Homework #4: Predicting the Presence of RNA Polymerase II

Anže Pečar (63060257)

December 15, 2012

1 Introduction

In the third homework we compare COX3 mitochindrial gene between multiple species and then build an evolutionary tree.

2 Data

Data was obtained from GenBank. We downloaded mitochondrial refrence sequences (Ref-Seq) for Gray Wolf (NC_002008.4), Goldfish (NC_006580.1), Chameleon (NC_012420.1), Daboia (NC_011391.1), Dolphin (NC_012061.1), Horse (NC_001640.1), Gorilla (NC_001645.1), Human (NC_012920.1), Neanderthal (NC_011137.1), Chimpanzee (NC_001643.1), Orangutan (NC_002083.1), Rat (NC_001665.2), Boar (NC_014692.1) and Pufferfish (NC_004299.1). We used common English names instead of Latin ones, because they are easier to understand.

3 Methods

We implemented the Needleman-Wunsch algorithm which performs a global alignment on two sequences. The algorithm is implemented using dynamic programming.

We used the $BLOSUM50$ amino acid substitution matrix. This is the default scoring used by FASTA and TFASTA for comparison of amino acid sequences.

Choosing different gap penalties didn't produce a significantly different result. As an example if we chose gap penalty of -50 instead of -5, most of the dendrogram remained and only one or two species switched positions (Neanderthal got pushed up a few levels). A gap penalty of -5 seemed to provide us with the best results.

4 Results

4.1 Table of alignment scores

Table 1 lists pairwise alignment scores for all the species.

Table 1: Alignment scores

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
1. Gray Wolf	1832	1556	1298	1381	1682	1688	1577	1608	1591	1603	1574	1672	1671	1555
2. Goldfish	1556	1824	1305	1417	1553	1564	1583	1585	1580	1600	1551	1552	1581	1743
3. Veiled Chameleon	1298	1305	1827	1290	1329	1312	1311	1304	1289	1296	1273	1322	1307	1305
4. Daboia	1381	1417	1290	1796	1395	1410	1422	1406	1401	1409	1386	1405	1400	1414
5. Dolphin	1682	1553	1329	1395	1826	1685	1612	1640	1621	1622	1585	1628	1704	1547
6. Horse	1688	1564	1312	1410	1685	1815	1608	1616	1610	1606	1585	1658	1717	1556
7. Gorilla	1577	1583	1311	1422	1612	1608	1814	1764	1759	1757	1727	1618	1602	1542
8. Human	1608	1585	1304	1406	1640	1616	1764	1823	1804	1777	1713	1625	1621	1550
9. Neanderthal	1591	1580	1289	1401	1621	1610	1759	1804	1816	1772	1708	1620	1602	1541
10. Chimpanzee	1603	1600	1296	1409	1622	1606	1757	1777	1772	1820	1703	1643	1612	1564
11. Orangutan	1574	1551	1273	1386	1585	1585	1727	1713	1708	1703	1804	1593	1588	1531
12. Rat	1672	1552	1322	1405	1628	1658	1618	1625	1620	1643	1593	1816	1645	1545
13. Boar	1671	1581	1307	1400	1704	1717	1602	1621	1602	1612	1588	1645	1814	1573
14. Pufferfish	1555	1743	1305	1414	1547	1556	1542	1550	1541	1564	1531	1545	1573	1826

4.2 The dendrogram

Figure 1 shows the calculated evolutionary tree. We used average linkage to calculate distances.

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.
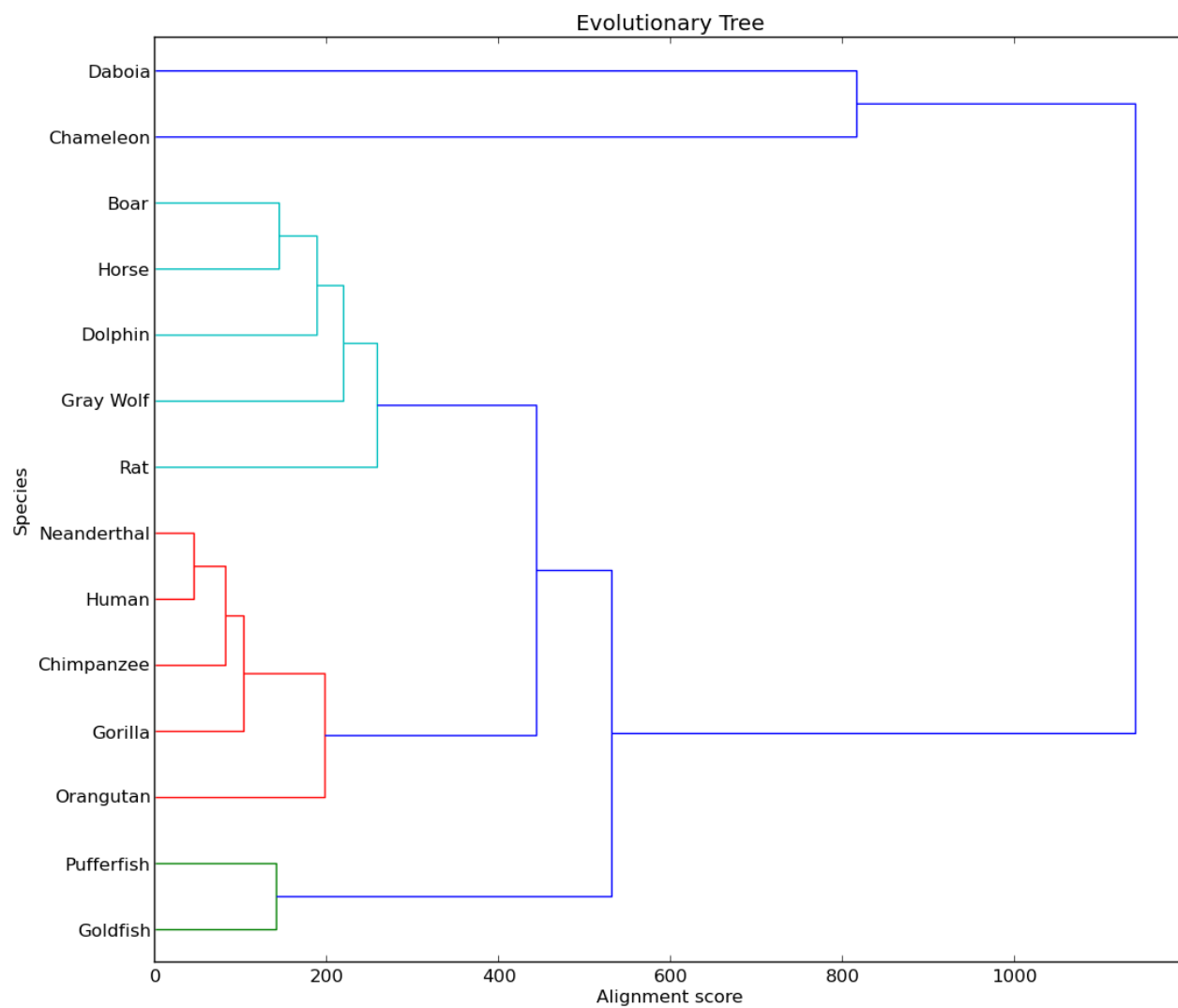
Figure 1: Dendrogram with avrage linkage