

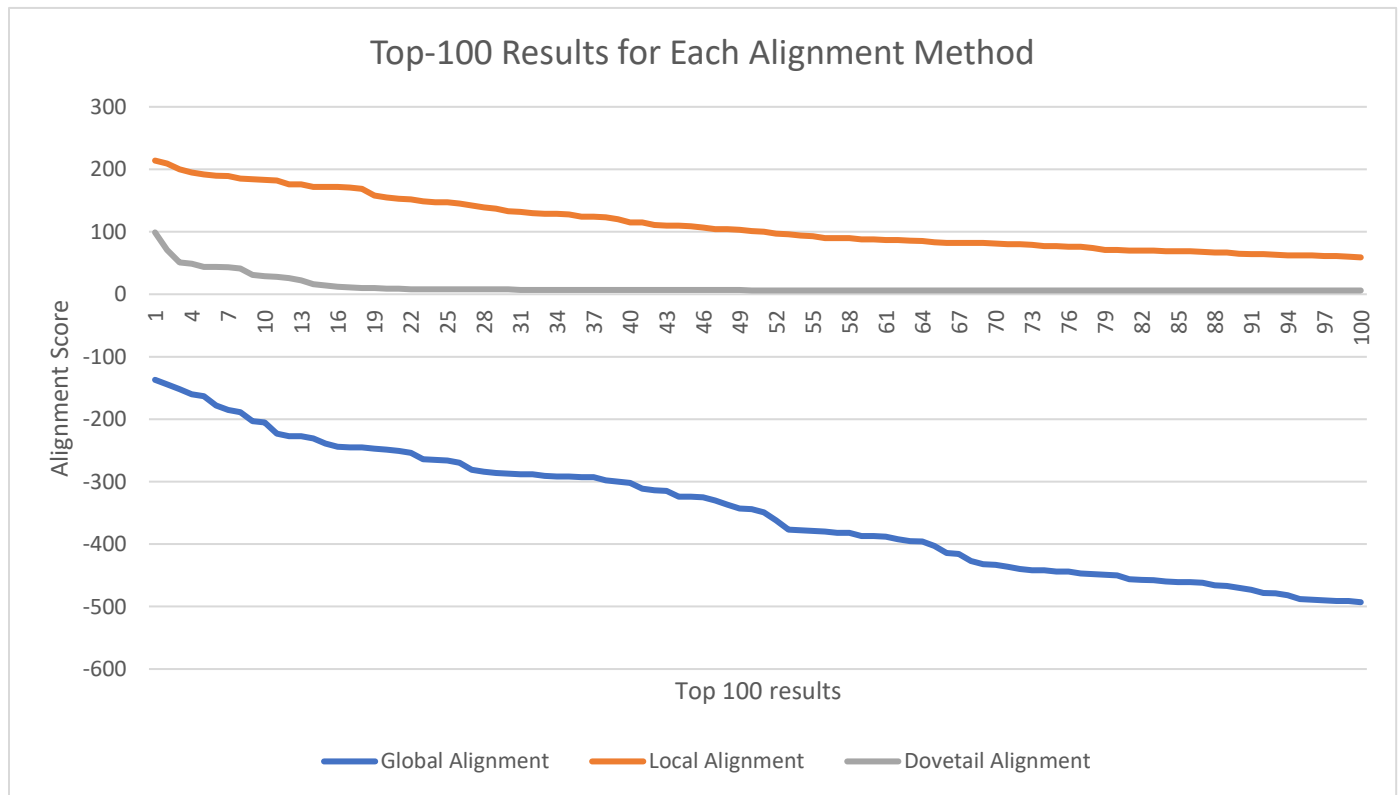
# CAP5510 Fall 2017 – Bioinformatics

## Homework – I

### Implementation of Various Sequence Alignment Methods

Jay Shah (1461-3930)

### Top hundred results using each sequence alignment method



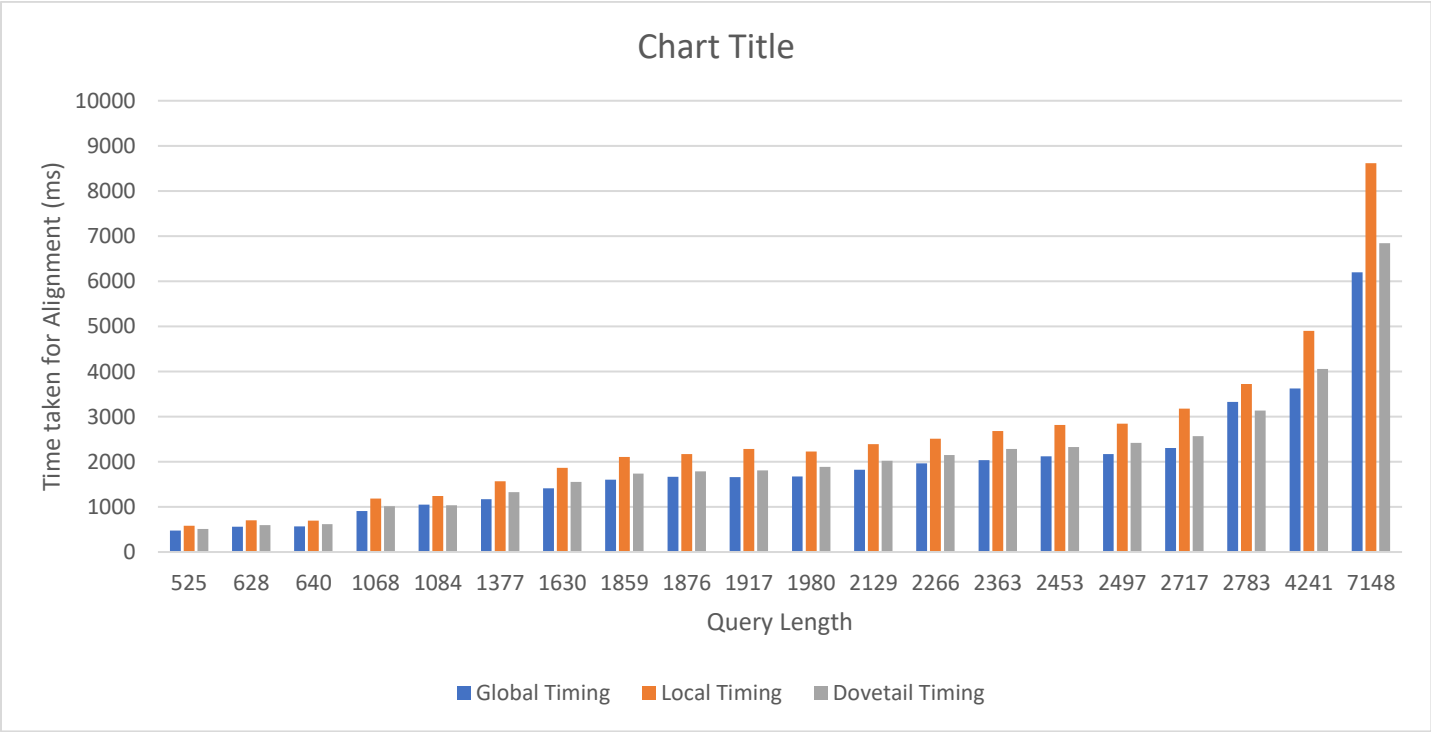
- Given top results are in the descending order.
- X-axis is result number and Y-axis shows the alignment score for each result.
- Observation: -
  - Similarity matrix results are positive in Local and Dovetail alignment because we only align specific chunk from the sequence while Global have negative results as we align entire sequence of a query with database sequence.
- Table for top-100 result is as below.

Number of Results	Global Alignment	Local Alignment	Dovetail Alignment
1	-137	214	99
2	-144	209	71
3	-152	200	51
4	-160	195	49
5	-163	192	44
6	-178	190	44
7	-185	189	43
8	-189	185	41
9	-203	184	31
10	-205	183	29
11	-223	182	28
12	-227	176	26
13	-227	176	22
14	-231	172	16
15	-239	172	14
16	-244	172	12
17	-245	171	11
18	-245	169	10
19	-247	158	10
20	-249	155	9
21	-251	153	9
22	-254	152	8
23	-264	149	8
24	-265	147	8
25	-266	147	8
26	-270	145	8
27	-281	142	8
28	-284	139	8
29	-286	137	8
30	-287	133	8
31	-288	132	7
32	-288	130	7
33	-291	129	7
34	-292	129	7
35	-292	128	7
36	-293	124	7
37	-293	124	7
38	-298	123	7
39	-300	120	7
40	-302	115	7
41	-311	115	7
42	-314	111	7

43	-315	110	7
44	-324	110	7
45	-324	109	7
46	-325	107	7
47	-330	104	7
48	-337	104	7
49	-343	103	7
50	-344	101	6
51	-349	100	6
52	-362	97	6
53	-377	96	6
54	-378	94	6
55	-379	93	6
56	-380	90	6
57	-382	90	6
58	-382	90	6
59	-387	88	6
60	-387	88	6
61	-388	87	6
62	-392	87	6
63	-395	86	6
64	-396	85	6
65	-403	83	6
66	-414	82	6
67	-416	82	6
68	-427	82	6
69	-432	82	6
70	-433	81	6
71	-436	80	6
72	-440	80	6
73	-442	79	6
74	-442	77	6
75	-444	77	6
76	-444	76	6
77	-447	76	6
78	-448	74	6
79	-449	71	6
80	-450	71	6
81	-456	70	6
82	-457	70	6
83	-458	70	6
84	-460	69	6
85	-461	69	6

86	-461	69	6
87	-462	68	6
88	-466	67	6
89	-467	67	6
90	-470	65	6
91	-473	64	6
92	-478	64	6
93	-479	63	6
94	-482	62	6
95	-488	62	6
96	-489	62	6
97	-490	61	6
98	-491	61	6
99	-491	60	6
100	-493	59	6

Running time of each method for each query



- Here, X-axis shows the length of each query (number of characters) and Y-axis represents total time in milliseconds taken by each query to align it with each database sequence.
- Table for query length and its running time is below.

Query Length	Global Timing	Local Timing	Dovetail Timing
525	474	585	509
628	563	700	597
640	569	696	615
1068	910	1183	1012
1084	1048	1245	1033
1377	1169	1570	1324
1630	1412	1865	1554
1859	1601	2108	1738
1876	1670	2170	1785
1917	1658	2282	1806
1980	1677	2225	1886
2129	1826	2387	2021
2266	1966	2509	2149
2363	2039	2681	2282
2453	2118	2818	2326
2497	2171	2841	2417
2717	2306	3181	2569
2783	3329	3727	3137
4241	3625	4901	4059
7148	6199	8614	6842

- Observation: -
  - Running time of an algorithm is proportional to size of the query. As query size increases, running time of each algorithm increases.
  - Among all alignment methods,
    - Local alignment runs longer than Dovetail and Global alignment.
    - Dovetail alignment runs longer than Global alignment.
    - Global alignment is the fastest algorithm in terms of running time among all three algorithms.