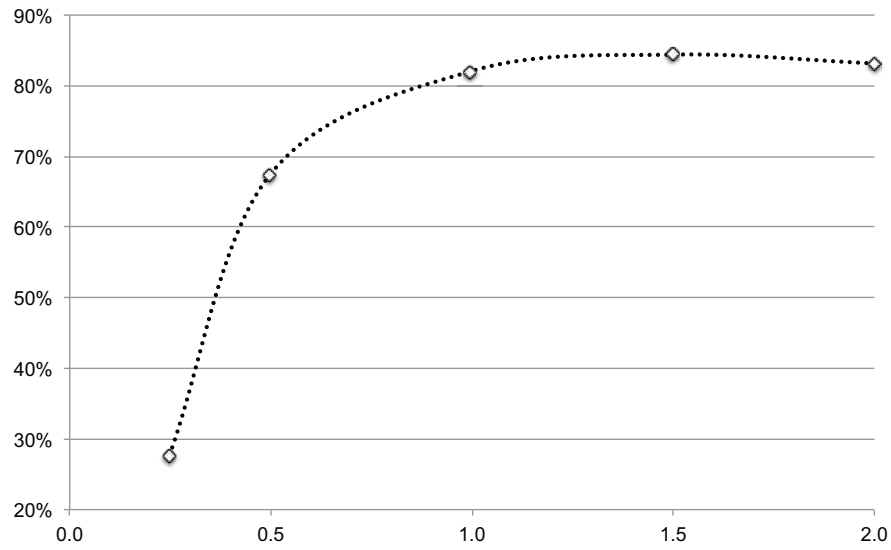
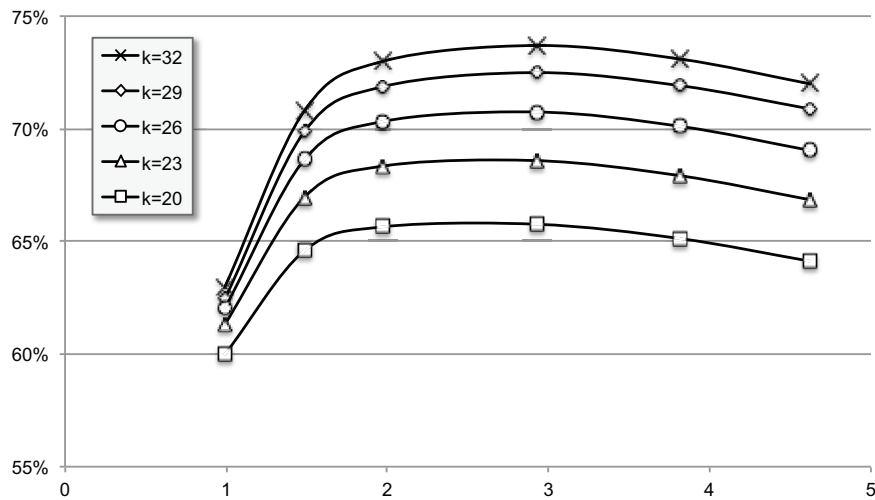


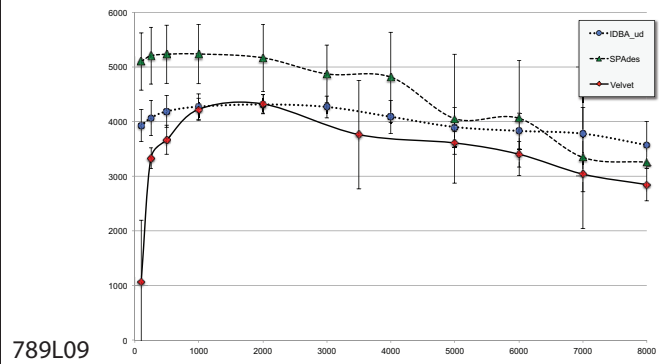
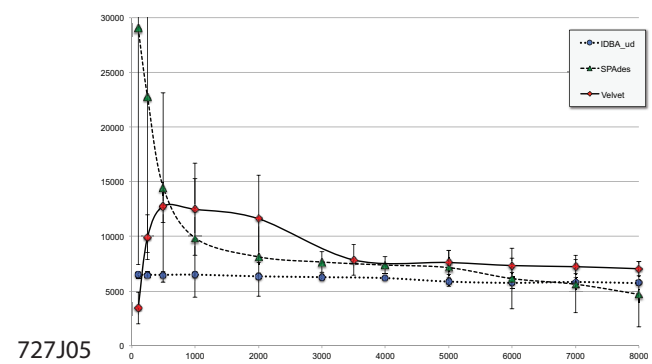
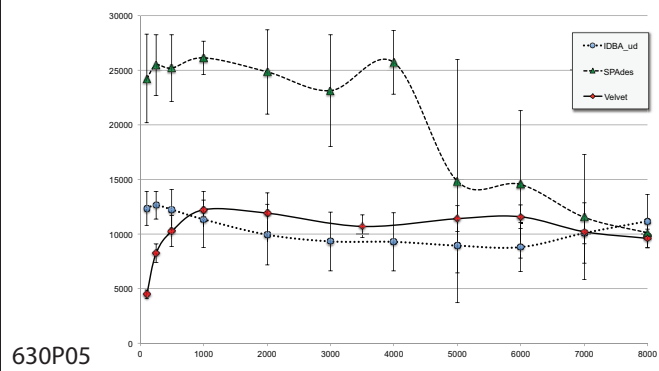
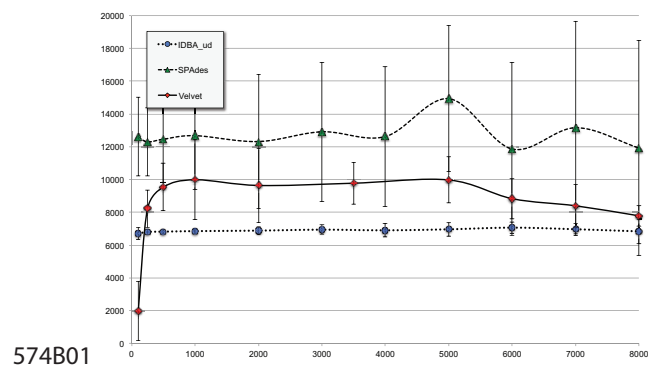
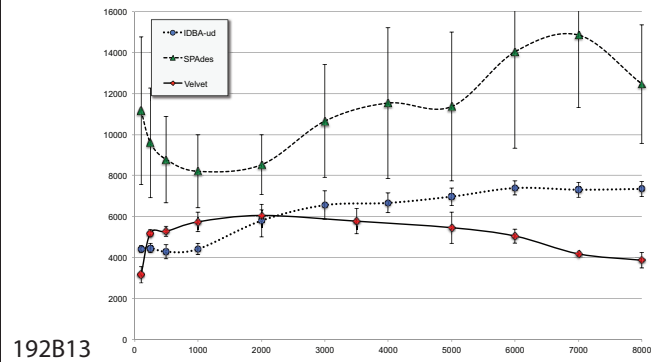
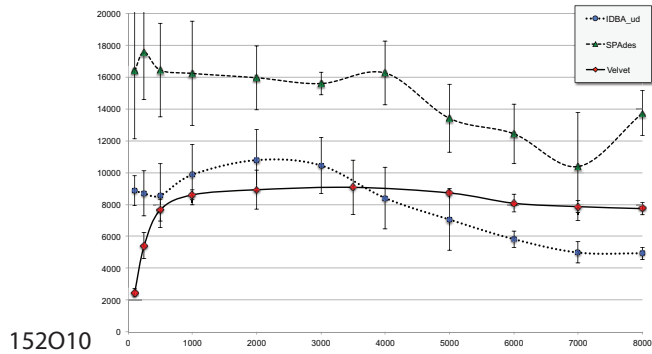
When Less is More: “Slicing” Sequencing Data Improves Read  
Decoding Accuracy and *De Novo* Assembly Quality  
**Supplemental Figures and Supplemental Tables**



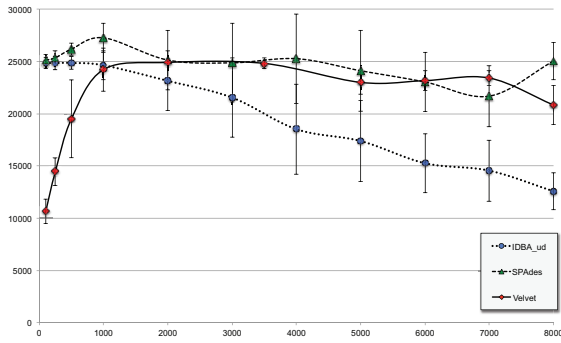
**Supplemental Figure S1:** The percentage of synthetic reads decoded by HASHFILTER on the rice genome as a function of the number of reads given in input ( $x$ : number of million of reads per pool)



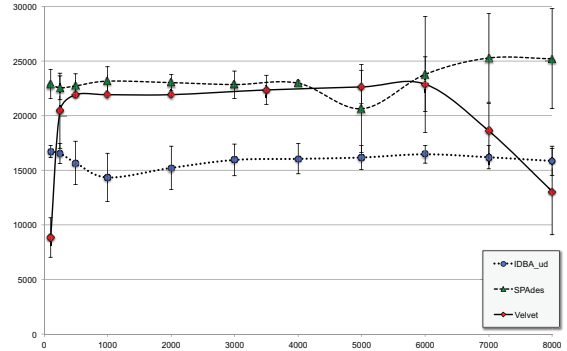
**Supplemental Figure S2:** The percentage of reads decoded by HASHFILTER on one of the barley datasets (Hv8) for several choices of the  $k$ -mer size ( $x$ : number of million of reads sampled per pool)



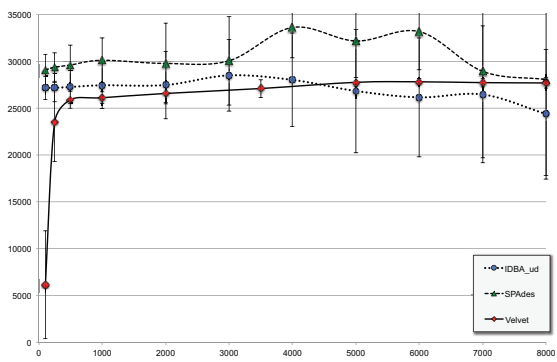
**Supplemental Figure S3:** n50 statistics (Y-axis) for the six ultra-deep coverage BACs, assembled with VELVET, SPADES and IDBA-UD for various levels of depth of sequencing (X-axis)



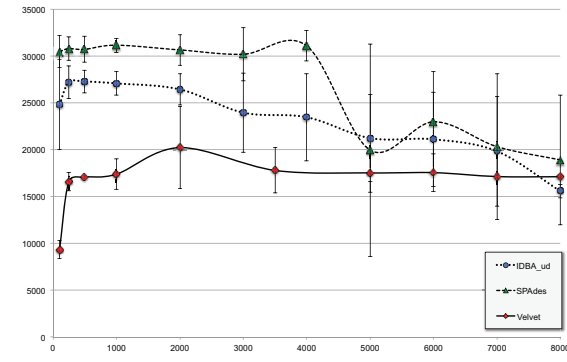
152O10



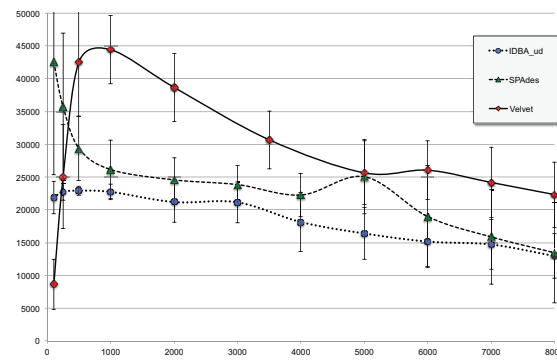
192B13



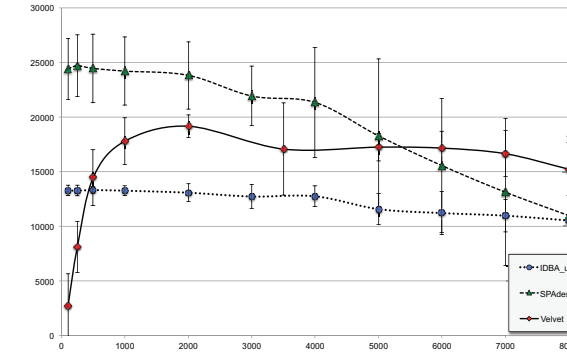
574B01



630P05

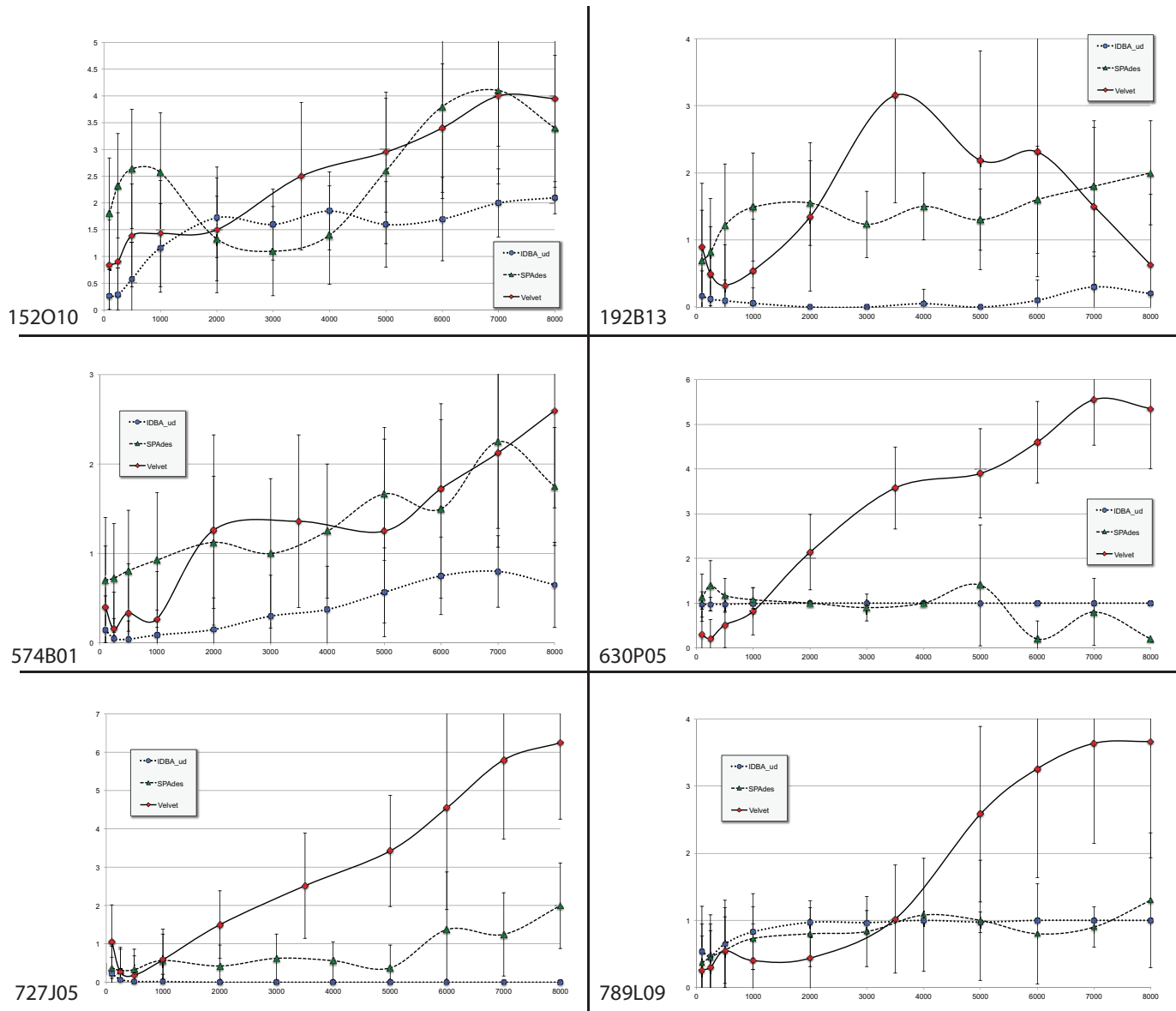


727J05

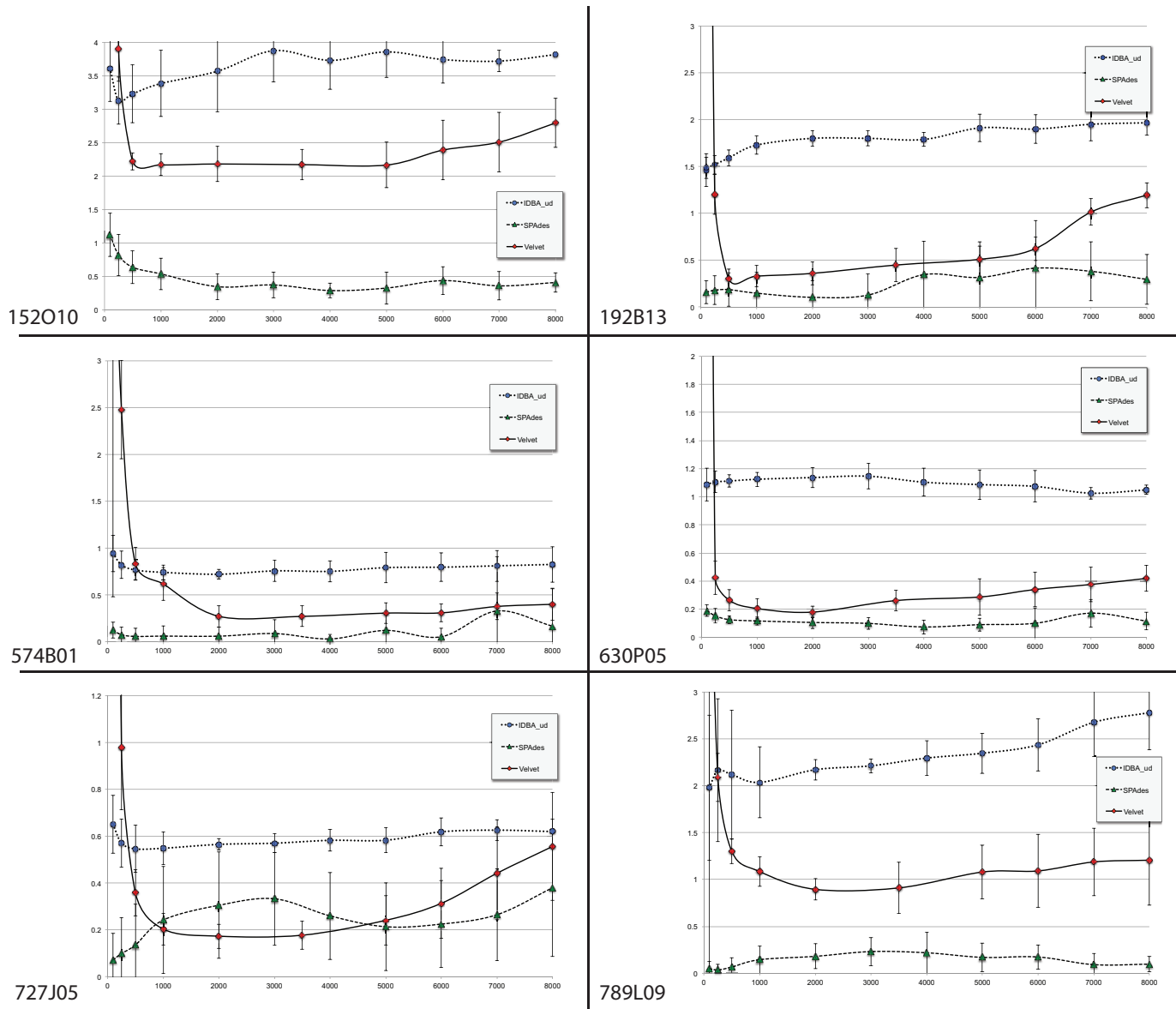


789L09

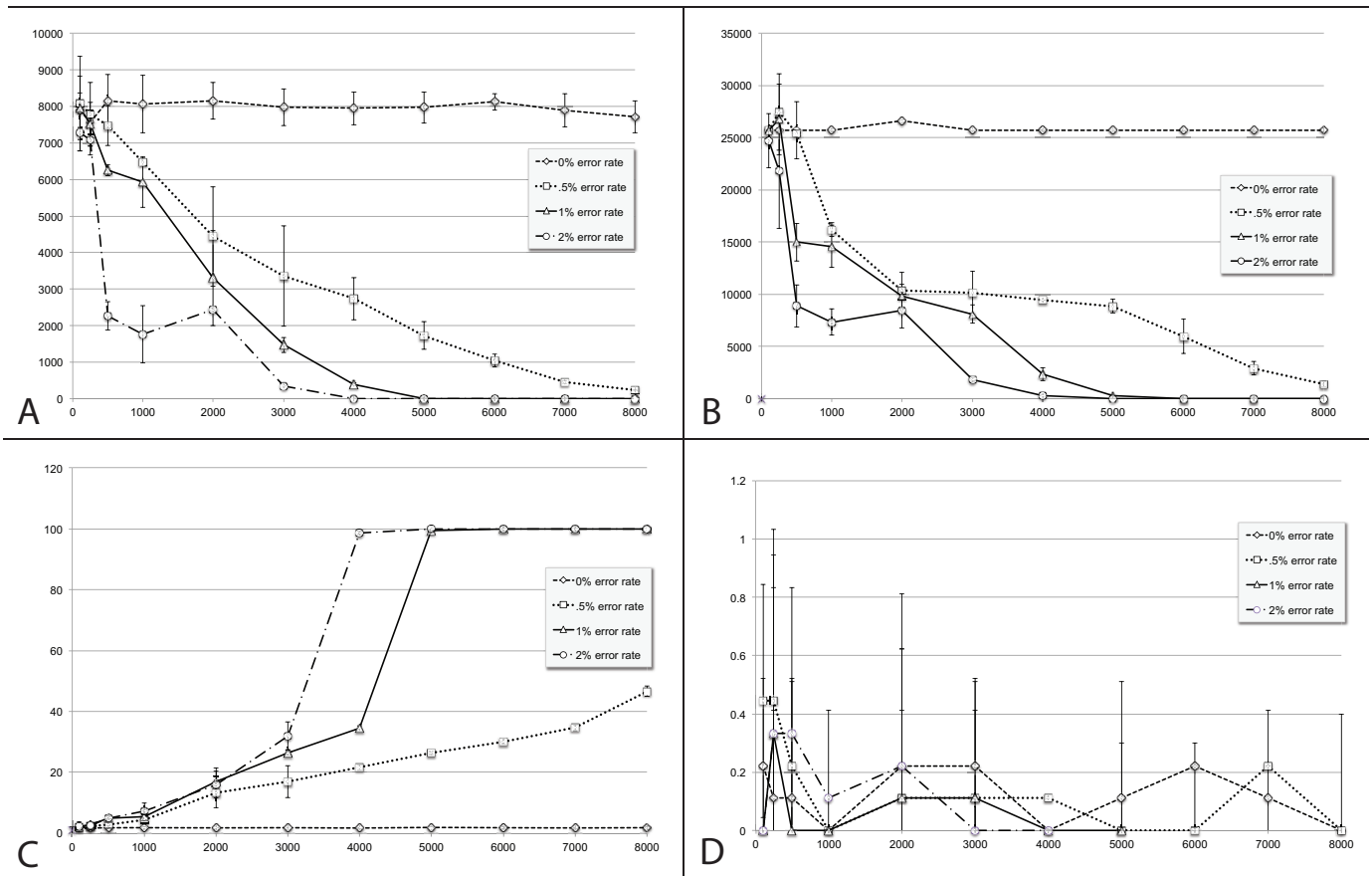
**Supplemental Figure S4:** Largest contig statistics (Y-axis) for the six ultra-deep coverage BACs, assembled with VELVET, SPADes and IDBA-UD for various levels of depth of sequencing (X-axis)



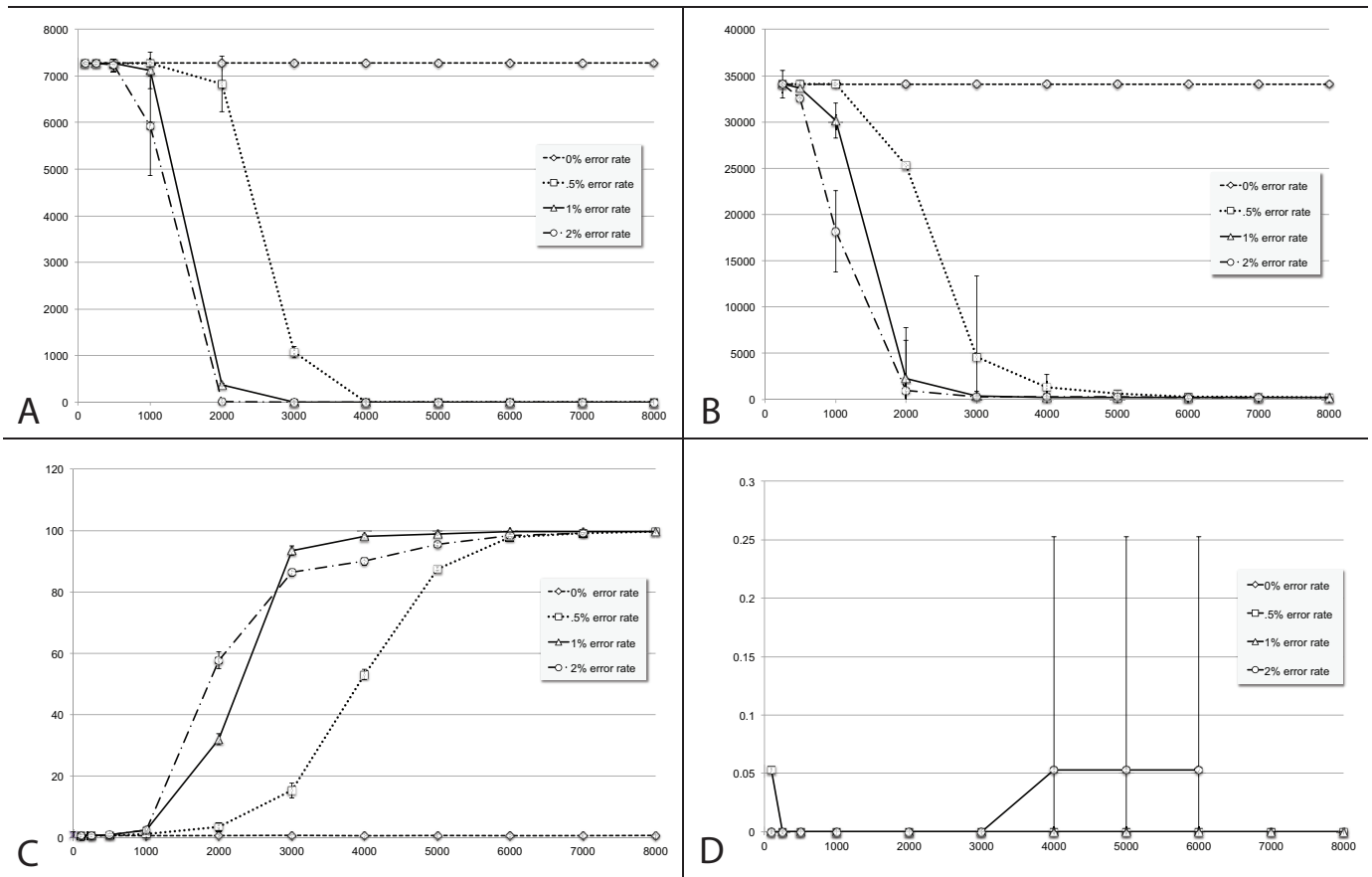
**Supplemental Figure S5:** Mis-assembly error statistics (Y-axis) for the six ultra-deep coverage BACs, assembled with VELVET, SPAdes and IDBA-UD for various levels of depth of sequencing (X-axis)



**Supplemental Figure S6:** Genome percentage missing (Y-axis) for the six ultra-deep coverage BACs, assembled with VELVET, SPADES and IDBA-UD for various levels of depth of sequencing (X-axis)

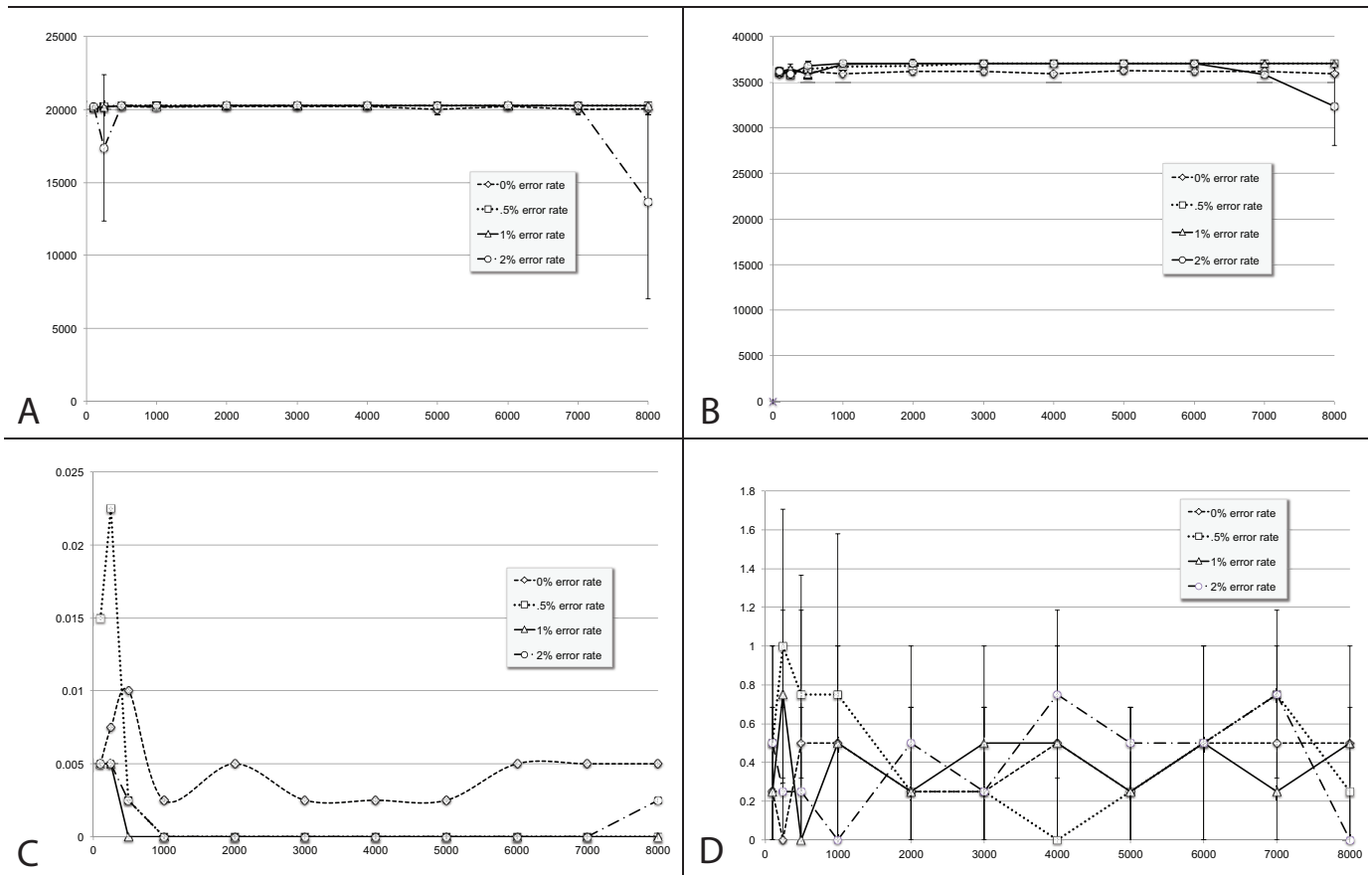


**Supplemental Figure S7:** VELVET assembly statistics (*Y*-axis) as a function of the depth of sequencing coverage (*X*-axis) for synthetic reads generated from BAC 574B01 for several choices of the sequencing error rate: (A) *n*50, (B) longest contig, (C) percentage of the target BAC not covered by the assembly, (D) number of assembly errors; each point is an average over twenty samples of the reads, errors bars indicate standard deviation among the samples



**Supplemental Figure S8:** IDBA-UD assembly statistics (Y-axis) as a function of the depth of sequencing coverage (X-axis) for synthetic reads generated from BAC 574B01 for several choices of the sequencing error rate: (A) n50, (B) longest contig, (C) percentage of the target BAC not covered by the assembly, (D) number of assembly errors; each point is an average over twenty samples of the reads, errors bars indicate standard deviation among the samples





**Supplemental Figure S9:** SPADES assembly statistics (*Y*-axis) as a function of the depth of sequencing coverage (*X*-axis) for synthetic reads generated from BAC 574B01 for several choices of the sequencing error rate: (A) n50, (B) longest contig, (C) percentage of the target BAC not covered by the assembly, (D) number of assembly errors; each point is an average over twenty samples of the reads, errors bars indicate standard deviation among the samples

	<i>after demultiplexing</i>			% <i>E.coli</i>	<i>after demultiplexing/cleaning/trimming</i>		
	<i>reads</i> (M)	<i>bases</i> (Mbp)	<i>read len</i> (bp)		<i>reads</i> (M)	<i>bases</i> (Mbp)	<i>read len</i> (bp)
Hv3	2,476	227,773	92.00	41.12%	1,240.2	110,056	88.74
Hv4	1,363	125,273	91.91	39.36%	713.4	63,384	88.85
Hv5	1,142	105,089	92.00	51.11%	505.1	45,088	89.27
Hv6	1,133	104,239	92.00	65.96%	282.4	24,970	88.42
Hv7	2,288	210,535	92.00	46.11%	928.9	82,503	88.82
Hv8	1,802	165,803	92.00	44.04%	730.8	64,651	88.46
Hv9	1,596	146,816	92.00	40.66%	736.2	65,697	89.24
Hv10	971	89,370	92.00	20.95%	748.2	67,600	90.36
Vu1	2,475	227,696	92.00	36.66%	1,208.1	108,666	89.95
Vu2	2,402	221,006	92.00	43.12%	1,144.6	103,026	90.01

**Supplemental Table S1:** Basic statistics on the ten sequenced read datasets (seven for barley, two for cowpeas) analyzed in this manuscript

<i>BAC</i>	<i>approx size</i> (bp)	<i>set</i>	<i>after demultiplexing</i>		% <i>E.coli</i>	<i>after demux/cleaning/trimming</i>		
			<i>reads</i> (M)	<i>bases</i> (Mbp)		<i>reads</i> (M)	<i>bases</i> (Mbp)	<i>coverage</i> (x)
<b>052L22</b>	105,788	Hv4	21.534	1,981	10.94%	16.795	1,488	14,065
<b>152O10</b>	117,543	Hv3, Hv9	18.575	1,709	16.93%	13.101	1,154	9,812
<b>192B13</b>	112,841	Hv4	19.581	1,801	12.20%	13.950	1,225	10,853
<b>574B01</b>	92,859	Hv3	32.952	3,032	15.27%	23.414	2,059	22,172
<b>630P05</b>	110,490	Hv3	16.224	1,493	15.45%	11.102	971	8,792
<b>727J05</b>	131,648	Hv8	22.613	2,080	14.80%	15.801	1,388	10,546
772L04	116,367	Hv3, Hv10	21.837	2,009	17.44%	14.407	1,255	10,784
773A02	185,718	Hv3, Hv10	21.756	2,002	15.86%	14.909	1,309	7,049
773F12	96,385	Hv3, Hv10	20.868	1,920	14.61%	14.661	1,286	13,341
773H21	71,701	Hv3, Hv10	23.679	2,178	16.38%	16.770	1,473	20,540
773L22	92,859	Hv3, Hv10	17.031	1,567	16.16%	12.282	1,081	11,639
774D07	117,543	Hv3, Hv10	13.077	1,203	20.51%	8.880	778	6,621
774G18	90,508	Hv3, Hv10	24.571	2,261	15.89%	16.763	1,469	16,229
774L04	103,437	Hv3, Hv10	22.053	2,029	17.13%	15.219	1,334	12,895
774O01	95,209	Hv3, Hv10	41.579	3,822	14.76%	29.877	2,642	27,754
<b>789L09</b>	84,631	Hv3	37.000	3,404	15.20%	25.730	2,264	26,754

**Supplemental Table S2:** Basic statistics on the read datasets for the 16 barley BACs sequenced individually

		average number of scaffolds of a given size						average number contigs of a given size										
		≥50	≥100	≥200	≥400	≥1K	≥10K	≥50	≥100	≥200	≥400	≥1K	≥10K	<i>n50</i>	<i>max</i>	<i>sum</i>	<i>sum/size</i>	<i>expected/observed</i>
Hv3	VELVET ( <i>l</i> = 49)	29.94	29.94	18.76	14.38	10.22	3.23	36.16	35.94	23.09	17.76	12.98	3.08	22,120	33,911	103,795	96.90%	1,471/1,435
	SPADES ( <i>l</i> = 49)	43.50	41.32	37.78	22.81	13.64	2.67	43.74	41.55	38.01	23.04	13.87	2.65	14,350	27,255	102,260	95.60%	1,471/1,435
	SPADES ( <i>l</i> = 31, 33, . . . , 79)	27.15	26.78	25.71	17.04	10.79	2.96	27.32	26.95	25.89	17.21	10.96	2.97	24,121	35,758	104,721	97.90%	1,471/1,434
	IDBA ( <i>l</i> = 31, 33, . . . , 79)	54.12	44.39	32.93	24.46	14.35	2.74	54.15	44.41	32.95	24.47	14.36	2.75	13,063	26,168	105,889	99.20%	1,471/1,435
Hv4	VELVET ( <i>l</i> = 49)	31.26	31.26	23.63	19.76	13.72	3.67	40.20	40.04	30.78	25.21	17.43	3.40	20,751	34,380	123,644	104.90%	473/319
	SPADES ( <i>l</i> = 31, 33, . . . , 79)	27.72	27.43	26.48	20.91	13.61	3.48	27.89	27.60	26.65	21.09	13.78	3.48	23,914	37,041	124,614	105.70%	473/320
Hv5	VELVET ( <i>l</i> = 49)	26.40	26.40	19.13	16.11	11.34	3.44	34.61	34.46	25.68	21.38	15.24	3.28	24,890	37,365	114,926	92.30%	226/196
	SPADES ( <i>l</i> = 31, 33, . . . , 79)	23.92	23.70	22.88	17.91	11.68	3.25	24.09	23.86	23.04	18.07	11.84	3.26	26,246	38,463	115,537	92.90%	226/196
Hv6	VELVET ( <i>l</i> = 49)	26.33	26.33	19.27	16.37	11.95	3.13	37.79	37.67	28.06	23.01	16.42	2.59	20,038	31,443	104,192	84.91%	244/206
	SPADES ( <i>l</i> = 31, 33, . . . , 79)	22.32	22.02	21.31	17.64	12.36	2.98	22.47	22.17	21.46	17.79	12.50	2.98	20,287	31,613	104,100	84.91%	244/205
Hv7	VELVET ( <i>l</i> = 49)	22.46	22.46	15.31	12.75	9.33	3.16	28.34	28.21	19.85	16.22	11.98	3.17	27,415	38,302	104,602	85.72%	228/204
	SPADES ( <i>l</i> = 31, 33, . . . , 79)	19.96	19.65	18.70	14.61	9.58	2.85	20.12	19.81	18.86	14.77	9.74	2.86	31,109	41,680	105,170	86.29%	228/203
Hv8	VELVET ( <i>l</i> = 49)	27.76	27.76	20.86	17.34	11.62	3.04	33.71	33.59	25.35	20.68	13.92	2.92	22,222	34,148	104,857	85.48%	207/188
	SPADES ( <i>l</i> = 31, 33, . . . , 79)	24.79	24.59	23.81	18.37	11.56	2.91	24.95	24.74	23.97	18.53	11.72	2.92	24,591	36,298	105,466	86.02%	207/188
Hv9	VELVET ( <i>l</i> = 49)	26.94	26.94	20.77	17.18	11.31	3.08	32.69	32.57	25.24	20.37	13.51	3.02	22,419	34,677	105,679	92.44%	361/310
	SPADES ( <i>l</i> = 31, 33, . . . , 79)	24.59	24.29	23.38	18.35	11.33	2.94	24.76	24.46	23.55	18.53	11.50	2.95	25,293	37,225	106,874	93.52%	361/310
Hv10	VELVET ( <i>l</i> = 49)	43.55	43.55	21.91	12.63	8.63	3.21	51.54	51.17	26.00	16.18	11.98	3.33	31,052	42,916	113,098	93.16%	27/20
	SPADES ( <i>l</i> = 31, 33, . . . , 79)	34.95	34.52	33.22	15.44	9.18	2.81	35.14	34.71	33.41	15.63	9.37	2.83	36,244	47,544	113,904	93.92%	27/20
Vu1	VELVET ( <i>l</i> = 49)	45.42	45.42	35.62	27.83	16.73	3.13	50.08	49.93	38.94	29.66	17.64	3.01	12,470	24,417	118,080	69.54%	612/519
	SPADES ( <i>l</i> = 31, 33, . . . , 79)	44.33	42.73	39.14	27.15	15.07	3.31	44.43	42.83	39.24	27.25	15.17	3.31	16,972	29,771	123,597	73.60%	612/536
Vu2	VELVET ( <i>l</i> = 49)	29.97	29.97	22.78	18.75	12.71	2.90	33.75	33.64	25.31	20.21	13.56	2.80	14,894	25,388	99,006	68.15%	503/449
	SPADES ( <i>l</i> = 31, 33, . . . , 79)	29.99	28.68	25.83	19.07	11.39	2.94	30.07	28.76	25.91	19.15	11.47	2.94	19,686	30,269	102,694	71.19%	503/450

**Supplemental Table S3:** Assembly statistics on all barley and cowpea datasets (sliced optimally as in Table 5 in the main manuscript, and decoded using  $k = 32$  in HASHFILTER) using VELVET, SPADES and IDBA for several choices of the hash size