Evolutionary Tree

Zidar Miha (63060317)

November 30, 2012

1 Introduction

The goal of this assignment was to download and compare a few mitochondrial sequences, based on *COX3* gene, with *Needleman-Wunsch* algorithm and to construct a dendrogram. The algorithm used a *Blosum50* table and a linear gap penalty.

2 Data

We used sequences of 14 different animals, show in the Animal table 1 Most of the homeworks will rely on some data. Describe what data was used, where did it come from, and report on some basic statistics (e.g., the number and size of the sequences).

Table 1: Table of animal species used.

Index	GeneBank id	English name	Latin name
1	NC_000845.1	pig	Sus scrofa
2	NC_004299.1	Fugu rubripes	Takifugu rubripes
3	AC_000022.2	Norway rat	Rattus norvegicus
4	NC_002083.1	Sumatran orangutan	Pongo abelii
5	NC_001643.1	chimpanzee	Pan troglodytes
6	NC_011137.1	Neandertal	Homo sapiens neanderthalensis
7	NC_012920.1	human	Homo sapiens
8	NC_001645.1	western gorilla	Gorilla gorilla
9	NC_002008.4	dog	Canis lupus familiaris
10	NC_006580.1	goldfish	Carassius auratus auratus
11	NC_012420.1	veiled chameleon	Chamaeleo calyptratus
12	NC_011391.1	Russell's viper	Daboia russellii
13	NC_012061.1	Longbeaked common dolphin	Delphinus capensis
14	NC_001640.1	horse	Equus caballus

1.	1814													
2.	1574	1826												
3.	1641	1545	1816											
4.	1592	1531	1586	1804										
5.	1613	1564	1643	1703	1820									
6.	1597	1535	1614	1702	1766	1816								
7.	1622	1546	1625	1713	1777	1798	1823							
8.	1603	1542	1611	1727	1757	1753	1764	1814						
9.	1665	1555	1672	1567	1603	1585	1608	1570	1832					
10.	1582	1743	1552	1551	1600	1574	1585	1583	1556	1824				
11.	1305	1305	1320	1264	1291	1278	1299	1307	1296	1305	1827			
12.	1401	1414	1405	1386	1409	1395	1406	1422	1381	1417	1290	1796		
13.	1698	1547	1628	1580	1622	1615	1640	1607	1682	1553	1329	1395	1826	
14.	1718	1556	1658	1585	1606	1604	1615	1608	1688	1564	1310	1410	1685	1815
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.

Table 2: Table comparison scores for all animal pairs. See Animal table 1 for index description

3 Methods

Report on the formal and computational methods that solve the problem from the homework. This can include the formal description of methods (equations), outline of the algorithm or description of its implementation in code. You can even include a short snippet that you find interesting and that is essential. If needed, you can illustrate your method in a figure or a diagram, like in Figure 1. Make sure you refer to all the figures and tables from the text in the report, as illustrated by an example in the previous sentence.

This part of the report can also include a short and illustrative code snippet. An example is provided below:

```
def fib(n):
    if n == 0:
        return 0
    elif n == 1:
        return 1
    else:
        return fib(n-1) + fib(n-2)
```

4 Results

Report on results and provide short, preferably one-paragraph long discussion. Depending on a homework, you can present quantitative results of your experiments in the table (see Table 3). Provide reference to every table from main text, just like we did in the previous sentence. Notice that Tables should not include vertical lines, and should include horizontal lines only to separate the header from the content and to indicate the end of the table.

4.1 Structure of the Report

If needed, this section could be structured. If the homework is composed of several tasks (or questions), the report on each of them could be presented in a separate section (use LaTeX's

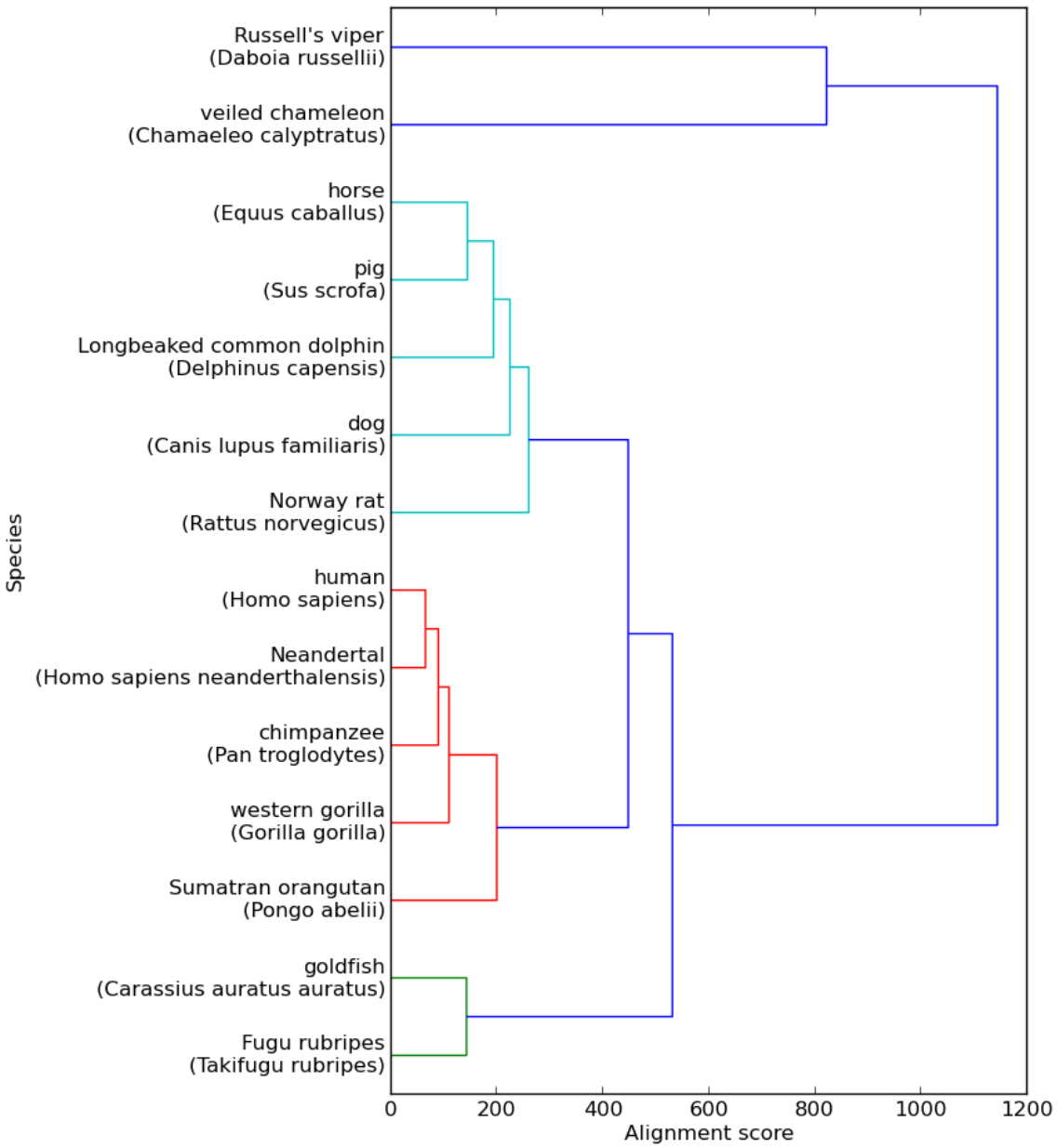

Figure 1: Every figure should include a caption with a figure description.

Table 3: Data attributes and their values.

name	domain	description
price	[0, 500]	item price in EUR
weight	[1, 1000]	item weight in dag
quality	[poor—acceptable—good]	quality of an item

`subsection` command). The report of the results should be brief. If really needed, any more extensive results of your homework should be reported in the Appendix (see Appendix A and B).

4.2 Paragraphs

A short note on the paragraphs: these are introduced in LaTeX by an empty line before the paragraph.

Honor Code

My answers to homework are my own work. I did not make solutions or code available to anyone else. I did not engage in any other activities that will dishonestly improve my results or dishonestly improve/hurt the results of others.

Appendices

A Detailed Results of Experiments

If you consider that report should include detailed results in the form of tables and figures, include them here. Essential results should be included in the main part of the report (section on results).

B Program Code

Each homework will require a development of specific implementation in Python. Most of the homeworks will request that this code is submitted as a separate document. Only if this is not the case, that is, the homework requests that you provide the entire code in the report, do so in this appendix. Following is an example, and includes a part of code from Orange (`http://www.biolab.si/orange`) that implements clustering.

```
import random
import Orange

data_names = ["iris", "housing", "vehicle"]
data_sets = [Orange.data.Table(name) for name in data_names]

print "%10s␣%3s␣%3s␣%3s" % ("", "Rnd", "Div", "HC")
for data, name in zip(data_sets, data_names):
    random.seed(42)
    km_random = Orange.clustering.kmeans.Clustering(data, centroids = 3)
    km_diversity = Orange.clustering.kmeans.Clustering(data, centroids = 3,
        initialization=Orange.clustering.kmeans.init_diversity)
    km_hc = Orange.clustering.kmeans.Clustering(data, centroids = 3,
        initialization=Orange.clustering.kmeans.init_hclustering(n=100))
    print "%10s␣%3d␣%3d␣%3d" % (name, km_random.iteration, \
    km_diversity.iteration, km_hc.iteration)
```