

cd exercise_6

```
#####
# show contents of folder
#
ls -lah
total 24K
drwxr-xr-x 2 dinkel gibson 4.0K Aug 20 10:46 .
drwxr-xr-x 7 dinkel gibson 4.0K Aug 31 11:17 ..
-rw-r--r-- 1 dinkel gibson 4.1K Jul 31 18:26 EPSINS.fasta
-rw-r--r-- 1 dinkel gibson 628 Aug 6 17:18 FYN_HUMAN.fasta
-rw-r--r-- 1 dinkel gibson 643 Aug 6 17:23 SRC_HUMAN.fasta
#####
```

```
#####
# cat (=catenate) the file SRC_HUMAN.fasta
#
cat SRC_HUMAN.fasta
>sp|P12931|SRC_HUMAN Proto-oncogene tyrosine-protein kinase Src OS=Homo sapiens GN=S
RC PE=1 SV=3
MGSNKS KPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPASADGHRGPSAAFAPAAAE
PKLFGGFNSSDVTVTSPQRAGPLAGGVTTFFVALYDYESRTE TDL SFKKGERLQIVNNTTEGD
WWLAHSLSTGQTGYI PSNYVAPSDSIQAE EWYFGKITRRESERLLLNAENPRGTF LVRES
ETTKGAYCLSVSDFDNAKGLNVKHYKIRKLD SGGFYITSRTQFN SLQQLVAYYSKHADGL
CHRLTTVCPTSKPQTQGLAKDAWEIPRESLRLEVKLGQCFGEVWMGTWNGTTRVAIKTL
KPGTMSPEAFLQEAQVMKKLRHEKLVQLYAVVSEEP IYIVTEYMSKGSLLDFLKGETGKY
LRLPQLVDMAAQIASGMAYVERMNYVHRDLRAANILVGENLVCKVADFG LARLIEDNEYT
ARQGA KFP I KWTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPYPGMVNREVL DQVER
GYRMP CPPECPE SLHDLMCQCWRKEPEERPTFEY LQAFLEDYFTSTEPQYQ PGENL
#####
```

```
#####
# cat the file SRC_HUMAN.fasta, showing line-numbers, Tabs, line-ends and non-printi
ng characters
#
cat -n -A SRC_HUMAN.fasta
 1 >sp|P12931|SRC_HUMAN Proto-oncogene tyrosine-protein kinase Src OS=Homo sapi
ens GN=SRC PE=1 SV=3$
 2 MGSNKS KPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPASADGHRGPSAAFAPAAAE$
 3 PKLFGGFNSSDVTVTSPQRAGPLAGGVTTFFVALYDYESRTE TDL SFKKGERLQIVNNTTEGD^I$
 4 WWLAHSLSTGQTGYI PSNYVAPSDSIQAE EWYFGKITRRESERLLLNAENPRGTF LVRES$
 5 ETTKGAYCLSVSDFDNAKGLNVKHYKIRKLD SGGFYITSRTQFN SLQQLVAYYSKHADGL$
 6 CHRLTTVCPTSKPQTQGLAKDAWEIPRESLRLEVKLGQCFGEVWMGTWNGTTRVAIKTL$
 7 KPGTMSPEAFLQEAQVMKKLRHEKLVQLYAVVSEEP IYIVTEYMSKGSLLDFLKGETGKY$
 8 LRLPQLVDMAAQIASGMAYVERMNYVHRDLRAANILVGENLVCKVADFG LARLIEDNEYT$
 9 ARQGA KFP I KWTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPYPGMVNREVL DQVER$
10 GYRMP CPPECPE SLHDLMCQCWRKEPEERPTFEY LQAFLEDYFTSTEPQYQ PGENL$
#####
```

```
#####
# cat the file SRC_HUMAN.fasta, showing line-numbers
#
cat -n SRC_HUMAN.fasta
 1 >sp|P12931|SRC_HUMAN Proto-oncogene tyrosine-protein kinase Src OS=Homo sapi
ens GN=SRC PE=1 SV=3
 2 MGSNKS KPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPASADGHRGPSAAFAPAAAE
 3 PKLFGGFNSSDVTVTSPQRAGPLAGGVTTFFVALYDYESRTE TDL SFKKGERLQIVNNTTEGD
 4 WWLAHSLSTGQTGYI PSNYVAPSDSIQAE EWYFGKITRRESERLLLNAENPRGTF LVRES
 5 ETTKGAYCLSVSDFDNAKGLNVKHYKIRKLD SGGFYITSRTQFN SLQQLVAYYSKHADGL
 6 CHRLTTVCPTSKPQTQGLAKDAWEIPRESLRLEVKLGQCFGEVWMGTWNGTTRVAIKTL
 7 KPGTMSPEAFLQEAQVMKKLRHEKLVQLYAVVSEEP IYIVTEYMSKGSLLDFLKGETGKY
 8 LRLPQLVDMAAQIASGMAYVERMNYVHRDLRAANILVGENLVCKVADFG LARLIEDNEYT
 9 ARQGA KFP I KWTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPYPGMVNREVL DQVER
10 GYRMP CPPECPE SLHDLMCQCWRKEPEERPTFEY LQAFLEDYFTSTEPQYQ PGENL
#####
```

```
#####
# cat two files
#
cat SRC_HUMAN.fasta FYN_HUMAN.fasta
>sp|P12931|SRC_HUMAN Proto-oncogene tyrosine-protein kinase Src OS=Homo sapiens GN=S
RC PE=1 SV=3
MGSNKS KPKDASQRRRSLEPAENVHGAGGGAFPASQTPSKPASADGHRGPSAAFAPAAAE
```

```

PKLFGGFNSSDVTVTSQORAGPLAGGVTTFFVALYDYESRTEETDLSFKKGERLQIVNNTTEGD
WLLAHSLSSTGQTGYIPSNYVAPSDSIQAEWEYFGKITRRESERLLLNAENPRGTFLLVRES
ETTKGAYCLSVSDFDNAKGLNVKHYKIRKLDSSGGFYITSRTOFNSLQQLVAYYSKHADGL
CHRLTTVCPTSKPQTOGLAKDAWEIPRESLRLEVKLGQGCFFGEVWMGTWNGTTRVAIKTL
KPGTMSPEAFLQEAQVMKCLRHEKLVQLYAVVSEEPYIVTEYMSKGSLLDFLKGETGKY
LRLPQLVDMAAQIASGMAYVERMNYVHRDLRAANILVGENLVCKVADDFGLARLIEDNEYT
ARQGAKFPIKWTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPYPGMVNREVLQDQVER
GYRMPCCPECPESLHDLMCQCWRKEPEERPTFEYLLQAFLEDYFTSTEPQYQPGENL
>sp|P06241|FYN_HUMAN Tyrosine-protein kinase Fyn OS=Homo sapiens GN=FYN PE=1 SV=3
MGCVQCKDKEATKLTEERDGLNQSSGYRYGTDPTPQHYPSTIPNYNMFHAAGGQG
LTVFGVNSSHTGTLRTRGGTGVTLFVALYDYEARTEDDLSFHKGEKFOILLNSSEGDWW
EARSLLTGETGYIPSNYVAPVDSIQAEWEYFGKLGKDAERQLLSFGNPRGTFLLIRESET
TKGAYLSLIRDWDDMKGDHVKHYKIRKLDNGGYYITTRAQFETLQQLVQHYSERAAAGLCC
RLVVPCHKGMPRLTDLVSVKTKDVWEIPRESLQLIKRLGNGQFGEVWMGTWNGNTKVAIKT
LKPGTMSPEFLEEAQIMKCLKHDKLVQLYAVVSEEPYIVTEYMNKGSLLDFLKDGEGR
ALKLPNLVDMAAQVAAGMAYIERMNYIHRDLRSANILVGNGLICKIADDFGLARLIEDNEY
TARQGAKFPIKWTAPEAALYGRFTIKSDVWSFGILLTELVTKGRVPYPGMNNREVLQVE
RGYRMPCPQDCPISLHELMIHCWKKDPEERPTFEYLLQSFLEDYFTTATEPQYQPGENL
#####

#####
# pause
#
sleep 2
#####

#####
# just view the first 10 lines of the file EPSINS.fasta
#
head EPSINS.fasta
>sp|Q96N21|AP4AT_HUMAN AP-4 complex accessory subunit tepsin OS=Homo sapiens GN=C17o
rf56 PE=1 SV=1
MAAAPPLRDRLSFLHRLPILLKGTSDDDVPCPGYLFEEIAKISHESPGSSQCLLEYLLSR
LHSSSGHGKLVKILLYLCSHGSSFFLLILKRNSAFIQEAFAAGPPDPLHGNSLYQKV
RAAAQDLGSTLFSDTVLPAPLPSQPLGTPPATGMGSQARPHSTLQGFYSGKEHGRITAVRHQ
PGQAGGGWDELDSGPSSQNSQNSDLRSVSDSGSHSGSDSHSGASREPGDLAERVEVVAL
SDCQQELSLVRTVTRGPRAFLSREEAQHFIAKACGLLNCEAVLQLLTCHLRGTSECTQLRA
LCAIASLGSDDLQPHEHILLRTRPWLQELSMGSPGPVTNKATKILRHFEASCGQLSPARG
TSAEPGPTAALPGPSDLLTDAVPLPGSQVFLQPLSSTPVSSRSPAPSSGMPSSPVPTPPP
DASPIPAPGDPSEAEARLAESRRWRPERIPGGTDSPKRGPSSCAWSRDSLFAGMELVACP
RLVGAGAAAGESCPDAPRAPQTSSQRTAAKEPPGSEPSAFAFLNA
#####

#####
# just view the last 10 lines of the file EPSINS.fasta
#
tail EPSINS.fasta
MNMLWSRMLKDNKKNWRRVYKSLLLLAYLIRNGSERVVTSAHEHIYDLRSLENYHFVDEH
GKDQGINIRQKVKELVEFAQDDRLREERKKAKKNKDKYVGVSSDSVGGFRYSERYDPEP
KSKWDEEWDKNKSAFPFSDKLGELSDKIGSTIDDTISKFRKDRDRESDPERCSDSDEEKKA
RRGRSPKGEFKDEEETVTTKHIHITQATETTTTRHKRTANPSKTIDLGAAHYTGDKASP
DQNASTHTPQSSVKTSPSSKSSGDLVDLFDGTSQSTGGSADLFGGFADFGSAAASGSP
SQVTATSGNGDFGDWSAFNQAPSGPVASSGEFFGSASQPAVELVSGSQSALGPPPAASNS
SDLFDLMGSSQATMTSSQSMNFSMMSTNTVGLGLPMSRSQNTDMVQKSVSKTLPSTWSDP
SVNISLDNLLPGMQPSKPPQPSLNTMIQQQNMQPMNVMTQSFQAVNLSPPSNMLPVRPQ
TNALIGGPMMPMSPNVMTGTMGMAPLGNTPMMNQSMGMNMNIGMSAAGMGLTGTMGGMGM
PNIAMTSGTVQPKQDAFANFANFSK
#####

#####
# view the first line of each fasta file in the current directory:
#
head -n1 *.fasta -q
>sp|Q96N21|AP4AT_HUMAN AP-4 complex accessory subunit tepsin OS=Homo sapiens GN=C17o
rf56 PE=1 SV=1
>sp|P06241|FYN_HUMAN Tyrosine-protein kinase Fyn OS=Homo sapiens GN=FYN PE=1 SV=3
>sp|P12931|SRC_HUMAN Proto-oncogene tyrosine-protein kinase Src OS=Homo sapiens GN=S
RC PE=1 SV=3
#####

```

```
#####
# pause
#
sleep 2
#####

#####
# Use two different ways to display only the headerline (first line, starting with ">")
# of the two files SRC_HUMAN.fasta FYN_HUMAN.fasta
#
head -n1 SRC_HUMAN.fasta FYN_HUMAN.fasta
==> SRC_HUMAN.fasta <==
>sp|P12931|SRC_HUMAN Proto-oncogene tyrosine-protein kinase Src OS=Homo sapiens GN=SRC PE=1 SV=3

==> FYN_HUMAN.fasta <==
>sp|P06241|FYN_HUMAN Tyrosine-protein kinase Fyn OS=Homo sapiens GN=FYN PE=1 SV=3
grep ">" SRC_HUMAN.fasta FYN_HUMAN.fasta
SRC_HUMAN.fasta:>sp|P12931|SRC_HUMAN Proto-oncogene tyrosine-protein kinase Src OS=Homo sapiens GN=SRC PE=1 SV=3
FYN_HUMAN.fasta:>sp|P06241|FYN_HUMAN Tyrosine-protein kinase Fyn OS=Homo sapiens GN=FYN PE=1 SV=3
cat SRC_HUMAN.fasta FYN_HUMAN.fasta | grep ">"
>sp|P12931|SRC_HUMAN Proto-oncogene tyrosine-protein kinase Src OS=Homo sapiens GN=SRC PE=1 SV=3
>sp|P06241|FYN_HUMAN Tyrosine-protein kinase Fyn OS=Homo sapiens GN=FYN PE=1 SV=3
#####

cd -
/home/dinkel/exercises
```