Name: Ali Eastman Oku
CSCI 652 Algorithmic Bioinformatics
Z-ID: Z1893417
Assignment 3

Programming Language Used: Python (3.9.6). I also used the version on the server to view the results (version 3.7.3).

Time Spent: I spent about 10 hours completing this assignment. I used about 8 hours to write the code and spent the remaining time writing documentation and the final report for project.

Dataset 01

Human and Chimp (Criteria = 5)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 3000, O = 300	52948	52944	52941	0.9999	0.9999
K = 3000, O = 400	52948	52944	52941	0.9999	0.9999
K = 3000, O = 500	52948	52944	52941	0.9999	0.9999
K = 5000, O = 300	52948	52944	52941	0.9999	0.9999
K = 5000, O = 400	52948	52944	52941	0.9999	0.9999
K = 5000, O = 500	52948	52944	52941	0.9999	0.9999
K = 2000, O = 300	52948	52944	52941	0.9999	0.9999
K = 2000, O = 400	52948	52944	52941	0.9999	0.9999
K = 2000, O = 500	52948	52944	52941	0.9999	0.9999
Average ± SD				0.9999 ± 0.0000	0.9999 ± 0.0000

Human and Chimp (Criteria = 0)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 3000, O = 300	52948	52944	52888	0.9989	0.9989
K = 3000, O = 400	52948	52944	52888	0.9989	0.9989
K = 3000, O = 500	52948	52944	52888	0.9989	0.9989
K = 5000, O = 300	52948	52944	52888	0.9989	0.9989
K = 5000, O = 400	52948	52944	52888	0.9989	0.9989
K = 5000, O = 500	52948	52944	52888	0.9989	0.9989
K = 2000, O = 300	52948	52944	52888	0.9989	0.9989
K = 2000, O = 400	52948	52944	52888	0.9989	0.9989
K = 2000, O = 500	52948	52944	52888	0.9989	0.9989
Average ± SD				0.9989 ± 0.0000	0.9989 ± 0.0000

Human and Dog (Criteria = 5)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 3000, O = 300	52948	50885	37800	0.7139	0.7429
K = 3000, O = 400	52948	44253	38703	0.7310	0.8746
K = 3000, O = 500	52948	42563	39317	0.7426	0.9237
K = 5000, O = 300	52948	50560	37800	0.7139	0.7476
K = 5000, O = 400	52948	44156	38606	0.7291	0.8743
K = 5000, O = 500	52948	40015	36871	0.6963	0.9214
K = 2000, O = 300	52948	52168	38365	0.7246	0.7354
K = 2000, O = 400	52548	44942	39371	0.7436	0.8760
K = 2000, O = 500	52948	43038	39785	0.7514	0.9244
Average ± SD				0.7274 ± 0.0174	0.8467 ± 0.0813

Human and Dog (Criteria = 0)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 3000, O = 300	52948	50885	33736	0.6372	0.6630
K = 3000, O = 400	52948	44253	34288	0.6476	0.7748
K = 3000, O = 500	52948	42563	34967	0.6604	0.8215
K = 5000, O = 300	52948	50560	33736	0.6372	0.6672
K = 5000, O = 400	52948	44156	34197	0.6459	0.7744
K = 5000, O = 500	52948	40015	32875	0.6209	0.8216
K = 2000, O = 300	52948	52168	34077	0.6436	0.6532
K = 2000, O = 400	52548	44942	34744	0.6562	0.7731
K = 2000, O = 500	52948	43038	35411	0.6688	0.8228
Average ± SD				0.6464 ± 0.0142	0.7524 ± 0.0716

Human and Mouse (Criteria = 5)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	52948	36484	22463	0.4242	0.6157
K = 2000, O = 400	52848	28563	22920	0.4329	0.8024
K = 2000, O = 500	52948	23704	20582	0.3887	0.8683
K = 3000, O = 300	52948	27327	17150	0.3239	0.6276
K = 3000, O = 400	52948	19277	15336	0.2896	0.7956
K = 3000, O = 500	52948	13787	1300	0.2323	0.8921
K = 5000, O = 300	52948	5554	2071	0.0391	0.3729
K = 5000, O = 400	52948	2268	1995	0.0377	0.8796
K = 5000, O = 500	52948	2268	2016	0.0381	0.8889
Average ± SD				0.2452 ± 0.1676	0.7492 ± 0.1769

Human and Mouse (Criteria = 0)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	52984	36484	17603	0.3324	0.4825
K = 2000, O = 400	52984	28563	18591	0.3511	0.6509
K = 2000, O = 500	52948	23704	16553	0.3126	0.6983
K = 3000, O = 300	52948	27327	13572	0.2563	0.4967
K = 3000, O = 400	52948	19277	12505	0.2362	0.6487
K = 3000, O = 500	52948	13787	9952	0.1880	0.7218
K = 5000, O = 300	52948	5554	1706	0.0322	0.3072
K = 5000, O = 400	52948	2268	1617	0.0305	0.7130
K = 5000, O = 500	52948	2268	1641	0.0310	0.7235
Average ± SD				0.1967 ± 0.1337	0.6047 ± 0.1447

Dataset 02

Human and Chimp (Criteria = 5)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	58012	58006	54792	0.9445	0.9445
K = 2000, O = 400	58012	58006	58004	0.9999	1.0000
K = 2000, O = 500	58012	58006	58004	0.9999	1. 0000
K = 3000, O = 300	58012	58006	54792	0.9445	0.9446
K = 3000, O = 400	58012	58006	58004	0.999	1.000
K = 3000, O = 500	58012	58006	0.58004	0.9999	1.000
K = 5000, O = 300	58012	58006	54792	0.9445	0.9446
K = 5000, O = 400	58012	58006	58004	0.9999	1.000
K = 5000, O = 500	58012	58006	58004	0.9999	1.000
Average ± SD				0.9813 ± 0.0276	0.9792 ± 0.0287

Human and Chimp (Criteria = 0)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	58012	58006	54776	0.9442	0.9443
K = 2000, O = 400	58012	58006	57986	0.9999	0.9997
K = 2000, O = 500	58012	58006	57986	0.9996	0.9997
K = 3000, O = 300	58012	58006	54776	0.9442	0.9443
K = 3000, O = 400	58012	58006	57986	0.9996	0.9997
K = 3000, O = 500	58012	58006	57986	0.9996	0.9997
K = 5000, O = 300	58012	58006	54776	0.9442	0.9443
K = 5000, O = 400	58012	58006	57986	0.9996	0.9997
K = 5000, O = 500	58012	58006	57986	0.9996	0.9997

Average ± SD				0.9812 ± 0.0277	0.9812 ± 0.0277

Human and Dog (Criteria = 5)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	58012	49090	35950	0.6197	0.7323
K = 2000, O = 400	58012	42761	38264	0.6596	0.8948
K = 2000, O = 500	58012	41640	38055	0.6560	0.9139
K = 3000, O = 300	58012	49090	35951	0.6197	0.7323
K = 3000, O = 400	58012	42761	38265	0.6596	0.8949
K = 3000, O = 500	58012	41525	37941	0.6540	0.9137
K = 5000, O = 300	58012	43091	30447	0.5248	0.7066
K = 5000, O = 400	58012	34730	31055	0.5353	0.8942
K = 5000, O = 500	58012	31373	28448	0.4904	0.9068
Average ± SD				0.6021 ± 0.0668	0.8433 ± 0.0903

Human and Dog (Criteria = 0)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	58012	49090	31418	0.5416	0.6400
K = 2000, O = 400	58012	42761	33815	0.5829	0.7908
K = 2000, O = 500	58012	41640	33543	0.5782	0.8055
K = 3000, O = 300	58012	49090	31372	0.5408	0.6391
K = 3000, O = 400	58012	42761	33777	0.5822	0.7899
K = 3000, O = 500	58012	41525	33387	0.5755	0.8040
K = 5000, O = 300	58012	43091	26389	0.4549	0.6124
K = 5000, O = 400	58012	34730	27263	0.4700	0.7850
K = 5000, O = 500	58012	31373	24889	0.4290	0.7933
Average ± SD				0.5283 ± 0.0608	0.7400 ± 0.0828

Human and Mouse (Criteria = 5)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	58012	35607	20739	0.3575	0.5824
K = 2000, O = 400	58012	29870	22793	0.3929	0.7631
K = 2000, O = 500	58012	26594	22053	0.3801	0.8292

K = 3000, O = 300	58012	30966	19337	0.333	0.6245
K = 3000, O = 400	58012	22247	17143	0.2955	0.7706
K = 3000, O = 500	58012	20395	16356	0.2819	0.8020
K = 5000, O = 300	58012	19611	12111	0.2088	0.6507
K = 5000, O = 400	58012	12840	10676	0.18403	0.8315
K = 5000, O = 500	58012	8568	7684	0.1325	0.8968
Average ± SD				0.2853 ± 0.0919	0.7501 ± 0.8968

Human and Mouse (Criteria = 0)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	58012	35607	16492	0.2843	0.4631
K = 2000, O = 400	58012	29870	17915	0.3088	0.5997
K = 2000, O = 500	58012	26594	17619	0.3037	0.6625
K = 3000, O = 300	58012	30966	15305	0.2638	0.4942
K = 3000, O = 400	58012	22247	13382	0.2306	0.6015
K = 3000, O = 500	58012	20395	13157	0.2268	0.6451
K = 5000, O = 300	58012	18611	9437	0.1627	0.5071
K = 5000, O = 400	58012	12840	8167	0.1408	0.6361
K = 5000, O = 500	58012	8568	6186	0.1066	0.7220
Average ± SD				0.2253 ± 0.0735	0.5924 ± 0.0867

Dataset 3

Human and Chimp (Criteria = 5)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	57116	57112	56121	0.9826	0.9826
K = 2000, O = 400	57116	57112	57106	0.9998	0.9999
K = 2000, O = 500	57116	57112	57106	0.9998	0.9999
K = 3000, O = 300	57116	57112	56121	0.9826	0.9826
K = 3000, O = 400	57116	57112	57106	0.9999	0.9999
K = 3000, O = 500	57116	57112	57106	0.9998	0.9999
K = 5000, O = 300	57116	57112	56121	0.9826	0.9826
K = 5000, O = 400	57116	57112	57106	0.9998	0.9999
K = 5000, O = 500	57116	57112	57106	0.9998	0.9999
Average ± SD				0.9941 ± 0.0086	0.9941 ± 0.0087

Human and Chimp (Criteria = 0)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	57116	57112	56070	0.9817	0.9818
K = 2000, O = 400	57116	57112	56954	0.9972	0.9972
K = 2000, O = 500	57116	57112	56954	0.9972	0.9972
K = 3000, O = 300	57116	57112	56070	0.9817	0.9818
K = 3000, O = 400	57116	57112	56954	0.9972	0.9972
K = 3000, O = 500	57116	57112	56954	0.9972	0.9972
K = 5000, O = 300	57116	57112	56070	0.9817	0.9818
K = 5000, O = 400	57116	57112	56954	0.9972	0.9972
K = 5000, O = 500	57116	57112	56954	0.9972	0.9972
Average ± SD				0.9920 ± 0.0078	0.9921 ± 0.0077

Human and Dog (Criteria = 5)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	57116	51333	40399	0.7073	0.7870
K = 2000, O = 400	57116	48222	42063	0.7364	0.8723
K = 2000, O = 500	57116	47631	42101	0.7371	0.8839
K = 3000, O = 300	57116	49711	39611	0.6935	0.7968
K = 3000, O = 400	57116	47318	41362	0.7242	0.8741
K = 3000, O = 500	57116	46727	41409	0.7250	0.8862
K = 5000, O = 300	57116	47128	37118	0.6499	0.7876
K = 5000, O = 400	57116	40551	35189	0.6161	0.8677
K = 5000, O = 500	57116	39960	35223	0.6167	0.8815
Average ± SD				0.6896 ± 0.0494	0.8486 ± 0.0440

Human and Dog (Criteria = 0)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	57116	51333	35243	0.6170	0.6866
K = 2000, O = 400	57116	48222	36765	0.6437	0.7624
K = 2000, O = 500	57116	47631	36807	0.6444	0.7728
K = 3000, O = 300	57116	49711	34470	0.6035	0.6934
K = 3000, O = 400	57116	47318	36116	0.6323	0.7633
K = 3000, O = 500	57116	46727	36346	0.6364	0.7778
K = 5000, O = 300	57116	47128	32209	0.5639	0.6834
K = 5000, O = 400	57116	40551	30667	0.5369	0.7563
K = 5000, O = 500	57116	39960	30749	0.5384	0.7694
Average ± SD				0.6018 ± 0.0442	0.7406 ± 0.0402

Human and Mouse (Criteria = 5)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	57116	29831	17718	0.3102	0.5939
K = 2000, O = 400	57116	25699	18672	0.3269	0.7266
K = 2000, O = 500	57116	20660	16857	0.2951	0.8159
K = 3000, O = 300	57116	26488	15302	0.2679	0.5777
K = 3000, O = 400	57116	17110	12085	0.2116	0.7063
K = 3000, O = 500	57116	11343	8657	0.1516	0.7632
K = 5000, O = 300	57116	0	0	0.0000	0.0000
K = 5000, O = 400	57116	0	0	0.0000	0.0000
K = 5000, O = 500	57116	0	0	0.0000	0.0000
Average ± SD				0.1737 ± 0.1406	0.4648 ± 0.3565

Human and Mouse (Criteria = 0)

Parameters	True Count	Computed Count	Correct	Sensitivity	Specificity
K = 2000, O = 300	57116	29831	13627	0.2386	0.4569
K = 2000, O = 400	57116	25699	14780	0.2588	0.5751
K = 2000, O = 500	57116	20660	13619	0.2384	0.6592
K = 3000, O = 300	57116	26488	11822	0.2070	0.4463
K = 3000, O = 400	57116	17110	9835	0.1722	0.5748
K = 3000, O = 500	57116	11343	7165	0.1254	0.6317
K = 5000, O = 300	57116	0	0	0.0000	0.0000
K = 5000, O = 400	57116	0	0	0.0000	0.0000
K = 5000, O = 500	57116	0	0	0.0000	0.0000
Average ± SD				0.1378 ± 0.1107	0.3716 ± 0.2873

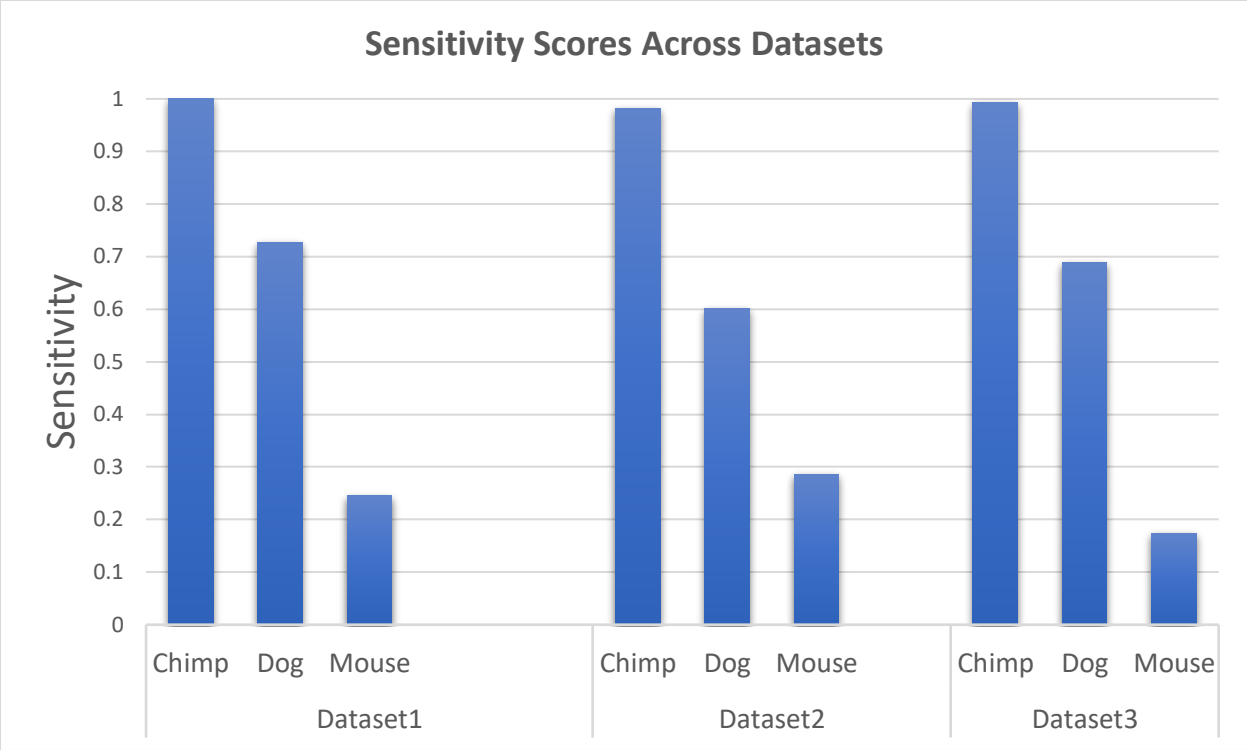

Fig 2: A bar chart showing sensitivity values between human and other species in three different datasets

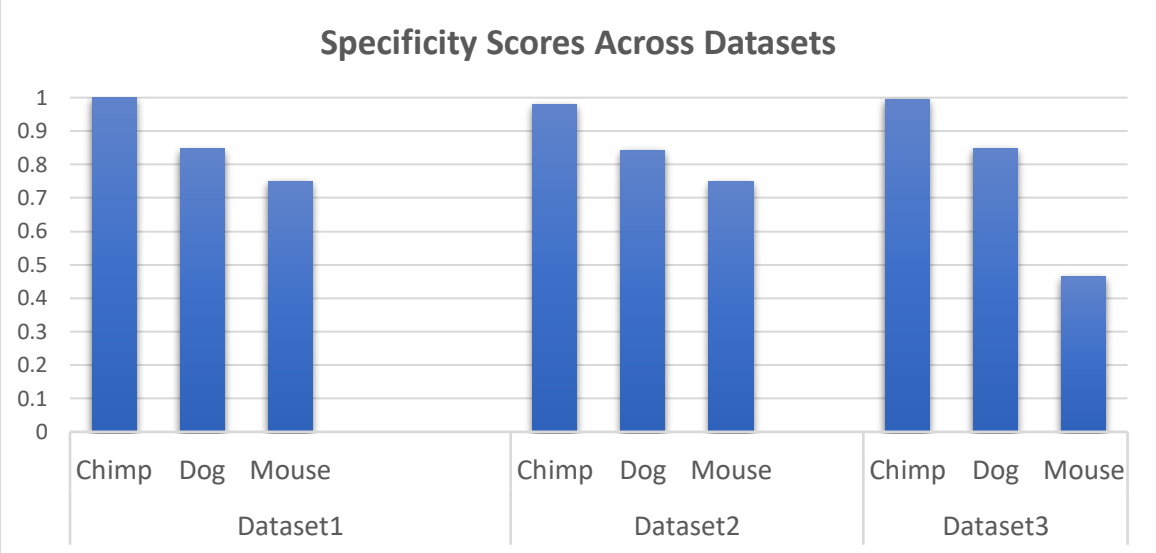

Fig 1: A bar chart showing sensitivity values between human and other species in three different datasets

Discussion:

The above results show that an increase in the “c” value generally leads to an increase in sensitivity and specificity metrics. A “c” value of 5 generally produced a higher specificity and sensitivity value than a “c” value of “0”. This is because an increase in the “c” value allows for the tolerance of more errors during alignment which increases the number of aligned sequences leading to an increase in sensitivity and specificity. The results also show that the value of “K” also affects the sensitivity. Since “K” corresponds to the alignment score cut off, a lower “K” value lowers the threshold for a true alignment which in turn would lead to an increase in the computed count and sensitivity. A greater “K” value indicates a greater threshold for true alignments which leads to a lower sensitivity score. For comparisons between human, dog and mice genomes, an increase in the gap penalty “O”, usually led to a decrease in computed count. This is because increasing the gap penalty leads to fewer alignment reaching the threshold score. The human and chimp genome comparison maintained a high sensitivity and specificity score regardless of the values for “K”, “O” and “C”. This is most likely because of the high similarity between human and chimpanzee genomes allowing for an easier alignment. In general, sensitivity and specificity between human and chimpanzee were higher than those between human-mouse and human-dog. This might indicate a closer genetic similarity between humans and chimpanzee in comparison to humans and dogs/mice.