

SUPPLEMENTARY MATERIAL

Supplementary methods

Benchmarks were done on a SMP server with 128 GB of memory and 32 AMD Opteron cores (8 quad-core processors). Open-MPI version 1.4.1 was utilized.

Ray is available and distributed as a source-code package, and can be downloaded on SourceForge.Net, a collaborative web site for open source software acting as a public repository for bioinformatics software.

Exact parameters for each assembler on each dataset are listed in the scripts directory inside Ray source-code distribution.

Availability: <http://denovoassembler.sf.net/>.

Ray license agreement (General Public License, version 3)

Availability: www.gnu.org/licenses/gpl.html.

TABLE S1. DATASETS DEFINED FOR COMPARISONS

<i>Dataset</i>	<i>Description</i>	<i>Reads accession key</i>
Simulations 1: <i>Streptococcus pneumoniae</i> R6 (nucore/NC_003098)		
sim1	Read length: 50, coverage: 50	—
sim2	Read length: 50, coverage: 50, error rate: 1%	—
sim3	Fragment length: 200, read length: 50, coverage: 50	—
Simulations 2: <i>Escherichia coli</i> K-12 MG1655 (nucore/NC_000913)		
sim4	Read length: 400, coverage: 50	—
sim5	Fragment length: 200, read length: 50, coverage: 50	—
sim6	Read length: 50, coverage: 50	—
Simulations 3: <i>Pseudomonas aeruginosa</i> PAO1 (nucore/NC_002516)		
sim7	Read length: 40, coverage: 50	—
sim8	Fragment length: 250, read length: 40, coverage: 50	—
Mixed set 1: <i>Escherichia coli</i> K-12 MG1655 (nucore/NC_000913)		
mix11	Illumina non-paired reads	sra/SRA001125
mix12	Illumina paired reads	sra/SRA001125
mix13	Roche/454 reads	sra/SRA001028
mix14	Roche/454+ Illumina paired reads	sra/SRA001125 ; sra/SRA001028
Mixed set 2: <i>Acinetobacter</i> sp. ADP1 (nucore/NC_005966)		
mix21	Illumina non-paired reads	sra/SRA003611
mix22	Roche/454 reads	sra/SRA003611
mix23	Roche/454+ Illumina paired reads	sra/SRA003611
Mixed set 3: <i>Cryptobacterium curtum</i> DSM 15641 (nucore/NC_013170)		
mix31	Illumina non-paired reads	sra/SRA008863
mix32	Roche/454 reads	sra/SRA008863
mix33	Roche/454+ Illumina paired reads	sra/SRA008863

TABLE S2. COMPARISONS ON SIMULATIONS

Assembler	Contigs	Bases	Average length	Genome					Running time	
				N50	Largest length	breadth of coverage	Misassembled contigs	Mismatches Indels		
sim1										
EULER-SR	261	1967594	7538	11621	61396	0.9419	6	68	123	7m22.779s
Velvet	280	1917129	6846	11279	44362	0.9437	1	23	8	2m15.931s
Ray	259	1954999	7548	11561	77867	0.9608	0	0	0	3m25.240s
ABySS	417	1898819	4553	7349	27222	0.9343	0	4	0	1m56.066s
sim2										
EULER-SR	267	1965104	7359	11477	61349	0.9413	6	79	237	11m15.383s
Velvet	290	1913682	6598	10302	42572	0.9423	2	27	11	2m40.792s
Ray	259	1939235	7487	11554	77853	0.9531	0	0	0	4m29.223s
ABySS	418	1898547	4541	7349	27222	0.9342	0	4	0	4m52.727s
sim3										
EULER-SR	235	1976831	8412	12383	61593	0.9458	13	69	187	9m59.464s
Velvet	113	1950222	17258	32111	123292	0.9565	30	382	140	2m15.371s
Ray	96	1964569	20464	36692	127906	0.9632	0	1	0	5m52.834s
ABySS	151	2019778	13376	22045	104182	0.9815	0	213	9	3m38.944s
sim4										
Newbler	75	4559811	60797	132967	326953	0.9833	0	0	2	16m56.911s
EULER-SR	136	5399988	39705	86536	280065	0.9668	58	106	903	19m16.247s
Velvet	693	4472973	6454	10495	53974	0.9638	0	4	0	3m42.483s
Ray	83	4595945	55372	122763	327771	0.9876	0	0	0	17m24.874s
ABySS	898	4450079	4955	8141	35440	0.9573	0	5	0	5m51.884s
sim5										
EULER-SR	248	4711107	18996	33519	131097	0.9728	23	84	1186	24m49.858s
Velvet	129	4527399	35096	68879	269606	0.9758	21	497	842	5m33.387s
Ray	110	4584885	41680	83042	269979	0.9841	0	0	0	12m40.040s
ABySS	147	4608010	31347	59697	174234	0.9880	0	289	8	7m39.983s
sim6										
EULER-SR	323	4713939	14594	27098	100573	0.9714	15	42	970	19m40.888s
Velvet	405	4504211	11121	18003	120868	0.9713	2	60	32	4m52.789s
Ray	317	4560682	14387	25607	120916	0.9780	0	136	0	9m39.849s
ABySS	898	4449989	4955	8141	35440	0.9573	0	11	0	4m34.519s
sim7										
EULER-SR	422	7066990	16746	35498	170764	0.9833	16	80	531	29m11.924s
Velvet	1264	6086321	4815	7720	49780	0.9716	5	136	93	8m10.649s
Ray	547	6143206	11230	22131	210111	0.9782	0	0	0	12m47.722s
ABySS	4057	4998428	1232	1383	14076	0.7933	0	17	0	6m5.597s
sim8										
EULER-SR	251	6305646	25122	55095	182061	0.9871	13	216	551	36m53.145s
Velvet	86	6189575	71971	289674	579752	0.9843	29	1079	2433	7m12.569s
Ray	122	6227165	51042	110638	346568	0.9919	0	0	0	17m3.064s
ABySS	152	6416064	42210	83309	368070	0.9954	0	455	11	11m35.355s

TABLE S3. COMPARISON ON MIXED DATASET 1

Assembler	Contigs	Bases	Genome						Running time	
			Average length	N50	Largest length	breadth of coverage	Misassembled contigs	Indels		
mix11										
EULER-SR	1937	4102206	2117	3001	19094	0.8656	12	1162	7360	135m2.904s
Velvet	528	4491661	8506	14236	60568	0.9672	26	136	100	28m55.106s
Ray	420	4522457	10767	16970	66556	0.9702	0	2	4	37m57.168s
ABySS	1026	4440595	4328	6898	34595	0.9550	0	5	4	23m3.949s
mix12										
EULER-SR	1761	4099614	2328	3429	19094	0.8708	13	1173	7355	143m27.647s
Velvet	83	4542631	54730	125611	311586	0.9657	28	456	895	31m51.199s
Ray	126	4591168	36437	72499	174569	0.9818	0	2	4	47m54.377s
ABySS	154	4661190	30267	56703	174288	0.9840	0	233	9	44m53.714s
mix13										
Newbler	874	4513335	5163	8771	42344	0.9731	9	64	247	29m53.841s
EULER-SR	1538	5056677	3287	4969	26316	0.9600	68	648	1199	7m49.404s
Velvet	2415	3942859	1632	2068	10286	0.8499	13	386	207	2m10.820s
Ray	1722	4737188	2750	3660	16250	0.9630	9	145	89	5m27.089s
ABySS	2662	4020201	1510	1899	12092	0.8630	0	63	41	3m45.703s
mix14										
EULER-SR	1813	3538992	1952	2632	15039	0.7598	598	1820	7077	157m46.099s
Velvet	754	4518284	5992	8787	29554	0.9730	26	267	250	35m18.761s
Ray	109	4579657	42015	87318	268385	0.9831	1	234	6	62m30.978s
ABySS	154	4667915	30311	56703	174286	0.9851	0	239	9	47m12.271s

TABLE S4. COMPARISON ON MIXED DATASET 2

Assembler	Contigs	Bases	Genome						Running time	
			Average length	N50	Largest length	breadth of coverage	Misassembled contigs	Indels		
mix21										
EULER-SR	1498	4706276	3141	4965	24129	0.9679	12	1115	210	87m18.839s
Velvet	430	4361792	10143	20490	85191	0.9773	10	381	15	16m3.742s
Ray	259	3677696	14199	25852	72730	0.9749	0	82	6	29m48.993s
ABySS	1105	4315007	3904	6485	30606	0.9697	0	172	6	28m59.212s
mix22										
Newbler	109	3547847	32549	61793	214173	0.9846	0	69	380	43m3.785s
EULER-SR	1715	6018217	3509	5511	29671	0.9722	26	1421	12480	18m3.905s
Velvet	1440	3356900	2331	3267	14690	0.9308	4	704	3166	3m54.495s
Ray	2482	3361598	1354	1554	7435	0.8571	1	58	305	5m22.559s
ABySS	2324	2991350	1287	1496	10227	0.8309	0	37	96	7m4.659s
mix23										
EULER-SR	2399	6932733	2889	4495	47460	0.9681	70	2149	8033	129m53.436s
Velvet	1165	4285023	3678	5820	21242	0.9672	0	753	3164	22m52.574s
Ray	91	3540404	38905	82891	215819	0.9804	1	7	1	36m27.635s
ABySS	1198	4294398	3584	5748	30606	0.9677	0	175	6	34m34.469s

TABLE S5. COMPARISON ON MIXED DATASET 3

<i>Assembler</i>	<i>Contigs</i>	<i>Bases</i>	<i>Average length</i>	<i>N50</i>	<i>Genome</i>			<i>Misassembled contigs</i>	<i>Mismatches</i>	<i>Indels</i>	<i>Running time</i>
					<i>length</i>	<i>length</i>	<i>of coverage</i>				
mix31											
EULER-SR	166	109935	662	651	1433	0.0608	1	741	620		79m50.304s
Velvet	208	1598265	7683	12698	54143	0.9890	11	133	60		13m2.283s
Ray	72	1606647	22314	36518	91303	0.9862	0	1	1		19m51.388s
ABySS	156	1596474	10233	19950	69217	0.9881	0	6	2		22m45.895s
mix32											
Newbler	30	1609423	53647	261125	477358	0.9904	0	0	8		21m24.064s
EULER-SR	153	3078576	20121	76729	258262	0.9861	3	104	430		13m55.961s
Velvet	58	1600395	27593	76702	191999	0.9914	4	26	176		3m29.796s
Ray	184	1573256	8550	14142	76631	0.9503	0	0	1		4m42.587s
ABySS	478	1566837	3277	5064	16718	0.9689	0	7	1		4m14.770s
mix33											
EULER-SR	45	27077	601	567	1383	0.0023	6	67	10		87m51.332s
Velvet	190	1593755	8388	13136	45625	0.9848	8	238	1701		16m37.093s
Ray	27	1602133	59338	116274	236544	0.9897	0	0	1		35m8.569s
ABySS	250	1590298	6361	10207	37072	0.9835	0	4	2		23m33.000s