Shasta assembly summary

Shasta version

Shasta Release 0.13.0

Reads used in this assembly

Read representation	1 (RLE)
Minimum read length	10000
Number of reads	2143541
Number of read bases	74555517468
Average read length	34781
Read N50	41845
Number of run-length encoded bases	52034665636
Average length ratio of run-length encoded sequence over raw sequence	0.6979
Number of reads flagged as palindromic by self alignment	527
Number of reads flagged as chimeric	3074

- Reads discarded on input are not included in the above table (see below).
- See ReadLengthHistogram.csv and Binned-ReadLengthHistogram.csv for details of the read length distribution of reads used in this assembly.

Reads discarded on input

	Reads	Bases
Reads discarded on input because they contained invalid bases	0	0
Reads discarded on input because they were too short	363962	2159212300
Reads discarded on input because they contained repeat counts greater than 255	2268	77057228
Reads discarded on input, total	366230	2236269528
Fraction of reads discarded on input over total present in input files	0.1459	0.02912

- Base counts in the above table are raw sequence bases.
- Here and elsewhere, "raw" refers to the original read sequence, as opposed to run-length encoded sequence.

Markers

Total number of markers on all reads, one strand	5173358648
Total number of markers on all reads, both strands	10346717296
Average number of markers per raw base	0.06939
Average number of markers per run-length encoded base	0.09942
Average base offset between markers in raw sequence	14.41
Average base offset between markers in run-length encoded sequence	10.06
Average base gap between markers in run-length encoded sequence	-3.942

Alignments

Number of alignment candidates found by the LowHash algorithm	48987287
Number of good alignments	48410886
Number of good alignments kept in the read graph	10666658

Alignment criteria actually used for creation of the read graph

minAlignedMarkerCount	105
minAlignedFraction	0.595
maxSkip	22
maxDrift	10
maxTrim	21

Read graph

Number of vertices	4287082
Number of edges	21333316

- The read graph contains both strands. Each read generates two vertices.
- Isolated reads in the read graph don't contribute to the assembly. See the table below for a summary of isolated reads in the read graph. Each isolated read corresponds to two isolated vertices in the read graph, one for each strand.

	Reads	Bases
Isolated reads	89226	2065796025
Non-isolated reads	2054315	72489721443
Isolated reads fraction	0.04163	0.02771
Non-isolated reads fraction	0.9584	0.9723

Marker graph

Total number of vertices	334108694
Total number of edges	482437222
Number of vertices that are not isolated after edge removal	331689162
Number of edges that were not removed	331685658

- The marker graph contains both strands.

Assembly graph

Number of vertices	7382
Number of edges	3870
Number of edges assembled	1935

- The assembly graph contains both strands.

Assembled segments ("contigs")

Number of segments assembled	1935
Total assembled segment length	2815321493
Longest assembled segment length	12905093
Assembled segments N_{50}	2758072

- Shasta uses GFA terminology (*segment* instead of the most common *contig*). A contiguous section of assembled sequence can consist of multiple segments, for example in the presence of heterozygous bubbles.
- See AssemblySummary.csv for lengths of assembled segments.

Performance

Elapsed time (seconds)	3889
Elapsed time (minutes)	64.82
Elapsed time (hours)	1.08
Average CPU utilization	0.106
Peak virtual memory utilization (bytes)	412349517824
Number of threads used	150
Total number of virtual CPUs available	255
Total physical memory available (bytes)	3921325645824