《Linux 生物信息基础》小组讨论总结报告

第 4 组，第 5 次讨论

组长：陈奕晗

执笔：朱瑾煜、邹济平、陈奕晗、高培翔

1 时间

2021 年 4 月 23 日，15:00 ~ 17:00

2 地点

3W 咖啡厅

3 人员

陈奕晗、邹济平、朱瑾煜、高培翔

4 方式

线下讨论

5 主题

5.1 上次课涉及的 BLAST 内容复习

（朱瑾煜主要执笔，陈奕晗补充）

5.2 vi 文本编辑

（邹济平执笔）

5.3 关于 MyNCBI

（高培翔执笔）

6 内容

6.1 上次课涉及的 BLAST 内容

6.1.1 建立 index—— makeblastdb 命令

使用 `makeblastdb` 命令可以读取 fasta 格式的序列文件，生成 *.phr、*.psq、*.pin 格式的索引文件。

例 1：将从 Swiss-Prot 上获取的人类蛋白质序列文件转换成本地 BLAST 可读取的索引文件

```
makeblastdb -dbtype prot -in sp_human.fasta -out sp_human
```

其中

`-dbtype [prot/nucl]` 即 database type，数据库类型。包括 `prot` (protein，蛋白质) 和 `nucl` (nucleic acid，核酸) 两种。

`-in [filename]` 输入的 fasta 格式文件。

`-out [filename]` 输出的索引文件的名称（不含扩展名）。默认是输入文件的名称。

执行上述命令后，可以看到生成了 `sp_human.phr`、`sp_human.pin`、`sp_human.psq` 三个索引文件。

```
<leb4c@bbt> 16:04:26 ~/MyBlastDB
$ makeblastdb -in sp_human.fasta -dbtype prot -out sp_human


Building a new DB, current time: 04/23/2021 16:11:16
New DB name:   /rd1/home/leb4c/MyBlastDB/sp_human
New DB title:  sp_human.fasta
Sequence type: Protein
Deleted existing Protein BLAST database named /rd1/home/leb4c/MyBlastDB/sp_human
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 20395 sequences in 0.432522 seconds.
<leb4c@bbt> 16:11:17 ~/MyBlastDB
$ ls
sp_human.fasta  sp_human.pin  uniprot_sprot.phr  uniprot_sprot.psq
sp_human.phr    sp_human.psq  uniprot_sprot.pin
<leb4c@bbt> 16:11:19 ~/MyBlastDB
$
```

image-20210423161149614

6.1.2 不使用 index，直接从目标序列 (subject) 中进行 BLAST 搜索—— -subject

例 2：以 17 个拟南芥 SBP (Squamosa promoter-binding-like protein) 转录因子搜索 19 个水稻 SBP 转录因子，输出结果到 ARATH-ORYSJ.xls 文件中。

```
blastp -query 17SPL_ARATH.FASTA -subject 19SPL_ORYSJ.FASTA -outfmt 7 -evalue 0.1 -out ARATH-ORYSJ.xls
```

其中：

`-query [filename]` 用于搜索的序列。

`-subject [filename]` 目标序列，即被搜索的序列。如果使用 index，则用 `-db [filename]` 取代。

`-outfmt [int]` 输出格式，详见 help 中的内容。这里使用 `-outfmt 7`，即带注释的表格格式。

`-evalue [float]` 允许的期望值 (expectation value) 阈值。期望值越低表示相似度越高，其阈值越低也就使得搜索结果越严格。默认值为 10，这里设成 0.1。

`-out [filename]` 输出文件名。

结果如下：

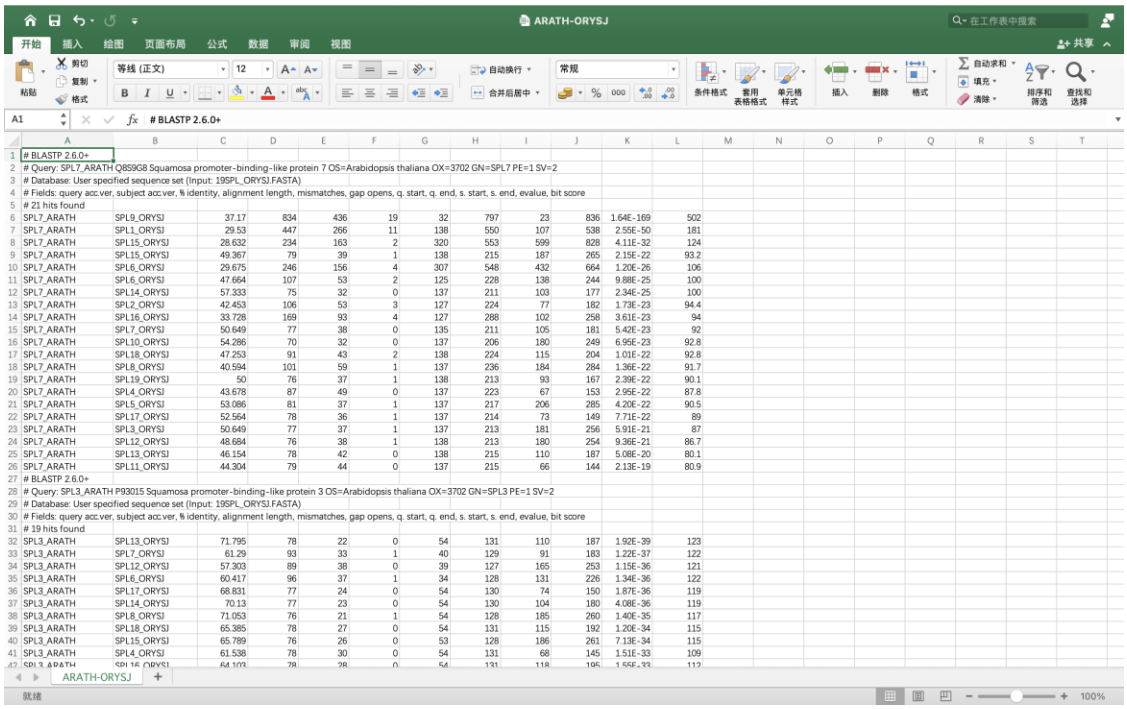
```
<leb4c@bbt> 16:40:28 ~/BLAST
$ ls
17SPL_ARATH.FASTA     HBA_SW_2.txt      SPL3D_CDS.FASTA    ZMTF_CDS.nin
19SPL_ORYSJ.FASTA     HBA_SW.txt        SPL3D_PEP.FASTA    ZMTF_CDS.nsq
CEA21_HUMAN_C2.FASTA  input             spl7_arath.needle  ZMTF_PEP.FASTA
CEA21_HUMAN.FASTA     output            SPL_BLAST.txt      ZMTF_PEP.phr
CEAM5_HUMAN.FASTA     out.txt           ZMTF_CDS.FASTA     ZMTF_PEP.pin
HBA_HUMAN.FASTA       SPL3_ARATH.FASTA  ZMTF_CDS.nhr       ZMTF_PEP.psq
<leb4c@bbt> 16:40:28 ~/BLAST
$ blastp -query 17SPL_ARATH.FASTA -subject 19SPL_ORYSJ.FASTA -outfmt 7 -evalue 0.1 -out ARAT
H-ORYSJ.xls
<leb4c@bbt> 16:40:41 ~/BLAST
$ ls
17SPL_ARATH.FASTA     HBA_SW_2.txt      SPL3D_PEP.FASTA    ZMTF_PEP.FASTA
19SPL_ORYSJ.FASTA     HBA_SW.txt        spl7_arath.needle  ZMTF_PEP.phr
ARATH-ORYSJ.xls       input             SPL_BLAST.txt      ZMTF_PEP.pin
CEA21_HUMAN_C2.FASTA  output            ZMTF_CDS.FASTA     ZMTF_PEP.psq
CEA21_HUMAN.FASTA     out.txt           ZMTF_CDS.nhr
CEAM5_HUMAN.FASTA     SPL3_ARATH.FASTA  ZMTF_CDS.nin
HBA_HUMAN.FASTA       SPL3D_CDS.FASTA   ZMTF_CDS.nsq
<leb4c@bbt> 16:40:43 ~/BLAST
$
```

image-20210423164139737

在本地终端输入如下命令，输入密码后，可以将这个 .xls 文件转移到本地，进而可以用 Excel 软件打开：

```
scp leb4c@117.78.18.116:~/BLAST/ARATH-ORYSJ.xls /Users/jinyuzhu/Desktop
```

用 Excel 查看结果如下：（仅显示部分）



	A	B	C	D	E	F	G	H	I	J	K	L
1	# BLASTP 2.6.0+											
2	# Query: SPL7_ARATH Q8S9G8 Squamosa promoter-binding-like protein 7 OS=Arabidopsis thaliana OX=3702 GN=SPL7 PE=1 SV=2											
3	# Database: User specified sequence set (Input: 19SPL_ORYSJ.FASTA)											
4	# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score											
5	# 21 hits found											
6	SPL7_ARATH	SPL9_ORYSJ	37.17	834	436	19	32	797	23	836	1.64E-169	502
7	SPL7_ARATH	SPL1_ORYSJ	29.53	447	266	11	138	550	107	538	2.55E-50	181
8	SPL7_ARATH	SPL15_ORYSJ	28.632	234	163	2	320	553	599	828	4.11E-32	124
9	SPL7_ARATH	SPL15_ORYSJ	49.367	79	39	1	138	215	187	265	2.15E-22	93.2
10	SPL7_ARATH	SPL6_ORYSJ	29.675	246	156	4	307	548	432	664	1.20E-26	106
11	SPL7_ARATH	SPL6_ORYSJ	47.664	107	53	2	125	228	138	244	9.88E-25	100
12	SPL7_ARATH	SPL14_ORYSJ	57.333	75	32	0	137	211	103	177	2.34E-25	100
13	SPL7_ARATH	SPL2_ORYSJ	42.453	106	53	3	127	224	77	182	1.73E-23	94.4
14	SPL7_ARATH	SPL16_ORYSJ	33.728	169	93	4	127	288	102	258	3.61E-23	94
15	SPL7_ARATH	SPL7_ORYSJ	50.649	77	38	0	135	211	105	181	5.42E-23	92
16	SPL7_ARATH	SPL10_ORYSJ	54.286	70	32	0	137	206	180	249	6.95E-23	92.8
17	SPL7_ARATH	SPL18_ORYSJ	47.253	91	43	2	138	224	115	204	1.01E-22	92.8
18	SPL7_ARATH	SPL8_ORYSJ	40.594	101	59	1	137	236	184	284	1.36E-22	91.7
19	SPL7_ARATH	SPL19_ORYSJ	50	76	37	1	138	213	93	167	2.39E-22	90.1
20	SPL7_ARATH	SPL4_ORYSJ	43.678	87	49	0	137	223	67	153	2.95E-22	87.8
21	SPL7_ARATH	SPL5_ORYSJ	53.086	81	37	1	137	217	206	285	4.20E-22	90.5
22	SPL7_ARATH	SPL17_ORYSJ	52.564	78	36	1	137	214	73	149	7.71E-22	89
23	SPL7_ARATH	SPL3_ORYSJ	50.649	77	37	1	137	213	181	256	5.91E-21	87
24	SPL7_ARATH	SPL12_ORYSJ	48.684	76	38	1	138	213	180	254	9.36E-21	86.7
25	SPL7_ARATH	SPL13_ORYSJ	46.154	78	42	0	138	215	110	187	5.08E-20	80.1
26	SPL7_ARATH	SPL11_ORYSJ	44.304	79	44	0	137	215	66	144	2.13E-19	80.9
27	# BLASTP 2.6.0+											
28	# Query: SPL3_ARATH P93015 Squamosa promoter-binding-like protein 3 OS=Arabidopsis thaliana OX=3702 GN=SPL3 PE=1 SV=2											
29	# Database: User specified sequence set (Input: 19SPL_ORYSJ.FASTA)											
30	# Fields: query acc.ver, subject acc.ver, % identity, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score											
31	# 19 hits found											
32	SPL3_ARATH	SPL13_ORYSJ	71.795	78	22	0	54	131	110	187	1.92E-39	123
33	SPL3_ARATH	SPL7_ORYSJ	61.29	93	33	1	40	129	91	183	1.22E-37	122
34	SPL3_ARATH	SPL12_ORYSJ	57.303	89	38	0	39	127	165	253	1.15E-36	121
35	SPL3_ARATH	SPL6_ORYSJ	60.417	96	37	1	34	128	131	226	1.34E-36	122
36	SPL3_ARATH	SPL17_ORYSJ	68.831	77	24	0	54	130	74	150	1.87E-36	119
37	SPL3_ARATH	SPL14_ORYSJ	70.13	77	23	0	54	130	104	180	4.08E-36	119
38	SPL3_ARATH	SPL8_ORYSJ	71.053	76	21	1	54	128	185	260	1.40E-35	117
39	SPL3_ARATH	SPL18_ORYSJ	65.385	78	27	0	54	131	115	192	1.20E-34	115
40	SPL3_ARATH	SPL15_ORYSJ	65.789	76	26	0	53	128	186	261	7.13E-34	115
41	SPL3_ARATH	SPL4_ORYSJ	61.538	78	30	0	54	131	68	145	1.51E-33	109

image-20210423164728585

6.1.3 指定搜索的起止位点—— -query_loc 和 -subject_loc

例 3：以人癌胚抗原 CEA21*HUMAN 中恒定结构域 (147～231 位) 搜索 CEAM5*HUMAN 中 6 个恒定结构域，将结果保存到 output 目录下的 CEA_BLAST.txt 中。（癌胚抗原的相关数据可以从癌胚抗原家族结构域信息网站 http://www.carcinoembryonic-antigen.de/human/index.html 中获得）

```
blastp -query CEA21_HUMAN.FASTA -query_loc 147-231 -subject CEAM5_HUMAN.
FASTA -subject_loc 145-675 -out output/CEA_BLAST.txt
```

其中：

`-query_loc [ini]-[ter]` 指定 query 序列参与比对的部分，从起始位点 [ini] 到终止位点 [ter]。

`-subject_loc [ini]-[ter]` 指定 subject 序列参与比对的部分，从起始位点 [ini] 到终止位点 [ter]。

使用 less 命令逐屏读取输出文件，可以看到序列比对的结果。一共搜索出 6 段。现截取一部分如下。

```
 Score = 61.2 bits (147),  Expect = 5e-17, Method: Compositional matrix adjust.
 Identities = 34/85 (40%), Positives = 49/85 (58%), Gaps = 1/85 (1%)

Query  147  PSIQAS-STTVTEKGSVVLTCHTNNTGTSFQWIFNNQRLQVTKRMKLSWFNHMLTIDPIR  205
            PSI ++ S  V +K +V  TC      T++ W  N Q L V+ R++LS  N  LT+  +
Sbjct  503  PSISSNNSKPVEDKDAVAFTCEPEAQNTTYLWWVNGQSLPVSPRLQLSNGNRTLTLFNVT  562

Query  206  QEDAGEYQCEVSNPVSSNRSDPLKL  230
            + DA  Y C + N VS+NRSDP+ L
Sbjct  563  RNDARAYVCGIQNSVSANRSDPVTL  587


 Score = 60.8 bits (146),  Expect = 7e-17, Method: Compositional matrix adjust.
 Identities = 33/85 (39%), Positives = 48/85 (56%), Gaps = 1/85 (1%)

Query  147  PSIQAS-STTVTEKGSVVLTCHTNNTGTSFQWIFNNQRLQVTKRMKLSWFNHMLTIDPIR  205
            PSI ++ S  V +K +V  TC       ++ W  NNQ L V+ R++LS  N  LT+  +
Sbjct  147  PSISSNNSKPVEDKDAVAFTCEPETQDATYLWWVNNQSLPVSPRLQLSNGNRTLTLFNVT  206

Query  206  QEDAGEYQCEVSNPVSSNRSDPLKL  230
            + D   Y+CE  NPVS+ RSD + L
Sbjct  207  RNDTASYKCETQNPVSARRSDSVIL  231


 Score = 57.8 bits (138),  Expect = 8e-16, Method: Compositional matrix adjust.
 Identities = 29/78 (37%), Positives = 45/78 (58%), Gaps = 0/78 (0%)

Query  153  STTVTEKGSVVLTCHTNNTGTSFQWIFNNQRLQVTKRMKLSWFNHMLTIDPIRQEDAGEY  212
            S  V ++ +V LTC      T++ W  NNQ L V+ R++LS  N  LT+  + + D G Y
Sbjct  332  SNPVEDEDAVALTCEPEIQNTTYLWWVNNQSLPVSPRLQLSNDNRTLTLLSVTRNDVGPY  391

Query  213  QCEVSNPVSSNRSDPLKL  230
            +C + N +S + SDP+ L
Sbjct  392  ECGIQNKLSVDHSDPVIL  409

:
```

image-20210423170031529

6.1.4 几种 BLAST 的区别和选择

BLAST 套件的 blastn、blastp、blastx、tblastn 和 tblastx 子工具的用途分别如下：

1. **blastn** 是将给定的核酸序列与核酸数据库中的序列进行比较。**（核酸搜核酸）**

2. **blastp** 是使用蛋白质序列与蛋白质数据库中的序列进行比较，可以寻找较远的关系。**搜蛋白质序列时，应首选 blastp。（蛋白搜蛋白）**

3. **blastx** 则是将给定的核酸序列按照六种阅读框架将其翻译成蛋白质与蛋白质数据库中的序列进行比对，**在转录组分析中**对分析新序列和 EST (expression sequence tag) 很有用。**（核酸搜蛋白）**

4. **tblastn** 将给定的氨基酸序列与核酸数据库中的序列（双链）按不同的阅读框进行比对，**在转录组分析中**对于寻找数据库中序列没有标注的新编码区很有用。**（蛋白搜核酸）**

5. **tblastx** 只在特殊情况下使用，它将 DNA 被检索的序列和核酸序列数据库中的序列按不同的阅读框全部翻译成蛋白质序列，然后进行蛋白质序列比对。**（核酸翻译之后搜核酸）**

在下面的例子中，我们分别采用这 5 种方法进行 BLAST。

例 4：以拟南芥 SPL3 转录因子搜索 764 个玉米转录因子数据集：

已有 `SPL3D_PEP.FASTA`、`SPL3D_CDS.FASTA` 和索引文件 `ZMTF_PEP.*`、`ZMTF_CDS.*`，期望值阈值设为 0.01。

1. **blastp** 以拟南芥 SPL3 转录因子 DNA 结合结构域搜索 764 个玉米转录因子数据集蛋白质序列

   ```
   blastp -query SPL3D_PEP.FASTA -db ZMTF_PEP -outfmt 6 -evalue 0.01
   ```

   在玉米转录因子数据集中搜出 4 个结果。

   ```
   <leb4c@bbt> 17:11:03 ~/BLAST
   $ blastp -query SPL3D_PEP.FASTA -db ZMTF_PEP -outfmt 6 -evalue 0.01
   AT2G33810.1     PTZm00608.1     64.865  74      26      0       3       76      171     2442.16e-31 107
   AT2G33810.1     PTZm00605.1     64.865  74      26      0       3       76      144     2173.88e-31 105
   AT2G33810.1     PTZm00606.1     56.757  37      16      0       3       39      180     2168.94e-13 55.5
   AT2G33810.1     PTZm00607.1     51.220  41      20      0       3       43      171     2111.26e-10 50.1
   ```

   image-20210423172102619

2. **blastn** 以拟南芥 SPL3 转录因子编码区核苷酸序列搜索 764 个玉米转录因子数据集编码区核苷酸序列

   ```
   blastn -query SPL3D_CDS.FASTA -db ZMTF_CDS -outfmt 6 -evalue 0.01
   ```

   没有搜出结果。说明此方法灵敏度不高。

3. **tblastn** 以拟南芥 SPL3 转录因子 DNA 结合结构域搜索 764 个玉米转录因子数据集编码区核苷酸序列

   ```
   tblastn -query SPL3D_PEP.FASTA -db ZMTF_CDS -outfmt 6 -evalue 0.0
   1
   ```

   在玉米转录因子数据集中搜出 4 个结果。

   ```
   <leb4c@bbt> 17:21:15 ~/BLAST
   $ tblastn -query SPL3D_PEP.FASTA -db ZMTF_CDS -outfmt 6 -evalue 0.01
   AT2G33810.1     PTZm00608.1     64.865  74      26      0       3       76      549     328     2.49e-30        106
   AT2G33810.1     PTZm00605.1     64.865  74      26      0       3       76      727     948     2.86e-30        105
   AT2G33810.1     PTZm00607.1     58.000  50      21      0       3       52      1279    1428    2.44e-17        69.7
   AT2G33810.1     PTZm00606.1     58.140  43      18      0       3       45      630     758     5.86e-16        65.1
   ```

   image-20210423172338501

4. **blastx** 以拟南芥 SPL3 转录因子编码区核苷酸序列搜索 764 个玉米转录因子数据集蛋白质序列

   ```
   blastx -query SPL3D_CDS.FASTA -db ZMTF_PEP -outfmt 6 -evalue 0.01
   ```

   在玉米转录因子数据集中搜出 4 个结果。

```
<leb4c@bbt> 17:22:55 ~/BLAST
$ blastx -query SPL3D_CDS.FASTA -db ZMTF_PEP -outfmt 6 -evalue 0.01
AT2G33810.1     PTZm00608.1     64.865  74      26      0       7       228     171     244     2.40e-31        107
AT2G33810.1     PTZm00605.1     64.865  74      26      0       7       228     144     217     4.31e-31        105
AT2G33810.1     PTZm00606.1     56.757  37      16      0       7       117     180     216     9.93e-13        55.5
AT2G33810.1     PTZm00607.1     51.220  41      20      0       7       129     171     211     1.40e-10        50.1
```

image-20210423172406342

5. **tblastx** 以拟南芥 SPL3 转录因子编码区核苷酸序列（翻译后）搜索 764 个玉米转录因子数据集编码区核苷酸序列

   ```
   tblastx -query SPL3D_CDS.FASTA -db ZMTF_CDS -outfmt 6 -evalue 0.0
   1
   ```

   在玉米转录因子数据集中搜出 5 个结果。最后两项来自同一条序列。

```
<leb4c@bbt> 17:24:24 ~/BLAST
$ tblastx -query SPL3D_CDS.FASTA -db ZMTF_CDS -outfmt 6 -evalue 0.01
AT2G33810.1     PTZm00608.1     64.865  74      26      0       7       228     549     328     4.31e-32        128
AT2G33810.1     PTZm00605.1     64.865  74      26      0       7       228     727     948     5.91e-32        128
AT2G33810.1     PTZm00607.1     58.000  50      21      0       7       156     1279    1428    5.05e-18        82.2
AT2G33810.1     PTZm00606.1     58.140  43      18      0       7       135     630     758     4.31e-16        75.8
AT2G33810.1     PTZm00606.1     56.818  44      19      0       137     6       760     629     1.39e-06        44.1
```

image-20210423172616898

BLAST 的主要特点就是：速度快，共线性输出结果简单易读。对于比较小的序列（如 cDNA 等）对大基因组的比对，BLAST 无疑是首选。
BLAST 虽然性能优异，但是它自身也存在着一定的局限性，对于特殊的任务需要注意选择合适的软件。例如 BLAST 用于远亲缘物种间的核酸序列比对时，比对精度就不够高，建议使用专门为此用途开发的 Blastz 软件。

6.2 vi 文本编辑

vi/vim 共分为三种模式，分别是**命令模式 (Command mode)**，**输入模式 (Insert mode)**和**底线命令模式 (Last line mode)**。

6.2.1 命令模式

此状态下敲击键盘动作会被 vim 识别为命令，而非输入字符。

以下是常用的几个命令：

i ——切换到输入模式，以输入字符。

x ——删除当前光标所在处的字符。

: ——切换到底线命令模式，以在最底一行输入命令。

命令模式只有一些最基本的命令，因此仍要依靠底线命令模式输入更多命令。

6.2.2 输入模式

在命令模式下按下 i 就进入了输入模式。

在输入模式中，可以使用以下按键：

字符按键以及 Shift 组合——输入字符

ENTER (回车键) ——换行

BACKSPACE (退格键) ——删除光标前一个字符

DEL (删除键) ——删除光标后一个字符

方向键——在文本中移动光标

HOME/END——移动光标到行首/行尾

Page Up/Page Down——上/下翻页

Insert——切换光标为输入/替换模式，光标将变成竖线/下划线

ESC——退出输入模式，切换到命令模式

6.2.3 底线命令模式

在命令模式下**按下:（英文冒号）**就进入了底线命令模式。

底线命令模式可以输入单个或多个字符的命令，可用的命令非常多。

在底线命令模式中，基本的命令有（已经省略了冒号）：

q——退出程序

w——保存文件

下面列举一些 vi 中常用的指令。

6.2.4 光标移动的方法

操作	说明
h 或 向左箭头键(←)	光标向左移动一个字符

j 或 向下箭头键(↓)	光标向下移动一个字符
k 或 向上箭头键(↑)	光标向上移动一个字符
l 或 向右箭头键(→)	光标向右移动一个字符
	如果你将右手放在键盘上的话，你会发现 hjkl 是排列在一起的，因此可以使用这四个按钮来移动光标。 如果想要进行多次移动的 话，例如向下移动 30 行，可以使用 "30j" 或 "30↓" 的组合按键， 亦即加上想要进行的次数(数字)后，按下动作。
[Ctrl] + [f]	屏幕『向下』移动一页，相当于 [Page Down]按键
[Ctrl] + [b]	屏幕『向上』移动一页，相当于 [Page Up] 按键
[Ctrl] + [d]	屏幕『向下』移动半页
[Ctrl] + [u]	屏幕『向上』移动半页
n	那个 n 表示『数字』，例如 20 。按下数字后再按空格键，光标会向右移动这一行的 n 个字符。例如 20 则光标会向后面移动 20 个字符距离。
0 或功能键[Home]	这是数字『 0 』：移动到这一行的最前面字符处
$ 或功能键[End]	移动到这一行的最后面字符处
H	光标移动到这个屏幕的最上方那一行的第一个字符
L	光标移动到这个屏幕的最下方那一行的第一个字符
G	移动到这个档案的最后一行
nG	n 为数字。移动到这个档案的第 n 行。例如 20G 则会移动到这个档案的第 20 行(可配合 :set nu)
gg	移动到这个档案的第一行，相当于 1G。
n	enter 为回车。n 为数字。光标向下移动 n 行

6.2.5 搜索替换的方法

命令	说明
/word	向光标之下寻找一个名称为 word 的字符串。
?word	向光标之上寻找一个字符串名称为 word 的字符串。
	使用 /word 配合 n 及 N 是非常有帮助的！可以让你重复的找到一些你搜寻的关键词！

:n1,n2s/word1/word2/g	n1 与 n2 为数字。在第 n1 与 n2 行之间寻找 word1 这个字符串，并将该字符串取代为 word2！举 例来说，在 100 到 200 行之间搜寻 line 并取代为 LINE 则： 『:100,200s/line/LINE/g』。
:1,$s/word1/word2/g 或 :%s/word1/word2/g	从第一行到最后一行寻找 word1 字符串，并将该字符串取代为 word2
:1,$s/word1/word2/gc 或 :%s/word1/word2/gc	从第一行到最后一行寻找 word1 字符串，并将该字符串取代为 word2！且在取代前显示提示字符给 用户确认 (confirm) 是否需要取代

6.2.6 删除、复制、粘贴的方法

命令	说明
x, X	在一行字当中，x 为向后删除一个字符 (相当于 [del] 按键)， X 为向前删除一个字符(相当于 [backspace] 亦即 是退格键)
nx	n 为数字，连续向后删除 n 个字符。举例来说，我要连续删除 10 个字符,『10x』。
dd	删除游标所在的那一整行
ndd	n 为数字。删除光标所在的向下 n 行，例如 20dd 则是删除 20 行
yy	复制游标所在的那一行(常用)
nyy	n 为数字。复制光标所在的向下 n 行，例如 20yy 则是复制 20 行(常用)
y0	复制光标所在的那个字符到该行行首的所有数据
y$	复制光标所在的那个字符到该行行尾的所有数据
p, P	p 为将已复制的数据在光标下一行贴上，P 则为贴在游标上一行！
J	将光标所在行与下一行的数据结合成同一行
c	重复删除多个数据，例如向下删除 10 行，[10cj]
u	复原前一个动作。(常用)
[Ctrl]+r	重做上一个动作。(常用)
.	小数点,意思是重复前一个动作的意思。 如果你想要重复删除、重复贴上等等动作，按下小数点『.』就好了

6.2.7 储存、离开等指令

命令	说明
:w	将编辑的数据写入硬盘档案中(常用)
:q	离开 vi (常用)

:q!	若曾修改过档案，又不想储存，使用 ! 为强制离开不储存档案。
:wq	储存后离开，若为 :wq! 则为强制储存后离开 (常用)
:w [filename]	将编辑的数据储存成另一个档案（类似另存新档）
:r [filename]	在编辑的数据中，读入另一个档案的数据。亦即将 『filename』 这个档案内容加到游标所在行后面
:n1,n2 w [filename]	将 n1 到 n2 的内容储存成 filename 这个档案。
:! command	暂时离开 vi 到指令行模式下执行 command 的显示结果！例如 『:! ls /home』即可在 vi 当中察看 /home 底下 以 ls 输出的档案信息

6.3 关于 MyNCBI

MyNCBI 的功能之一：保存的搜索策略 (Saved Search Strategies)

保存方法：在使用个人账户登录到 NCBI 并进行 BLAST 之后，采取以下两种操作之一：

1) 在想要保存的 BLAST 报告页面点击 Save Search 按钮；

2) 在 Recent Results 页面，找到想要保存的检索条目，点击对应的 Save 按钮。

用途：保存一次检索使用的程序类型，输入序列和各种参数，以后可以随时查看此次检索的结果。

7 问题与建议

无。