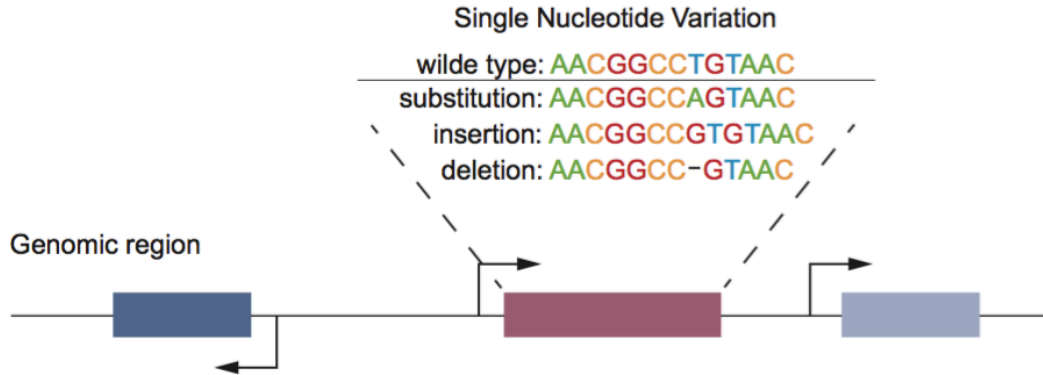


SESSION 6. PRACTICE

Human disease Cancer as a result of aberrant proteins

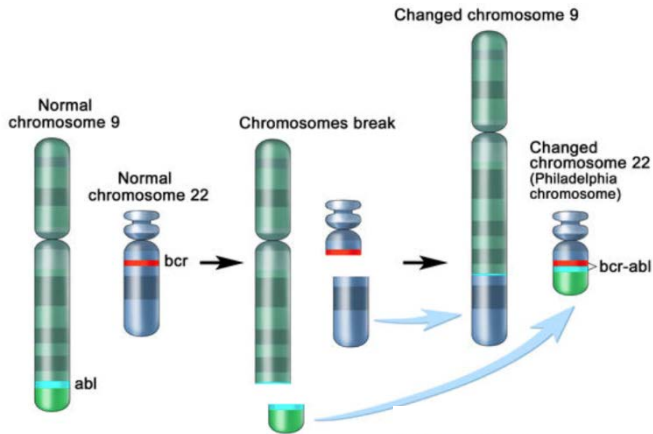
Cancer genomics

2

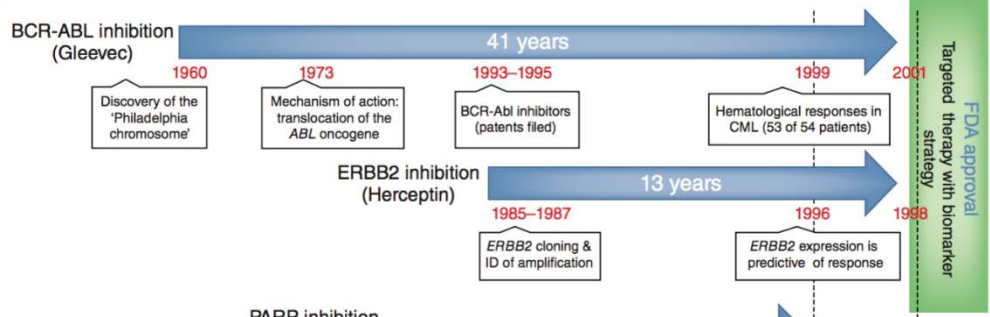


Philadelphia chromosome

3



Tyrosine kinase that is “always on” in CML (Chronic myeloid leukemia) → Uncontrolled division of cells



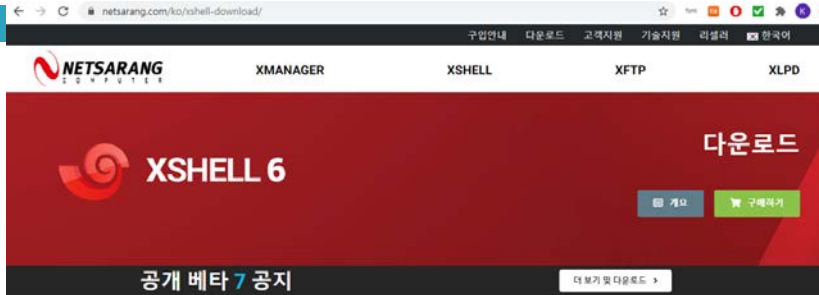
CDAP inhibition

What is Xshell6

- It's ssh (Secure Shell Protocol) client that enables users from remote place to connect to host server through secured way
- It needs either private public key
- You can use Xshell6, one of ssh client that is free of charge, to connect to our server in BIGLAB (CentOS)

Get access for downloading the software

(<https://www.netsarang.com/ko/xshell-download/>)



라이선스 종류를 선택하세요

기재하신 정보 중 이름, 회사(기관), 이메일 정보는 계층을 올리시나, 업데이트 정보 등을 신속하게 제공하기 위해 사용됩니다. 이 정보는 서비스 기간 동안 보관하고 있으며, 이외의 다른 목적으로 사용되지 않습니다.



정품 사용자

제품 키(필수)

버전을 선택하세요(필수)

버전 6 버전 5 버전 4

다음으로



30일 평가판

이름(필수)

이메일(필수)

다운로드 링크가 이메일로 보내집니다.

회사

향후 제품 안내나 프로모션 등과 관련한 이메일 수신에 동의합니다.

평가 시작

가정 및 학교 내 사용자를 위한 무료 라이선스

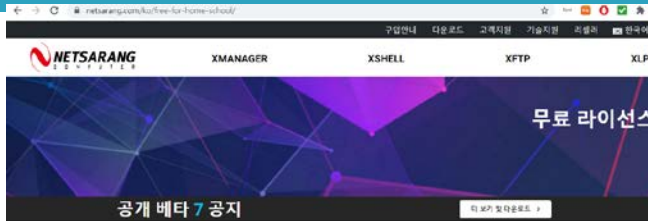
▶ 무료 라이선스 페이지

관련 다운로드

- ▶ 소프트웨어 사용권 및 개인 정보 이용 동의서
- ▶ 사용자 매뉴얼
- ▶ 데이터서트

Click

Get access for downloading the software



무료 사용자의 최고의 선택



최고의 성능

Xshell과 Xftp 무료 라이선스는 성능이나 기능적인 면에서 이례적일 만한 것이 없습니다. 어떤 점도 따라올 수 없습니다.



가정 내 사용

가정에서 개인 서버를 관리하거나 관리 기술을 연마할 수 있습니다. 저가 무료 라이선스는 개인적이고 비상업적 용도에 사용 가능합니다.



학교 내 사용

무료 라이선스는 일기된 교육 기관이 학생, 교직원에게 사용 가능합니다. Xshell과 Xftp는 카드를 받고 배우고 관리하는 데 사용 가능합니다.

  **무료 사용 약관**

비상업적 사용에 한하여 무료

이름(필수)

이메일(필수)

다운로드 Xshell만 Xftp만

주의: 중요한 이메일 주소가 필요합니다. 다운로드 링크가 귀하의 이메일로 보내집니다.

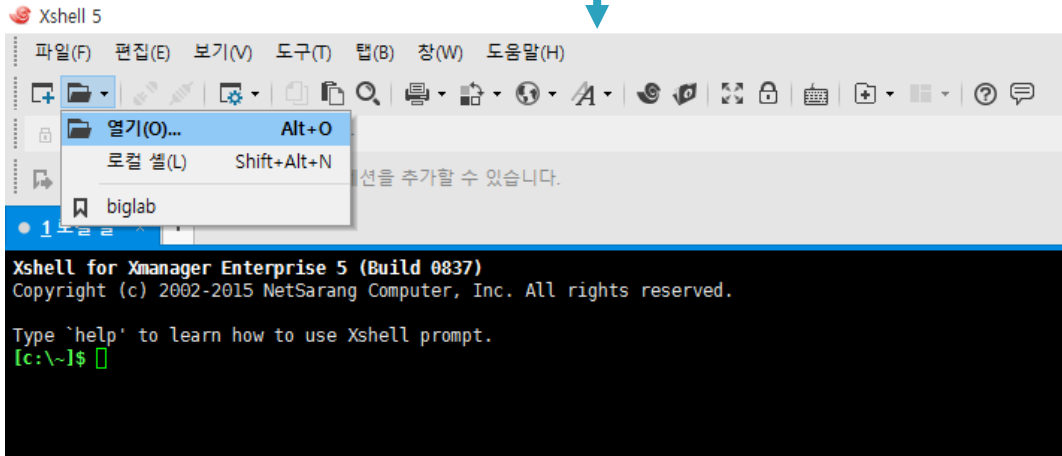
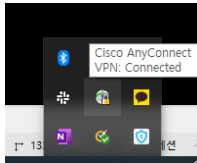
이제 시작할 준비는 완료되었습니다. Xshell과 Xftp를 클라우드에서 번개 그룹을 직접 연결한 무료로 제공되는 것에 대해 축하를 드립니다. 저가 무료 라이선스는 먼저 비용을 요구하지 않는 것이 아니라, 광고, 스파이웨어 또는 사용자에 대한 정보도 공유하지 않습니다. 저가는 어떤 유형이나 조건도 사용하지도 않으며 개인 및 소규모 50명 이하의 사용자 그룹을 대상으로 배포됩니다. 개인용이며, 일차 및 다시 복제에 사용되지 않습니다.

이제 시작할 준비는 완료되었습니다. Xshell과 Xftp는 무료 라이선스는 비상업적 사용으로 제한됩니다. 일부 용도로는 저가 무료 라이선스를 사용하는 것은 저가 무료 라이선스 사용처에 대한 라이선스를 구매하기를 권장합니다. 이는 저가가 저가 제품을 향상시키는 것을 돕는 것입니다.

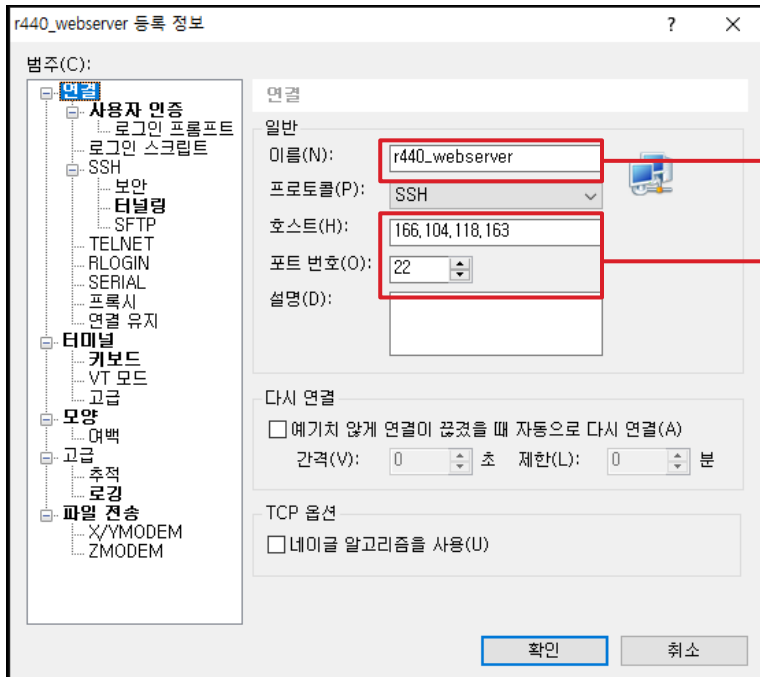
저가 무료 라이선스 제품을 다운로드하는 것은 무료로 발송되는 판매 촉진용 위한 할인 정보나 프로모션 코드 안내와 같은 특별한 이벤트와 관련하여 이메일 수신에 동의한 것을 의미합니다. 대량 이메일 마케팅 수신을 방지하지 않을 경우 해당 이메일 주소의 수신 설정을 변경하여 수신용 제한할 수 있습니다. 저가는 '발대부 크리에이티브'를 3개월간 제공하지 않습니다.

- Fill-in the information
- Download link will be sent the email specified
- Download the software and install

A First Step to Linux Shell



A First Step to Linux Shell




이름은 임의로 지정 가능
= 임의의 이름으로 지정 가능

호스트와 포트 번호는 왼쪽과 동일하게 할 것!
= 호스트와 포트 번호는 왼쪽과 동일하게 할 것!

A First Step to Linux Shell

SSH 보안 경고


 알 수 없는 호스트 키

166.104.138.232(포트: 22) 의 호스트 키가 로컬 호스트 키 데이터베이스에 등록되지 않았습니다. 다음에 이 호스트를 인증하려면 호스트 키를 저장해야 합니다.

호스트 키 손도장(MD5 체크섬):
ssh-rsa 2048 09:ed:33:d9:80:87:e8:7c:63:82:b4:25:ce:e3:6d:4e

이 호스트 키를 수락하시겠습니까?

SSH 사용자 이름

원격 호스트: **166.104.138.232:22 (새 세션)** 


서버 종류: **SSH2, OpenSSH_5.3**

로그인 할 사용자 이름을 입력하십시오(E):

사용자 이름 기억(R)

User: biguser
Password: 2022bio

SSH 사용자 인증

원격 호스트: **166.104.138.232:22 (새 세션)** 

로그인 이름: **biguser**

서버 종류: **SSH2, OpenSSH_5.3**

아래에서 적절한 사용자 인증 방법을 선택하고 로그인하기 위한 정보를 입력하십시오.

Password(P)

암호(W):

Public Key(U)

사용자 키(K):

암호(H):

Keyboard Interactive(I)

사용자 인증에 키보드 입력을 사용합니다.

암호 저장(R)

A First Step to Linux Shell

10

```
Connecting to 166.104.118.163:22...
Connection established.
To escape to local shell, press 'Ctrl+Alt+]'.

Last login: Wed Sep 29 22:24:43 2021 from 166.104.118.183
(py27) [biguser@R440 ~]$ █
```

성공적으로 connection 되면 위와 같은 메시지가 나옴

Basic Shell Commands

ls or ll – **list** directory contents
pwd – **print** name of **working** (current) **d**irectory
cd – **change** **d**irectory

```
(py27) [biguser@R440 ~]$ pwd
/home/biguser
(py27) [biguser@R440 ~]$ ll
total 8
-rw-rw-r-- 1 biguser biguser  0 Sep 27 11:36 jupyter.log
drwxrwxr-x 3 biguser biguser  87 Sep 29 22:19 programs
-rw-rw-r-- 1 biguser biguser  64 Aug 31 17:35 run_jupyter.sh
drwxrwxr-x 26 biguser biguser 4096 Sep 16 11:19 students
drwxrwxr-x 9 biguser biguser  142 Sep 29 22:27 tutor
(py27) [biguser@R440 ~]$ cd /home/biguser/students/
(py27) [biguser@R440 students]$ pwd
/home/biguser/students
(py27) [biguser@R440 students]$ ll
total 4
drwxrwxr-x 7 biguser biguser 126 Sep 30 09:51 2014020664_JunYoungKim
drwxrwxr-x 6 biguser biguser  91 Sep 23 09:34 2016040528_Bangjiyoung
drwxrwxr-x 8 biguser biguser 125 Oct  5 14:47 2016040664_LeeJaeSeung
drwxrwxr-x 5 biguser biguser  74 Sep 16 10:21 2017026108_KimSeoJung
drwxrwxr-x 7 biguser biguser 134 Oct  3 05:57 2017044193_HyunSeokSong
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:09 2017044320_JaeholLee
drwxrwxr-x 7 biguser biguser 187 Sep 30 10:10 2018021976_ChoiSooMin
drwxrwxr-x 4 biguser biguser  57 Sep  9 10:26 2018023390_SeoyeonLee
drwxrwxr-x 7 biguser biguser 134 Sep 30 10:00 2018075996_LeeHaJung
drwxrwxr-x 5 biguser biguser 226 Sep 16 10:23 2019011249_dongjuchang
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:00 2019013527_Jobayeong
drwxrwxr-x 7 biguser biguser 112 Sep 30 10:12 2019020382_SitChoiSee
drwxrwxr-x 7 biguser biguser 130 Sep 30 10:10 2019037510_ChangSungwoo
drwxrwxr-x 4 biguser biguser 154 Oct  5 12:54 2019079425_LeeSera
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:10 2019083372_JangYeoun
drwxrwxr-x 7 biguser biguser 130 Sep 30 10:02 2019086990_ChunJongHyun
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:12 2019090710_ChoelIwon
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:02 2021118974_jihoonkim
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:10 2021151522_Hansoll_Na
drwxrwxr-x 7 biguser biguser 122 Sep 30 09:03 2021178423_SangjinLee
drwxrwxr-x 7 biguser biguser 112 Sep 30 09:04 2021206755_gayoungkim
-rw-rw-r-- 1 biguser biguser 1459 Sep  8 23:02 5.translation_code.ipynb
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:05 9143920218_KimMinjoo
drwxrwxr-x 3 biguser biguser  40 Sep  9 10:36 mhchoi
(py27) [biguser@R440 students]$ cd ..
(py27) [biguser@R440 ~]$ pwd
/home/biguser
(py27) [biguser@R440 ~]$ ll
total 8
-rw-rw-r-- 1 biguser biguser  0 Sep 27 11:36 jupyter.log
drwxrwxr-x 3 biguser biguser  87 Sep 29 22:19 programs
-rw-rw-r-- 1 biguser biguser  64 Aug 31 17:35 run_jupyter.sh
drwxrwxr-x 26 biguser biguser 4096 Sep 16 11:19 students
drwxrwxr-x 9 biguser biguser  142 Sep 29 22:27 tutor
(py27) [biguser@R440 ~]$
```

Basic Shell Commands

ls or ll – **list** directory contents
pwd – **print** name of **working** (current) **directory**
cd – **change directory**
mkdir – **make** new **directory**

```
(py27) [biguser@R440 ~]$ ll
total 8
-rw-rw-r-- 1 biguser biguser 0 Sep 27 11:36 jupyter.log
drwxrwxr-x 3 biguser biguser 87 Sep 29 22:19 programs
-rw-rw-r-- 1 biguser biguser 64 Aug 31 17:35 run_jupyter.sh
drwxrwxr-x 26 biguser biguser 4096 Sep 16 11:19 students
drwxrwxr-x 9 biguser biguser 142 Sep 29 22:27 tutor
(py27) [biguser@R440 ~]$ cd /home/biguser/students/
(py27) [biguser@R440 students]$ ll
total 4
drwxrwxr-x 7 biguser biguser 126 Sep 30 09:51 2014020664_JunYoungKim
drwxrwxr-x 6 biguser biguser 91 Sep 23 09:34 2016040528_BangJiyoung
drwxrwxr-x 8 biguser biguser 125 Oct 5 14:47 2016040664_LeeJaeSeung
drwxrwxr-x 5 biguser biguser 74 Sep 16 10:21 2017026108_KimSeoJung
drwxrwxr-x 7 biguser biguser 134 Oct 3 05:57 2017044193_HyunSeokSung
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:09 2017044320_JaehoLee
drwxrwxr-x 7 biguser biguser 187 Sep 30 10:10 2018021976_ChoiSooMin
drwxrwxr-x 4 biguser biguser 57 Sep 9 10:26 2018023390_SeoyeonLee
drwxrwxr-x 7 biguser biguser 134 Sep 30 10:00 2018075996_LeeHaJung
drwxrwxr-x 5 biguser biguser 226 Sep 16 10:23 2019011249_dongjuchang
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:00 2019013527_JoDayeong
drwxrwxr-x 7 biguser biguser 112 Sep 30 10:12 2019020382_SitChoiSee
drwxrwxr-x 7 biguser biguser 130 Sep 30 10:10 2019037510_ChangSungwoo
drwxrwxr-x 4 biguser biguser 154 Oct 5 12:54 2019079425_LeeSera
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:10 2019083372_JangYewon
drwxrwxr-x 7 biguser biguser 130 Sep 30 10:02 2019086999_ChunJongHyun
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:12 2019090719_ChoEJiwon
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:02 2021110974_jihoonkim
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:10 2021151522_Hansoll_Na
drwxrwxr-x 7 biguser biguser 122 Sep 30 09:03 2021178423_SangjinLee
drwxrwxr-x 7 biguser biguser 112 Sep 30 09:04 2021206755_gayoungkim
-rw-rw-r-- 1 biguser biguser 1459 Sep 8 23:02 5.translation_code.ipynb
drwxrwxr-x 7 biguser biguser 108 Sep 30 10:05 9143920218_KimMinjoo
drwxrwxr-x 3 biguser biguser 40 Sep 9 10:36 mhchoi
(py27) [biguser@R440 students]$ cd mhchoi/
(py27) [biguser@R440 mhchoi]$ pwd
/home/biguser/students/mhchoi
(py27) [biguser@R440 mhchoi]$ ll
total 0
(py27) [biguser@R440 mhchoi]$ mkdir week6
(py27) [biguser@R440 mhchoi]$ cd week6/
(py27) [biguser@R440 week6]$ pwd
/home/biguser/students/mhchoi/week6
(py27) [biguser@R440 week6]$
```

Basic Shell Commands

```
$ ll  
$ vi print_script.py
```

```
[In Vim]  
i: insert
```

```
print "Today is Wednesday!"
```

```
[In Vim] "ESC"  
:wq save and quit
```

Basic Shell Commands

```
$ ll  
$ python print_script.py
```

[Output] Today is Wednesday!

```
$ less print_script.py  
$ cat print_script.py
```

[Output] print "Today is Wednesday!"

Basic Shell Commands

```
$ mv print_script.py new_script.py
$ cp new_script.py new_script_copy.py
$ ll
```

mv : **move file**

cp : **c**opy and **p**aste

```
$ rm new_script_copy.py
$ ll
```

rm : **r**emove

Basic Shell Commands

16

```
$ cd <Your Directory>  
$ mkdir session6  
$ cd session6
```


Using BLAST+ to examine the BCR-ABL

Workflow

17

1. Database indexing
-> **makeblastdb**
2. Download or generate sequence data (fasta format)
-> **blastdbcmd**
3. Global alignment of protein sequences
-> **blastp**
4. Evaluation

Using BLAST+ to examine the BCR-ABL

Workflow

18

1. Database indexing
-> **makeblastdb**
2. Download or generate sequence data (fasta format)
-> **blastdbcmd**
3. Global alignment of protein sequences
-> **blastp**
4. Evaluation

Using BLAST+ to examine the BCR-ABL (**makeblastdb**)

19

```
$ ln -s /home/biguser/tutor/Session6/swissprot .  
(의 명령어는 반드시 여러분의 working director에 Session6 directory에서 실행하세요!!)  
$ ls  
$ less swissprot  
$ makeblastdb -help
```

-in <File_In>

Input file/database name

Default = `-'

-dbtype <String, `nucl', `prot'>

Molecule type of target db

-parse_seqids

Option to parse seqid for FASTA input if set, for all other input types seqids are parsed automatically

Removing corrupted soft-link file

20

```
(py27) [biguser@R440 tutor]$ ln -s /home/biguser/tutor/week6/abl1.fa
(py27) [biguser@R440 tutor]$ ll
total 4
lrwxrwxrwx 1 biguser biguser  33 Oct  6 23:25 abl1.fa -> 
drwxrwxr-x 2 biguser biguser  69 Sep  8 22:09 Week1
drwxrwxr-x 3 biguser biguser 285 Sep  9 10:46 Week2
drwxrwxr-x 3 biguser biguser  97 Sep 16 11:20 Week3
drwxrwxr-x 3 biguser biguser 4096 Sep 29 23:53 Week4
drwxrwxr-x 3 biguser biguser  163 Sep 29 23:46 Week5
drwxrwxr-x 2 biguser biguser  221 Oct  6 23:24 Week6
(py27) [biguser@R440 tutor]$ unlink abl1.fa
(py27) [biguser@R440 tutor]$ ll
total 4
drwxrwxr-x 2 biguser biguser  69 Sep  8 22:09 Week1
drwxrwxr-x 3 biguser biguser 285 Sep  9 10:46 Week2
drwxrwxr-x 3 biguser biguser  97 Sep 16 11:20 Week3
drwxrwxr-x 3 biguser biguser 4096 Sep 29 23:53 Week4
drwxrwxr-x 3 biguser biguser  163 Sep 29 23:46 Week5
drwxrwxr-x 2 biguser biguser  221 Oct  6 23:24 Week6
(py27) [biguser@R440 tutor]$
```

절대 rm 명령어로 지우지 마세요! unlink 명령어로 제거할 것!

Using BLAST+ to examine the BCR-ABL (**makeblastdb**)

21

```
$ makeblastdb -in swissprot -dbtype prot -parse_seqids  
$ ll
```

```
-rw-rw-r-- 1 biguser biguser 280362561 Oct  6 23:04 swissprot  
-rw-rw-r-- 1 biguser biguser 104599070 Oct  6 23:23 swissprot.phr  
-rw-rw-r-- 1 biguser biguser  4522104 Oct  6 23:23 swissprot.pin  
-rw-rw-r-- 1 biguser biguser  2261048 Oct  6 23:23 swissprot.pog  
-rw-rw-r-- 1 biguser biguser 17902158 Oct  6 23:23 swissprot.psd  
-rw-rw-r-- 1 biguser biguser   421163 Oct  6 23:23 swissprot.psi  
-rw-rw-r-- 1 biguser biguser 204416076 Oct  6 23:23 swissprot.psq
```

Using BLAST+ to examine the BCR-ABL

Workflow

22

1. Database indexing
-> `makeblastdb`
2. Download or generate sequence data (fasta format)
-> `blastdbcmd`
3. Global alignment of protein sequences
-> `blastp`
4. Evaluation

Using BLAST+ to examine the BCR-ABL (**blastdbcmd**)

23

```
$ blastdbcmd -help
```

-entry <String>

Comma-delimited search string(s) of sequence identifiers:

e.g.: 555, AC147927, 'gnl|dbname|tag', or 'all' to select all sequences in the database

* Incompatible with: entry_batch, ipg, ipg_batch, taxids, taxidlist, info, tax_info, list, recursive, remove_redundant_dbs, list_outfmt, show_blastdb_search_path

-db <String>

BLAST database name

Default = `nr`

* Incompatible with: list, recursive, remove_redundant_dbs, list_outfmt, show_blastdb_search_path

-long_seqids

Use long seq id for fasta deflines

* Incompatible with: info

Using BLAST to examine the BCR-ABL (**blastdbcmd**)

24

```
$ blastdbcmd -entry ABL1_HUMAN -db swissprot -long_seqids > abl1.fa  
$ less abl1.fa
```

```
>gi|85681908|sp|P00519.4|ABL1_HUMAN RecName: Full=Tyrosine-protein kinase ABL1; AltName: Full  
=Abelson murine leukemia viral oncogene homolog 1; AltName: Full=Proto-oncogene c-Abl; AltNam  
e: Full=p150  
MLEICLKLVGCKSKKGLSSSSCYLEEALQRPVASFPEQGLSEAAWNSKENLLAGPSENDPNLFVALYDFVASGDNTL  
SITKGEKLRVLGYNHNGEWCEAQTKNQGQWVPSNYITPVNSLEKHSWYHGPVSRNAAEYLLSSGINGSFLVRESESSPGQ  
RSISLRYEGRVYHYRINTASDGKLYVSSESRFNTLAE LVHHSTVADGLITTLHYPAKRNKPTVYGVSPNYDKWEMERT  
DITMKHKLGGGQYGEVYEGVWKKYSLTVAVKTLKEDTMEVEEFLEKAAVMKEIKHPNLVQLLGVCTREPPFYIITEFMTY  
GNLLDYLR CNRQEVNAVLLYMATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTAHAGAK  
FPIKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLSQVYELLEKDYRMERPEGCEKVYELMRACWQWNP  
SDRPSFAEIHQAFETMFQESSISDEVEKE LGKQGVRGAVSTLLQAPELPTKTRTSRRAAEHRDITDVPMPHSGQGQESD  
PLDHEPAVSPLLPRKERGPPEGLNEDERLLPKDKKTNLFSALIKKKKKTAPTPPKRSSSFREMDGQPERRGAGEEEGRD  
ISNGALAFPLDTADPAKSPKPSNGAGVPNGALRESGSGFRSPHLWKKSSLTSSRLATGEEEGGSSSRFLRSCSAS  
CVPHGAKDTEWRSVTLPRDLQSTGRQFDSSTFGGHKSEKPALPRKRAGENRSQVTRGTVTPPRLLVKKNEEADEVFKD  
IMESSPGSSPNNLTPKPLRRQVTVAPASGLPHKEEAGKGSALGTPAAAEPVTPTSKAGSGAPGGTSKGPAAESRVRHKKH  
SSESPGRDKGKLSRLKPAPPPPAASAGKAGGKPSQSPSQEAAGEAVLGAKTKATSLVDVNSDAAKPSQPGGLKPKVPL  
PATPKPQSAKPSGTPISPAPVPSTLPSASSALAGDQPSSTAFIPLISTRVSLRKRTRQPPERIASGAITKGVVLDSTEALC  
LAISRNSQMASHSAVLEAGKNLYTFCVSYVDSIQMRNKFAFREAINKLENNLRELQICPATAGSGPAATQDFSKLLSS  
VKEISDIVQR
```


Using BLAST+ to examine the BCR-ABL

Workflow

25

1. Database indexing
-> `makeblastdb`
2. Download or generate sequence data (fasta format)
-> `blastdbcmd`
3. Global alignment of protein sequences
-> `blastp`
4. Evaluation

Using BLAST+ to examine the BCR-ABL (blastp)

26

```
$ ln -s /home/biguser/tutor/Week6/bcrabl.fa .  
$ less bcrabl.fa
```

```
>tr|A9UF02|A9UF02_HUMAN BCR/ABL fusion protein isoform X9 OS=Homo sapiens GN=BCR/ABL fusion P  
E=2 SV=1  
MVPDPVGF AEAWKAQFPDSEPPRME LRSVGDIEQELERCKASIRRLQE VNVQERFRMIY LQ  
TLLAKEKKS YDRQRWGFRRAAQAPD GASEPRASASRPQPAPADGADPPPAEPEEARPDGE  
GSPGKARPTARRPGAAA SGERDDRGPPASVAALRSNFERIRKGHGQPGADA EKPFYVNV  
EFHHERGLVKVNDKEVSDRISLGSQAMQMERKKSQHAGAGSSVGDASRPPYGRSSSESSC  
GVDGDYEDAE LNPRFLKDNLIDANGGSRPPWPPLEYQPYQSIYVGGMMEGEGKGPLLRSQ  
STSEQEKRLTWPRRSYSPRSFEDCGGGYTPDCSSNENLTSSEEDFSSGQSSRVSPSPTTY  
RMFRDKSRSPQNSQQSFDSSSPPTQCHKRHRHCPVVVSEATIVGVRKTGQIWPNDGEG  
AFHGADAGSFGTPPGYGCAADRAEEQRRHQDGLPYIDDS PSSSPHLSKGRGSRDALVSG  
ALESTKASELDLEKGLMERKVVLSGILASEETYLSHLQMLTNSCVKLQTVHSHPLTINKE  
EALQRPVADSFEPQGLSE AARWNSKENLLAGPSENDPNLFVALYDFVASGDNTLSITKGE  
KLRVLYGNHNGEWCEAQTKNQGQWVPSNYITPVNSLEKHSWYHG PVS RNAEYLLSSGIN  
GSFLVRESESSPQRSISLRYEGRVYHYRINTASDGKLYVSSSESRFNTLAE LVHHHSTVA  
DGLITTLHYPAKRNKPTVYGVSPNYDKWEMERTDITMKHKLGGGQYGEVYEGWVKYSL  
TVAVKTLKEDTMEVEEFLKEAAVMKEIKHPNLVQLGVCTREPPFYIITFMTYGNLLDY  
LRECNQDEVNAVLLYMATQISSAMEYLEKKNF IHRDLAARNCLVGENHLVKVADFLGLSR  
LMTGDTYTAHAGAKFP IKWTAPE SLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLS  
QVYELLEKDYRMERPEGCEKPYVELMRACWQWNP SDRPSFAEIHQAFETMFQESSISDEV  
EKELGKQGVRGAVSTLLQAPELPTKTRTSRRAAEHRD TTDVPEMPHSKGQGESDPLDHEP  
AVSPLLPKPKERGPEGGLNEDERLLPKDKKTNLFSALIKKKKKTAPT PPKRSSSFREMDG  
QPERRAGEEEGRDISNGALFTPLDTADPAKSPKPSNGAGVPNGALRESGGSGFRSPHL  
WKKSSTLTSSRLATGEEEGGSSSKRFLRSCSASC VPHGAKDTEWRSVTLPRDLQSTGRQ  
FDSSTFGGHKSEKPALPKRKRAGENRSDQVTRGTVTPPRLVKKNEEAADEVFKDIMESSP  
GSSPNLTPKPLRRQVTVAPASGLPHKEEAGKGSALGTPAAAEPVPTPSKAGSGAPGGTS  
KGAPEESRVRRHKHSSESPGRDKGLSRLKPAPPPPPAASAGKAGGKPSQSPQEAAGEA  
VLGAKTKATSLVD AVNSDAAKPSQPEGGLKPKVLPATPKPQSAKPSGTPISPAPVSTLP  
SASSALAGDQPSSTAFIPLISTRVLRKTRQPPERIASGAIITKGVVLDSTEALCLATSRN  
SEQMASHSAVLEAGKNLYTFCVSYVDSIQMRNKF AFREAINKLENNLRELQICPATAGS  
GPAATQDFSKLLSSVKEISDIVQR
```

Using BLAST+ to examine the BCR-ABL (blastp)














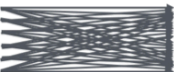

27

```
$ blastp -help
```

```
-db <String>  
  BLAST database name  
  * Incompatible with:  subject, subject_loc  
  
-query <File_In>  
  Input file name  
  Default = '-'  
  
-out <File_Out>  
  Output file name  
  Default = '-'  
  
-show_gis  
  Show NCBI GIs in deflines?
```

Using BLAST+ to examine the BCR-ABL (blastp)

28

Program	Query Type	Subject Type	Computation
blastn	N 	  N	~ 1X
blastp	P 	  P	~ 1X
blastx	N 	  P	~ 6X
tblastn	P 	  N	~ 6X
tblastx	N 	  N	~36X

Using BLAST+ to examine the BCR-ABL (blastp)

29

```
$ blastp -db swissprot -query bcrabl.fa -out bcrabl.blastp -show_gis  
$ ls
```

```
bcrabl.blastp  
bcrabl.fa  
swissprot  
swissprot.phr  
swissprot.pin  
swissprot.pnd  
swissprot.pni  
swissprot.pog  
swissprot.psd  
swissprot.psi  
swissprot.psq
```

Using BLAST to examine the BCR-ABL

Workflow

30

1. Database indexing
-> `makeblastdb`
2. Download or generate sequence data (fasta format)
-> `blastdbcmd`
3. Global alignment of protein sequences
-> `blastp`
4. Evaluation

Using BLAST+ to examine the BCR-ABL (blastp)

31

```
$ less bcrabl.blastp
```

```
Length=1644
Sequences producing significant alignments:
                                Score      E
                                (Bits)    Value
sp|P00519|ABL1_HUMAN|Tyrosine-protein kinase ABL1 OS=Homo sapiens ... 2268    0.0
sp|P00520|abl-1_MOUSE|Tyrosine-protein kinase ABL1 OS=Mus musculus ... 1885    0.0
sp|P11274|BCR_HUMAN|Breakpoint cluster region protein OS=Homo sapiens ... 1084    0.0
sp|P42684|ABL2_HUMAN|Tyrosine-protein kinase ABL2 OS=Homo sapiens ... 991     0.0
sp|Q4JIM5|ABL2_MOUSE|Tyrosine-protein kinase ABL2 OS=Mus musculus ... 963     0.0
sp|P10447|ABL_FSVH|Tyrosine-protein kinase transforming protein A... 932     0.0
sp|P00521|ABL_MLVAB|Tyrosine-protein kinase transforming protein A... 895     0.0
sp|Q6PAJ1|BCR_MOUSE|Breakpoint cluster region protein OS=Mus muscu... 867     0.0
sp|P00522|ABL_DROME|Tyrosine-protein kinase Abl OS=Drosophila mela... 780     0.0
sp|P03949|ABL1_CAEEL|Tyrosine-protein kinase abl-1 OS=Caenorhabdit... 619     0.0
sp|F1LXF1|BCR_RAT|Breakpoint cluster region protein OS=Rattus norv... 563     1e-176
sp|P42686|SRK1_SPOLA|Tyrosine-protein kinase isoform SRK1 OS=Spong... 419     3e-130
sp|P42690|SRK4_SPOLA|Tyrosine-protein kinase isoform SRK4 OS=Spong... 410     5e-127
sp|P17713|STK_HYDVI|Tyrosine-protein kinase STK OS=Hydra vulgaris ... 400     2e-123
sp|Q9V9J3|SRC42_DROME|Tyrosine-protein kinase Src42A OS=Drosophila... 399     6e-123
sp|Q922K9|FRK_MOUSE|Tyrosine-protein kinase FRK OS=Mus musculus OX... 392     2e-120
sp|Q62662|FRK_RATT|Tyrosine-protein kinase FRK OS=Rattus norvegicus... 390     5e-120
sp|P42685|FRK_HUMAN|Tyrosine-protein kinase FRK OS=Homo sapiens OX... 383     2e-117
sp|Q02977|YRK_CHICK|Proto-oncogene tyrosine-protein kinase Yrk OS=... 381     3e-116
sp|A1A5H8|YES_DANRE|Tyrosine-protein kinase yes OS=Danio rerio OX=... 379     2e-115
sp|Q6EWH2|FYNA_DANRE|Tyrosine-protein kinase fyn OS=Danio rerio O... 379     2e-115
sp|P09324|YES_CHICK|Tyrosine-protein kinase Yes OS=Gallus gallus O... 378     3e-115
sp|P07947|YES_HUMAN|Tyrosine-protein kinase Yes OS=Homo sapiens OX... 378     4e-115
sp|Q04736|YES_MOUSE|Tyrosine-protein kinase Yes OS=Mus musculus OX... 378     6e-115
sp|P27446|FYN_XIPH|Tyrosine-protein kinase Fyn OS=Xiphophorus hel... 376     1e-114
```

Using BLAST+ to examine the BCR-ABL (blastp)

32

```
>gi|85681908|sp|P00519.4|ABL1_HUMAN RecName: Full=Tyrosine-protein kinase ABL1; AltName: Full=Abelson murine leukemia viral oncogene homolog 1; AltName: Full=Proto-oncogene c-Abl; AltName: Full=p150 Length=1130
```

The sequence we inserted

```
Score = 2268 bits (5876), Expect = 0.0, Method: Compositional matrix adjust.  
Identities = 1105/1105 (100%), Positives = 1105/1105 (100%), Gaps = 0/1105 (0%)
```

```
Query 540 EEALQRPVASFEPQGLSEAAARWNSKENLLAGPSENDPNLFVALYDFVASGDNTLSITKG 599  
Sbjct 26 EEALQRPVASFEPQGLSEAAARWNSKENLLAGPSENDPNLFVALYDFVASGDNTLSITKG 85  
Query 600 EKLRVLGYNHNGEWCEAQTKNQGQGWVPSNYITPVNSLEKHSWYHGPVSRNAEYLLSSGI 659  
Sbjct 86 EKLRVLGYNHNGEWCEAQTKNQGQGWVPSNYITPVNSLEKHSWYHGPVSRNAEYLLSSGI 145  
Query 660 NGSFLVRESESSPGQRSISLRYEGRVYHYRINTASDGKLYVSSESRFNTLAELVHHHSTV 719  
Sbjct 146 NGSFLVRESESSPGQRSISLRYEGRVYHYRINTASDGKLYVSSESRFNTLAELVHHHSTV 205  
Query 720 ADGLITTLHYAPKRNKPTVYGVSPNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKKYS 779  
Sbjct 206 ADGLITTLHYAPKRNKPTVYGVSPNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKKYS 265  
Query 780 LTVAVKTLKEDTMEVEEFLKEAAMKEIKHPNLVQLLGVCTREPPFYIITEFMTYGNLLD 839  
Sbjct 266 LTVAVKTLKEDTMEVEEFLKEAAMKEIKHPNLVQLLGVCTREPPFYIITEFMTYGNLLD 325
```

A sequence in DB

Using BLAST+ to examine the BCR-ABL (blastp)

33

```
>gi|143811366|sp|P11274.2|BCR_HUMAN RecName: Full=Breakpoint cluster region protein; AltName:  
Full=Renal  
carcinoma antigen NY-REN-26  
Length=1271
```

```
Score = 1084 bits (2804), Expect = 0.0, Method: Compositional matrix adjust.  
Identities = 518/520 (99%), Positives = 519/520 (99%), Gaps = 0/520 (0%)
```

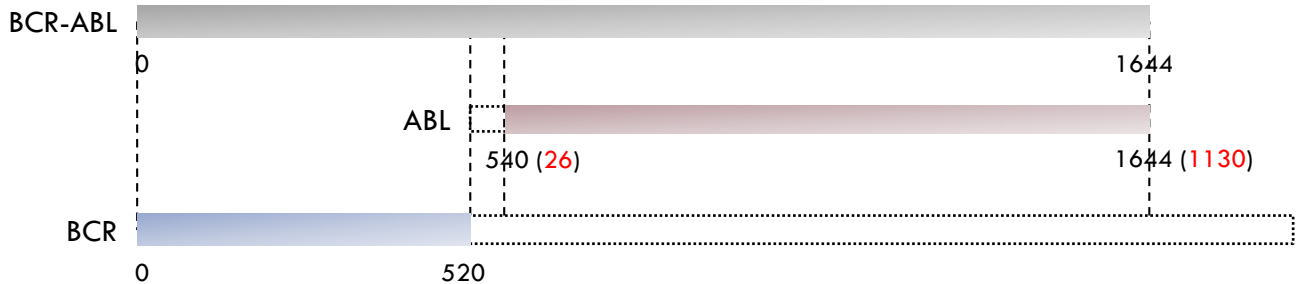
```
Query 1 MVDPVGF AEAWKAQFPDSEPPRMELRSVGDIEQELERCKASIRRLEQEVNQERFRMIYLQ 60  
Sbjct 1 MVDPVGF AEAWKAQFPDSEPPRMELRSVGDIEQELERCKASIRRLEQEVNQERFRMIYLQ 60  
  
Query 61 TLLAKEKKS YDRQRWGFRRAAQAPDGASEPRASASRPQAPADGADPPPAEEPEARPDGE 120  
Sbjct 61 TLLAKEKKS YDRQRWGFRRAAQAPDGASEPRASASRPQAPADGADPPPAEEPEARPDGE 120  
  
Query 121 GSPGKARPGTARRPGAAASGERDDRGPPASVAALRSNFERIRKGGHGQPGADAEKPFYVNV 180  
Sbjct 121 GSPGKARPGTARRPGAAASGERDDRGPPASVAALRSNFERIRKGGHGQPGADAEKPFYVNV 180  
  
Query 181 EFHHERGLVKVNDKEVSDRISLGSQAMQMERKKSQHGAGSSVGDASRPPYRGRSSESSC 240  
Sbjct 181 EFHHERGLVKVNDKEVSDRISLGSQAMQMERKKSQHGAGSSVGDASRPPYRGRSSESSC 240  
  
Query 481 ALESTKASELDLEKGLEMRKWVLSGILASEETYLSHLQML 520  
Sbjct 481 ALESTKASELDLEKGLEMRKWVLSGILASEETYLSHL+ L 520
```

Conservative substitution

←

Using BLAST+ to examine the BCR-ABL (blastp)

34



Exercise

35

- A typical application of BLAST is to identify sequences that are homologous to a query sequence. Modify the BLAST search of this chapter so you will be able to identify homologous of the human BCR protein (BCR_HUMAN) from Swiss-Prot.

- Which portion of BCR_HUMAN is similar to ABR_HUMAN?

More about ABR;

<https://www.wikigenes.org/e/gene/e/29.html>

<https://www.genecards.org/cgi-bin/carddisp.pl?gene=ABR&keywords=ABR>

Exercise

36

```
$ blastdbcmd -entry BCR_HUMAN -db swissprot -long_seqids > bcr.fa
$ less bcr.fa
```

```
>gi|143811366|sp|P11274.2|BCR_HUMAN RecName: Full=Breakpoint cluster region protein; AltName:
  Full=Renal carcinoma antigen NY-REN-26
MVDPVGFAEAWKAQFPDSEPPRMELRSVGDIEQE LERCKASIRRLQE V NQERFRMIY LQTLLAKEKKS YDRQRWGFRA
AQAPDGASEPRASASRPQAPADGADPPP AEEPE ARPDGEGSPGKARPGTARRPGAAASGERDDRGPPASVAALRSNFER
IRKGGHQPGADA EKFYVNVF FHHERGLV KVN DKEVSDRISSLGSQAMQMERKKSQHGAGSSVGDASRPPYGRSSESSC
GVDGDYEDAELNPRFLKDNLIDANGSRPPWPPLE YQPYQSIYVGGMMEGEGKGPLLR SQTSEQE KRLTWP RRSYSPRS
FEDCGGGYTPDCSSNENLTSSEDFSSGQSSRVSPSP TTYRMFRDKSRSPSQNSQQSFDSSSPPTPQCHKRHRHCPVVVS
EATIVGV RKTGQIWPNDGEGAFHG DADG SFGTPPGYCAADRAEEQRRHQDGLPYID DSSPHLSSKGRGRSDALVSG
ALESTKASELDLEKGLEMRKWVLSGILASEE TYLSHLEALLPMKPLKAAATTSQPVLTSQQIETIFFKVP ELYEIHKEF
YDGLFPRVQVQWSHQQRVGD L FQKLASQLGVYRAFVDNYGVAMEMA EKCQANAQFAEISENL RARSNKDAKDPTTKNSLE
TLLYKPVDRVTRSTLVLHDL LKHTPASHPDHPLLQDALRISQNF LSSINEEITPRRQ SMTVKKGEHRQLLKDSFMV E LVE
GARKLRHVFLFTDLLLCTKLKKQSGGKTQYDCKWYIPLTDL SFQMVDELEAVPN IPLVPDEE LDALKIKISQIKNDIQ
EKRANKGSKATERLKKLSEQES LLLMSPSMAFRVHSRNGKSYTFLISSDYERA EWRENIREQQKCFRSFSLTSVELQ
MLTNSCVKLQTVHSIPLTINKEDDESPLYGFLNVIHVSATGFKQSSNLYCTLEVDSFGYFVNKAKTRVRYRDTAEPN WNE
EFEIELEGSQTLRILCYEKYCNKTKIPKEDGESTDRLMGKGQVQLDPQALQDRDWQRTVIAMNGIEVKLSVKFNSRE FSL
KRMP SRKQTGVFGVKIAVVTKRERSKVPIYIVRQCVEEIERGMEEVGIYRVSGVATDIQALKA AFDVNNKDV SVMSEM
VNAIAGTLKLYFRELP ELPFTDFEYFPNFAEGIALSDPVAKESCMLNLLLSLPEANLLTFLFLLDHLKRVAEKEAVNKMSL
HNLATVFGPTLLRPSEKESKLPANPSQPI TMTDSWSLEVMSQVQLLYFLQL EAIAPAPDSKRQSI L FSTEV
```

Exercise

37

```
$ blastp -db swissprot -query bcr.fa -out bcr.blastp -show_gis  
$ less bcr.blastp
```

Sequences producing significant alignments:	Score (Bits)	E Value
gi 143811366 sp P11274.2 BCR_HUMAN RecName: Full=Breakpoint clus...	2638	0.0
gi 124007119 sp Q6PAJ1.2 BCR_MOUSE RecName: Full=Breakpoint clus...	2343	0.0
gi 5915668 sp Q12979.1 ABR_HUMAN RecName: Full=Active breakpoint...	1172	0.0
gi 215275190 sp A6QNS3.1 ABR_BOVIN RecName: Full=Active breakpoi...	1172	0.0
gi 81910107 sp Q5SSL4.1 ABR_MOUSE RecName: Full=Active breakpoin...	1171	0.0

Exercise

38

```
>gi|5915668|sp|Q12979.1|ABR_HUMAN RecName: Full=Active breakpoint cluster region-related protein  
Length=859
```

```
Score = 1172 bits (3032), Expect = 0.0, Method: Compositional matrix adjust.  
Identities = 573/832 (69%), Positives = 685/832 (82%), Gaps = 11/832 (1%)
```

```
Query 444 EEQRRHQDG---LPYIDDSPSSPHLSSKGRGSRDALVSGALESTKASELDLEKGLEMRK 500  
EEQ+ +G +PYID+SP+ SP LS++ +G D VS A ++ KGLEMRK  
Sbjct 35 EEQKGPPEGSETMPYIDESPTMSPQLSARSQGRGDG-VSPTPPEGLAPGVEAGKGLEMRK 93  
  
Query 501 WVLSGILASEETYLSHLEALLLPMKPLKAAATTSQPVLTSQQIETIFFKVPELYEIHKEF 560  
VLSG LASEE Y++ LEALLLPMKPLKA ATTSQPVLTTQQIETIF+K+ ++YEIHKEF  
Sbjct 94 LVLSGFLASEEIIYINQLEALLLPMKPLKATATTSQPVLTIQQIETIFFYKIQDIYEIHKEF 153  
  
Query 1160 KESCLNLLLSLPEANLLTFLFLLDHLKRVAEKEAVNKMSLHNLATVFGPTLLRPSEKES 1219  
KE+CM++LL SLP+ NL+TFLFLL+HLKRVAEKE +NKMSLHNLATVFGPTLLRPSE ES  
Sbjct 753 KENCMMHLLRSLDPNLITFLFLEHLKRVAEKEPINKMSLHNLATVFGPTLLRPSEVES 812  
  
Query 1220 KLPANPSQPITMTDSWSLEVMSQVQLLYFLQLEAIPAPDSKRQSILFSTEV 1271  
K + + D WS +VM+QVQVLLY+LQ I + KR ++ FST+V  
Sbjct 813 K-----AHLTSAADIWSDVMAQVQVLLYYLQHPPISFAELKRNTLYFSTDV 859
```

□ 이번주 과제는 어렵습니다.