**Name:** Ali Eastman Oku

**CSCI 652 Algorithmic Bioinformatics**

**Z-ID: Z1893417**

**Assignment 2**

**Programming Language Used**: Python (3.9.6). I also used the version on the server to view the results (version 2.7.16).

**Time Spent:** I spent about 8 hours completing this assignment. I used about 6 hours to write the code and spent the remaining time writing documentation and the final report for project.

**Human and Chimpanzee**

Matches**:** 13766918

Mismatches**:** 189115

Gap Rate: 0.00126

Gap Count: 17589

|  |  |  |
| --- | --- | --- |
| **Gap Length (bases)** | **Gap Count** | **Gap Frequency** |
| 1 | 8228 | 0.4678 |
| 2 | 2600 | 0.1478 |
| 3 | 1489 | 0.0847 |
| 4 | 1278 | 0.0727 |
| 5 | 521 | 0.0296 |
| 6 | 419 | 0.0238 |
| 7 | 314 | 0.0179 |
| 8 | 292 | 0.0166 |
| 9 | 218 | 0.0124 |
| … | … | … |
| Total | 17589 | 1 |

**Application

Description automatically generated with medium confidence**

**Human and Mouse**

**Matches:** 4441074

**Mismatches:** 2203624

**Gap Rate:** 0.0389

**Gap Count:** 269242

|  |  |  |
| --- | --- | --- |
| **Gap Length (bases)** | **Gap Count** | **Gap Frequency** |
| 1 | 82442 | 0.3062 |
| 2 | 40019 | 0.1486 |
| 3 | 26836 | 0.0997 |
| 4 | 20479 | 0.0761 |
| 5 | 13972 | 0.0519 |
| 6 | 11196 | 0.0416 |
| 7 | 9055 | 0.0336 |
| 8 | 7693 | 0.0286 |
| 9 | 6522 | 0.0242 |
| … | .. | … |
| Total | 269242 | 1 |

Graphical user interface, text, application, email

Description automatically generated

**Human and Dog**

**Matches:** 12655647

**Mismatches:** 4455165

**Gap Rate:** 0.02796

**Gap Count**: 492234

|  |  |  |
| --- | --- | --- |
| **Gap Length (bases)** | **Gap Count** | **Gap Frequency** |
| 1 | 161478 | 0.3281 |
| 2 | 71950 | 0.1462 |
| 3 | 47915 | 0.0973 |
| 4 | 37365 | 0.0759 |
| 5 | 24485 | 0.0497 |
| 6 | 18824 | 0.0382 |
| 7 | 15030 | 0.0305 |
| 8 | 12845 | 0.0261 |
| 9 | 11044 | 0.0224 |
| … | … | … |
| Total | 492234 | 1 |

A picture containing graphical user interface

Description automatically generated

**Discussion:**

From the above results, the gap rate between human and chimpanzee was the lowest among all other species compared above. A lower gap rate indicates a lower occurrence of indels between the two species. This low gap rate might also indicate that humans are more closely related to chimps than to the other two species compared (dog and mouse). This is because a more recent common ancestor would mean less time to accumulate indels. The results also show that there is a higher gap rate between human and dog in comparison to human and mouse. This also indicates that humans might share a more recent common ancestor with mice in comparison to dogs. The three graphs above show a decrease in gap frequency with increasing gap lengths. This indicates that shorter gaps or shorter indels are more common that larger gaps/larger indels. A gap length of one was with the gap length with the highest gap frequency. This indicates that most indels/gaps in nature typically involve one base.