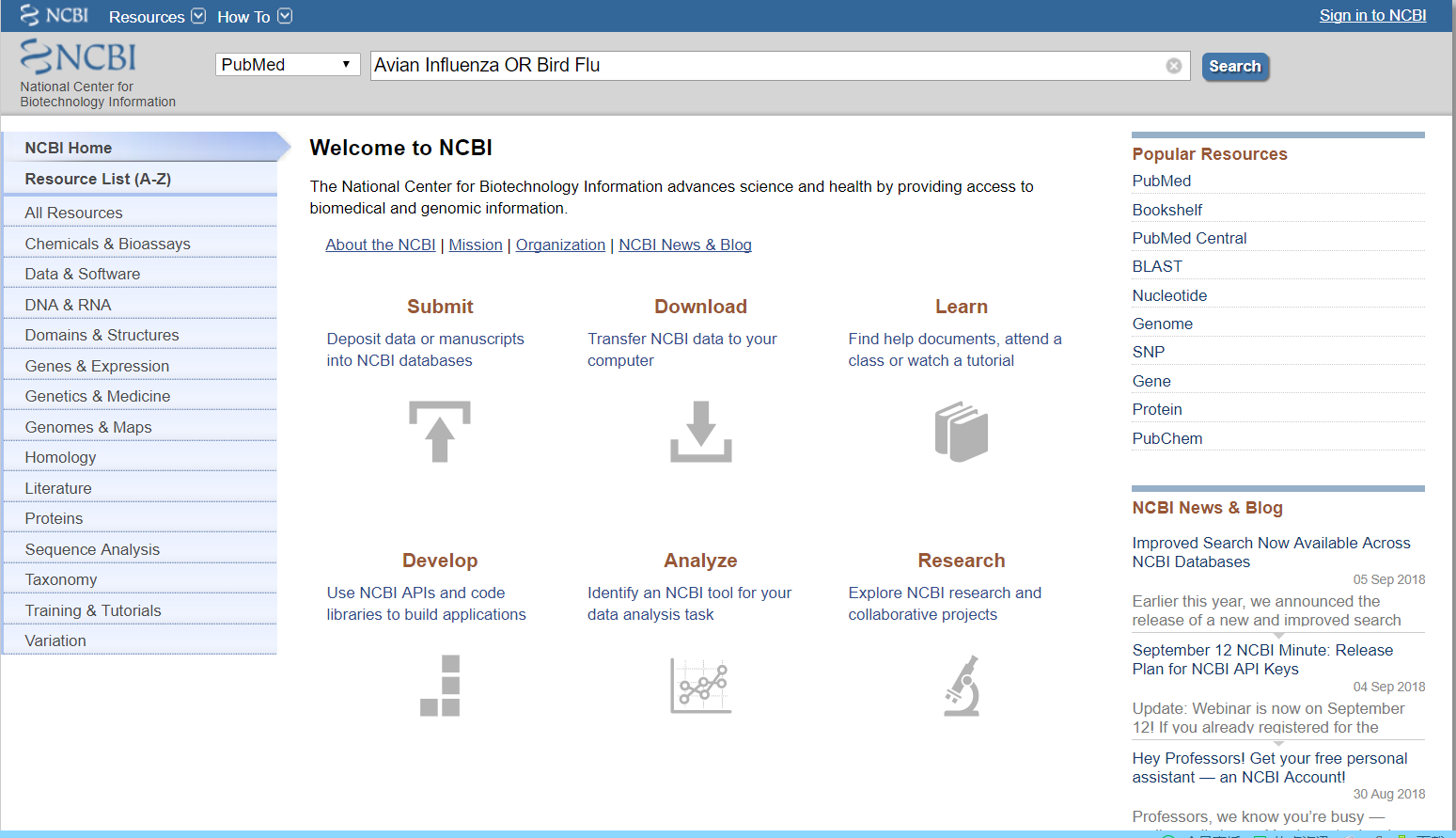
ENA(EMAL): European Nucleotide Archive;

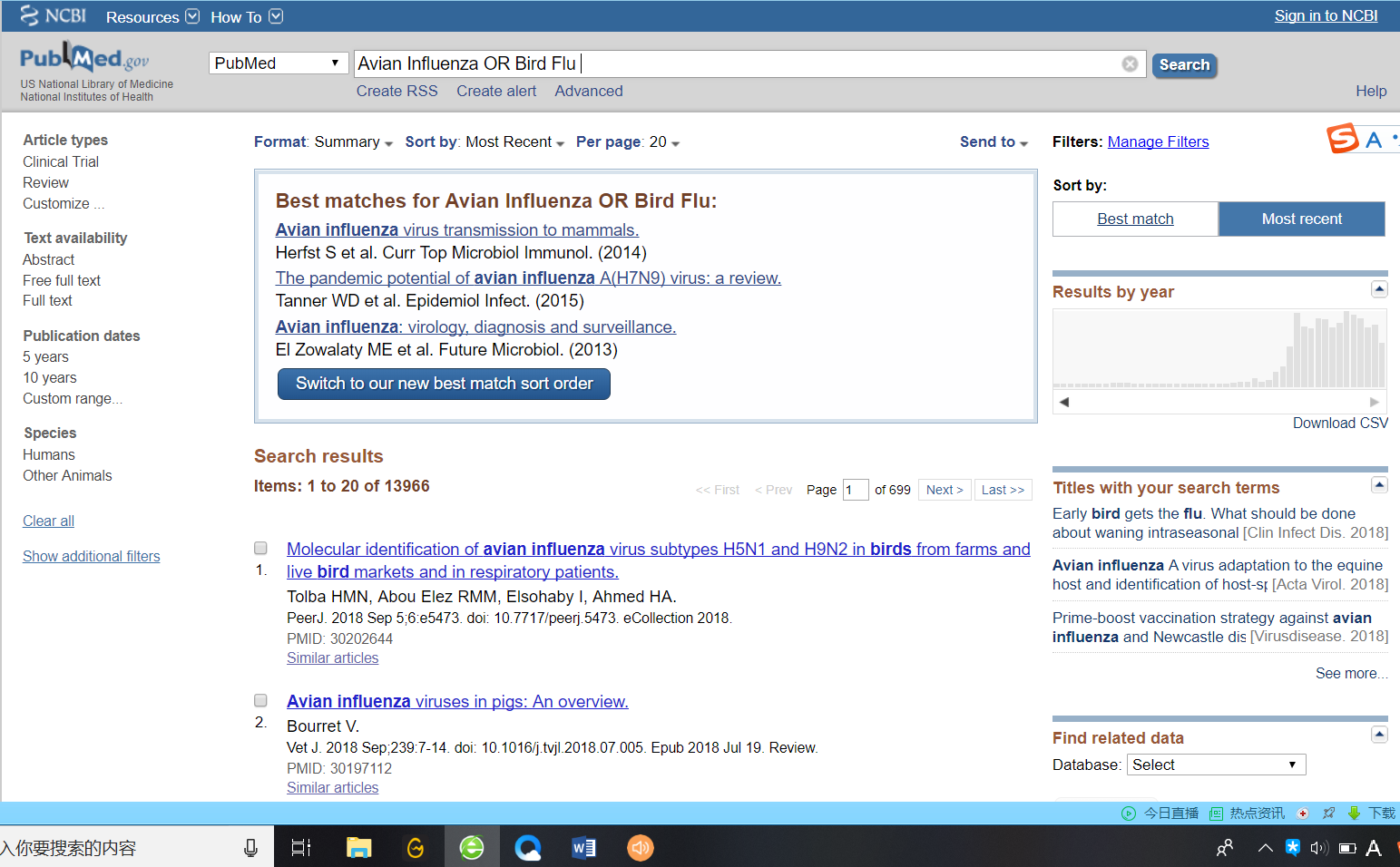
DDBJ: DNA Data Bank of Japan  
3）针对Genbank数据库，寻找其最新发布的版本（日期）、目前可获得的核酸记录条数等信息。https://www.ncbi.nlm.nih.gov/genbank/release/  
答：1.最新版本：

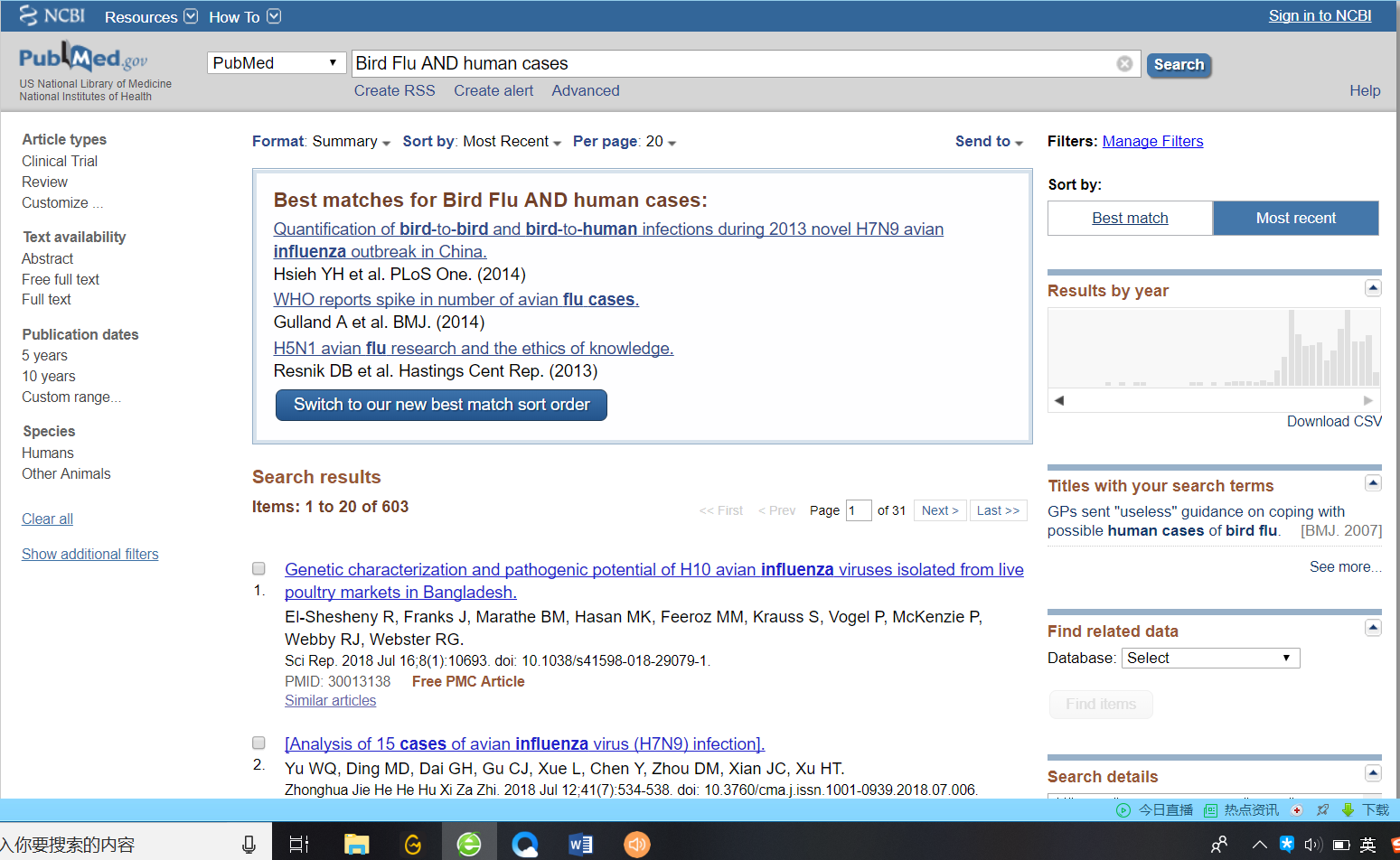
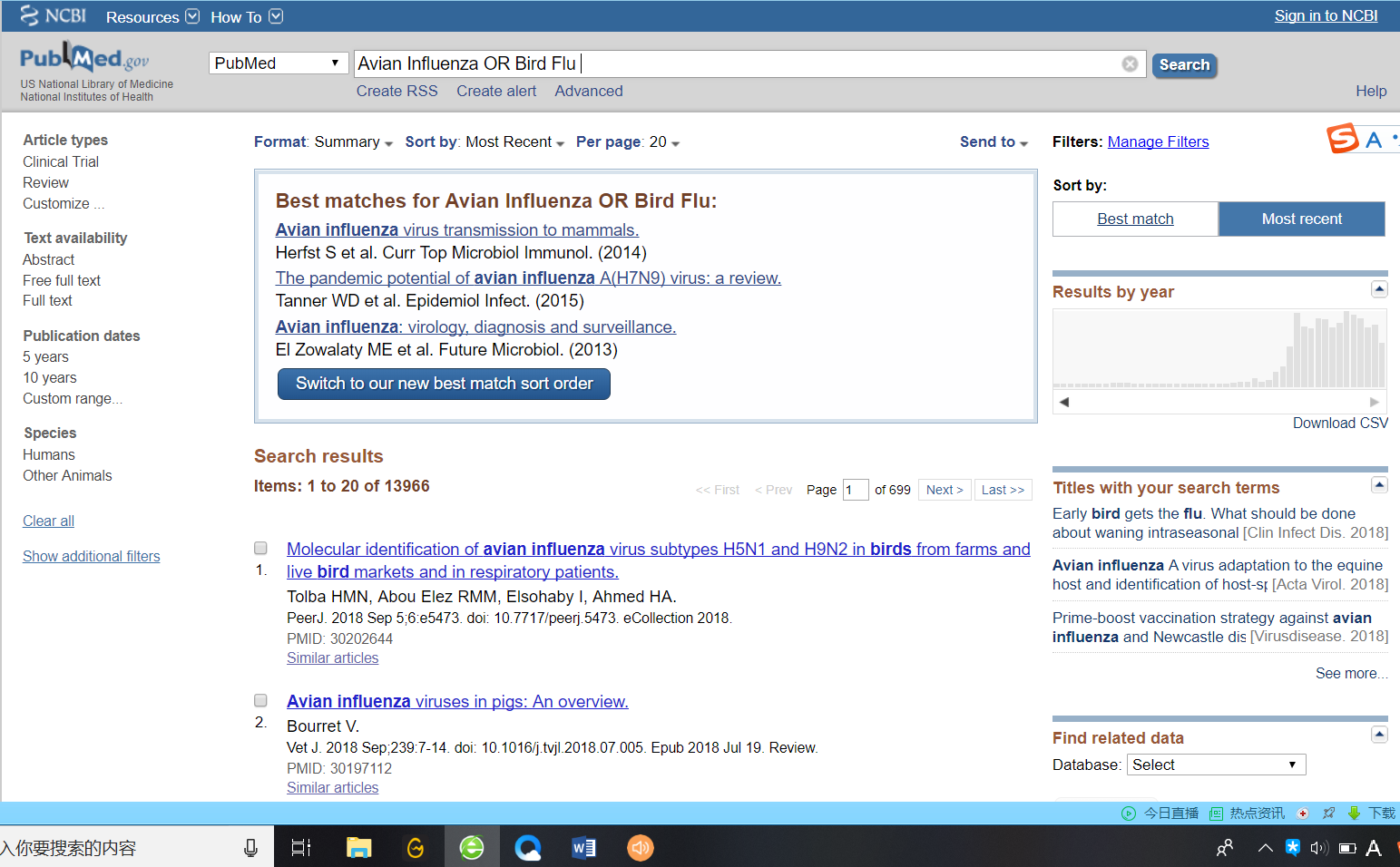
2.记录条数

1. 学习使用Pubmed：选择一关键词，查询文献。
2. 在输入栏内输入关键词[Avian Influenza/Bird Flu](http://www.cdc.gov/flu/avian/)
3. 统计查询结果，并阅读感兴趣文献的摘要或全文。
4. 练习使用AND, OR, BUT逻辑词来限定关键词，如Bird Flu AND human cases,[Avian Influenza OR Bird Flu](http://www.cdc.gov/flu/avian/)等查询人感染禽流感的相关记录，比较查询结果。

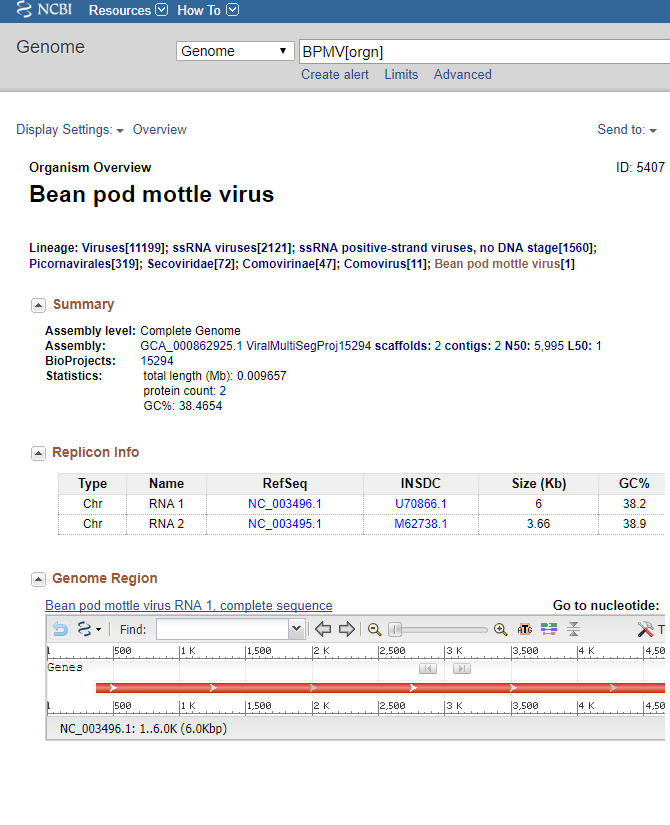
答：1)

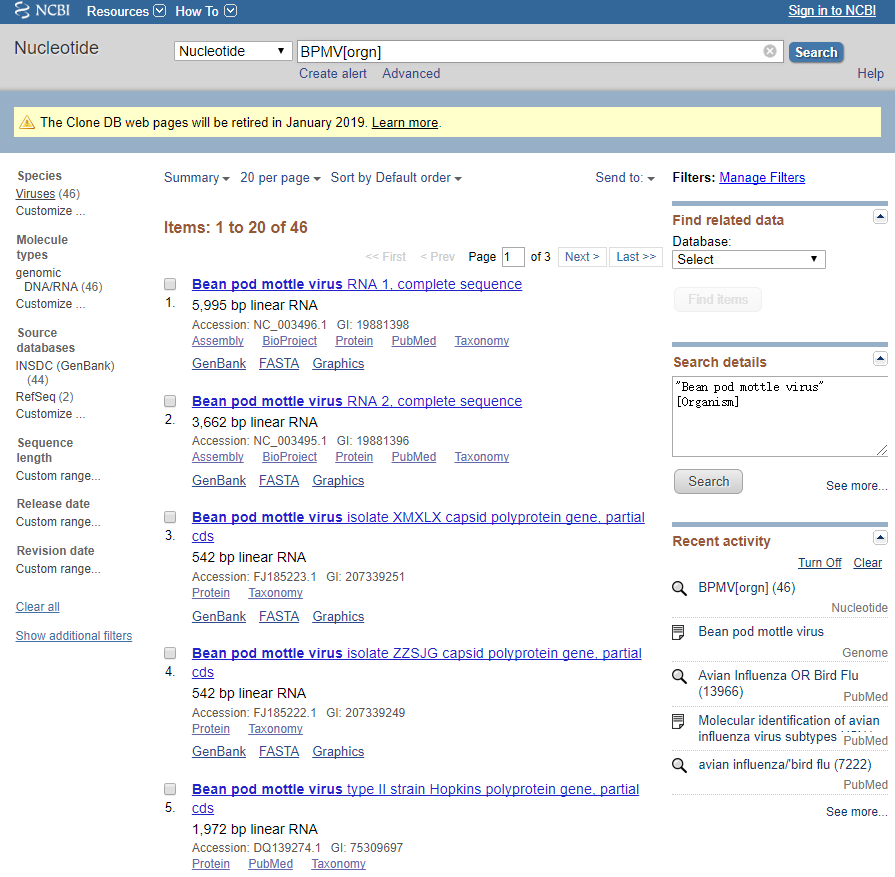


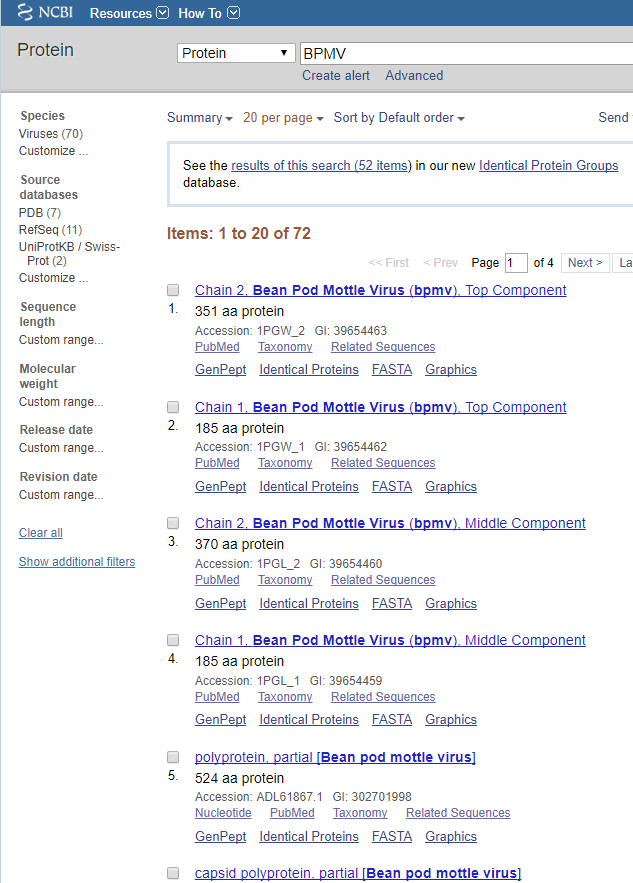
2)

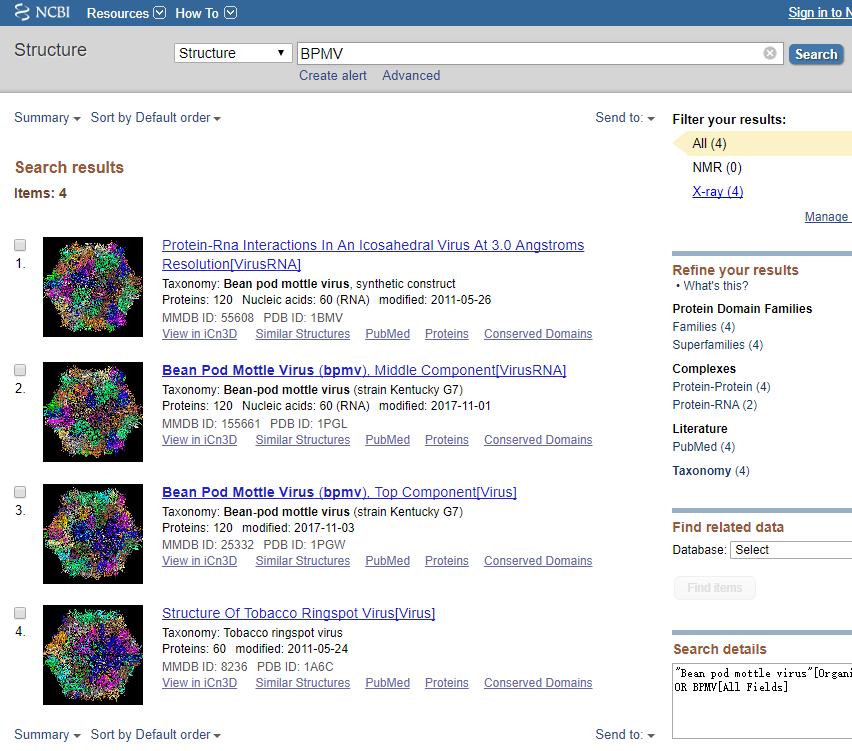
3)

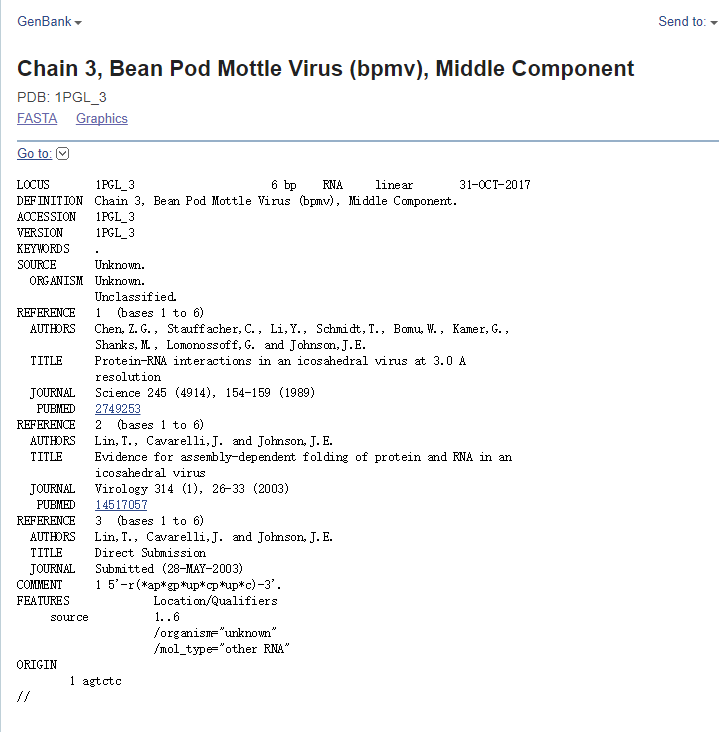
1. 使用Entrez信息查询系统。
2. 已知菜豆斑驳病毒(bean pod mottle virus, BPMV)的名字,查询BPMV的基因组信息,核酸序列信息,蛋白序列信息和结构信息。并对比和书中结果的差异。
3. 阅读查询结果，选择一条感兴趣的核酸序列，点击该序列与数据库的超链接，阅读序列格式的解释，理解其含义。(序列文件太长就部分截图)

答：1) 基因组信息：

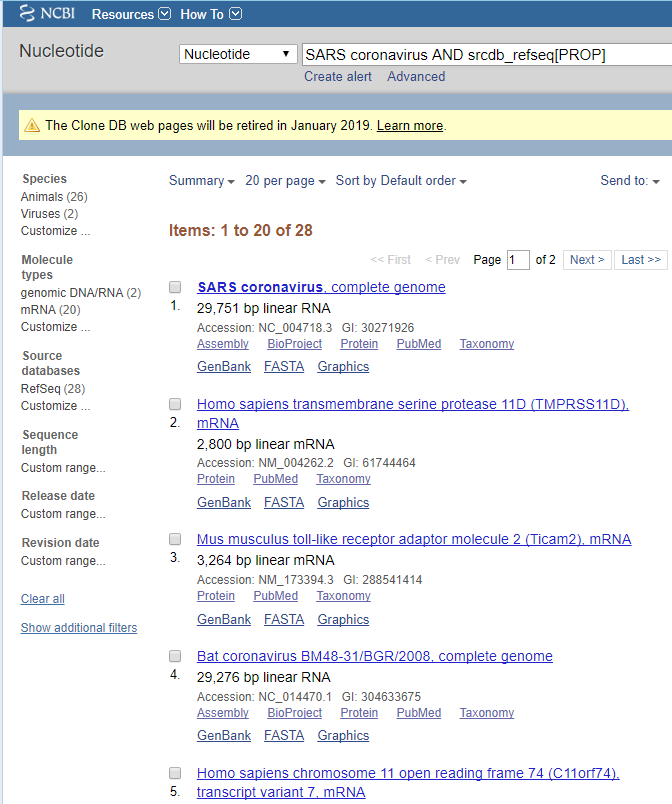
 核酸序列信息：

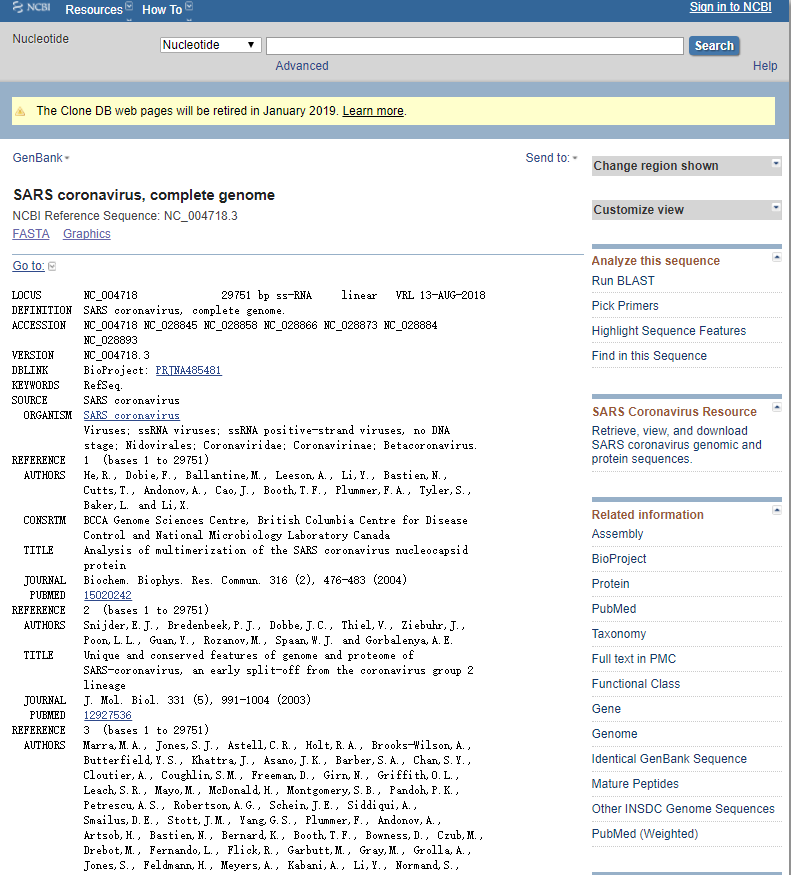
 蛋白序列信息：

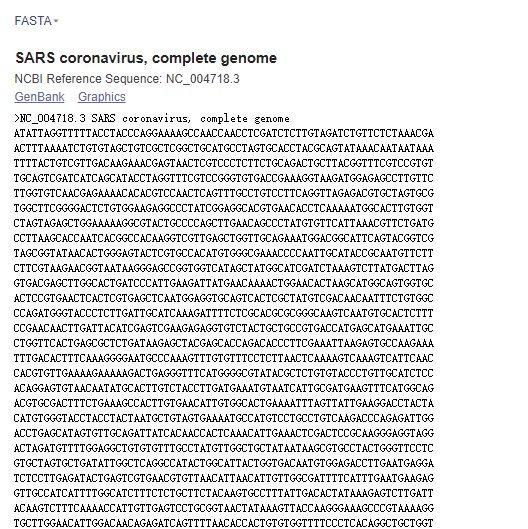
 结构信息：

2）

1. 访问核酸序列数据库（Refseq），进入其查询系统。查询与SARS coronavirus（Severe Acute Respiratory Syndromescoronavirus）相关的核酸序列，链接提取其中一条感兴趣的序列内容，阅读序列格式的解释，理解其含义。并请分别保存两种格式：GBFF和FASTA。https://www.ncbi.nlm.nih.gov/refseq/

答：

GBFF:

FASTA:

1. 补充习题：（选做）
2. 了解蛋白质序列数据库： Uniprot。参考书P29-30练习。
3. 了解Gene 和Genome数据库。参考书P16-17练习。
4. 了解UCSC数据库，自学如何使用（参考书P20-25）。