

Supplementary material

to the paper

“Genome-scale *de novo* assembly using ALGA”

Sylwester Swat ¹, Artur Laskowski ¹, Jan Badura ¹, Wojciech Frohmberg ¹, Pawel Wojciechowski ^{1,2}, Aleksandra Swiercz ^{1,2}, Marta Kasprzak ¹, Jacek Blazewicz ^{1,2}

1) Poznan University of Technology, Institute of Computing Science, Piotrowo 2, 60-965 Poznan, Poland

2) Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland

1. Data sets and reference genomes

In the computational experiment, data sets from Illumina paired-end sequencing of six organisms were used. Reference genomes of the organisms were used in the evaluation of the results by QUASt. Links to the files are specified below.

a) *Homo sapiens* chromosome 14

- genome:
<https://www.ncbi.nlm.nih.gov/nucleotide/568815584>
- reads:
http://gage.cbcb.umd.edu/data/Hg_chr14/Data.original/frag_1.fastq.gz
http://gage.cbcb.umd.edu/data/Hg_chr14/Data.original/frag_2.fastq.gz

b) *Caenorhabditis elegans*

- genome:
ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/985/GCF_000002985.6_WBcel235/GCF_000002985.6_WBcel235_genomic.fna.gz
- reads:
ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/DRA000/DRA000967/DRX007632/DRR008444_1.fastq.bz2
ftp://ftp.ddbj.nig.ac.jp/ddbj_database/dra/fastq/DRA000/DRA000967/DRX007632/DRR008444_2.fastq.bz2

c) *Arabidopsis thaliana*

- genome:
https://www.ncbi.nlm.nih.gov/assembly/GCA_902460285.1
- reads:
<https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR3624574>

d) *Chlorella sorokiniana*

- genome:
<ftp://ftp.cngb.org/pub/CNSA/data1/CNP0000212/CNS0018477/CNA0000826/>
- reads:
<http://ftp.cngb.org/pub/CNSA/data1/CNP0000212/CNS0018477/CNX0022250/CNR0026866/>

e) *Escherichia coli*

- genome:
ftp://ftp.ensemblgenomes.org/pub/bacteria/release-45/fasta/bacteria_0_collection/escherichia_coli_str_k_12_substr_mg1655/dna/Escherichia_coli_str_k_12_substr_mg1655.ASM584v2.dna.chromosome.Chromosome.fa.gz
- reads:
ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR022/ERR022075/ERR022075_1.fastq.gz
ftp://ftp.sra.ebi.ac.uk/vol1/fastq/ERR022/ERR022075/ERR022075_2.fastq.gz

f) *Microthrix parvicella*

- genome:
https://sra-download.ncbi.nlm.nih.gov/traces/wgs03/wgs_aux/AM/PG/AMPG01/AMPG01.1.fsa_nt.gz
- reads:
<https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos/sra-pub-run-5/SRR576810/SRR576810.1>

g) *Homo sapiens* whole genome

- genome:
https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/GCA_000001405.15_GRCh38_genomic.fna.gz
- reads:
<https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=ERR3587301>

2. Software

For all the programs from the computational experiment, their versions and web pages with source codes are given.

a) ALGA

- version: 1.0.3 (access date 25 Nov 2020)
- source: <http://alga.put.poznan.pl>

b) GRASShopper

- version: – (access date 10 Jan 2020)
- source: <https://sourceforge.net/projects/grasshopper-assembler>

c) MEGAHIT

- version: 1.2.9 (access date 12 May 2020)
- source: https://github.com/voutcn/megahit/releases/download/v1.2.9/MEGAHIT-1.2.9-Linux-x86_64-static.tar.gz

d) Platanus

- version: 1.2.4 (access date 14 Apr 2020)
- source: <http://platanus.bio.titech.ac.jp/platanus-assembler/platanus-1-2-4>

e) Readjoiner

- version: 1.6.1 (access date 3 Mar 2020)
- source: <http://genometools.org/pub>

f) SAGE2

- version: 1.0.0 (access date 22 Apr 2020)
- source: <https://github.com/lucian-ilie/SAGE2>

g) SGA

- version: 0.10.14 (access date 3 Mar 2020)
- source: downloaded with command “docker pull ljdursi/sga-docker”

h) SOAPdenovo2

- version: 2.04 (access date 17 Apr 2020)
- source: <https://github.com/aquaskyline/SOAPdenovo2>

i) SPAdes

- version: 3.14.0 (access date 26 Dec 2019)
- source: <http://cab.spbu.ru/software/spades>

j) Velvet

- version: 1.2.10 (access date 22 Jan 2020)
- source: <https://www.ebi.ac.uk/~zerbino/velvet>

k) Musket

- version: 1.1 (access date 2 Jan 2020)
- source: <http://musket.sourceforge.net>

l) QUAST

- version: 5.0.2 (access date 22 Nov 2019)
- source: <http://quast.sourceforge.net/quast.html>

m) BUSCO

- version: 4.0.5 (access date 20 May 2020)
- source: downloaded with command “`docker pull ezlabgva/busco:v4.0.5_cv1`”

3. Program parameters

The following parameters were specified while the programs were executed. Parameters pointing to the input and output files, as well as parameters used for an additional description of a particular data set (e.g. expected depth of coverage, expected genome size, or average insert size), are omitted. Most of the settings were recommended by the authors of the software. Some of the programs did not require a parameter for the number of threads and automatically took 16 available cores.

a) ALGA

- "--threads=16" (16 threads)

b) GRASShopPER

- "-ws=50", "-sli=31", "-ss=14", "-awa=TRUE" (program-specific parameters steering the level of exactness)
- "-e=4", "-fs=2" (tolerance to errors)
- "-minconf=7", "-maxrefs=7" (tolerance to distant paired-end reads)
- "-sc=140", "-ps=140", "-trimparams="ILLUMINACLIP:\${TRIMMOMATIC_PATH}/adapters/TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:30:20 LEADING:15 TRAILING:15 MINLEN:85 CROP:210"" (parameters set for *C. sorokiniana* because of longer reads)

c) MEGAHIT

- "-t 16" (16 threads)

d) Platanus

- "-m 128" (memory limit)
- "-t 16" (16 threads)

e) Readjoinder

- "-l 60" (minimum allowed length of overlaps of reads)
- "-j 8" (8 threads, 16 didn't work)
- "-spmfiles 8" (8 threads)
- parameters characterizing the input data set

f) SAGE2

- "-k 50" (minimum allowed length of overlaps of reads)

g) SGA

- "--pe-mode 1" (a mode of pairing reads)
- "-a ropebwt", "--no-reverse", "-k 41", "--discard", "--learn" (parameters for error correction)
- "-x 2", "--homopolymer-check", "--low-complexity-check" (parameters for filtering reads)
- "-m 55" (minimum allowed length of overlaps of reads)
- "-m 75" (at the assembly stage, minimum allowed length of overlaps of merged paths)
- "-g 0", "-r 10" (other parameters for the assembly stage)
- "-t 16" (16 threads)

h) SOAPdenovo2

- “-K 63” (length of k-mers)
- “-R” (resolving repeats by reads)
- “asm_flags=3”, “rank=1”, “pair_num_cutoff=3”, “map_len=32” (program-specific parameters)
- parameters characterizing the input data set

i) SPAdes

- “-k 21,33,55,77” (lengths of k-mers)
- “--careful” (a parameter for error reduction)

j) Velvet

- parameters characterizing the input data set

k) Musket

- “-omulti” (produces multiple output files)
- “-p 18” (18 threads)

l) QUASt

- “--min-contig 250” (or “--min-contig 500” for *C. sorokiniana* because of longer reads)
- “--ambiguity-usage one” (takes one best alignment of a contig)
- “--space-efficient” (creates only primary output)
- “--threads 16” (16 threads)

m) BUSCO

- “-m genome” (sets the assessment mode to genome)
- “-l <data set>” (specifies the BUSCO lineage data set)

4. How to use ALGA

For more information about ALGA please visit <http://alga.put.poznan.pl>.

a) REQUIREMENTS

- CMake version 2.8.7 or higher
- C++ 17 or higher

b) INSTALLATION

- Download the archive with code of ALGA and unpack it, or clone the ALGA repository. Use CMake to obtain the binary file. For example, in Linux, in the main directory of ALGA, you can use the following commands:
 - `mkdir build`
 - `cd build`
 - `cmake ..`
 - `make`
- After this, the executable file named "ALGA" is in directory "build"

c) USAGE TIPS

- It is strongly recommended to use Musket, a tool for read correction based on a k-mer analysis, before running ALGA
- A typical usage of ALGA consists in specifying one or two input files (both with .fastq or .fasta extension), the number of threads and the output file name for contigs. An example:

```
./ALGA --file1=path1/corrected-reads_1.fastq
      --file2=path2/corrected-reads_2.fastq
      --threads=8
      --output=contigs.fasta
```
- Only one input file may be specified, just remove the argument
`--file2=path2/corrected-reads_2.fastq`
- The number of threads is an optional parameter and can be removed, it is set to 6 by default

d) ADDITIONAL PARAMETER

- If you suspect that the input data are for some reason of very poor quality and may – even after the read correction – still contain a large number of errors, you can additionally use the option
`--error-rate=0.02`
(the specified value, here 0.02, denotes the average expected fraction of errors).

5. Results of computations

On the subsequent pages, detailed results of the computational experiment are shown. Tables S1–S6 present reports from QUAST for all assemblers and data sets used in the assemblers' comparison. Tables S7–S12 present results of evaluation by BUSCO. Tables S13–S15 show the usage of time and memory. Table S16 presents a QUAST report for tests with a whole human genome data set. Table S17 contains results of ALGA executed with the additional parameter "--error-rate=0.05".

Table S1: HOMO SAPIENS CHROMOSOME 14

REPORT FROM QUAST FOR ASSEMBLY RESULTS

Genome statistics	ALGA	GRASShopper	MEGAHIT	Platanus	Readjoiner	SAGE2	SGA	SOAPdenovo2	SPAdes	Velvet
Genome fraction (%)	92.963	91.353	90.59	76.156	56.802	78.717	90.859	92.643	94.459	90.857
Duplication ratio	1.003	1.04	1.009	1.005	1.083	1.002	1.007	1.01	1.011	1.012
Largest alignment	64709	29556	99626	15459	3233	44201	63564	61616	103374	41413
Total aligned length	84443371	85870994	82525950	69304095	55711845	71351973	82856749	84664412	86468768	83025542
NG50	5811	1860	10448	1113	300	1357	5092	3895	5846	2751
NG75	2505	765	4467	281	-	380	2205	1738	2455	1265
NA50	6360	2000	10409	1615	431	1954	5698	4248	6121	3051
NA75	3143	945	5201	860	323	982	2896	2126	2804	1583
NGA50	5798	1855	9231	1112	300	1355	5077	3889	5743	2723
NGA75	2496	757	3872	280	-	377	2191	1732	2412	1235
LG50	4271	12867	2416	20540	93285	15568	4971	6351	3825	9148
LG75	10164	31564	5692	58978	-	45762	11679	15038	9864	21231
LA50	3737	11603	2270	12421	43782	9525	4226	5581	3529	7882
LA75	8433	27285	5085	27165	81405	22565	9307	12612	8759	17386
LGA50	4277	12887	2689	20544	93340	15584	4980	6355	3908	9215
LGA75	10186	31666	6452	59037	-	45912	11714	15061	10053	21497
Misassemblies										
# misassemblies	60	213	2109	1	3	5	56	17	340	818
# relocations	56	213	2097	1	3	5	56	16	339	817
# translocations	0	0	0	0	0	0	0	0	0	0
# inversions	4	0	12	0	0	0	0	1	1	1
# misassembled contigs	59	210	1884	1	3	5	56	17	334	751
Misassembled contigs length	273648	172376	17843346	786	1103	9643	303731	37087	3012555	1384361
# local misassemblies	16	342	143	5	1	6	17	11	51	292
# scaffold gap ext. mis.	0	0	0	0	0	0	0	0	0	0
# scaffold gap loc. mis.	0	0	0	0	0	0	0	0	0	0
# unaligned mis. contigs	3	1	5	0	2	2	2	3	2	42
Unaligned										
# fully unaligned contigs	34	52	12	34	51	34	12	28	47	40
Fully unaligned length	24666	27052	12388	14595	20105	21032	16989	20852	20154	23462
# partially unaligned contigs	6	7	27	2	2	8	10	7	13	104
Partially unaligned length	8948	7956	40748	1123	1170	7841	14386	11467	26513	98449
Mismatches										
# mismatches	78129	115496	112232	52032	30310	43139	71946	72964	80807	81455
# indels	11529	15289	18777	2227	2560	8210	8139	9964	17295	11921
Indels length	40495	54275	75283	6344	7491	24806	30188	30580	71312	53138
# mismatches per 100 kbp	92.8	139.59	136.79	75.44	58.92	60.51	87.43	86.96	94.46	98.99
# indels per 100 kbp	13.69	18.48	22.89	3.23	4.98	11.52	9.89	11.88	20.22	14.49
# indels (<= 5 bp)	9891	13642	16156	2037	2313	7292	6934	8797	14457	10231
# indels (> 5 bp)	1638	1647	2621	190	247	918	1205	1167	2838	1690
# N's	0	0	0	0	0	0	0	0	1	0
# N's per 100 kbp	0	0	0	0	0	0	0	0	0	0
Statistics without reference										
# contigs	26484	75294	13091	62936	130777	55888	27458	36581	28037	44441
# contigs (>= 0 bp)	30207	87970	13152	626436	1341049	129096	115956	173124	74279	44441
# contigs (>= 1000 bp)	16913	26059	11236	23207	3032	22177	17874	21423	17680	25092
# contigs (>= 5000 bp)	5210	1946	5224	775	0	1713	5090	4337	4682	3171
# contigs (>= 10000 bp)	1743	206	2546	26	0	234	1415	980	1672	400
# contigs (>= 25000 bp)	102	3	424	0	0	2	62	31	235	3
# contigs (>= 50000 bp)	5	0	34	0	0	0	2	1	24	0
Largest contig	64709	29556	99626	15460	3233	44201	63564	61616	103374	41413
Total length	84492543	86076655	82843300	69330571	55750155	71428656	82915135	84761155	86553965	83362754
Total length (>= 0 bp)	85234924	88032101	82856672	99891985	226219328	80715479	93744225	97855849	93222662	83362754
Total length (>= 1000 bp)	79546977	63435124	81570455	48335261	3759105	53217765	77935439	76823136	80946498	72638496
Total length (>= 5000 bp)	50590884	13979749	66072398	4980567	0	12836011	46122967	36663757	50156448	23689097
Total length (>= 10000 bp)	26507413	2582469	46854598	297078	0	3017164	20588597	13641978	29349091	5170444
Total length (>= 25000 bp)	3290094	86168	14767556	0	0	83324	1942291	964962	8278743	95628
Total length (>= 50000 bp)	308375	0	2205049	0	0	0	122798	61616	1482475	0
N50	6375	2000	11701	1615	431	1956	5723	4251	6258	3083
N75	3149	950	5979	860	323	984	2905	2132	2850	1614
L50	3732	11586	2045	12419	43765	9518	4219	5578	3453	7829
L75	8417	27215	4506	27158	81361	22534	9283	12596	8591	17208
GC (%)	40.73	40.31	40.77	39.42	37.38	40.15	40.54	40.68	40.79	40.49

Table S2: CAENORHABDITIS ELEGANS

REPORT FROM QUAST FOR ASSEMBLY RESULTS

Genome statistics	ALGA	GRASShopper	MEGAHIT	Platanus	Readjoiner	SAGE2	SGA	SOAPdenovo2	SPAdes	Velvet
Genome fraction (%)	95.335	95.57	95.653	93.93		88.305	96.146	94.427	95.798	93.34
Duplication ratio	1.004	1.019	1.009	1.005		1.002	1.01	1.004	1.003	1.008
Largest alignment	147732	96243	218906	125773		50722	164617	127853	383483	35134
Total aligned length	95905046	97543907	96582293	94672992		88664856	97326423	95042222	96271132	94263994
NG50	15255	8676	24752	11441		4976	14980	11208	38172	4615
NG75	5792	3097	10637	4158		1829	6007	4367	17492	2226
NA50	16229	9019	23394	12607		5939	15567	12155	39195	4914
NA75	6781	3398	10594	5109		2816	6644	5219	18999	2593
NGA50	15245	8655	22215	11428		4970	14967	11202	37084	4601
NGA75	5767	3067	9522	4148		1823	5985	4355	16880	2211
LG50	1684	2871	1087	2162		5184	1753	2284	703	6260
LG75	4333	7728	2608	5807		13345	4380	5840	1663	14032
LA50	1550	2733	1119	1933		4126	1661	2066	679	5652
LA75	3830	7167	2646	4914		9559	4050	5044	1564	12231
LGA50	1686	2876	1193	2163		5187	1754	2286	730	6274
LGA75	4343	7754	2896	5815		13364	4390	5850	1726	14079
Misassemblies										
# misassemblies	44	202	1097	34		9	54	29	189	133
# relocations	20	86	381	14		5	23	10	79	88
# translocations	22	100	663	20		4	26	19	102	34
# inversions	2	16	53	0		0	5	0	8	11
# misassembled contigs	43	196	982	33		9	52	27	179	131
Misassembled contigs length	230643	372887	18951683	107797		46987	217060	117660	6823210	423098
# local misassemblies	92	436	745	52		29	95	56	246	168
# scaffold gap ext. mis.	0	0	0	0		0	0	0	0	0
# scaffold gap loc. mis.	0	0	0	0		0	0	0	0	0
# unaligned mis. contigs	16	17	14	18		1	15	11	10	9
Unaligned										
# fully unaligned contigs	25	48	21	20		41	18	29	36	33
Fully unaligned length	10651	24715	9812	14106		31378	9586	19034	19072	28275
# partially unaligned contigs	41	36	62	28		12	43	32	43	27
Partially unaligned length	51732	44082	76579	37075		16805	52955	38793	67046	28592
Mismatches										
# mismatches	6959	30904	34050	1989		1609	4237	2805	9812	9664
# indels	2023	5882	6815	1152		972	1431	1214	3318	9866
Indels length	8726	52982	85102	4186		2458	6691	3740	35037	20361
# mismatches per 100 kbp	7.28	32.24	35.5	2.11		1.82	4.39	2.96	10.21	10.32
# indels per 100 kbp	2.12	6.14	7.1	1.22		1.1	1.48	1.28	3.45	10.54
# indels (<= 5 bp)	1824	4863	5163	1069		917	1290	1136	2662	9564
# indels (> 5 bp)	199	1019	1652	83		55	141	78	656	302
# N's	0	0	0	0		0	0	0	0	0
# N's per 100 kbp	0	0	0	0		0	0	0	0	0
Statistics without reference										
# contigs	19063	35848	10951	21927		30321	20634	20858	7986	33681
# contigs (>= 0 bp)	25680	88966	11563	136921		52795	331218	209935	57574	33681
# contigs (>= 1000 bp)	10524	15222	7321	12654		18696	10813	12835	4924	22409
# contigs (>= 5000 bp)	4851	5122	4196	5005		5166	5020	5258	3214	5497
# contigs (>= 10000 bp)	2714	2385	2757	2539		1694	2782	2632	2355	1240
# contigs (>= 25000 bp)	773	453	1075	627		112	769	582	1189	26
# contigs (>= 50000 bp)	172	47	304	103		1	156	84	469	0
Largest contig	147732	96243	218906	125773		50722	164617	127853	383483	35134
Total length	96006321	97758741	96923847	94757369		88745523	97428088	95149169	96410279	94382960
Total length (>= 0 bp)	97237655	106979488	97061141	106169149		92043573	134763311	110902367	101625212	94382960
Total length (>= 1000 bp)	92062461	88950869	95083419	90291892		82531881	93056145	91237251	95115694	88231995
Total length (>= 5000 bp)	78008376	64944282	87126307	71562781		50057082	78722041	72499821	90706462	46483505
Total length (>= 10000 bp)	62841160	45619590	76758094	54172437		25830676	62781454	53816723	84513267	17143007
Total length (>= 25000 bp)	32469497	16316123	49854487	24628989		3500935	31456758	22345962	65238878	747033
Total length (>= 50000 bp)	11830162	2946779	22867086	6789331		50722	10680298	5504420	39866132	0
N50	16265	9033	25917	12616		5950	15590	12169	40738	4927
N75	6796	3420	12186	5120		2823	6665	5228	19710	2608
L50	1548	2729	1021	1932		4124	1659	2064	654	5640
L75	3822	7145	2387	4908		9550	4042	5036	1508	12194
GC (%)	35.36	35.39	35.42	35.36		35.3	35.39	35.37	35.38	35.34

Readjoinder finished computations, but all its contigs were shorter than 250 and QUAST did not included these results.

Table S3: ARABIDOPSIS THALIANA

REPORT FROM QUAST FOR ASSEMBLY RESULTS

Genome statistics	ALGA	GRASShopper	MEGAHIT	Platanus	Readjoiner	SAGE2	SGA	SOAPdenovo2	SPAdes	Velvet
Genome fraction (%)	94.655		95.75	91.449	84.334	87.178	95.851	93.985	95.462	92.016
Duplication ratio	1.008		1.018	1.007	1.132	1.004	1.013	1.007	1.004	1.011
Largest alignment	177159		392309	101718	3315	31479	159831	105557	731281	26666
Total aligned length	114674522		116911228	110734440	114863433	105309017	116689301	113798844	115274330	111813891
NG50	15039		50273	10711	487	3887	19877	12070	91778	4486
NG75	5615		19274	3786	345	1523	7750	4902	37172	2203
NA50	14889		-	11603	493	3273	19042	11896	-	3254
NA75	5446		-	4796	350	581	6955	4710	-	-
NGA50	14804		46100	10554	484	3834	19503	11927	88072	4435
NGA75	5353		17091	3617	341	1423	7488	4754	34640	2089
LG50	2148		648	3006	83339	8702	1665	2803	342	8280
LG75	5321		1617	7644	157038	20717	4058	6626	849	17623
LA50	2178		-	2720	80621	10663	1741	2844	-	12372
LA75	5402		-	6458	151570	31936	4334	6787	-	-
LGA50	2190		691	3065	83683	8777	1693	2831	358	8332
LGA75	5450		1759	7829	158064	21213	4139	6737	894	17923
Misassemblies										
# misassemblies	106		854	26	10	6	109	48	189	106
# relocations	42		243	9	2	0	42	21	70	65
# translocations	53		603	17	7	6	66	27	116	41
# inversions	11		8	0	1	0	1	0	3	0
# misassembled contigs	103		788	25	10	6	103	46	173	102
Misassembled contigs length	583896		13255971	8201	3147	20460	83601	41609	7726132	400206
# local misassemblies	23		258	6	3	7	50	15	85	72
# scaffold gap ext. mis.	0		0	0	0	0	0	0	0	0
# scaffold gap loc. mis.	0		0	0	0	0	0	0	0	0
# unaligned mis. contigs	30		52	12	5	1	47	28	38	10
Unaligned										
# fully unaligned contigs	9596		208158	2015	7066	81936	8140	15435	239172	107203
Fully unaligned length	5192711		121335023	1914284	2447441	28385413	5360850	6747238	147056431	39233939
# partially unaligned contigs	35		177	19	5	3	68	30	288	20
Partially unaligned length	48510		330988	20282	4174	4174	82145	36783	357368	24630
Mismatches										
# mismatches	17157		76153	7258	2411	3002	14742	8857	22386	13412
# indels	5547		11951	4098	1666	3951	5077	4280	7403	5657
Indels length	8997		53960	5607	2128	6452	7980	5768	23579	13671
# mismatches per 100 kbp	15.07		66.1	6.6	2.38	2.86	12.78	7.83	19.49	12.11
# indels per 100 kbp	4.87		10.37	3.72	1.64	3.77	4.4	3.78	6.45	5.11
# indels (<= 5 bp)	5411		10639	4033	1656	3864	4965	4233	6837	5401
# indels (> 5 bp)	136		1312	65	10	87	112	47	566	256
# N's	0		0	0	0	0	0	0	13	0
# N's per 100 kbp	0		0	0	0	0	0	0	0	0
Statistics without reference										
# contigs	40253		224076	32617	249705	127066	35438	43665	248486	152488
# contigs (>= 0 bp)	57267		228980	311472	1418620	231460	754044	511100	407580	152488
# contigs (>= 1000 bp)	12490		19525	14583	9413	26081	11244	14364	32049	27361
# contigs (>= 5000 bp)	5701		3555	6262	0	5876	5304	6540	2549	6920
# contigs (>= 10000 bp)	3412		2379	3262	0	1182	3338	3536	1725	1106
# contigs (>= 25000 bp)	989		1326	613	0	13	1172	699	1108	1
# contigs (>= 50000 bp)	140		653	63	0	0	262	55	673	0
Largest contig	178991		643625	635011	3315	31479	347434	105557	731281	26666
Total length	119997086		238907422	112701156	117346509	133724486	122218175	120647466	262758271	151134178
Total length (>= 0 bp)	123471608		240010301	137436278	284865355	154001638	203821058	165403051	282111111	151134178
Total length (>= 1000 bp)	108262050		135209380	104449522	11667358	96923001	111714069	108859258	158775459	105060267
Total length (>= 5000 bp)	92273086		109299251	84226871	0	47735119	98043456	89829764	113724808	53744388
Total length (>= 10000 bp)	75789942		101055156	62823889	0	15663454	83915903	68256314	108014997	13936005
Total length (>= 25000 bp)	37778697		83919084	22203231	0	365318	49218349	24672850	98129990	26666
Total length (>= 50000 bp)	9226212		60423816	4386585	0	0	18157798	3389167	82695776	0
N50	15102		1884	11798	496	3337	19342	12043	1656	3324
N75	5703		562	4943	354	765	7237	4862	613	553
L50	2137		6935	2667	80296	10557	1713	2816	9933	12259
L75	5276		81093	6323	150625	29935	4246	6674	81730	38880
GC (%)	35.79		32.62	35.84	35.77	34.49	35.72	35.64	32.25	33.98

GRASShopper did not finish computations because of too large memory requirements.

Table S4: CHLORELLA SOROKINIANA

REPORT FROM QUAST FOR ASSEMBLY RESULTS

Genome statistics	ALGA	GRASShopper	MEGAHIT	Platanus	Readjoinder	SAGE2	SGA	SOAPdenovo2	SPAdes	Velvet
Genome fraction (%)	96.038	93.242	96.761	96.181	50.267	13.913	95.181	74.077	96.601	25.189
Duplication ratio	1.005	1.005	1.062	1.011	1.467	1.037	1.012	1.024	1.006	1.011
Largest alignment	117281	87200	142903	117347	1004	2573	67064	6617	220875	2192
Total aligned length	55533431	53854122	57523451	55993828	42279820	8124274	55393979	43621372	55733343	14589415
NG50	16566	14109	22674	10263	502	-	10108	841	41564	-
NG75	8470	6615	11809	5043	-	-	5002	-	21185	-
NA50	17293	15420	21338	10713	503	641	10650	1058	42121	679
NA75	9517	8417	10494	5684	500	519	5729	747	22260	566
NGA50	16529	14066	21889	10250	501	-	10053	837	40286	-
NGA75	8370	6517	11189	5040	-	-	4907	-	19988	-
LG50	1061	1207	774	1717	56945	-	1766	21944	412	-
LG75	2306	2680	1656	3738	-	-	3804	-	908	-
LA50	978	1054	827	1592	40963	4855	1620	14062	395	8753
LA75	2070	2229	1813	3373	62158	8483	3386	26424	855	14817
LGA50	1066	1210	793	1718	57082	-	1769	22014	424	-
LGA75	2319	2695	1712	3740	-	-	3822	-	941	-
Misassemblies										
# misassemblies	227	159	1644	197	17846	3868	218	47	324	66
# relocations	25	10	211	21	2098	439	19	6	53	13
# translocations	201	145	1215	174	15737	3426	199	41	269	53
# inversions	1	4	218	2	11	3	0	0	2	0
# misassembled contigs	192	154	1455	172	17818	2998	199	46	288	63
Misassembled contigs length	1115863	827193	5528596	175974	8974226	2438874	805757	63941	5997919	48077
# local misassemblies	172	113	1117	44	237	233	61	8	292	112
# scaffold gap ext. mis.	0	0	0	0	0	0	0	0	0	0
# scaffold gap loc. mis.	0	0	0	0	0	0	0	0	0	0
# unaligned mis. contigs	5	3	818	1	4	3	2	2	5	32
Unaligned										
# fully unaligned contigs	76	60	1160	16	68	34	93	90	177	207
Fully unaligned length	59940	54283	690910	14138	36343	21305	70441	61936	130737	302329
# partially unaligned contigs	48	60	318	9	1	6	63	23	158	34
Partially unaligned length	41910	48338	246363	8975	511	3870	53866	16755	164992	40617
Mismatches										
# mismatches	22234	1866	63080	2847	115591	18202	5187	15888	3122	5402
# indels	1224	451	3435	313	17764	5002	295	434	1028	1232
Indels length	12112	9269	55475	3363	33316	10957	2892	1251	31047	6508
# mismatches per 100 kbp	40.19	3.47	113.17	5.14	399.17	227.1	9.46	37.23	5.61	37.23
# indels per 100 kbp	2.21	0.84	6.16	0.56	61.34	62.41	0.54	1.02	1.85	8.49
# indels (<= 5 bp)	1024	312	2378	208	17438	4893	211	414	458	1075
# indels (> 5 bp)	200	139	1057	105	326	109	84	20	570	157
# N's	0	0	0	0	0	0	0	0	7	0
# N's per 100 kbp	0	0	0	0	0	0	0	0	0.01	0
Statistics without reference										
# contigs	6005	6776	10949	9328	83239	11228	9696	44157	2809	21356
# contigs (>= 0 bp)	6771	33616	66942	15355	1487787	52184	359727	981869	6209	80485
# contigs (>= 1000 bp)	5233	5559	4064	7967	1	1507	7661	15864	2349	1536
# contigs (>= 5000 bp)	3179	3229	2719	3763	0	0	3806	3	1776	15
# contigs (>= 10000 bp)	1977	1910	1885	1782	0	0	1790	0	1440	1
# contigs (>= 25000 bp)	473	412	649	193	0	0	187	0	767	0
# contigs (>= 50000 bp)	63	36	132	10	0	0	3	0	314	0
Largest contig	117283	87200	143048	117347	1004	2699	67064	6617	221020	12515
Total length	55710250	54075102	60152684	56064282	42513120	8336452	55598660	43772315	56267820	15008742
Total length (>= 0 bp)	55971947	59129338	80072816	57805725	452219063	20755301	149957336	218045116	57070487	36045104
Total length (>= 1000 bp)	55160341	53266876	56102500	55070004	1004	1810759	54245155	23781774	55962497	1944579
Total length (>= 5000 bp)	49750326	47189609	52633346	44145328	0	0	44026993	17115	54492168	114397
Total length (>= 10000 bp)	40985508	37709211	46516814	30011025	0	0	29593144	0	52018964	12515
Total length (>= 25000 bp)	17583823	14760405	26389123	6349065	0	0	5857416	0	40778832	0
Total length (>= 50000 bp)	3991182	2122477	8815108	693372	0	0	177887	0	24899338	0
N50	17318	15543	22058	10715	503	745	10667	1062	43837	693
N75	9600	8474	11029	5687	501	598	5781	751	23386	580
L50	973	1051	807	1591	40852	4260	1617	14025	384	8417
L75	2059	2218	1752	3371	62009	7401	3373	26332	826	14349
GC (%)	63.73	63.87	63.71	63.8	64.05	64.71	63.86	63.91	63.79	64.07

Table S5: ESCHERICHIA COLI

REPORT FROM QUAST FOR ASSEMBLY RESULTS

Genome statistics	ALGA	GRASShopper	MEGAHIT	Platanus	Readjoiner	SAGE2	SGA	SOAPdenovo2	SPAdes	Velvet
Genome fraction (%)	98.128	53.26	98.458	97.459	3.336	93.944	97.793	97.706	98.427	91.905
Duplication ratio	1.001	1.193	1.001	1.01	1.008	1	1.001	1.001	1	1.041
Largest alignment	166847	3200	284869	58133	538	24062	174176	97152	224028	3832
Total aligned length	4558176	2935200	4574946	4569691	155555	4360778	4545297	4539527	4570446	4438082
NG50	60754	291	124164	12333	-	3350	51987	20370	133230	751
NG75	35258	-	64626	6257	-	1744	30906	12246	86559	489
NA50	60754	371	124164	12541	270	3619	52586	20760	133230	776
NA75	36049	298	64427	6525	259	2019	31674	12689	87059	522
NGA50	60754	288	124164	12333	-	3350	51987	20370	133230	750
NGA75	35258	-	64427	6257	-	1744	30906	12246	86559	489
LG50	24	5371	13	104	-	414	28	68	13	2060
LG75	49	-	25	233	-	889	58	140	24	3964
LA50	24	2710	13	101	261	374	27	65	13	1929
LA75	47	5025	25	225	409	776	55	134	23	3668
LGA50	24	5398	13	104	-	414	28	68	13	2060
LGA75	49	-	25	233	-	889	58	140	24	3966
Misassemblies										
# misassemblies	0	7	1	1	1	0	0	0	1	1
# relocations	0	5	1	1	1	0	0	0	1	1
# translocations	0	0	0	0	0	0	0	0	0	0
# inversions	0	2	0	0	0	0	0	0	0	0
# misassembled contigs	0	7	1	1	1	0	0	0	1	1
Misassembled contigs length	0	2577	2910	1811	297	0	0	0	107859	641
# local misassemblies	3	11	17	0	0	0	0	0	2	3
# scaffold gap ext. mis.	0	0	0	0	0	0	0	0	0	0
# scaffold gap loc. mis.	0	0	0	0	0	0	0	0	0	0
# unaligned mis. contigs	0	0	0	0	0	0	0	0	0	0
Unaligned										
# fully unaligned contigs	0	0	0	0	0	0	0	0	5	2
Fully unaligned length	0	0	0	0	0	0	0	0	1608	1123
# partially unaligned contigs	0	0	0	0	0	0	0	0	0	1
Partially unaligned length	0	0	0	0	0	0	0	0	0	530
Mismatches										
# mismatches	340	3341	157	2	3	4	42	15	62	782
# indels	2	1300	17	1	1	0	1	1	3	40
Indels length	2	2884	319	2	1	0	2	2	4	126
# mismatches per 100 kbp	7.46	135.15	3.44	0.04	1.94	0.09	0.93	0.33	1.36	18.33
# indels per 100 kbp	0.04	52.59	0.37	0.02	0.65	0	0.02	0.02	0.07	0.94
# indels (<= 5 bp)	2	1246	14	1	1	0	1	1	3	39
# indels (> 5 bp)	0	54	3	0	0	0	0	0	0	1
# N's	0	0	0	0	0	0	0	0	0	0
# N's per 100 kbp	0	0	0	0	0	0	0	0	0	0
Statistics without reference										
# contigs	194	7715	119	808	561	1863	218	388	94	6606
# contigs (>= 0 bp)	241	52293	133	1987	159220	3577	2533	2615	175	6606
# contigs (>= 1000 bp)	145	155	87	560	0	1273	160	328	67	1044
# contigs (>= 5000 bp)	103	0	59	271	0	186	117	235	51	0
# contigs (>= 10000 bp)	90	0	55	145	0	17	101	158	47	0
# contigs (>= 25000 bp)	62	0	47	31	0	0	65	53	44	0
# contigs (>= 50000 bp)	34	0	29	3	0	0	29	8	32	0
Largest contig	166847	3200	285131	58133	538	24062	174176	97152	224028	3832
Total length	4558417	2948969	4576125	4569796	156095	4360798	4545481	4539731	4572315	4440582
Total length (>= 0 bp)	4567042	8640784	4579403	4717384	20517020	4552367	4858776	4757989	4583360	4440582
Total length (>= 1000 bp)	4536286	228049	4560973	4454044	0	4003167	4522922	4512499	4560245	1444347
Total length (>= 5000 bp)	4441637	0	4492803	3697112	0	1380809	4419156	4268349	4523388	0
Total length (>= 10000 bp)	4336567	0	4462929	2780857	0	222980	4293954	3681996	4493604	0
Total length (>= 25000 bp)	3890167	0	4345203	1110586	0	0	3702484	2000445	4441633	0
Total length (>= 50000 bp)	2883228	0	3724522	162627	0	0	2382690	537084	4016107	0
N50	60754	374	124164	12541	270	3619	52586	20760	133230	776
N75	36049	300	64626	6525	259	2019	31674	12689	87059	524
L50	24	2701	13	101	260	374	27	65	13	1928
L75	47	5002	25	225	408	776	55	134	23	3666
GC (%)	50.74	48.91	50.74	50.74	51.87	50.71	50.72	50.72	50.74	50.77

Table S6: MICROTHRIX PARVICELLA

REPORT FROM QUAST FOR ASSEMBLY RESULTS

Genome statistics	ALGA	GRASShopper	MEGAHIT	Platanus	Readjoinder	SAGE2	SGA	SOAPdenovo2	SPAdes	Velvet
Genome fraction (%)	98.733	98.777	98.768	98.632	72.471	97.834	98.764	98.491	98.967	98.269
Duplication ratio	1.001	1.007	1.001	1.01	1.107	1.001	1.002	1.001	1.001	1.006
Largest alignment	209112	127818	408073	78680	1662	104005	295212	156108	740450	32170
Total aligned length	4152364	4176105	4152278	4184941	3369597	4114581	4157976	4143440	4161687	4154946
NG50	80267	32094	160873	9301	332	18362	69463	49036	156137	7980
NG75	44053	17844	105714	5263	262	8851	40405	29040	104284	4112
NA50	87496	32079	153944	9305	368	18743	69463	49560	150746	8083
NA75	44059	17526	92835	5300	300	9281	40459	30583	73687	4201
NGA50	80267	32079	150423	9301	332	18362	69463	49036	151220	7980
NGA75	44053	17081	92835	5263	261	8851	40405	29040	88932	4112
LG50	18	37	9	136	4575	71	19	27	9	163
LG75	35	81	17	283	8156	153	40	54	17	341
LA50	17	37	10	135	3386	69	19	26	10	160
LA75	34	82	19	281	5933	147	39	52	20	334
LGA50	18	37	11	136	4578	71	19	27	9	163
LGA75	35	83	19	283	8162	154	40	54	18	341
Misassemblies										
# misassemblies	0	9	8	2	0	1	2	1	4	3
# relocations	0	2	2	0	0	0	0	0	0	1
# translocations	0	7	6	2	0	1	2	1	4	2
# inversions	0	0	0	0	0	0	0	0	0	0
# misassembled contigs	0	9	7	2	0	1	2	1	4	3
Misassembled contigs length	0	270985	812186	26171	0	42886	36426	23430	368501	4801
# local misassemblies	2	8	11	0	0	0	0	0	1	0
# scaffold gap ext. mis.	0	0	0	0	0	0	0	0	0	0
# scaffold gap loc. mis.	0	0	0	0	0	0	0	0	0	0
# unaligned mis. contigs	0	0	0	0	0	0	0	0	0	0
Unaligned										
# fully unaligned contigs	0	5	81	0	0	0	0	1	730	14
Fully unaligned length	0	2021	34613	0	0	0	0	257	233698	4876
# partially unaligned contigs	0	0	0	0	0	0	0	0	0	0
Partially unaligned length	0	0	0	0	0	0	0	0	0	0
Mismatches										
# mismatches	101	359	278	20	23	18	431	33	62	329
# indels	12	75	82	1	6	2	2	3	8	18
Indels length	156	523	1976	2	6	68	3	70	334	142
# mismatches per 100 kbp	2.43	8.65	6.7	0.48	0.76	0.44	10.38	0.8	1.49	7.97
# indels per 100 kbp	0.29	1.81	1.98	0.02	0.2	0.05	0.05	0.07	0.19	0.44
# indels (<= 5 bp)	3	63	16	1	6	1	2	2	2	8
# indels (> 5 bp)	9	12	66	0	0	1	0	1	6	10
# N's	0	0	0	0	0	0	0	0	19	0
# N's per 100 kbp	0	0	0	0	0	0	0	0	0.43	0
Statistics without reference										
# contigs	174	353	153	764	9015	480	176	201	809	927
# contigs (>= 0 bp)	202	462	158	1426	37923	671	994	1434	1300	927
# contigs (>= 1000 bp)	133	236	61	624	33	385	132	161	64	730
# contigs (>= 5000 bp)	89	161	46	296	0	220	94	116	49	290
# contigs (>= 10000 bp)	69	111	40	112	0	135	76	91	44	115
# contigs (>= 25000 bp)	44	52	29	13	0	39	52	57	30	2
# contigs (>= 50000 bp)	29	17	23	1	0	3	29	25	23	0
Largest contig	209112	138392	408073	78680	1662	104005	295212	156108	740450	32170
Total length	4152931	4181438	4190056	4185158	3371541	4114795	4158186	4144018	4395905	4160128
Total length (>= 0 bp)	4158288	4199749	4191117	4255877	7796775	4138567	4263726	4255657	4495775	4160128
Total length (>= 1000 bp)	4133413	4131448	4154833	4104486	39030	4060481	4139201	4125032	4160739	4048842
Total length (>= 5000 bp)	4038154	3939942	4123843	3222573	0	3625586	4059303	4020902	4130144	2921428
Total length (>= 10000 bp)	3890025	3576427	4079954	1880361	0	2984399	3928341	3830466	4096247	1677011
Total length (>= 25000 bp)	3502288	2550857	3907613	442483	0	1432183	3544422	3246701	3877039	58068
Total length (>= 50000 bp)	2901848	1356310	3664257	78680	0	220427	2692240	2032528	3608506	0
N50	87496	32450	160873	9305	369	19052	69463	49560	151220	8083
N75	44064	17850	105714	5304	300	9281	40459	30583	82635	4201
L50	17	36	9	135	3384	68	19	26	10	160
L75	34	80	17	280	5930	146	39	52	19	334
GC (%)	66.42	66.4	66.31	66.46	66.29	66.43	66.43	66.42	65.87	66.43

Table S7: HOMO SAPIENS CHROMOSOME 14

BUSCO MEASURE FOR ASSEMBLY RESULTS

	Dataset	Complete BUSCOs	Complete and single-copy BUSCOs	Complete and duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCO groups searched
reference genome	primates_odb10	435	435	0	16	13329	13780
ALGA	primates_odb10	141	141	0	56	13583	13780
GRASShopPER	primates_odb10	43	43	0	22	13715	13780
Platanus	primates_odb10	40	40	0	22	13718	13780
SAGE2	primates_odb10	65	65	0	41	13674	13780
SGA	primates_odb10	109	109	0	56	13615	13780
SOAPdenovo2	primates_odb10	115	115	0	49	13616	13780
Velvet	primates_odb10	86	86	0	46	13648	13780

Table S8: CAENORHABDITIS ELEGANS

	Dataset	Complete BUSCOs	Complete and single-copy BUSCOs	Complete and duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCO groups searched
reference genome	nematoda_odb10	3113	3096	17	3	15	3131
ALGA	nematoda_odb10	2820	2808	12	174	137	3131
GRASShopPER	nematoda_odb10	2678	2667	11	239	214	3131
Platanus	nematoda_odb10	2751	2740	11	203	177	3131
SAGE2	nematoda_odb10	2569	2547	22	250	312	3131
SGA	nematoda_odb10	2825	2812	13	179	127	3131
SOAPdenovo2	nematoda_odb10	2751	2737	14	206	174	3131
Velvet	nematoda_odb10	2466	2446	20	324	341	3131

Table S9: ARABIDOPSIS THALIANA

	Dataset	Complete BUSCOs	Complete and single-copy BUSCOs	Complete and duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCO groups searched
reference genome	brassicales_odb10	4533	4495	38	6	57	4596
ALGA	brassicales_odb10	4388	4336	52	74	134	4596
GRASShopPER	brassicales_odb10						4596
Platanus	brassicales_odb10	4339	4301	38	108	149	4596
SAGE2	brassicales_odb10	4428	4382	46	68	100	4596
SGA	brassicales_odb10	4451	4413	38	45	100	4596
SOAPdenovo2	brassicales_odb10	4331	4288	43	119	146	4596
Velvet	brassicales_odb10	3695	3654	41	417	484	4596

Table S10: CHLORELLA SOROKINIANA

	Dataset	Complete BUSCOs	Complete and single-copy BUSCOs	Complete and duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCO groups searched
reference genome	chlorophyta_odb10	1488	1475	13	9	22	1519
ALGA	chlorophyta_odb10	1429	1421	8	52	38	1519
GRASShopPER	chlorophyta_odb10	1388	1379	9	76	55	1519
Platanus	chlorophyta_odb10	1374	1364	10	85	60	1519
SAGE2	chlorophyta_odb10	87	86	1	162	1270	1519
SGA	chlorophyta_odb10	1298	1290	8	125	96	1519
SOAPdenovo2	chlorophyta_odb10	301	297	4	390	828	1519
Velvet	chlorophyta_odb10	32	32	0	188	1299	1519

Table S11: ESCHERICHIA COLI

	Dataset	Complete BUSCOs	Complete and single-copy BUSCOs	Complete and duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCO groups searched
reference genome	enterobacterales_odb10	440	438	2	0	0	440
ALGA	enterobacterales_odb10	439	437	2	1	0	440
GRASShopPER	enterobacterales_odb10	27	24	3	108	305	440
Platanus	enterobacterales_odb10	416	414	2	17	7	440
SAGE2	enterobacterales_odb10	137	137	0	150	153	440
SGA	enterobacterales_odb10	440	438	2	0	0	440
SOAPdenovo2	enterobacterales_odb10	434	433	1	6	0	440
Velvet	enterobacterales_odb10	147	146	1	179	114	440

Table S12: MICROTHRIX PARVICELLA

	Dataset	Complete BUSCOs	Complete and single-copy BUSCOs	Complete and duplicated BUSCOs	Fragmented BUSCOs	Missing BUSCOs	Total BUSCO groups searched
reference genome	bacteria_odb10	116	116	0	5	3	124
ALGA	bacteria_odb10	116	116	0	5	3	124
GRASShopPER	bacteria_odb10	115	115	0	6	3	124
Platanus	bacteria_odb10	108	108	0	12	4	124
SAGE2	bacteria_odb10	20	19	1	63	41	124
SGA	bacteria_odb10	116	116	0	5	3	124
SOAPdenovo2	bacteria_odb10	113	113	0	8	3	124
Velvet	bacteria_odb10	108	108	0	13	3	124

TIME AND MEMORY CONSUMPTION

Table S13: MEMORY PEAK [GB]

	H. sapiens chr. 14	C. elegans	A. thaliana	C. sorokiniana	E. coli	M. parvicella	H. sapiens
ALGA	11.2	19.3	19.9	27.9	7.8	1.8	247.3
GRASShopPER	85.4	361.6		638.9	192.0	17.6	
MEGAHIT	3.0	6.1	8.0	18.9	3.7	0.8	87.6
Platanus	67.3	122.1	106.1	120.0	117.6	117.6	
Readjoiner	1.1	2.3	3.0	6.4	1.7	0.3	
SAGE2	13.4	21.7	28.1	18.3	4.6	1.5	
SGA	1.7	3.3	4.2	7.7	1.8	0.3	43.5
SOAPdenovo2	7.1	7.3	25.5	16.3	1.8	2.5	269.3
SPAdes	15.9	14.4	*	49.2	10.7	10.6	
Velvet	29.9	21.0	127.6	107.6	10.1	9.3	

Table S14: ELAPSED (WALL CLOCK) TIME [hh:mm:ss]

	H. sapiens chr. 14	C. elegans	A. thaliana	C. sorokiniana	E. coli	M. parvicella	H. sapiens
ALGA	00:11:09	00:24:48	00:31:53	00:48:11	00:04:06	00:01:29	15:31:50
GRASShopPER	06:31:44	57:12:58		53:33:33	15:29:26	02:02:28	
MEGAHIT	00:16:50	00:31:59	00:59:18	04:30:26	00:10:16	00:02:38	15:43:34
Platanus	00:13:47	01:03:06	01:41:31	03:50:25	00:37:55	00:16:39	
Readjoiner	00:07:26	00:07:29	00:12:02	00:52:28	00:04:55	00:00:41	
SAGE2	03:27:03	04:46:44	04:01:31	02:34:26	00:18:14	00:04:03	
SGA	01:18:11	02:33:15	03:52:03	09:54:32	01:12:16	00:11:47	98:58:49
SOAPdenovo2	00:15:11	00:27:02	00:37:58	01:21:05	00:09:05	00:02:13	15:46:12
SPAdes	11:11:27	12:14:42	*	22:22:05	01:10:39	00:26:50	
Velvet	01:18:53	02:05:00	04:09:57	14:42:04	00:44:50	00:08:52	

**) SPAdes finished computations, but there was a technical problem with the measurement.*

Table S15: MUSKET CORRECTION

	H. sapiens chr. 14	C. elegans	A. thaliana	C. sorokiniana	E. coli	M. parvicella	H. sapiens
Memory peak [GB]	3.1	1.7	7.2	12.2	0.5	0.5	72.3
Time [hh:mm:ss]	00:17:02	00:18:58	00:40:50	01:28:05	00:11:52	00:05:38	07:15:14

Table S16: HOMO SAPIENS WHOLE GENOME

REPORT FROM QUAST

Genome statistics	ALGA	MEGAHIT	SGA	SOAPdenovo2
Genome fraction (%)	90.297	91.707	90.538	85.723
Duplication ratio	1.009	1.04	1.108	1.018
Largest alignment	140579	453369	67917	47254
Total aligned length	2775879684	2892744943	3055094153	2658682524
NG50	11495	41177	4481	2495
NG75	3686	14935	1468	687
NA50	13834	39249	4753	3264
NA75	6648	18170	1813	1544
NGA50	11453	35275	4471	2490
NGA75	3648	12734	1452	677
LG50	74181	21702	186133	318442
LG75	191808	52682	486999	903374
LA50	57752	21670	172243	223748
LA75	129950	48872	428400	520190
LGA50	74402	25104	186503	318873
LGA75	192694	61377	489004	907520
Misassemblies				
# misassemblies	2230	30456	3688	739
# relocations	1161	5815	1747	397
# translocations	1034	23354	1863	299
# inversions	35	1287	78	43
# misassembled contigs	2090	27715	3560	705
Misassembled contigs length	13097797	705421771	6953129	1447334
# local misassemblies	4209	15849	6296	2026
# scaffold gap ext. mis.	0	0	0	0
# scaffold gap loc. mis.	0	0	0	0
# unaligned mis. contigs	1189	2398	1079	339
Unaligned				
# fully unaligned contigs	17690	115368	44272	10928
Fully unaligned length	8777301	40941253	15838074	5297029
# partially unaligned contigs	2031	2067	1684	752
Partially unaligned length	3193085	3325133	2313312	939636
Mismatches				
# mismatches	2959935	5229732	3441365	2349688
# indels	569648	859258	802806	365301
Indels length	2228894	4226723	2889390	1217634
# mismatches per 100 kbp	107.5	187.01	124.65	89.89
# indels per 100 kbp	20.69	30.73	29.08	13.97
# indels (<= 5 bp)	483698	716539	692550	320088
# indels (> 5 bp)	85950	142719	110256	45213
# N's	0	0	0	0
# N's per 100 kbp	0	0	0	0
Statistics without reference				
# contigs	582040	728920	2111636	1559273
# contigs (>= 0 bp)	647131	1050311	8029560	12376038
# contigs (>= 1000 bp)	313815	114854	588579	724792
# contigs (>= 5000 bp)	160051	78635	161134	110106
# contigs (>= 10000 bp)	87580	63511	46655	20157
# contigs (>= 25000 bp)	19100	37014	2473	358
# contigs (>= 50000 bp)	2048	16185	33	0
Largest contig	140579	453909	67917	47254
Total length	2790359536	2953804967	3077843557	2667281219
Total length (>= 0 bp)	2803904187	3025191016	4050122556	3654318968
Total length (>= 1000 bp)	2659625172	2734266794	2530880678	2258048882
Total length (>= 5000 bp)	2269990638	2651336854	1486353131	878393530
Total length (>= 10000 bp)	1748347814	2540616224	693958318	271525858
Total length (>= 25000 bp)	689782714	2099381482	76228873	10337952
Total length (>= 50000 bp)	126578797	1354606311	1874179	0
N50	13881	45488	4764	3269
N75	6683	21314	1827	1551
L50	57586	18751	171908	223496
L75	129511	42025	426954	519141
GC (%)	40.85	40.95	40.82	40.63

Table S17: PERFORMANCE OF ALGA WITH PARAMETER "--error-rate=0.05"

	H. sapiens chr. 14	C. elegans	A. thaliana	C. sorokiniana	E. coli	M. parvicella
Genome fraction (%)	92.443	95.396	94.711	96.083	96.865	98.183
Duplication ratio	1.004	1.004	1.010	1.005	1.004	1.002
Largest alignment	75282	147600	141405	117363	70255	127169
NG50	4609	14768	14682	16448	21625	40321
Misassembled contigs	173376	349183	655323	1166215	9248	10128
Fully unaligned length	20966	10843	5584528	59742	0	0
Partially unaligned ler	10081	50351	43376	42066	0	0
# contigs	32466	19732	42628	6015	391	277
Memory peak [GB]	14.1	24.1	26.7	32.5	7.8	2.2
Time [hh:mm:ss]	00:14:53	00:51:21	00:40:13	01:33:20	00:04:35	00:01:38