

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 `caspl3.txt`.

```

1 >H0974 048503/048504, , subunit 1, 72 residues;
2 MSYDYSSLLGKITEKCGTQYNFAIAMGLSERTVSLKLNKDKVTWKDDEILKAVHVLELNPQ
3 DIPKYFFNAKVH
4 >H0974 048503/048504, , subunit 2, 95 residues;
5 MQTDTSNRLKQIMAERNLKQVDILNLSIPFQKKFGIKLSKSTLSQYVNSVQSPDQNR IYL
6 LAKTLGVSEAWLMGFDVPMVESSKIENDSHHHHHH
7 >H0980 Q3KP22-3 Q8NHR7, Human, subunit 1, 111 residues;
8 SLKPFTYFPETRFHAGPNVYKFKIRYGKSIRGEEIENKEVITQELEDVSRVVLGNLDN
9 LQPFATEHFIVFPYKSKWERVSHLKFKHGEIILIPYPFVFTLYVEMKWFHE
10 >H0980 Q3KP22-3 Q8NHR7, Human, subunit 2, 52 residues;
11 VNNMVTGYISIDAMKKFLGELHDFIPGTSGLAYHVQNEINMSAIKNKLKRK
12 >H0993 MlaFA, E. coli, subunit 1, 269 residues;
13 MEQSVANLVDMRDVSFTRGNRCIFDNISLTVPRGKITAIMGPSGIGKTTLLRLIGGQIAP
14 DHGEILFDGENIPAMSRSLYTVRKRMSMLFQSGALFTDMNVFDNVAAYPLREHTQLPAPL
15 LHSTVMMKLEAVGLRGAAKLMPSELSSGMARRAALARAI ALEPDLIMFDEPFVGGDPITM
16 GVLVKLISELNSALGVTCVVVSHDVPVELSIADHAWILADKKIVA HGSQAALQANPDPRV
17 RQFLDGIADGPPFRYPAGDYHADLLPGS

```

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

(a) `cat caspl3.txt | sort | cut -c 1-7 | sort -u | head -2 | wc -l`

(b) `cat caspl3.txt | sort | cut -c 1-7 | head -2 | sort -u | wc -l`

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

`cat caspl3.txt | grep ">" | grep 4 | wc -l`

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

`cat caspl3.txt | awk -F ',,' '$0 ~ />/ {print $1}' | sort -u | wc -l`

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

`cat caspl3.txt | sed '/>/d' | sed '/^M/d' | wc -l`