

Shell scripting - 21/05/2020

Programming for Bioinformatics - Module 3

Laurea Magistrale in Bioinformatics

A.Y. 2019-2020

Consider the following data, saved into the space-separated file `casp13.txt`.

```
1 >H0974 048503/048504, , subunit 1, 72 residues;
2 MSYDYSSLLGKITEKCGTQYNFAIAMGLSERTVSLKLNDKVTWKDDEILKAVHVLELNPQ
3 DIPKYFFNAKVH
4 >H0974 048503/048504, , subunit 2, 95 residues;
5 MQTDTSNRLKQIMAERNLKQVDILNLSIPFQKKFGIKLSKSTLSQYVNSVQSPDQNRILY
6 LAKTLGVSEAWLMGFDVPMVESSKIENDSHHHHHH
7 >H0980 Q3KP22-3 Q8NHR7, Human, subunit 1, 111 residues;
8 SLKPFTYFPETRFHAGPNVYKFKIRYGKSIRGEEIENKEVITQELEDSEVRVVLGNLDN
9 LQPFATEHFIVFPYKSKWERVSHLKFHGEIILIPYPFVFTLYVEMKWFHE
10 >H0980 Q3KP22-3 Q8NHR7, Human, subunit 2, 52 residues;
11 VNNMVTGYISIDAMKKFLGELHDFIPGTSGYLAYHVQNEINMSAIKNKLKRK
12 >H0993 MlaFA, E. coli, subunit 1, 269 residues;
13 MEQSVANLVDMDRVSFTRGNRCIFDNISLTVPRGKITAIMGPSGIGKTTLLRLIGGQIAP
14 DHGEILFDGENIPAMSRSLYTVRKRMSMLFQSGALFTDMNVFDNVAYPLREHTQLPAPL
15 LHSTVMMKLEAVGLRGAAKLMPSELSSGGMARRAALARAIALEPDLIMFDEPFVGGDPITM
16 GVLVKLISELNSALGVTCVVVSHDVPVELSLIADHAWILADKKIVAHGSAQALQANPDPRV
17 RQFLDGIADGPVPFRYPAGDYHADLLPGS
```

1. (2 points) What is the output of the following two commands?

(a) `cat casp13.txt | sort | cut -c 1-5 | sort -u | head -2 | wc -l`

(b) `cat casp13.txt | sort | cut -c 1-5 | head -2 | sort -u | wc -l`

2. (2 points) What is the output of the following command?

```
cat casp13.txt | grep ";$" | grep E | wc -l
```

3. (3 points) What is the output of the following command?

```
cat casp13.txt | awk -F ',' '($0 ~ /;$/ {if($2 ~ /E/) print $2})' | sort -u | wc -l
```

4. (3 points) What is the output of the following command?

```
cat casp13.txt | sed '/;$/d' | sed '/^D/d' | wc -l
```