Damler, Lucas

Z1761739

CSCI 490

1.)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Human vs Chimp Species 2 | | | | | |
| Species 1 |  | A | C | G | T |
| A | 3433807 | 7472 | 32003 | 4523 |
| C | 7222 | 3443935 | 9920 | 32987 |
| G | 32981 | 10012 | 3457527 | 6961 |
| T | 4622 | 31679 | 7528 | 3431492 |

Substitution Rate: 0.013562 ~ 1.4%

Ti/Tv: 2.17468 ~ 2.2%

Gap Rate: 0.00125873 ~ 0.1%

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Human vs Mouse Species 2 | | | | | |
| Species 1 |  | A | C | G | T |
| A | 1091232 | 118519 | 304301 | 106300 |
| C | 118694 | 1125759 | 111566 | 342378 |
| G | 343010 | 112616 | 1134315 | 119675 |
| T | 105952 | 301722 | 118890 | 1089768 |

Substitution Rate: 0.331636 ~ 33%

Ti/Tv: 1.41569 ~ 1.4%

Gap Rate: 0.0389419 ~ 3.9%

|  |  |  |
| --- | --- | --- |
| Human vs Chimp  Gap Rate | | |
| Gap Length | Gap Count | Gap Frequency |
| 1 | 8228 | 46.78% |
| 2 | 2600 | 14.78% |
| 3 | 1489 | 8.47% |
| 4 | 1278 | 7.27% |
| 5 | 521 | 2.96% |
| 6 | 419 | 2.38% |
| 7 | 314 | 1.79% |
| 8 | 292 | 1.66% |
| 9 | 218 | 1.24% |
| 10 | 184 | 1.05% |
| 11 | 155 | 0.88% |
| 12 | 134 | 0.76% |
| 13 | 124 | 0.70% |
| 14 | 140 | 0.80% |
| 15 | 100 | 0.57% |
| 16 | 127 | 0.72% |
| 17 | 93 | 0.53% |
| 18 | 108 | 0.61% |
| 19 | 85 | 0.48% |
| 20 | 88 | 0.50% |
| 21 | 74 | 0.42% |
| 22 | 59 | 0.34% |
| 23 | 49 | 0.28% |
| 24 | 57 | 0.32% |
| 25 | 50 | 0.28% |
| 26 | 31 | 0.18% |
| 27 | 37 | 0.21% |
| 28 | 23 | 0.13% |
| 29 | 27 | 0.15% |
| 30 | 38 | 0.22% |
| 31 | 29 | 0.16% |
| 32 | 26 | 0.15% |
| 33 | 20 | 0.11% |
| 34 | 12 | 0.07% |
| 35 | 13 | 0.07% |
| 36 | 17 | 0.10% |
| 37 | 11 | 0.06% |
| 38 | 11 | 0.06% |
| 39 | 15 | 0.09% |
| 40 | 7 | 0.04% |
| 41 | 10 | 0.06% |
| 42 | 11 | 0.06% |
| 43 | 12 | 0.07% |
| 44 | 16 | 0.09% |
| 45 | 10 | 0.06% |
| 46 | 16 | 0.09% |
| 47 | 9 | 0.05% |
| 48 | 4 | 0.02% |
| 49 | 6 | 0.03% |
| 50 | 7 | 0.04% |
| 51 | 11 | 0.06% |
| 52 | 6 | 0.03% |
| 53 | 4 | 0.02% |
| 54 | 8 | 0.05% |
| 55 | 8 | 0.05% |
| 56 | 7 | 0.04% |
| 57 | 5 | 0.03% |
| 58 | 7 | 0.04% |
| 59 | 3 | 0.02% |
| 60 | 7 | 0.04% |
| 61 | 5 | 0.03% |
| 62 | 5 | 0.03% |
| 63 | 3 | 0.02% |
| 64 | 6 | 0.03% |
| 65 | 4 | 0.02% |
| 66 | 5 | 0.03% |
| 67 | 4 | 0.02% |
| 68 | 5 | 0.03% |
| 69 | 4 | 0.02% |
| 70 | 2 | 0.01% |
| 71 | 5 | 0.03% |
| 72 | 3 | 0.02% |
| 73 | 4 | 0.02% |
| 74 | 7 | 0.04% |
| 75 | 1 | 0.01% |
| 76 | 3 | 0.02% |
| 77 | 2 | 0.01% |
| 78 | 4 | 0.02% |
| 79 | 1 | 0.01% |
| 80 | 7 | 0.04% |
| 81 | 2 | 0.01% |
| 82 | 1 | 0.01% |
| 83 | 1 | 0.01% |
| 84 | 1 | 0.01% |
| 85 | 1 | 0.01% |
| 86 | 4 | 0.02% |
| 87 | 1 | 0.01% |
| 88 | 1 | 0.01% |
| 89 | 5 | 0.03% |
| 90 | 4 | 0.02% |
| 92 | 5 | 0.03% |
| 93 | 2 | 0.01% |
| 94 | 1 | 0.01% |
| 95 | 2 | 0.01% |
| 96 | 1 | 0.01% |
| 97 | 2 | 0.01% |
| 98 | 1 | 0.01% |
| 99 | 1 | 0.01% |
| 100 | 3 | 0.02% |

|  |  |  |
| --- | --- | --- |
| Human vs Mouse  Gap Rate | | |
| Gap Length | Gap Count | Gap Frequency |
| 1 | 82442 | 30.62% |
| 2 | 40019 | 14.86% |
| 3 | 26836 | 9.97% |
| 4 | 20479 | 7.61% |
| 5 | 13972 | 5.19% |
| 6 | 11196 | 4.16% |
| 7 | 9055 | 3.36% |
| 8 | 7693 | 2.86% |
| 9 | 6522 | 2.42% |
| 10 | 5587 | 2.08% |
| 11 | 5142 | 1.91% |
| 12 | 4423 | 1.64% |
| 13 | 3827 | 1.42% |
| 14 | 3308 | 1.23% |
| 15 | 2963 | 1.10% |
| 16 | 2622 | 0.97% |
| 17 | 2267 | 0.84% |
| 18 | 2059 | 0.76% |
| 19 | 1822 | 0.68% |
| 20 | 1672 | 0.62% |
| 21 | 1486 | 0.55% |
| 22 | 1302 | 0.48% |
| 23 | 1224 | 0.45% |
| 24 | 1055 | 0.39% |
| 25 | 946 | 0.35% |
| 26 | 851 | 0.32% |
| 27 | 820 | 0.30% |
| 28 | 693 | 0.26% |
| 29 | 658 | 0.24% |
| 30 | 562 | 0.21% |
| 31 | 504 | 0.19% |
| 32 | 422 | 0.16% |
| 33 | 352 | 0.13% |
| 34 | 317 | 0.12% |
| 35 | 312 | 0.12% |
| 36 | 318 | 0.12% |
| 37 | 262 | 0.10% |
| 38 | 240 | 0.09% |
| 39 | 192 | 0.07% |
| 40 | 205 | 0.08% |
| 41 | 187 | 0.07% |
| 42 | 179 | 0.07% |
| 43 | 160 | 0.06% |
| 44 | 138 | 0.05% |
| 45 | 149 | 0.06% |
| 46 | 129 | 0.05% |
| 47 | 125 | 0.05% |
| 48 | 108 | 0.04% |
| 49 | 93 | 0.03% |
| 50 | 88 | 0.03% |
| 51 | 69 | 0.03% |
| 52 | 95 | 0.04% |
| 53 | 67 | 0.02% |
| 54 | 63 | 0.02% |
| 55 | 68 | 0.03% |
| 56 | 62 | 0.02% |
| 57 | 56 | 0.02% |
| 58 | 49 | 0.02% |
| 59 | 32 | 0.01% |
| 60 | 34 | 0.01% |
| 61 | 38 | 0.01% |
| 62 | 31 | 0.01% |
| 63 | 37 | 0.01% |
| 64 | 34 | 0.01% |
| 65 | 24 | 0.01% |
| 66 | 45 | 0.02% |
| 67 | 21 | 0.01% |
| 68 | 25 | 0.01% |
| 69 | 24 | 0.01% |
| 70 | 30 | 0.01% |
| 71 | 18 | 0.01% |
| 72 | 27 | 0.01% |
| 73 | 17 | 0.01% |
| 74 | 20 | 0.01% |
| 75 | 17 | 0.01% |
| 76 | 12 | 0.00% |
| 77 | 18 | 0.01% |
| 78 | 13 | 0.00% |
| 79 | 12 | 0.00% |
| 80 | 19 | 0.01% |
| 81 | 7 | 0.00% |
| 82 | 5 | 0.00% |
| 83 | 11 | 0.00% |
| 84 | 9 | 0.00% |
| 85 | 12 | 0.00% |
| 86 | 6 | 0.00% |
| 87 | 5 | 0.00% |
| 88 | 7 | 0.00% |
| 89 | 8 | 0.00% |
| 90 | 6 | 0.00% |
| 91 | 5 | 0.00% |
| 92 | 11 | 0.00% |
| 93 | 13 | 0.00% |
| 94 | 4 | 0.00% |
| 95 | 2 | 0.00% |
| 96 | 6 | 0.00% |
| 97 | 3 | 0.00% |
| 98 | 4 | 0.00% |
| 99 | 3 | 0.00% |
| 100 | 5 | 0.00% |
| 101 | 4 | 0.00% |
| 102 | 3 | 0.00% |
| 103 | 5 | 0.00% |
| 104 | 6 | 0.00% |
| 105 | 1 | 0.00% |
| 106 | 1 | 0.00% |
| 107 | 2 | 0.00% |
| 110 | 1 | 0.00% |
| 111 | 2 | 0.00% |
| 113 | 2 | 0.00% |
| 114 | 1 | 0.00% |
| 115 | 1 | 0.00% |
| 116 | 1 | 0.00% |
| 118 | 2 | 0.00% |
| 119 | 1 | 0.00% |
| 120 | 1 | 0.00% |
| 121 | 1 | 0.00% |
| 123 | 1 | 0.00% |
| 124 | 1 | 0.00% |
| 130 | 3 | 0.00% |
| 135 | 2 | 0.00% |
| 136 | 1 | 0.00% |
| 137 | 1 | 0.00% |
| 138 | 1 | 0.00% |
| 140 | 1 | 0.00% |
| 141 | 1 | 0.00% |
| 168 | 1 | 0.00% |
| 178 | 1 | 0.00% |
| 227 | 1 | 0.00% |

Damler, Lucas - Z1761739

Assignment 1

Project Report

This gene comparator program was written in C++. To write the basic logic to compare two files of gene blocks took no more than 4 hours. The reason for this length was not the difficulty in programming but understanding the requirements. Writing the logic was for the most part fundamental C++ any computer science student who has been through CSCI 241 could do. The difficult part was understanding the biology and how to extract the exact required information from the maf file. Knowing the basics of the biology was the hardest part but once it was understood what was required to produce the desired output the rest of the programming went smoothly. There was a point when I thought the program was completed but little did I know the output was wrong. It was hard to design a program to do something when I had no idea what the desired output was supposed to be. Exact numbers for the comparison would have been helpful to know if my program is producing what it is intended to. The program took little time to write but closer to 7 hours of tweaking and debugging.

As for the comparison itself between the human and chimp genes there were a lot more bases than in the mouse, but the human vs chimp also had a higher Ti/Tv percentage at 2.2% vs a 1.4% I the human vs mouse. The human/chimp substitution rate was only 1.4% while the human/mouse was around 33% despite the human/chimp higher Ti/Tv. This is to be expected considering the similarities between the two species. There were more shorter gap lengths in the human/chimp comparison than the human/mouse, but the human/mouse also had more gap lengths overall. It is interesting to see the exact computed results from the comparison, but the numbers are not too surprising. Humans and chimps are more similar and should have a higher transition/transversion rate with a significantly lower gap rate. Just comparing a human and a mouse at a glance a 33% gap rate shouldn’t be surprising. In the end the results are as expected between the human/chimp and human/mouse comparisons.