

Supplemental Table S4. Assembly quality assessment (Quast; all assemblies)

Assembly*	Comment	Contig number	Largest contig	Total length	N50	NG50	N75	NG75	L50	LG50	L75	LG75	Misassemblies > 1kb †	Mismatches/100kb	Indels/100kb
ecoli_M_30x	low coverage	26	916042	4577644	356518	356518	155650	155650	5	5	10	10	3	2.72	139.79
ecoli_L_30x	low coverage	11	1363117	4588137	969401	969401	358060	358060	2	2	5	5	3	1.50	107.10
ecoli_LH_30x	low coverage	13	1366089	4611905	969380	969380	358054	358054	2	2	5	5	3	1.70	112.37
yeast_M_128	small sketch	91	983833	12461105	457297	457297	213780	237877	10	10	20	19	123	76.94	58.05
yeast_L_128	small sketch	43	1537479	12579864	745960	777025	446753	455202	7	6	12	11	126	82.75	31.19
yeast_LH_128	small sketch	37	1530874	12487180	751107	751107	572465	572465	7	7	11	11	122	83.50	27.63
ecoli_M		2	4654122	4685061	4654122	4654122	4654122	4654122	1	1	1	1	5	0.09	14.02
ecoli_L		2	4654191	4685139	4654191	4654191	4654191	4654191	1	1	1	1	5	0.35	14.26
ecoli_LH		2	4654137	4685085	4654137	4654137	4654137	4654137	1	1	1	1	5	0.22	13.72
yeast_M		39	1533267	12569074	818055	818055	457460	600310	6	6	11	10	130	80.73	28.99
yeast_L		36	1537666	12460414	678320	750938	446799	468828	7	6	12	11	123	81.07	29.89
yeast_LH		41	1530583	12644445	818185	818185	462140	468595	6	6	11	10	125	86.08	26.30
yeast_M_p	pbdagcon	45	1530555	12599740	815596	815596	474596	474596	6	6	11	11	130	78.57	18.19
yeast_L_p	pbdagcon	37	1535936	12463610	815424	815424	582263	594116	6	6	11	10	128	79.11	19.16
yeast_LH_p	pbdagcon	40	1537667	12538181	818104	818104	574548	602184	6	6	11	10	121	77.59	17.36
cel_M		153	5285091	104406335	2119841	2220855	1332283	1392101	17	16	31	29	1749	15.64	46.44
cel_L		153	4763590	104240890	1967842	2031208	1188901	1226313	17	16	33	31	1546	14.40	46.53
cel_LH		108	7255918	103011904	2689323	2838280	1232606	1549290	13	12	26	24	1649	15.26	47.12
ara_M		727	15819004	134469351	9063257	9587932	6095039	6717465	6	5	10	9	6019‡	600.43‡	162.15‡
ara_L		620	17168897	133073544	8753656	8919426	4697047	6094458	6	5	11	9	5880‡	598.11‡	164.15‡
ara_LH		633	18788518	133270108	9610620	13499602	6095028	6717020	5	4	9	8	5767‡	602.62‡	168.38‡
dros_M		1072	21678627	169543188	4879969	7157936	511201	1674008	7	5	38	15	11136	14.57	93.37
dros_L		963	18648553	167735167	5370285	7147503	550370	1849454	7	5	30	14	10278	14.40	88.27
dros_LH		1019	25756195	169542479	7911769	13654652	604262	2159549	5	4	27	10	11161	14.60	110.28
human_chr15_M	sorted reads	343	7309895	83890337	2647243	1602477	1201816	213215	10	15	23	45	78	86.32	95.86
human_chr15_L	sorted reads	330	8894105	83396444	2800843	2152643	947681	191831	9	13	21	48	73	83.96	94.97
human_chr15_LH	sorted reads	322	9276182	83558937	3188709	1897981	1278137	385258	8	12	20	38	77	82.60	90.37
human_chr17_M	sorted reads	414	5412440	83732343	2081827	2081827	885562	885562	14	14	29	29	182	73.84	103.51
human_chr17_L	sorted reads	388	6783205	82743477	1606756	1606756	841796	841796	14	14	32	32	190	72.78	104.28
human_chr17_LH	sorted reads	361	8323604	82720089	2081188	2081188	1070300	1070300	11	11	25	25	152	67.11	96.96

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

† As reported by Quast (Gurevich et al. 2013): misassemblies > 1kb, or joining different chromosomes. The large values in *Drosophila* probably were caused by the fragmentation of the reference sequence.

‡ The reported assembly errors in *Arabidopsis* are unreliable because the reference genome came from a different strain.