

Introduction to Bioinformatics

Course NR. 22111

Plain text files and Geany & Taxonomy databases



Stefan Olevinskiy
s246026

Polina Krasikova
s245850

Plain text files and Geany

Question 1:

Note down the FILE SIZE (in bytes) of each of the three file:

alpha_globin_OldMac.fsa – 453 bytes

alpha_globin_Unix.fsa – 453 bytes

alpha_globin_Windows.fsa – 461 bytes

Why? *Because of different line ending representation standards.*

UNIX standard: 10 - LF ("Line feed" char).

Old Mac (System 9 and before): 13 - CR ("Carriage Return" char).

DOS/Windows: 13, 10 - both CR and LF, a.k.a. CRLF

A byte is 8 bits. There are 8 newlines in each document, hence 8 for LF and CR (each represented with 1 byte), and 16 bytes for CRLF, hence the difference of 8 between OldMac/Unix and Windows

Question 2:

We can find the same information in each of the files, but formatting is different.

FILE SIZE (in bytes):

alpha_globin.doc – 28.5 KB = 28,500 bytes

alpha_globin.html – 667 bytes

alpha_globin.rtf – 855 bytes

File sizes differ a lot, but the DNA sequence is the same in all three; the difference comes from the file format, which determines how "heavy" the file is.

Question 3:

Note that one of the files cannot be opened in Geany. Can you figure out why not?

The .doc file could not be opened in Geany. This is because Geany supports only plain text file formats, not file formats with formatting, images, etc. So it tries to interpret all binary as text, which does not work for .doc files.

What kind of extra information has been added to the files than can be opened in Geany? (Is it "Human readable"?)

Geany can read only and interpret plain text – so any text file that is encoded in ASCII, UTF-8 or any other common plain text encoding. HTML and RTF are written in ASCII, so can be opened by Geany, but not interpreted, so the markup tags used in both markup languages are shown as plain text. This cannot be done with .doc files as they're written in binary format that does not denote ASCII characters, but instead needs a dedicated interpreter software, like Microsoft Word to open it.

Question 4:

```
AACGGGCACGGGACGCATGTAGCTGGAACAGTGGCAGCCGTAAATAATAATGGTATCGGA
GTTGCCGGGGTTGCAGGAGGAAACGGCTCTACCAATAGTGGAGCAAGGTTAATGTCCACA
CAAATTTTAAATAGTGATGGGGATTATACAAATAGCGAAACTCTTGTGTACAGAGCCATT
GTTTATGGTGCAGATAACGGAGCTGTGATCTCGCAAATAGCTGGGGTAGTCAGTCTCTG
ACTATTAAGGAGTTGCAGAAAGCTGCGATCGACTATTTCAATTGATTATGCAGGAATGGAC
GAAACAGGAGAAATACAGACAGGCCCTATGAGGGGAGGTATATTTATAGCTGCCGCCGGA
AACGATAACGTTTCCACTCCAAATATGCCTTCAGCTTATGAACGGGTTTTAGCTGTGGCC
TCAATGGGACCAGATTTTACTAAGGCAAGCTATAGCACTTTTGGAACATGGACTGATATT
ACTGCTCCTGGCGGAGATATTGACAAATTTGATTTGTCAGAATACGGAGTTCTCAGCACT
TATGCCGATAATTATTATGCTTATGGAGAGGGAACATCCATGGCTTGTCCACATGTCGCC
GGCGCCGCC
```

Taxonomy databases

"Top-Down" task

Question 1a:

Do you encounter any extinct animal groups along your route to the felines? (Look for a cross-like symbol at the end of a branch).

Yes, too many to count.

Question 1b:

How many species are listed within the genus *Felis*?

7 species, namely:

- Felis bieti (Chinese mountain cat)
- Felis catus (domestic cat)
- Felis chaus (jungle cat)
- Felis manul (Pallas' cat)
- Felis margarita (sand cat)
- Felis nigripes (black-footed cat)
- Felis silvestris (wildcat)

"Bottom-up" task

Question 2a:

What is the name of the first higher-level group? Does this makes sense considering the scientific naming scheme (the binominal names in Latin)?

The name of the first higher-level group is *Sus*. This does make sense in the context of binomial nomenclature, as the genus name is the first part of the species name.

Question 2b:

Continue navigating "backwards", until you encounter a the first taxonomical group that includes animals that are clearly not pigs. What is the name of this group?

The name is Artiodactyla, and it includes even-toes ungulates (mammals that have hooves – ungula is latin for hoof).

Question 2c:

Navigate all the way back to Eutheria (Placental mammals). Which (surprising?) group is the "sister group" to the one containing the pigs? (A sister group is the neighboring group in the tree - the most closely related group).

Cetacea is the sister group to Artiodactyla, and it contains whales, dolphins, and porpoises. The closest relative of both is the hippo.

Question 3a:

There will be three sub-groups within *Dinosauria*. Are they all extinct? During the walk through the tree notice what kinds of animals are included at the various levels of taxonomy - especially notice which groups are extinct and which are not.

Of the three sub-groups, only Theropoda is extant.

Tyrannosauridae is close to Aves (birds). All the other animals are extinct, and only Aves are the closest living relatives.

Question 3b:

Based on your observations: are all dinosaurs extinct?

No, birds are extant.

Question 3c:

Is the Chicken a dinosaur - in the taxonomical sense?

Yes, as it is contained within Theropoda (bipedal predatory dinosaurs, including birds).

Question 4:

What is the TaxID of "Metazoa"?

33208.

Question 5:

What is the family that contains humans?

Hominidae.

Question 6:

What is the name and what is the rank?

Euarchontoglires, superorder that is in the Eutheria (placentals) clade.

Question 7:

Which ranked group do connect Human and Fruit fly (ignore "no rank" groups)? Which rank?

Kingdom Metazoa (Animals).

QUESTION 8a:

**In this selection of species, what is the sister group (nearest neighbour) to the Zebrafish?
What is the sister group to the lungfish?**

If we are only considering this selection of species, the nearest neighbour or sister group to the Zebrafish is Atlantic cod. The nearest neighbour or sister group to the lungfish is latimeria chalumnae or Blue fish.

Now try to add yourself (i.e. Human) to the tree, using either the Latin name or the TaxID. Any surprises?

Lepidosiren Paradoxa and Homo sapiens are both grouped within the clade Dipnotetrapodomorpha and are sister groups in this selection of species.

QUESTION 8b:

What is now the sister group to the lungfish?

Gadus Morhua – Atlantic Cod

QUESTION 8c:

Which of the following is most closely related to the "Blue fish": the cod, or you?

I am.

QUESTION 8d:

Which of the following is most closely related to the cod: the shark, or you?

I am.

QUESTION 8e:

Which of the following is most closely related to the shark: the lamprey, or you?

The lamprey.

QUESTION 8f:

Does the category "fish" make any scientific sense?

No.

QUESTION 9a:

Are there any differences in the branching pattern between the gene tree and the Common tree from the previous question?

Yes.

QUESTION 9b:

Can the gene tree be made to comply with the Common tree by swapping two species? If so, which two?

Yes, *Lepidosirena paradoxa* and *Latimeria chalumnae*