

Pivot Table_jellyfish_performance_1

Sum - Wall clock time, in se		k		Threads		13					15					17					19					21					Total Result	
Jellyfish versi	Genome size	11	1	2	4	8	16	1	2	4	8	16	1	2	4	8	16	1	2	4	8	16	1	2	4	8	16	1	2	4		8
1.1.2	11743	0.028	0.019	0.018	0.018	0.022	0.131	0.104	0.08	0.098	0.08	10.94	5.782	3.125	2.621	2.661	11.06	5.833	3.399	3.276	3.333	11.58	6.214	3.403	3.014	2.982	12	6.485	3.693	3.201	3.29	108.4819291
	4031080	0.264	0.177	0.134	0.098	0.098	0.446	0.331	0.208	0.18	0.177	12.17	6.547	3.447	2.725	2.676	12.38	6.564	3.588	2.815	2.788	12.65	6.645	4.041	3.148	3.025	12.88	6.79	3.732	3.129	3.151	117.0010624
	280072263	12.89	6.875	3.286	3.2	2.778	16.17	8.635	5.094	4.723	4.44	58.32	27.82	14.28	10.17	9.471	54.45	31.35	17.67	13.86	13.2	55.27	27.88	15.13	10.11	10.69	53.73	28.97	14.96	10.25	11.63	557.2876005
2.2.10	11743	0.074	0.036	0.024	0.023	0.029	0.854	0.448	0.438	0.266	0.285	13.66	7.394	6.158	4.99	5.288	14.42	8.669	6.94	4.67	5.091	16.02	8.746	7.324	4.895	5.078	14.56	7.918	6.49	4.973	4.826	160.585674
	4031080	0.697	0.364	0.197	0.152	0.192	1.869	0.974	0.739	0.462	0.49	11.71	6.417	5.64	3.589	3.8	12.5	8.299	6.589	5.42	5.097	16.96	9.603	7.674	5.365	5.54	17.51	10.33	10.15	4.809	5.431	168.5695405
	280072263	26.72	15.65	9.729	7.267	6.53	34.22	16.75	8.997	7.485	7.958	58.34	28.92	18.3	16.76	15.33	60.71	29.05	20.14	15.9	16.07	58.34	33.28	22.1	16.11	16.29	61.84	34.45	22.57	15.97	16.22	717.9716806
2.3.0	11743	0.097	2.035	2.367	2.382	2.424	1.53	33.83	38.33	38.64	39.15	26.17	531.1	622.7	659	672.6	16.44	8.464	6.827	5.937	6.263	16.49	9.091	7.714	6.48	7.153	17.22	10.38	8.002	6.377	6.729	2811.873609
	4031080	0.702	0.886	1.115	1.078	1.083	2.658	12.82	15.24	15.35	15.63	26.56	570.5	636.1	665.9	648.8	14.16	7.389	7.022	6.983	7.287	18.48	11.7	10.11	6.702	9.377	20.54	9.19	7.387	6.802	6.681	2764.255971
	280072263	22.07	13.45	8.851	6.919	7.022	37.45	44.9	41.2	39.8	38.38	115.6	456.3	502.9	509.8	515.8	43.45	22.07	13.17	13.45	15.09	43.99	22.8	13.33	12.53	16.42	44.44	22.93	13.8	13.23	15.25	2686.332221
Total Result		63.54	39.49	25.72	21.14	20.18	95.32	118.8	110.3	107	106.6	333.5	1641	1813	1875	1876	239.6	127.7	85.35	72.3	74.22	249.8	136	90.83	68.35	76.56	254.7	137.4	90.78	68.74	73.2	10092.35929

Cells are the wall-clock time in seconds of a single run.

Rows correspond to a given jellyfish version and genome size

Columns correspond to a given k-mer length (k) and number of threads passed to Jellyfish in -m and -t arguments (respectively).

Colors are to highlight performance; Green indicates completion in under one second, yellow in under 100 seconds, and dark red in under 1000 seconds.

System tested on is an 8-core laptop [Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz] with 32GB of RAM.

No concessions were made to ensure accuracy (i.e. I didn't isolate the system from other unrelated activity), but the scale of the issue and repeatability moot that challenge.

Original data is available, and I am open to performing additional tests if desired.

Example command:

```
jellyfish count -C -o fsplit0.Jelly -m 11 -s 1000000000 -t 1 EEE_BeAr436087.fasta
```

Jellyfish versions are as downloaded from official sources. 1.1.2 is a legacy version found in a kSNP3 package, 2.2.10 on GitHub, and 2.3.0 distributed as a Debian package from the Debian repo.

Genome sizes are file sizes for the FASTA files, not in number of base-pairs.

Specific genomes: (original data can be found in kSNP3 Examples.zip, with caveat that ERR579925 was generated using kSNP3's merge_fasta_reads.pl on the original FASTA file)

File name: Notes:

EEE_BeAr436087.fasta This is the smallest genome.

ERR579925.fasta.merged This file has significant redundancy; it consists of many 49-base-pair sections that were merged, and counts for most k-mers are much greater than one. It is the largest.

VcMS6.fa

This suggests there is a clear (orders of magnitude) regression in 2.3.0 performance for kmer length of 13 and 15, and that it is associated with threading (single-threaded performs significantly better than even 2 threads)

Noteworthy: Relative performance of 2.3.0 improves with larger amounts of data / inputs.