

# Trabalho 1 BD parte 2

Integrantes

Maycon Douglas Batista dos Santos 11921BSI209

&

Matheus Costa Monteiro 12111BSI281

## Prática de Laboratório – Comandos Linux

### Questão 1)

1.1 → Acione o comando wc sobre o arquivo sequence.fasta:

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ wc sequence.fasta  
13 12 681 sequence.fasta  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

1.2 → Compare o resultado acima com o retornado pelo comando: cat sequence.fasta:

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ wc sequence.fasta  
13 12 681 sequence.fasta  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ cat sequence.fasta  
>Cp1002_0126a  
MHFKTRMSLFCTATTAATSLAVASLQPAAVEQPSNTIVSTIMLPTKATVTKFTVSSTKGTARADYSSN  
SITVQPGDTISVKIHSQGGYTEFSELTFFVPSVGRHLTESITFKEGDSGPHPLKVAGWNATSQADRVTFR  
TNDGKPKAITLDTTLEYTYTVGVRATGDPSTRFQLSSSDSNTVFTSASGPKIHVKKTLPSWLSGAFFGAI  
FDSLTLNLLSPILRALNIL  
>Cp1002_1802  
MLFPSRFQGTFLKPLITAALAVFCVGFTPATAQVIPYTDPDGFYTSIPSAENTTPGTVLSQRDVPMPVLD  
VLVKMKRIAYTSTHPNGFSTPVTGAVLLPTAPWRGPGPRPVALLAPGTQGAGDSCAPSKLLTMGGEYEMF  
SAAALLNRGWTVAVDYQGLGTPGNHTYMNRAQGAALLDLGRAITTLNLPDVNNHTPIIPWGYSSQGGGA  
SAAAAEMHRAYAPDVNVVLAYAGGVANLLSVSSSLEGTALTGALGYVITGMYEITYPEIREPIHNFLNTR  
GQVWLDQTSRDCLPESLLTMPLPDTSLTVSGQRLTSLISDDVFQRAISEQQIGLTAPDIPVFVAQGLND  
GIIPAEQARIMVNGWLSQGADVTYWEDPSPALDKLSGHIHVLASSFLPAVEWAEQRLAALGQPTP  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

### Questão 2)

Então podemos usar os comandos head em um pipeline com o wc. Veja o resultado deste comando:

2.1

head -n 1 sequence.fasta | wc

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ head -n 1 sequence.fasta | wc  
1      1      14  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

2.2 → Agora digite apenas `head -n 1 sequence.fasta` e compare com o que foi retornado pelo pipeline.

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ head -n 1 sequence.fasta | wc  
1      1      14  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ head -n 1 sequence.fasta  
>Cp1002_0126a  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

2.3 → Qualquer formato textual gerado pelos comandos pode ser direcionada como entrada para o processamento do comando `wc`. Por exemplo, como saber quantos arquivos existem dentro do diretório `proteinas`?

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find . -maxdepth 1 -type f | wc -l  
6  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

### Questão 3)

3.1 → `grep AKAP proteinas/*`

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample

maycon@maycon-Inspiron-3584:~/Downloads/sample$ grep AKAP proteínas/*
proteínas/Rv0338c:CQSQCPAWNTGKPLSPKLVIMDLRDHWMAKAPYILGQKDASAGGEAGHQEHHPESGFG
proteínas/Rv0425c:AAGLLSGYLLARKVVDAQAPRPAPAHWEHMSVEQVRKALPSPDEQAPAKAPPSPYPARA
proteínas/Rv0540:MSCLPVSVLVAKAPEPGRVKTRLAAAIIGDKVAADIAAAALLDTLDAVAAAPVTARAVAl
proteínas/Rv0706:PAKDQRSASRRARTEASKAASKVGATAPAKKAAAKAPAKKAPASSGVKKTPAKKAPAK
proteínas/Rv1425:KADVGNQVSSMTASLATHIEDPAKRLAAIHSTLSAKEMAKAPSAHQIMGLTETTPPGLL
proteínas/Rv1426c:ATLPTEPMRSRGRNLPRLTAMARRRYVETTNVVCYGPYGRANLADIWRRDLPRDAKAPV
proteínas/Rv1484:MTGLLDGKRILVSGIITDSSIAFHIAQVQEQGAQLVLTGFDRRLRIQRITDRLPAKAPL
proteínas/Rv1530:VAALTEQTGGLADVVDVTAKAPAAFAQAIALARPAGTVVAGTRGVGSGAPGFSPDVVV
proteínas/Rv1985c:GFTAAAAAKAPSLAWNRRDGLQDMLVRKAFFRAITRPTHFVPTTEGFTAAARAGLGWGMF
proteínas/Rv2164c:QTSPMLSPFDRPAPAKNTSQAKARAKARKAKAPKLVRPTMERLAARLTSIDLRPRLAN
proteínas/Rv2215:APYVTPLVRLASENNIDLAVTGTGVGGIRKQDVLAAAEQKKRAKAPAPAAQAAAAPA
proteínas/Rv2448c:DFLAKAPDAVIAKIRDRQVAQQETERITTRLAALQ
proteínas/Rv2536:VFLVGIVGVAVGRWLVDRLAKAPVRHHGLAAEHERAADTVFSAVRADDSPTEMQVAQ
proteínas/Rv2703:ASAPQDTTSTIPKRKTRAAKSAAAKAPSARGHATKPRAPKDAQHEAADPEDALDSVE
proteínas/Rv2780:VPGAKAPKLVSNSLVAHMKPGAVLVDAIDQGGCFEGSRPTTYDHTFVAVHDTLFYCVAN
proteínas/Rv2973c:LGMVVVDEQHRFGVEQRDQLAKAPAGITPHLLVMTATPIPRVALTVYGDLETSTLREL
proteínas/Rv3099c:LDEIAATMGFVVGKLAKAPDGSRVLELTGPLSR SIRVSDGRARVDDFGGPAPTATIR
proteínas/Rv3269:MAIQVFLAKATTTVITGLAGVTAYEILKAAAKAPLRQTAVSAAALGLRGRKAEAAES
proteínas/Rv3456c:VTSEANRARRVAAAQAKAKKAAAMPTEESEAKPAEEGDVVGASEPDAKAPEEPPAEAPEN
proteínas/Rv3691:LAKAPGDLLLVAPTSRTALTLPQLRIAAASPNSQPNCITLREANRAGSVQWGPSDTYQA
proteínas/Rv3808c:YEALKNTDCQILFMDDDIRLEPDSILRVLAMHRFAKAPMLVGGQMLNLQEPShLHIMGE
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

3.2 → R: O comando acima retornou o documento e a linha de cada documento onde a palavra-chave AKAP foi encontrada.

## Questão 4)

O comando grep também possibilita fazer a contagem da quantidade de resultados retornados. Ao invés de você ficar contando na tela a quantidade de motivos retornados (destacados em vermelho) é muito mais prático contá-los com o uso do próprio grep. Acesse o manual do grep e descubra qual parâmetro o configura para fazer a contagem, ao invés de listar o motivo buscado na tela. Em seguida:

4.1 → Use este parâmetro no último comando linux para contar quantos motivos existem.

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample
maycon@maycon-Inspiron-3584:~/Downloads/sample$ grep -c AKAP proteínas/*
proteínas/Rv0001:0
proteínas/Rv0002:0
proteínas/Rv0003:0
proteínas/Rv0004:0
proteínas/Rv0005:0
proteínas/Rv0006:0
proteínas/Rv0007:0
proteínas/Rv0008c:0
proteínas/Rv0009:0
proteínas/Rv0010c:0
proteínas/Rv0011c:0
proteínas/Rv0012:0
proteínas/Rv0013:0
proteínas/Rv0014c:0
proteínas/Rv0015c:0
proteínas/Rv0016c:0
proteínas/Rv0017c:0
proteínas/Rv0018c:0
proteínas/Rv0019c:0
proteínas/Rv0020c:0
proteínas/Rv0021c:0
proteínas/Rv0022c:0
proteínas/Rv0023:0
proteínas/Rv0024:0
proteínas/Rv0025:0
proteínas/Rv0026:0
proteínas/Rv0027:0
proteínas/Rv0028:0
proteínas/Rv0029:0
proteínas/Rv0030:0
proteínas/Rv0031:0
proteínas/Rv0032:0
proteínas/Rv0033:0
proteínas/Rv0034:0
proteínas/Rv0035:0
proteínas/Rv0036c:0
proteínas/Rv0037c:0
proteínas/Rv0038:0
```

## Questão 5)

Agora façamos a busca no arquivo mt2.fasta por um motivo/padrão que é apenas o sinal de maior (>). Porquê? É simples: o sinal de maior em um arquivo de proteínas ou nucleotídeos sempre indica o início de um cabeçalho - a parte que dá nome - a uma sequência de aminoácidos (proteína) ou nucleotídeos (DNA). Dessa forma se contarmos quantos sinais de maior existem em um arquivo fasta significa que estamos contando quantas sequências existem em um arquivo multi-fasta (um fasta com várias sequências).

### 5.1

Posicione sua janela de comandos no diretório onde está o seu arquivo mt.fasta ou mt2.fasta. Se achar conveniente pode criar o arquivo mt.fasta novamente, lembra-se de como fazê-lo? Uma dica: cat isso > aquilo.

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ ls proteínas/* > mt.fasta  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ grep -c '>' mt.fasta  
0  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

5.2 → O comando `grep '>' mt.fasta` apenas lista todas as linhas que possuem o sinal de maior. Verifique executando este comando. Perceba que não é isso o que queremos. Dessa forma utilize o parâmetro que você descobriu no item anterior para contar quantas sequências existem dentro do arquivo. Compare o resultado com o retornado pelo item 2.3.

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find . -maxdepth 1 -type f | wc -l  
6  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ grep -c '>' mt.fasta  
0  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

## Questão 6)

6.1 → Encontre os arquivos modificados a mais de dois dias dentro do diretório `sample` com o comando `find . -mtime +2`

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find . -mtime +2  
./pan-genome-size.pdf  
./SOAPdenovo2.pdf  
./sequence.fasta  
./andromeda.jpg  
./proteinas  
./proteinas/Rv0214  
./proteinas/Rv0773c  
./proteinas/Rv1028c  
./proteinas/Rv3908  
./proteinas/Rv2380c  
./proteinas/Rv2313c  
./proteinas/Rv3474  
./proteinas/Rv2753c  
./proteinas/Rv1470  
./proteinas/Rv3808c  
./proteinas/Rv3196A  
./proteinas/Rv2732c  
./proteinas/Rv2733c  
./proteinas/Rv2277c  
./proteinas/Rv3920c  
./proteinas/Rv0152c  
./proteinas/Rv2979c  
./proteinas/Rv3688c  
./proteinas/Rv3827c  
./proteinas/Rv0258c  
./proteinas/Rv0755A  
./proteinas/Rv2084  
./proteinas/Rv0668  
./proteinas/Rv1693  
./proteinas/Rv3023c  
./proteinas/Rv2082  
./proteinas/Rv2688c  
./proteinas/Rv1047  
./proteinas/Rv2856  
./proteinas/Rv0793  
./proteinas/Rv3433c  
./proteinas/Rv1082
```

## 6.2

Encontre os arquivos modificados a menos de um dia dentro do diretório sample

com o comando `find . -mtime -1`

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find . -mtime -2  
.  
./mt.fasta  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

## 6.3

Encontre os arquivos filhos que possuem um tamanho maior que 500 kilobytes

com o comando `find . -size +500k`

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find . -size +500k  
./Poster.pdf  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

6.4

Encontre os arquivos filhos que possuem um tamanho maior que 500 kilobytes e

liste as propriedades básicas destes arquivos com o comando `find . -size +500k -ls`

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find . -size +500k -ls  
27546927    600 -rw-rw-r--  1 maycon  maycon    611453 mai  5  2014 ./Poster.pdf  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

6.5

Encontre todos os arquivos cujo nome termine com pdf com o comando

`find . -name '*.pdf' -ls`

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find . -name '*.pdf' -ls  
27546928    316 -rw-rw-r--  1 maycon  maycon    322808 fev 24  2014 ./pan-genome-size.pdf  
27546521    376 -rw-rw-r--  1 maycon  maycon    382965 jan  7  2014 ./SOAPdenovo2.pdf  
27546927    600 -rw-rw-r--  1 maycon  maycon    611453 mai  5  2014 ./Poster.pdf  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

6.6

Encontre arquivos ou diretórios, a partir do diretório sample, que possuem a

permissão de execução para o grupo 'outros'. Lembre-se que existem os grupos de

permissões: usuário proprietário, grupo proprietário e outros, que utilizam as letras 'u', 'g'

e 'o' para significar cada grupo, respectivamente, sendo que as permissões podem ser

r=read, w=write e x=executar. Este comando gera o relatório: `find . -perm -o=x`

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find . -perm -o=x  
./proteinas  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

6.7

Encontre os diretórios a partir da pasta sample que possuem permissão de escrita



para o grupo proprietário: find . -perm -g=w -type d

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find . -perm -g=w -type d  
.  
./proteinas  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

## 6.8

Encontre e liste os arquivos de proteínas, da pasta proteínas, apenas com arquivos cujas sequencias de proteínas forem maiores do que 1 kilobyte (1k):

find ./proteinas -size +1k -type f -ls

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find ./proteinas -size +1k -type f -ls  
27550311 4 -rw-rw-r-- 1 maycon maycon 1720 mai 5 2014 ./proteinas/Rv2380c  
27547650 4 -rw-rw-r-- 1 maycon maycon 1346 mai 5 2014 ./proteinas/Rv0668  
27550155 4 -rw-rw-r-- 1 maycon maycon 1221 mai 5 2014 ./proteinas/Rv2124c  
27548794 4 -rw-rw-r-- 1 maycon maycon 1572 mai 5 2014 ./proteinas/Rv2932  
27547638 4 -rw-rw-r-- 1 maycon maycon 1244 mai 5 2014 ./proteinas/Rv0648  
27551176 4 -rw-rw-r-- 1 maycon maycon 1771 mai 5 2014 ./proteinas/Rv3800c  
27548796 4 -rw-rw-r-- 1 maycon maycon 1866 mai 5 2014 ./proteinas/Rv2934  
27549932 4 -rw-rw-r-- 1 maycon maycon 1201 mai 5 2014 ./proteinas/Rv1640c  
27548795 4 -rw-rw-r-- 1 maycon maycon 2233 mai 5 2014 ./proteinas/Rv2933  
27550679 4 -rw-rw-r-- 1 maycon maycon 1652 mai 5 2014 ./proteinas/Rv2946c  
27548165 4 -rw-rw-r-- 1 maycon maycon 1067 mai 5 2014 ./proteinas/Rv1536  
27549501 4 -rw-rw-r-- 1 maycon maycon 1122 mai 5 2014 ./proteinas/Rv0630c  
27549050 4 -rw-rw-r-- 1 maycon maycon 1105 mai 5 2014 ./proteinas/Rv3512  
27547649 4 -rw-rw-r-- 1 maycon maycon 1200 mai 5 2014 ./proteinas/Rv0667  
27549969 4 -rw-rw-r-- 1 maycon maycon 1080 mai 5 2014 ./proteinas/Rv1753c  
27550414 4 -rw-rw-r-- 1 maycon maycon 3130 mai 5 2014 ./proteinas/Rv2524c  
27548663 4 -rw-rw-r-- 1 maycon maycon 1167 mai 5 2014 ./proteinas/Rv2566  
27549480 4 -rw-rw-r-- 1 maycon maycon 1337 mai 5 2014 ./proteinas/Rv0578c  
27550314 4 -rw-rw-r-- 1 maycon maycon 1447 mai 5 2014 ./proteinas/Rv2383c  
27548443 4 -rw-rw-r-- 1 maycon maycon 1358 mai 5 2014 ./proteinas/Rv2006  
27548963 4 -rw-rw-r-- 1 maycon maycon 1547 mai 5 2014 ./proteinas/Rv3296  
27549176 4 -rw-rw-r-- 1 maycon maycon 1125 mai 5 2014 ./proteinas/Rv3795  
27549140 4 -rw-rw-r-- 1 maycon maycon 1091 mai 5 2014 ./proteinas/Rv3728  
27549350 4 -rw-rw-r-- 1 maycon maycon 2250 mai 5 2014 ./proteinas/Rv0304c  
27549886 4 -rw-rw-r-- 1 maycon maycon 1175 mai 5 2014 ./proteinas/Rv1522c  
27547328 4 -rw-rw-r-- 1 maycon maycon 2562 mai 5 2014 ./proteinas/Rv0101  
27550770 4 -rw-rw-r-- 1 maycon maycon 1138 mai 5 2014 ./proteinas/Rv3080c  
27549368 4 -rw-rw-r-- 1 maycon maycon 3364 mai 5 2014 ./proteinas/Rv0355c  
27551191 4 -rw-rw-r-- 1 maycon maycon 1117 mai 5 2014 ./proteinas/Rv3823c  
27548797 4 -rw-rw-r-- 1 maycon maycon 1521 mai 5 2014 ./proteinas/Rv2935  
27549774 4 -rw-rw-r-- 1 maycon maycon 1261 mai 5 2014 ./proteinas/Rv1248c  
27551225 4 -rw-rw-r-- 1 maycon maycon 1429 mai 5 2014 ./proteinas/Rv3894c  
27550310 4 -rw-rw-r-- 1 maycon maycon 1495 mai 5 2014 ./proteinas/Rv2379c  
27547937 4 -rw-rw-r-- 1 maycon maycon 1261 mai 5 2014 ./proteinas/Rv1161  
27550911 4 -rw-rw-r-- 1 maycon maycon 3787 mai 5 2014 ./proteinas/Rv3350c  
27550909 4 -rw-rw-r-- 1 maycon maycon 3219 mai 5 2014 ./proteinas/Rv3347c  
27549048 4 -rw-rw-r-- 1 maycon maycon 1941 mai 5 2014 ./proteinas/Rv3508
```

## 6.9

Crie um arquivo multi-fasta denominado maiores.fasta, a partir da pasta proteínas, contendo apenas os arquivos com sequencias de proteínas que forem maiores do que 1 kilobyte (1k): find ./proteinas -size +1k -type f -exec cat > maiores.fasta { } \;



```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find ./proteinas -size +1k -type f -exec cat > maiores.fasta {} \;  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

## 6.10

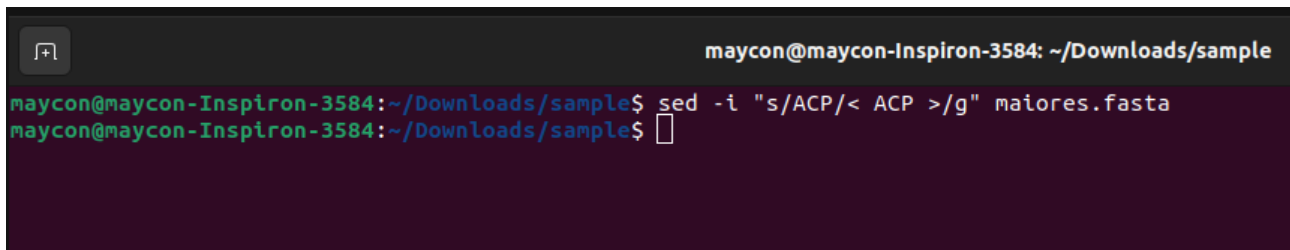
Encontre os mesmos arquivos do item anterior, mas ao invés criar um arquivo com as sequencias, execute um comando grep para procurar pelo motivo conservado AKAP:  
AKAP: find ./proteinas -size +1k -type f -ls -exec grep AKAP {} \;

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ find ./proteinas -size +1k -type f -ls -exec grep AKAP {} \;  
27550311 4 -rw-rw-r-- 1 maycon maycon 1720 mai 5 2014 ./proteinas/Rv2380c  
27547650 4 -rw-rw-r-- 1 maycon maycon 1346 mai 5 2014 ./proteinas/Rv0668  
27550155 4 -rw-rw-r-- 1 maycon maycon 1221 mai 5 2014 ./proteinas/Rv2124c  
27548794 4 -rw-rw-r-- 1 maycon maycon 1572 mai 5 2014 ./proteinas/Rv2932  
27547638 4 -rw-rw-r-- 1 maycon maycon 1244 mai 5 2014 ./proteinas/Rv0648  
27551176 4 -rw-rw-r-- 1 maycon maycon 1771 mai 5 2014 ./proteinas/Rv3800c  
27548796 4 -rw-rw-r-- 1 maycon maycon 1866 mai 5 2014 ./proteinas/Rv2934  
27549932 4 -rw-rw-r-- 1 maycon maycon 1201 mai 5 2014 ./proteinas/Rv1640c  
27548795 4 -rw-rw-r-- 1 maycon maycon 2233 mai 5 2014 ./proteinas/Rv2933  
27550679 4 -rw-rw-r-- 1 maycon maycon 1652 mai 5 2014 ./proteinas/Rv2946c  
27548165 4 -rw-rw-r-- 1 maycon maycon 1067 mai 5 2014 ./proteinas/Rv1536  
27549501 4 -rw-rw-r-- 1 maycon maycon 1122 mai 5 2014 ./proteinas/Rv0630c  
27549050 4 -rw-rw-r-- 1 maycon maycon 1105 mai 5 2014 ./proteinas/Rv3512  
27547649 4 -rw-rw-r-- 1 maycon maycon 1200 mai 5 2014 ./proteinas/Rv0667  
27549969 4 -rw-rw-r-- 1 maycon maycon 1080 mai 5 2014 ./proteinas/Rv1753c  
27550414 4 -rw-rw-r-- 1 maycon maycon 3130 mai 5 2014 ./proteinas/Rv2524c  
27548663 4 -rw-rw-r-- 1 maycon maycon 1167 mai 5 2014 ./proteinas/Rv2566  
27549480 4 -rw-rw-r-- 1 maycon maycon 1337 mai 5 2014 ./proteinas/Rv0578c  
27550314 4 -rw-rw-r-- 1 maycon maycon 1447 mai 5 2014 ./proteinas/Rv2383c  
27548443 4 -rw-rw-r-- 1 maycon maycon 1358 mai 5 2014 ./proteinas/Rv2006  
27548963 4 -rw-rw-r-- 1 maycon maycon 1547 mai 5 2014 ./proteinas/Rv3296  
27549176 4 -rw-rw-r-- 1 maycon maycon 1125 mai 5 2014 ./proteinas/Rv3795  
27549140 4 -rw-rw-r-- 1 maycon maycon 1091 mai 5 2014 ./proteinas/Rv3728  
27549350 4 -rw-rw-r-- 1 maycon maycon 2250 mai 5 2014 ./proteinas/Rv0304c  
27549886 4 -rw-rw-r-- 1 maycon maycon 1175 mai 5 2014 ./proteinas/Rv1522c  
27547328 4 -rw-rw-r-- 1 maycon maycon 2562 mai 5 2014 ./proteinas/Rv0101  
27550770 4 -rw-rw-r-- 1 maycon maycon 1138 mai 5 2014 ./proteinas/Rv3080c  
27549368 4 -rw-rw-r-- 1 maycon maycon 3364 mai 5 2014 ./proteinas/Rv0355c  
27551191 4 -rw-rw-r-- 1 maycon maycon 1117 mai 5 2014 ./proteinas/Rv3823c  
27548797 4 -rw-rw-r-- 1 maycon maycon 1521 mai 5 2014 ./proteinas/Rv2935  
27549774 4 -rw-rw-r-- 1 maycon maycon 1261 mai 5 2014 ./proteinas/Rv1248c  
27551225 4 -rw-rw-r-- 1 maycon maycon 1429 mai 5 2014 ./proteinas/Rv3894c  
27550310 4 -rw-rw-r-- 1 maycon maycon 1495 mai 5 2014 ./proteinas/Rv2379c  
27547937 4 -rw-rw-r-- 1 maycon maycon 1261 mai 5 2014 ./proteinas/Rv1161  
27550911 4 -rw-rw-r-- 1 maycon maycon 3787 mai 5 2014 ./proteinas/Rv3350c  
27550909 4 -rw-rw-r-- 1 maycon maycon 3219 mai 5 2014 ./proteinas/Rv3347c  
27549048 4 -rw-rw-r-- 1 maycon maycon 1941 mai 5 2014 ./proteinas/Rv3508
```

## Questão 7)

O comando sed nos auxilia realizando edições em lote em diversos arquivos. Vejamos um exemplo simples de uso do sed, os exemplos mais interessantes são criados quando conjugamos o sed com expressões regulares. Suponha que uma novo motivo proteico está sendo estudado, o motivo ACP. Deseja-se editar o arquivo maiores.fasta, criado nos item anterior, para delimitar todas as ocorrências de ACP com um tipo de tag, de modo a que este

motivo não fique oculto em possíveis análises visuais do arquivo. Para tanto onde houver “ACP” deve ser substituído por “< ACP >”. Execute este comando para realizar tal operação: `sed -i "s/ACP/< ACP >/g" maiores.fasta`

A terminal window with a dark background. The title bar at the top reads "maycon@maycon-Inspiron-3584: ~/Downloads/sample". The prompt is "maycon@maycon-Inspiron-3584:~/Downloads/sample\$". The command "sed -i "s/ACP/< ACP >/g" maiores.fasta" has been entered. The prompt is now "maycon@maycon-Inspiron-3584:~/Downloads/sample\$" followed by a cursor.

```
maycon@maycon-Inspiron-3584:~/Downloads/sample$ sed -i "s/ACP/< ACP >/g" maiores.fasta
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

## Questão 8)

8.1

Os comandos `grep` e `sed` se tornam muito poderosos quando expressões regulares são utilizadas nos parâmetros de busca. Um exemplo simples pode ser a busca de um motivo conservado no qual existe um aminoácido que pode variar entre  $n$  letras, digamos  $n=3$  letras. Suponha o motivo AKAP no qual a segunda letra A pode ser substituída pelos aminoácidos R e P. Este comando gera o relatório que precisamos:

`grep "AK[ARP]P" proteínas/*`

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample
proteinas/Rv1425:KADVGNQVSSMTASLATHIEDPAKRLAAIHESLTSAKEMAKAPSAHQIMGLTETTPPGLL
proteinas/Rv1426c:ATLPTEPMRSRGRNPLRTAMARRRYVETTNVVCYGPYGRANLADIWRRDLPRDAKAPV
proteinas/Rv1483:MTATATEGAKPPFVSRSVLVTGGNRIGLAIQRLAADGHKVAVTHRGSGAPKGLFGVEC
proteinas/Rv1484:MTGLLDGKRILVSGIITDSSIAFHIAQRLAADGHKVAVTHRGSGAPKGLFGVEC
proteinas/Rv1530:VAALTEQTTGGLADVVDVTAKAPAFAQAIALARPAGTVVAGTRGVGSGAPGFSPDVVV
proteinas/Rv1660:TWRSLEIGNLSSASVLHVLRTIAKPPPSGSPGLMIAMGPGFCSELVLLRWH
proteinas/Rv1731:TDWAKRPVIERAIVIRRYRDLVIENREFLMDLLQAEAGKARWAAQEEIVDLIANANYAR
proteinas/Rv1985c:GFTAAAAAKAPSLAWNRRDGLQDMLVRKAFFRAITRPTHFVPTTEGFTAAARAGLGWGMF
proteinas/Rv2164c:QTSPMLSPFDRPAPAKNTSQAKARAKARKAKAPKLVRPTPMERLAARLTSIDLRPRTLAN
proteinas/Rv2183c:EQHPLLTVIADHSLALLEVIRAIIVDDIDRSAKPPPEGPPGGGTGASGGENTNGEGSMKS
proteinas/Rv2215:APYVTPLVRLASENNIDLAVTGTGVGGRIRKQDVLAAAEQKKRAKAPAPAAQAAAAAPA
proteinas/Rv2448c:DFLAKAPDAVIAKIRDQRVAQQETERITRRLAALQ
proteinas/Rv2482c:LYAKRPLSMDAMLRVFFEAIVEIVADVLRDAPPDIGPEELTELALGLGRQFVAQGRVRSSE
proteinas/Rv2528c:VDATAKRPSGDDVQVSPEDLIDAALAEENRAAVEGEILKKALTSPTGFEDLVIRLLEAMG
proteinas/Rv2536:VFLVGIVGAVGRWLVDRLAKAPVRHHGLAAEHERAADTVFSAVRADDSPGEMQVAQ
proteinas/Rv2703:ASAPQDTTSTIPKRKTRAAKSAAAKAPSARGHATKPRAPKDAQHEATDPEDALDSVE
proteinas/Rv2780:VPGAKAPKLVSNSLVAHMKGAVLVDIAIDQGGCFEGSRPTYDHPFVAVHDTLFCVAN
proteinas/Rv2796c:MRWPTAWLLALVCVMATGCGPSGHGTRAGEEGLSPEKVAELENPLRAKPPLEDAKDQYR
proteinas/Rv2895c:PLTLDSEADLPTAKRPVRTMTVRHVDAAAREIAVDIVLHGEHGVAGPWAAGAQRGQPIY
proteinas/Rv2935:LGGDSLMAISIAMAAANEGLTITPDLYEYPTLASLTAAVDASFASSGLAKPPEAQANPA
proteinas/Rv2973c:LGMVVVDEQHRFGVEQRDQLRAKAPAGITPHLLVMTATPIRPTVALTVYGDLETSTLREL
proteinas/Rv2986c:VRKAATKAPAKKAAAKRPATKAPAKKATARRGRK
proteinas/Rv3099c:LDEIAATMGFVVGKLAKAPDGSRVLELTGPLSRIRSVSDGRARVVDDFGGPAPTATIR
proteinas/Rv3212:IQRIVPEPGVRPGSGARVLVVSQNNATVYLPARSGAQPRVDVIDETGATVSSTLLAKPPS
proteinas/Rv3240c:DGSAQVQRNGGAHKTPAGVPAGASRRERREAAARRQGRGAKPPKS VKKR
proteinas/Rv3254:GTWVLTTFGVADAKPPPTFDEMRLADKLLPARFTAALAAQAPIGCPAFHAFPASRWRRY
proteinas/Rv3269:MAIQVFLAKATTTVITGLAGVTAYEILKAAAKAPLRQTAVSAAALGLRGRKAEAAES
proteinas/Rv3424c:RHGLSRVVQCLGGCENKRHAYFRSISKTPHIRARGVPSVCAVRTVGVDGAKRPPKPIPVQ
proteinas/Rv3456c:VTSEANRRARRVAAQAKAKKAAAMPTEESEAKPAEEDVVGASEPDAKAPEEPPAEAPEN
proteinas/Rv3609c:RLVDLYAKRPQVQERLTSQIADALMKKLDPRGVIVVIEAEHLCMAMRGVRKPGSVTTTSA
proteinas/Rv3646c:AAELLADRRARGPAKRPARKAARKVPAKKAAKRD
proteinas/Rv3691:LAKAPGDLLLVAPTSRTALTPLRLIAAASPNSQPNCTLREANRAGSVQWGPSDTYQA
proteinas/Rv3731:VIAKPPHITYQPDKRVFMFKIKHLRTADCVVAGYRVHKS GSDAIGSLLLGLYQEDGQLASV
proteinas/Rv3774:RYVAAWNAAFLPSKDLTEGISATFAKRPPQFTGE
proteinas/Rv3808c:YEALKNTDCQQILFMDDDRLEPDSILRVLAMHRFAKAPMLVGGQMLNLQEPSHLHIMGE
proteinas/Rv3811:DLAGTSARVRAKRPDGDWGPWYQTEYETEPDRPAGTDGSVELGGLNPGPRSTDPVFGTT
proteinas/Rv3854c:AKRPAPVGV
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

## 8.2

Outra variação do comando anterior seria a possibilidade de que qualquer outro aminoácido pudesse entrar no lugar da segunda letra A. Neste caso o ponto '.' vem ao nosso favor: `grep "AK.P" proteinas/*`

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample
maycon@maycon-Inspiron-3584:~/Downloads/sample$ grep "AK.P" proteinas/*
proteinas/Rv0014c: EGLSADLDAVLKALAKNPENRYQTAAEMRADLVRVHNGEPPEAPKVLTAERTSLLSSA
proteinas/Rv0060: ELLASTHWVATREGAKEPATAAAVRKWTKRKRIYSDDRIGVALDRILMTA
proteinas/Rv0061: LLAVAHGQVAKTPSATRAIAFRHVRLMRVRWICAGNRGRKHRRCTTQYRSTQASKLQLH
proteinas/Rv0072: PRVSEGRSPSKPDEVAASSTMGRHLGDTVEVGARRLRVVGIVPNSTALAKIPNVFLTTEG
proteinas/Rv0126: ERDYMYAEYAKDPRMKANVGIIRRLAPLLDNRNQIELFTALLLSLPGSPVLVYGDEIGM
proteinas/Rv0147: AVRERVIREVPAGGMMVNHAFQVSTAKLPFGGVGASGMGAYHGRWGFEESHRSVLTk
proteinas/Rv0156: LVLQVILFVAVVFGTLNVIGGFIVTDRMLGMFKAKKPAVPAKPDREALR
proteinas/Rv0197: RVVLNEIAKDPGRSMIVIDPVVTDATKMADFHLRVQPGCDAWCLAALAAVLVQENLCNEA
proteinas/Rv0206c: VGGTPALELDSIHGLFAKMPLMVVILLTTTIVLMFLAFGSVVLPKATLMSALTGSTMG
proteinas/Rv0213c: PGTPLYQAYS DAGYLTAKWPLGQWEFVDPEASRVYADVAVKAPDVGISFDEAEAYFLSR
proteinas/Rv0283: SSYALKDSGKTISDTVQYYAVLPDGLQQISPVLAAILRNNNSYGLQQPPRLGADEVAKLP
proteinas/Rv0309: PNHWWSGDDNSPTFNSMQKSCQCFSTADSENLIQIPQYKHSVVMGVNKAkVPGKGSF
proteinas/Rv0338c: CQSQC PAWNTGKPLSPKLVIMDLRDHWMKAPYILGQK DASAGGEAGHQEHHPESGFG
proteinas/Rv0338c: MAAGAKRPGAKKAAPTAPAAPAPVKG LGIAAGAKRPGAKKT PPPAPGLAEPAAPQP
proteinas/Rv0353: MAKNP KDGESRTFLISVAAELAGMHAQTLRTYDRLGLVSPRRTSGGRRYS LHDVELLRQ
proteinas/Rv0363c: SGLGVKDMVTGAVALAEFTHVIAAKYFVNVALHTDHC PKDKLDSYVRPLLAISAQRVSKG
proteinas/Rv0370c: GVAETIDWVAALVALGVADLTAADSSPALASLGALAKTPDDRTQIRDAYQAFTECSHA
proteinas/Rv0372c: LVASTVRGASILDSDLSDAERARVHTPVGLAIGAKTPAEIAVSIAELIATLRGGGPRG
proteinas/Rv0373c: TGFARDYIMVGEIAANRDGKILAIRSNVLADHGAFNAQAAPAKYFAGFFGVFTGSYDIEA
proteinas/Rv0375c: CTVLDAVCLAKGPSGEREIAIDDFLVGPYETALAHNEVLI EVRIPLRHNTSSAYAKVERR
proteinas/Rv0425c: AAGLLSGYLLARKVVDAQAPRPAPAH EWHAMSVEQVRKALPSPDEQAPAKAPSPYPARA
proteinas/Rv0444c: AEQVLTAPDVRTVSRPLGAGTATVVF SRDRNTGLLVMNNVAPPSRGTVYQMWLLGGAKGP
proteinas/Rv0447c: MTVETSQTPSAAIDSDRPAVAKVPRGPLAAASAAIANRLRLRTATHLPLRLVYSDGTAT
proteinas/Rv0462: TEQQARNEG YD VVVAKFPFTANAKAHGVGDPSGFVKLVADAKHGELLGGHLVGHDVAELL
proteinas/Rv0469: LEPGMTLLDIGCGWGGGLQRAIENYD VNVIGITLSRNQFEYSKAKLAKIPTERSVQVRLQ
proteinas/Rv0486: GVDLDVFRPGDRRAARAALGLPVD ERVAVFVGRIQPLKAPDIVLRAAAKLPGVRIIVAGG
proteinas/Rv0540: MSCLPVSVLVVAKAPEGRVKTRLAAAI GDKVAADIAAAALD TLDAVAAAPVTARAVAL
proteinas/Rv0548c: TVGLQWAAEINAKSPQAQRMLKFAFNLL DDGLVGQQLFAGEATRLAYMTDEAVEGRDAFL
proteinas/Rv0573c: DTGYSYPVAVSDRIVGELARLRHADTAEAHPGSNVVGAKAKRP
proteinas/Rv0581: MDKTTVYLPDELKAAVKRAARQRGVSE AQVIRESI RAAVGGAKPPRGGLYAGSEPIARR
proteinas/Rv0586: TLAALTTMTAAAKRPSDYRK LADICSGDPTGAKKAAQDLLELANTSLMAVLVSQASRQ
proteinas/Rv0663: MPQPRTHLPIPSAARTGLITYDAKDPDSTYPPIEQLRPPAGAPNVLLILLDDVGFGASSA
proteinas/Rv0687: VALTNTLAIELGEFGIRVNSIHPYSVDTPMIEPEAMIQTFAKHPGYVHSFPPMLQPKGF
proteinas/Rv0706: PAKDQSAKSSRAR RTEASKAASKVGATAPAKKAAAKAPAKKAPASSGVKKTPAKKAPAK
proteinas/Rv0713: RVEVTGSLEGWDISAVRVGESSQTS DRPDNVIITLKRAGKPLVFDLGICLVLITLPTLAL
proteinas/Rv0753c: MTTQISHFIDGQRTAGQSTRSADV FPDNTGQIQAKVPMAGKSDIDA AVASAVEAQKGWAA
proteinas/Rv0754: IDTAVPGPLTALGQQQAQAIANALAKGPYAGIFDSQLIRTQQTAAPLANLLGMAPQVL
```

### 8.3

Mais uma variação do comando inicial. Suponha que um outro motivo possua uma ou mais letras A no local onde ocorre a segunda letra A. Neste caso o operador de repetição de ER's resolve a nossa consulta. Nessa caso o operador '+' precisa ser precedido por uma contra-barras: `grep "AKA\+P" proteinas/*`



```
maycon@maycon-Inspiron-3584: ~/Downloads/sample
maycon@maycon-Inspiron-3584:~/Downloads/sample$ grep "AKA\+P" proteinas/*
proteinas/Rv0338c:CQSQC PAWNTGKPLSPKLVIMDLRDHWMAKAPYILGQKDASAGGEAGHQEHHPVESGFG
proteinas/Rv0425c:AAGLLSGYLLARKVVDAQAPRPAPAHWEHMSVEQVRKALPSPDEQAPAKAPSPYPARA
proteinas/Rv0540:MSCLPVSVLVVAKAPEPGRVKTRLAAAIGDKVAADIAAAALLDTLDAVAAAPVTARAVAl
proteinas/Rv0706:PAKDQRS AKSSRRRTEASKAASKVGATAPAKKAAAKAPAKKAPASSGVKKTPAKKAPAK
proteinas/Rv0710:MMAEAKTGAKAAPRVAKAAKAAPKKAAPNDAAEIGAANAANVGPKHTPRTPKPRGRRKT
proteinas/Rv1425:KADVGNQVSSMTASLATHIEDPAKRLAAIHSTLSAKEMAKAPSAHQIMGLTETTPGLL
proteinas/Rv1426c:ATLPT EPMRSGRNPLRLTAMARRRYVETTNVVCYGPYGRANLADIWRRRDLPRDAKAPV
proteinas/Rv1484:MTGLLDGKRILVSGIITDSSIAFHIA RVAQEQAQLVLTGFDRRLIQRITDRLPAKAPL
proteinas/Rv1530:VAALTEQTGGLADVVDVTAKAPAAFAQAIALARPAGTVVAGTRGVGSGAPGFSPDVVV
proteinas/Rv1657:MSRAKAAPVAGPEVAANRAGRQARIVAILSSAQVRSQNELAALLAAEGIEVTQATLSRDL
proteinas/Rv1985c:GFTAAAAAKAPSLAWNRRDGLQDMLVRKAFRRAITRPTHFVPTTEGFTAAARAGLGWGMF
proteinas/Rv2164c:QTSPLSPFDRPAPAKNTSQA KARAKARKAKAPKLVRPTPMERLAARLTSIDLRPRTLAN
proteinas/Rv2215:APYVTPLVKRKLA SENNIDLAVTGTGVGGRIRKQDVLAAAEQKKRAKAPAPAAQAAAAAPA
proteinas/Rv2448c:DFLAKAPDAVIAKIRDRQ RVAQQETERITTRLAALQ
proteinas/Rv2536:VFLVGIVGVAVGRWLVDRLAKAPVRHHGLAAEHERAADTDVFSAVRADDSP TGEMQVAQ
proteinas/Rv2703:ASAPQDTTSTIPKRKTRAAAKSAAAKAPSARGHATKPRAPKDAQHEAADPEDALDSVE
proteinas/Rv2780:VPGAKAPKLVSNSLVAHMKGAVLVDIAIDQGGCFEGSRPTTYDHPTFAVHDTLFYCVAN
proteinas/Rv2973c:LGMVVVDEQHRFGVEQRDQLRAKAPAGITPHLLVMTATPIPTVALTVYGDLETSTLREL
proteinas/Rv3014c:DEGE LFALTE RDLLRTDLFRTKAGELSANGKRLLVNLDKAKAAPLWRVLVALSIRHVGPT
proteinas/Rv3099c:LDEIAATMGFVVGKLAKAPDGSRVLELTGPLSR SIRVSDGRARVVDDFGGPAPTATIR
proteinas/Rv3144c:TESARAPEPAASAPPEAVVEVPELVPAMGVLP TVDPKVAAKAAPLSTTRVQSAGSGIP
proteinas/Rv3269:MAIQVFLAKATTVITGLAGVTAYEILKAAAKAPLRQTAVSAAALGLRGRKAEAAES
proteinas/Rv3456c:VTSEANRRRVAAAQAKAKKAAAMPTEESEAKPAEEGDVVGASEPDAKAPEEPPAEAPEN
proteinas/Rv3691:LAKAPGDLLL VAPTSRTRTALTPQLRIAAASPFNSQPNCTLREANRAGSVQWGPSDTYQA
proteinas/Rv3803c:MKGRSALLRALWIAALS FGLGGVAVAAEPTAKAAPYENLMVPSPSMGRDIPVAFLAGGPH
proteinas/Rv3808c:YEALKNTDCQQILFMDDDIRLEPDSILRVLAMHRFAKAPMLVGGQMLNLQEP SHLHIMGE
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

#### 8.4

Uma variação do último comando. Suponha que no lugar do segundo A possa existir zero ou mais ocorrências de A: `grep "AKA*P" proteinas/*`

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample
maycon@maycon-Inspiron-3584:~/Downloads/sample$ grep "AKA*P" proteinas/*
proteinas/Rv0006:LIELLDIDEIQAQAILDMQLRRLAALERQRIIDDLAKIEAEIADLEDILAKPERQRGIVR
proteinas/Rv0033:TVGDVYTSVAVWFPETAKPAPLGKGT
proteinas/Rv0038:ETAVYNVLPQWAKLAAKPKTMFIGGPVKRDAALCLAVLRVGADPEGVPGLRHVAGRLVMV
proteinas/Rv0048c:TSDPGAKPDGIAPVVLTPPRLHSLGGLTGLLEQTRKRFGDTMGYRLVIYPEYASLDRV
proteinas/Rv0090:DLPVLRAAGFKVDQMTVIHALEKALAAKPKSTLALITGMLAAYAVLQAVEGVGLWLLKRWG
proteinas/Rv0095c:RITKPEAGRRAEAKP
proteinas/Rv0102:FVSVAEPVGGFFAASLAGALCLGALIHVMTAKPEPDGLIDAAAARIHLAERVSGLWLGL
proteinas/Rv0103c:EAKEPVRAYPAAASVVGTVMDGRLVIEATAVGADTQFAAMVRLVEQAQTQKARAQRLADH
proteinas/Rv0130:GTVQATVSTTVEVEGSAKPAACVAESIVRYVA
proteinas/Rv0156:LVLQVILFVAVVFGTLNVIGGFIVTDRMLGMFKAKKPAVPAPKPDREALR
proteinas/Rv0165c:MIKHVVVWTLWPERPNNKPPSPRQVPGNPGPTLKVLAHVNAPLSAKPRSQPLRRAQ
proteinas/Rv0171:MRTLEPPNRMRIGLMGIVVALLVAVGQSFTSVPMFLAKPSYGGQFTDSGGLHKGDRVRI
proteinas/Rv0189c:MPQTTDEAASVSTVADIKPRSRDVTGDLKAAARGMLRAVGMDDDFAKPQIGVASSWNE
proteinas/Rv0192:AGQDPTSFVGPPFRPPTFNPVDGAMVGVAKPVIINFAVPIADRAMAESAIHISSIPPVP
proteinas/Rv0199:ATSVSENAGAKPQTVHWNLRDVSVDGKLMISRLESIR
proteinas/Rv0207c:GFAVFAPKPKVDESDVDRLMAHIDERYREGLAALVVASADGQAFRQPLEAVARSGTPVQ
proteinas/Rv0242c:VVGGTPEAAASTNERIAQRALEGFTRSLGKELRRGATTALVYLSPDAPKPAATGLESTMR
proteinas/Rv0269c:LLMLAEELGPPQKAQSAKPLIEIARAKTRAEAMAALDIWRDRYPGAAALLRPADVLDGM
proteinas/Rv0291:GLNATEVVRRLTATAHARGRESSNIVGAGNLDAVAALTWQLPAEPGGGAAPAKPVADPPV
proteinas/Rv0338c:CQSQCPAWNTGKPLSPKLIVMDLRDHWMAKAPYILGQKASAGGEAGHQEHHPVESGFG
proteinas/Rv0338c:EAKEPQPEPAAPPKPQTDGDPAAAPVKGGLIARGARPPGKR
proteinas/Rv0357c:QRDLCAKPVYEELPGWVEDISGAREFDDLPAKARDYVLRLEQLAGAPVSCIGVGPGR
proteinas/Rv0363c:PADAKEPFDVFHGGSGSLKSEIEEALRYGVVKNVDTDTQYAFTRPIAGHMTNYDGV
proteinas/Rv0381c:VLIDVDASPTDPHAPHAKPEDIRFLYGALAHPMESAALGSCGDAMAHVRDVVSPLSAPA
proteinas/Rv0418:GGPNGVPMADTRHLIAKP
proteinas/Rv0419:DASGNPETGEVGINEETLSTLMELMGKIFSPKNPPTLSYQAGCPDAKPSPPAAYCPATN
proteinas/Rv0425c:AAGLLSGYLLARKVVDAQAPRPAPAEHWHAMSVEQVRKALPSPDEQAPAKAPSPYPARA
proteinas/Rv0479c:FVTSSVTHPADGTVELKGLNNITAKPIVAGKGLELQIINFNTLGFSLPKETVQSTLNE
proteinas/Rv0483:INLSTENAEQYRSVYGDPEVTGSSIQLSYADGDIWDWAVDWDTWVMSALPPPAKP
proteinas/Rv0533c:ELLTTGAKEPGDLALLIGYGAGLSYAAQVVRMPKG
proteinas/Rv0540:MSCLPVSVLVVAKAPEGRVKTRLAAAIIGDKVAADIAAAALDLDLDAVAAAPVTARAV
proteinas/Rv0581:MDKTTVYLPDELKAAVKRAARQGVSEAQVIREIRAAGVGAKEPPRGGLYAGSEPIARR
proteinas/Rv0606:TAKEPSLRVLRKRWNTVKDEVCVNAETGTVWPECSKEAYADGIAGAVDAYWNWQRRAGK
proteinas/Rv0610c:DGREYSGLELVRDIAPLVGGLSFDPEPWGSAFRPGALPELQNWARTVELEDAVAKPGPG
proteinas/Rv0629c:LHRLLGAKPGARFRQDRQNRLPHNVIVDETSMVSLTLMARLAEAVRPGARLILVGDA
proteinas/Rv0631c:LDALDQTTSAPVRERILVTHPLQPFDRKNVTPGALLGAKPFTFDPAALAAAQAAAGKRC
proteinas/Rv0677c:SGSGSYVNIYLDLDAPKQRIDGAALPWSLTLKTTPSAAPNILAQGDGTSITCRITVDG
```

## Questão 9)

9.1

Liste os cabeçalhos dos arquivos presentes no diretório proteinas com o

comando: `grep ">" proteinas/*`



maycon@maycon-Inspiron-3584: ~/Downloads/sample

```
maycon@maycon-Inspiron-3584:~/Downloads/sample$ grep ">" proteinas/*
proteinas/Rv0001:>Mt_Rv0001
proteinas/Rv0002:>Mt_Rv0002
proteinas/Rv0003:>Mt_Rv0003
proteinas/Rv0004:>Mt_Rv0004
proteinas/Rv0005:>Mt_Rv0005
proteinas/Rv0006:>Mt_Rv0006
proteinas/Rv0007:>Mt_Rv0007
proteinas/Rv0008c:>Mt_Rv0008c
proteinas/Rv0009:>Mt_Rv0009
proteinas/Rv0010c:>Mt_Rv0010c
proteinas/Rv0011c:>Mt_Rv0011c
proteinas/Rv0012:>Mt_Rv0012
proteinas/Rv0013:>Mt_Rv0013
proteinas/Rv0014c:>Mt_Rv0014c
proteinas/Rv0015c:>Mt_Rv0015c
proteinas/Rv0016c:>Mt_Rv0016c
proteinas/Rv0017c:>Mt_Rv0017c
proteinas/Rv0018c:>Mt_Rv0018c
proteinas/Rv0019c:>Mt_Rv0019c
proteinas/Rv0020c:>Mt_Rv0020c
proteinas/Rv0021c:>Mt_Rv0021c
proteinas/Rv0022c:>Mt_Rv0022c
proteinas/Rv0023:>Mt_Rv0023
proteinas/Rv0024:>Mt_Rv0024
proteinas/Rv0025:>Mt_Rv0025
proteinas/Rv0026:>Mt_Rv0026
proteinas/Rv0027:>Mt_Rv0027
proteinas/Rv0028:>Mt_Rv0028
proteinas/Rv0029:>Mt_Rv0029
proteinas/Rv0030:>Mt_Rv0030
proteinas/Rv0031:>Mt_Rv0031
proteinas/Rv0032:>Mt_Rv0032
proteinas/Rv0033:>Mt_Rv0033
proteinas/Rv0034:>Mt_Rv0034
proteinas/Rv0035:>Mt_Rv0035
proteinas/Rv0036c:>Mt_Rv0036c
proteinas/Rv0037c:>Mt_Rv0037c
```



```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ grep ">" proteínas/*  
proteínas/Rv0001:>Rv0001  
proteínas/Rv0002:>Rv0002  
proteínas/Rv0003:>Rv0003  
proteínas/Rv0004:>Rv0004  
proteínas/Rv0005:>Rv0005  
proteínas/Rv0006:>Rv0006  
proteínas/Rv0007:>Rv0007  
proteínas/Rv0008c:>Rv0008c  
proteínas/Rv0009:>Rv0009  
proteínas/Rv0010c:>Rv0010c  
proteínas/Rv0011c:>Rv0011c  
proteínas/Rv0012:>Rv0012  
proteínas/Rv0013:>Rv0013  
proteínas/Rv0014c:>Rv0014c  
proteínas/Rv0015c:>Rv0015c  
proteínas/Rv0016c:>Rv0016c  
proteínas/Rv0017c:>Rv0017c  
proteínas/Rv0018c:>Rv0018c  
proteínas/Rv0019c:>Rv0019c  
proteínas/Rv0020c:>Rv0020c  
proteínas/Rv0021c:>Rv0021c  
proteínas/Rv0022c:>Rv0022c  
proteínas/Rv0023:>Rv0023  
proteínas/Rv0024:>Rv0024  
proteínas/Rv0025:>Rv0025  
proteínas/Rv0026:>Rv0026  
proteínas/Rv0027:>Rv0027  
proteínas/Rv0028:>Rv0028  
proteínas/Rv0029:>Rv0029  
proteínas/Rv0030:>Rv0030  
proteínas/Rv0031:>Rv0031  
proteínas/Rv0032:>Rv0032  
proteínas/Rv0033:>Rv0033  
proteínas/Rv0034:>Rv0034  
proteínas/Rv0035:>Rv0035  
proteínas/Rv0036c:>Rv0036c  
proteínas/Rv0037c:>Rv0037c
```

## 9.2

Modifique o cabeçalho de todos os arquivos dentro da pasta proteínas para incluir o prefixo Mt assim adaptar os arquivos para utilização com o tal programa:

```
sed -i "s/>(\(Rv[0-9]\|+c\?\)\)/>Mt_\1/g" proteínas/*
```

```
maycon@maycon-Inspiron-3584: ~/Downloads/sample  
maycon@maycon-Inspiron-3584:~/Downloads/sample$ sed -i "s/>(\(Rv[0-9]\|+c\?\)\)/>Mt_\1/g" proteínas/*  
maycon@maycon-Inspiron-3584:~/Downloads/sample$
```

## 9.3

Execute o passo 9.1 novamente.