

BIOINFORMATICS

SESSION 6. PRACTICE

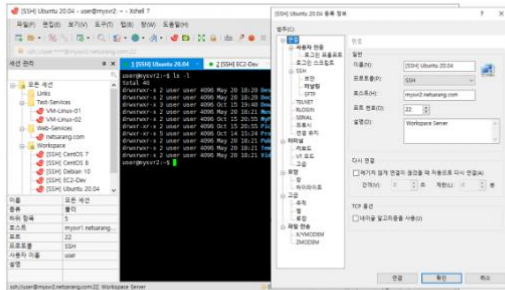
2023-10-09

Iron imbalance and the iron responsive element

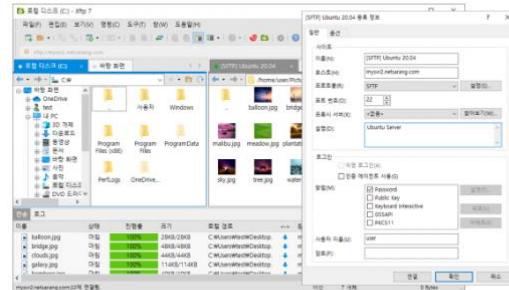
Get access for downloading the software



업계 최고의 강력한 SSH 클라이언트



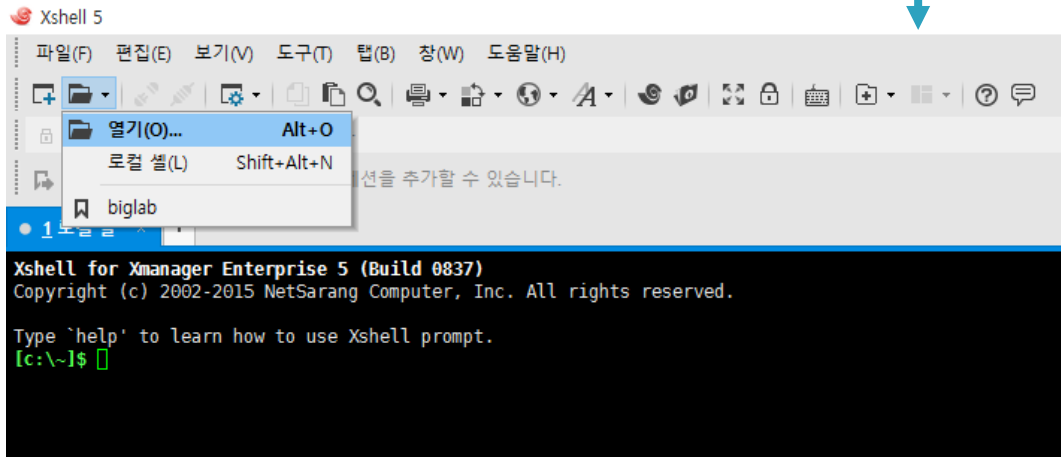
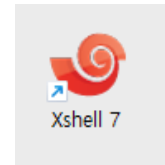
네트워크를 통한 수고스럽지 않은 파일 전송



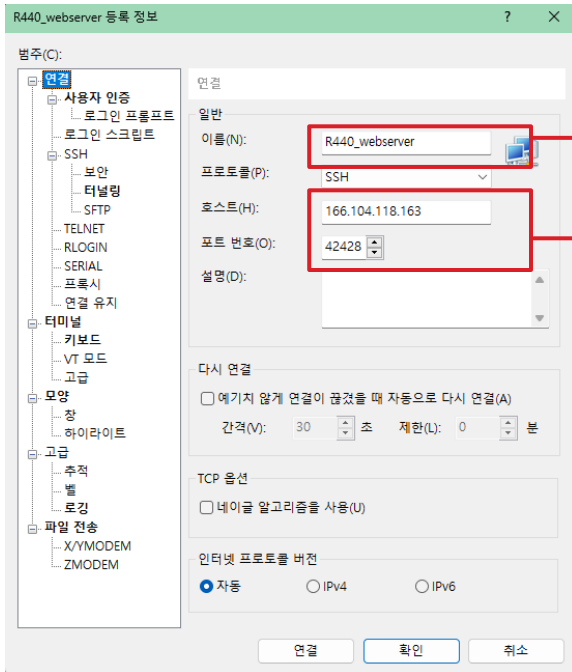
What is Xshell

- 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 Xshell, one of ssh client that is free of charge, to connect to our server in BIGLAB (CentOS)

A First Step to Linux Shell



A First Step to Linux Shell



이름은 임의로 지정 가능

호스트와 포트 번호는 사진과 동일하게 할 것!

A First Step to Linux Shell

SSH 보안 경고

? X



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:f3:82:b4:25:ce:e3:6d:4e

이 호스트 키를 수락하시겠습니까?

한 번 수락(O)

수락 및 저장(S)

취소

SSH 사용자 이름

X

원격 호스트: 166.104.138.232:22 (새 세션)

서버 종류: SSH2, OpenSSH_5.3



로그인 할 사용자 이름을 입력하십시오(E):

biguser

사용자 이름 기억(R)

확인

취소

User: biguser
Password: 2023bio

SSH 사용자 인증

? X

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):

▼

찾아보기(B)...▼

암호(H):

Keyboard Interactive(I)

사용자 인증에 키보드 입력을 사용합니다.

암호 저장(R)

확인

취소

For mac users



iTerm2

iTerm2 is a terminal emulator for macOS that does amazing things.



Terminal

```
$ ssh biguser@166.104.118.163 -p 42428
```

```
biguser@166.104.118.163's password:
```

```
Last login: Mon Oct 2 10:14:33 2023 from 166.104.160.100
```

```
[biguser@R440 ~]$
```

A First Step to Linux Shell

```
Connecting to 166.104.118.163:42428...  
Connection established.  
To escape to local shell, press 'Ctrl+Alt+]'.  
  
Last login: Mon Sep 25 10:21:34 2023 from 166.104.138.104  
[biguser@R440 ~]$ █
```

성공적으로 connection 되면 위와 같은 메시지가 나옴

Basic commands

Editing command line

- Delete character
 - Backspace

```
# ls -a /home/  
# ls -a /home
```

- Delete word
 - Ctrl + w

```
# ls -a /home/  
# ls -a
```

- Delete sentence
 - Ctrl + U

```
# ls -a /home/  
#
```

Basic commands Reusing command line

- Up arrow
 - ↳ Last command
- Down arrow
 - ↳ Next command

```
[biguser@R440 ~]$ cd tutor/  
[biguser@R440 tutor]$ cd tutor/█
```

```
[biguser@R440 ~]$ cd tutor/  
[biguser@R440 tutor]$ █
```

Basic Shell Commands

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

```
[biguser@R440 ~]$ ll
total 9352
drwx----- 8 root      root          200 Sep 25 08:29 materials
-rw----- 1 biguser  biguser    5449128 Sep 25 10:19 nohup.out
drwxrwxr-x 7 biguser  biguser     297 Oct 20 2021 programs
drwxrwxr-x 3 biguser  biguser      53 Dec  8 2021 R
-rw-rw-r-- 1 biguser  biguser      64 Aug 22 14:13 run_jupyter.sh
drwxrwxr-x 27 biguser biguser    4096 Sep 24 16:40 students
drwxrwxr-x 3 biguser  biguser      88 Sep  4 15:21 trashcan
drwxrwxr-x 5 biguser  biguser     80 Sep 22 18:42 tutor
[biguser@R440 ~]$ cd students/
[biguser@R440 students]$ ls
2016040819_hojungjeong 2019045641_federicolopez 2019093181_hanjiyoon 2021024975_byeonggwon 2021092951_jiyoungahn  refseq_human.txt          Yongil_Cho_attendant
2018023427_seunghyunLee 2019053036_sangwooklee 2020018040_yeonjunkim 2021042960_baeyoonbin 2023161186_SeongjunLee  session3_sequence.ipynb
2018023545_wonguJi 2019062579_naraepark 2020046199_JanghyunKIM 2021057983_hyunminchoi 9222520232_leeyoojin  session4_dna.ipynb
2018023563_jaehyeokchoi 2019065141_choitaihwan 2020064693_seoyeonglee 2021076244_hyungho  codon_table.ipynb      Untitled1.ipynb
2019025532_dongkyulee 2019088668_taeawkim 2021019489_joseonmin 2021089043_sangeuncho  exercise1.ipynb       Untitled.ipynb
[biguser@R440 students]$ pwd
/home/biguser/students
[biguser@R440 students]$
```

Basic Shell Commands

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

```
[biguser@R440 tutor]$ ll
total 4
drwxrwxr-x 3 biguser biguser 4096 Sep 18 14:22 session2
drwxrwxr-x 3 biguser biguser  331 Sep 25 10:07 session3
[biguser@R440 tutor]$ mkdir session6
[biguser@R440 tutor]$ ls
session2 session3 session6
[biguser@R440 tutor]$ █
```

Vim editor - Vim: text editor for Linux

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

[In Vim]
i: insert

```
print("Today is Wednesday!")
```

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

```
VIM - Vi IMproved  
  
version 8.2.2072  
by Bram Moolenaar et al.  
Modified by <bugzilla@redhat.com>  
Vim is open source and freely distributable  
  
Help poor children in Uganda!  
type :help iccf<Enter>      for information  
  
type :q<Enter>              to exit  
type :help<Enter> or <F1>   for on-line help  
type :help version8<Enter> for version info
```

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 : **copy** and **paste**

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

rm : **remove**

Basic Shell Commands

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


Using BLAST+ to examine the BCR-ABL Workflow

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

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

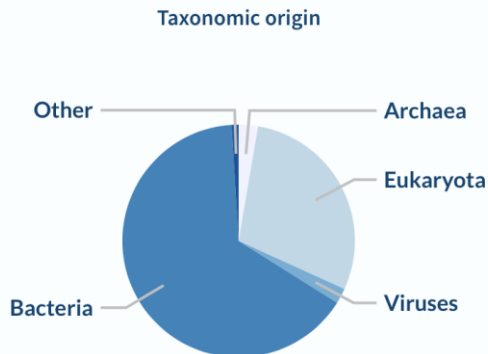
Using BLAST+ to examine the BCR-ABL (**makeblastdb**)

UniProtKB





The **UniProt Knowledgebase** (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added.

The UniProt Knowledgebase consists of two sections: a section containing manually-annotated records with information extracted from literature and curator-evaluated computational analysis (UniProtKB/Swiss-Prot), and a section with computationally analyzed records that await full manual annotation (UniProtKB/TrEMBL).



Number of Entries

 Reviewed (Swiss-Prot)
570,157 entries

 Unreviewed (TrEMBL)
251,600,768 entries

[Explore the 2023_04 release »](#)

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

```
$ ln -s /home/biguser/tutor/session6/swissprot .  
(위 명령어는 반드시 여러분의 working directory 아래 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

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

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

```
[biguser@R440 session6]$ ll  
total 600096  
-rwxr-xr-x 1 biguser biguser 280362561 Sep  7 14:48 swissprot  
-rw-rw-r-- 1 biguser biguser 104599070 Oct  8 14:40 swissprot.phr  
-rw-rw-r-- 1 biguser biguser  4522104 Oct  8 14:40 swissprot.pin  
-rw-rw-r-- 1 biguser biguser  2261048 Oct  8 14:40 swissprot.pog  
-rw-rw-r-- 1 biguser biguser 17902158 Oct  8 14:40 swissprot.psd  
-rw-rw-r-- 1 biguser biguser  421163 Oct  8 14:40 swissprot.psi  
-rw-rw-r-- 1 biguser biguser 204416076 Oct  8 14:40 swissprot.psq
```

Using BLAST+ to examine the BCR-ABL Workflow

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**)

```
$ 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**)

```
$ blastdbcmd -entry ABL1_HUMAN -db swissprot -long_seqs > 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; AltName: Full=p150  
MLEICLKLVGCKSKKGLSSSSSCYLEEALQRPVASFEPQGLSEAARWNSKENLLAGPSENDPNLFVALYDFVASGDNTL  
SITKGEKLRVLYGNHNGEWCEAQTKNQGQWVPSNYITPVNSLEKHSWYHGPVSRNAAEYLLSSGINGSFLVRESESSPGQ  
RSISLRYEGRVYHYRINTASDGKLYVSSERFNTLAE LVHHSTVADGLITTLHYPAKRNKPTVYGVSPNYDKWEMERT  
DITMKHKLGGGQYGEVYEGVWKYSLTVAVKTLKEDTMEVEEFLKEAAMKEIKHPNLVQLLVGCTREPPFYIITEFMTY  
GNLLDYLRNCRQEVNAVLLYMATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDYTAHAGAK  
FP IKWTAPESLAYNKFSIKSDVMAFGVLLWEIATYGMSPPYGGIDLSQVYELLEKDYRMERPEGCPEKVYELMRACWQWNP  
SDRPSFAEIHQAFETMFQESSISDEVEKELGKQGVRAVSTLLQAPLPTKTRTSRRAAEHRD TTDVPEMPHSGKQGESD  
PLDHEPAVSPLLPRKERGPPEGLNEDERLLPKDKKTNLFSALIKKKKTAPTTPKRSSSFREMDGQPERRGAGEEEGRD  
ISNGALAFPTLDTADPAKSPKPSNGAGVPNGALRESGSGFRSPHLWKKSSLTSSRLATGEEEGGGSSSKRFLRSCSAS  
CVPHGAKDTEWRSVTLPRDLQSTGRQFDSSTFGGHKSEKPALPRKRAGENRSDQVTRGTVTPPPRLVKKNEEADEVFKD  
IMESSPGSSPNLTPKPLRRQVTVAPASGLPHKEEAGKGSALGTPAAAEPVTPTSKAGSGAPGGTSKGAEE SRVRRHKH  
SSESPGRDKGKLSRLKPAPPPPPAASAGKAGGKPSQSPSQEAAGEAVLGAKTKATSLVDVNSDAAKPSQPGEGLKKPVL  
PATPKPQSAKPSGTPISPAPVPSTLPSASSALAGDQPSSTAFIPLISTRVSLRKRTRQPPERIASGAIKGVLDSTEALC  
LAISRNSEQMASHSAVLEAGKNLYTFCVSYVDSIQQMRNKFAFREAINKLENNLRELQICPATAGSGPAATQDFSKLLSS  
VKEISDIVQR
```


Using BLAST+ to examine the BCR-ABL Workflow

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**)

```
$ 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  
MVDPVGF AEAWKAQFPDSEPPRMLRSVGDIEQELERCKASIRRLQE VNVQERFRMIYLO  
TLLAKEKSYDRQRWGFRRAAQAPD GASEPRASASRPQAPADGADPPPAEPEEARPDGE  
GSPGKARPTARRPGAAASGERDDRGPPASVAALRSNFERIRKGGHQP GADAEKPFYVNV  
EFHHERGLVKVNDKEVSDRISLGSQAMQMERKKSQHAGAGSSVGDASRPPYGRSSSESSC  
GVDGDYEDAE LNPRFLKDNLIDANGGSRPPWPPLEYQPYQSIYVGGMMEGEGKGPLLRSQ  
STSEQEKRLTWPRRSYSPRSFEDCGGGYTPDCSSNENLTSSEEDFSSGQSSRVSPSPTTY  
RMFRDKSRSPSQNSQQSFDSSSPPTQCHKRHRHCPVVVSEATIVGVRKTGQIWPNDGEG  
AFHGADAGSFGTTPPYGCAADRAEEQRRHQDGLPYIDDSPPSSPHLSSKGRGSRDALVSG  
ALESTKASELDLEKGLMERKVVLSGILASEETYLSHLQMLTNSCVKLQTVHSIPLTINKE  
EALQRPVASDFEPQGLSE AARWNSKENLLAGPSENDPNLFVALYDFVASGDNTLSITKGE  
KLRVLYGNHNGEWCEAQTKNQGQWVPSNYITPVNSLEKHSWYHG PVSRNAEYLLSSGIN  
GSFLVRESESSPGQRSISLRYEGRVYHYRINTASDGKLYVSSSESRFNTLAE LVHHSTVA  
DGLITTLHYPAKRNKPTVYGVSPNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKYSL  
TVAVKTLKEDTMEVEEFLKEAAVMKEIKHPNLVQLGVCTREPPFYIITFMTYGNLLDY  
LRECNQEVNAVVLLYMATQISSAMEYLEKKNF IHRDLAARNCLVGENHLVKVADFLSR  
LMTGDTYTAHAGAKFP IKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLS  
QVYELLEKDYRMERPEGCEKPVYELMRACWQWNP SDRPSFAEIHQAFETMFQESSISDEV  
EKELGKQGVIRGAVSTLLQAPELPTKTRTSRRAAEHRD TTDVPEMPHSKGGQESDPLDHEP  
AVSPLLPKPKERGPEGGLNEDERLLPKDKKTNLFSALIKKKKTAPT PPKRSSSFREMDG  
QPERRAGEEEEGRDISNGALFTPLDTADPAKSPKPSNGAGVVPNGALRESGGSGFRSPHL  
WKKSSLTSSRLATGEEEGGGSSSKRFLRSCSASC VPHGAKDTEWRSVTLPRDLQSTGRQ  
FDSSTFGGKSEKPALPKRKRAGENRSDQVTRGTVTPPRLVKKNEEAADEVFKDIMESSP  
GSSPNLTPKPLRRQVTVAPASGLPHKEEAGKGSALGTPAAAEPVPTSKAGSGAPGGTS  
KGAPEESRVRRHKHSSESPGRDKGKLSRLKPAPPPPPAASAGKAGGKPSQSPQEAAGEA  
VLGAKTKATSLVD AVNSDAAKPSQPEGGLKPVLPATPKPQSAKPSGTPISPAPVSTLP  
SASSALAGDQPSSTAFIPLISTRVLRKTRQPPERIASGAIITKGVVLDSTEALCLATSRN  
SEQMASHSAVLEAGKNLYTFCVSYVDSIQMRNKFAFREAINKLENNLRELQICPATAGS  
GPAATQDFSKLLSSVKEISDIVQR
```

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

```
$ 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**)

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**)

```
$ 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

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**)

```
$ 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|ABL1_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_FSHV|Tyrosine-protein kinase transforming protein A... 932     0.0
sp|P00521|ABL_MLVA|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|P17113|STK_HYDVL|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|A1ASH8|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_XIPHET|Tyrosine-protein kinase Fyn OS=Xiphophorus hel... 376     1e-114
```

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

```
>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
```

```
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	EKLRVLGYNHNGEWCEAQTNGQGWPVSNYITPVNSLEKHSWYHGPVSRNAEYLLSSGI	659
Sbjct	86	EKLRVLGYNHNGEWCEAQTNGQGWPVSNYITPVNSLEKHSWYHGPVSRNAEYLLSSGI	145
Query	660	NGSFLVRESESSPGQRSISLRYEGRVYHYRINTASDGKLYSSSERFNTLAELVHHHSTV	719
Sbjct	146	NGSFLVRESESSPGQRSISLRYEGRVYHYRINTASDGKLYSSSERFNTLAELVHHHSTV	205
Query	720	ADGLITTLHYAPKRNKPTVYGVSPNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKKYS	779
Sbjct	206	ADGLITTLHYAPKRNKPTVYGVSPNYDKWEMERTDITMKHKLGGGQYGEVYEGVWKKYS	265
Query	780	LTVAVKTLKEDTMEVEEFLKEAAVMKEIKHPNLVQLLGVCTREPPFYIITEFMTYGNLLD	839
Sbjct	266	LTVAVKTLKEDTMEVEEFLKEAAVMKEIKHPNLVQLLGVCTREPPFYIITEFMTYGNLLD	325

Query

The sequence we inserted

sequence identity

Sbjct

A sequence in DB

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

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

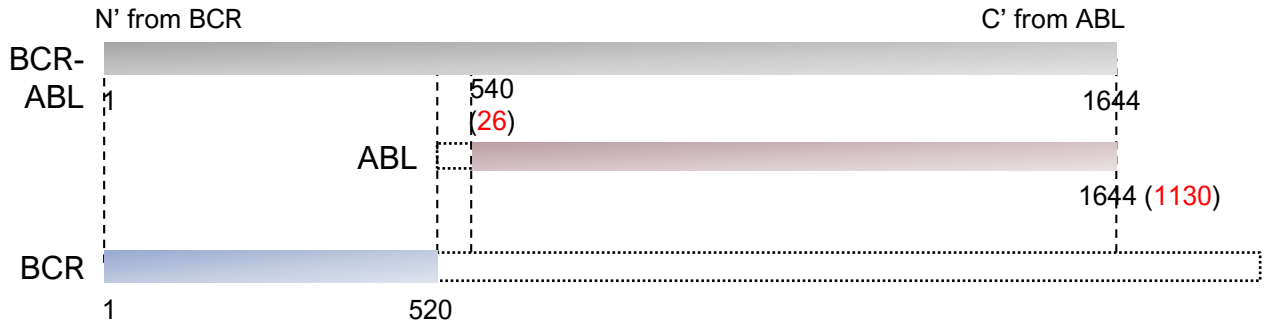
```
Score = 1084 bits (2804), Expect = 0.0, Method: Compositional matrix adjust.  
Identities = 518/520 (99%), Positives = 519/520 (99%), Gaps = 0/520 (0%)
```

```
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 EFHHERGLVKVNDKEVSDRISLGSQAMQMERKKSQHGAGSSVGDASRPPYGRSSESSC 240  
Sbjct 181 EFHHERGLVKVNDKEVSDRISLGSQAMQMERKKSQHGAGSSVGDASRPPYGRSSESSC 240  
  
Query 481 ALESTKASELDLEKGLEMRKWVLSGILASEETYLSHLQML 520  
Sbjct 481 ALESTKASELDLEKGLEMRKWVLSGILASEETYLSHL+ L 520
```

Conservative substitution

*blank refers to low score match based on BLOSUM substitution score

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



Exercise

- 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

```
$ 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  
MVDPVGFAEAWKAQFPDSEPPRMELRSVGDIEQELERCKASIRRLQE V NQERFRMIY LQTLLAKEKKS YDRQRWGFRA  
AQAPDGASEPRASASRPQAPADGADPPP AEEPEARPDGEGSPGKARPGTARRPGAAASGERDDRGPPASVAALRSNFER  
IRKGGHQPGADAEKPFYVNVEFHHERGLVKVNDKEVSDRISSLGSQAMQMERKKSQHGAGSSVGDASRPPYGRSSESSC  
GVDGDYEDAELNPRFLKDNLIDANGSRPPWPPLE YQPYQSIYVGGMMEGEGKGPLLRQSQTSEQEKRLLTWP RRSYSPRS  
FEDCGGGYTPDCSSNENLTSSEEDFSSGQSSRVSPSP TTYRMFRDKSRSPSQNSQQSF DSSSPPTPQCHKRHRHCPVVVS  
EATIVGVRKTGGIWPNDGEGAFHGADGSGFTPPG YGCAADRAEEQRRHQDGLPYIDDSPSSPHLSSKGRGRSDALVSG  
ALESTKASELDLEKGLEMRKWWLSGILASEE TYLSHLEALLPMKPLKAAATTSQPVLTSQQIETIFFKVP ELYEIHKEF  
YDGLFPRVQQWSHQQRVGD L FQKLASQLGVYRAFVDNYGVAMEMA EKC CQANAQFAEISENL RARSNKDAKDPTTKNSLE  
TLLYKPVDRVTRSTLVLHDLKHTPASHPDHPLLQ DALRISQNFLSSINEEITPRRQSMTVKKGEHRQLLKDSFMV E LVE  
GARKLRHVFLFTDLLLCTKLKKQSGGKTQYDCKWYI PLTDL SFQMVDELEAVPN I LVPDEE LDALKIKISQIKNDIQG  
EKRANKGSKATERLKKLSEQESLLL MSPSMAFRVHSRNGKSYTFLISSDYERAEWRENIREQQKCFRSFSLTSVELQ  
MLTNSCVKLQTVHSIPLTINKEDDESPLYGFLNVIHVSATGFKQSSNLYCTLEVDSFGYFVNKAKTRVRYRDTAEPN WNE  
EFEIELEGSQTLRILCYEKCYNKTKIPKEDGESTDR LMGKGQVQLDQPALQDRDWQRTVIAMNGIEVKLSVKFNSRE FSL  
KRMP SRKQTGVGFVKIAVVTKRERSKVPIYIVRQCVEE IERRGMEEVGIYRVSGVATDIQALKA AFDVNNKDV SVMSEM D  
VNAIAGTLKLYFRELP EPLFTDFEYFPNFAEGIALSDP VAKESCMLNLLL SLPEANLLTFLFLLDHLKRVAEKEAVNKMSL  
HNLATVFGPTLLRPSEKESKLPANPSQPI TMTDSWSLEVMSQVQLLYFLQLEAIPAPDSKRQSI L FSTEV
```

Exercise

```
$ 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

>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---LPYIDSPSSPHLSSKGRGSRDALVSGALESTKASELDLEKGLEMRK 500
          EEQ+ +G +PYID+SP+ SP LS++ +G D VS A ++ KGLEMRK
Sbjct 35 EEQKGPPEGSETMPYIDESPTMSPQLSARSQGRGDG-VSPTPEGLAPGVEAGKGLEMRK 93

Query 501 WVLSGILASEETYLSHLEALLLPMKPLKAAATTSQPVLTSQQIETIFFKVPELYEIHKEF 560
          VLSG LASEE Y++ LEALLLPMKPLKA ATTSQPVLTSQQIETIF+K+ ++YEIHKEF
Sbjct 94 LVLSGFLASEEIIYINQLEALLLPMKPLKATATTSQPVLTIQQIETIFYKIQDIYEIHKEF 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
```