Supplemental Table S5. Assembly quality assessment: Illumina-derived versus genome-derived list of valid k-mers (Quast).

Organism	Method*	Valid k-mer source†	Contig number	Largest contig	Total length	N50	NG50	N75	NG75	L50	LG50	L75	LG75	Misassemblies > 1kb	Mismatches/100kb	Indels/100kb
C. elegans	M	-	153	5285091	104406335	2119841	2220855	1332283	1392101	17	16	31	29	1749	15.64	46.44
C. elegans	L	Illumina (low frequency removed)	153	4763590	104240890	1967842	2031208	1188901	1226313	17	16	33	31	1546	14.40	46.53
C. elegans	L	genome (all k-mers)	153	5266065	104190476	2118433	2435154	1447476	1499931	16	15	31	29	1661	15.93	46.77
C. elegans	LH	Illumina (low and high frequency removed)	108	7255918	103011904	2689323	2838280	1232606	1549290	13	12	26	24	1649	15.26	47.12
C. elegans	LH	genome (single-copy k-mers)	159	4955962	104445248	2281043	2340439	1341813	1629514	17	16	31	28	1657	16.21	47.45
E. coli	M	-	2	4654123	4685062	4654123	4654123	4654123	4654123	1	1	1	1	5	0.09	14.02
E. coli	L	Illumina (low frequency removed)	2	4654192	4685140	4654192	4654192	4654192	4654192	1	1	1	1	5	0.35	14.28
E. coli	L	genome (all k-mers)	2	4654189	4685137	4654189	4654189	4654189	4654189	1	1	1	1	5	0.35	14.24
E. coli	LH	Illumina (low and high frequency removed)	2	4654133	4685081	4654133	4654133	4654133	4654133	1	1	1	1	5	0.19	13.42
E. coli	LH	genome (single-copy k-mers)	2	4654144	4685092	4654144	4654144	4654144	4654144	1	1	1	1	5	0.32	13.44

* "M", standard MHAP; "L", MHAP with low-frequency k-mer masking; "LH", MHAP with low and high frequency k-mer masking.

† Note that a genome-derived list of all k-mers is similar to an Illumina derived list with low frequency k-mers removed (both were called "L-masking" method here). Analogously, a genome-derived list of single-copy k-mers is similar to an Illumina derived list with low frequency and high-frequency k-mers removed (both were called "LH-masking" method).