

# Supplementary figures and tables

## MetaVelvet : An extension of Velvet assembler to *de novo* metagenome assembly from short sequence reads

Toshiaki Namiki <sup>1,2,\*</sup>, Tsuyoshi Hachiya <sup>1†</sup>, Hideaki Tanaka <sup>1</sup>, and Yasubumi Sakakibara <sup>1‡</sup>

<sup>1</sup>Department of Biosciences and Informatics, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223-8522, Japan,  
and <sup>2</sup>Japan Post Network Co.,Ltd.

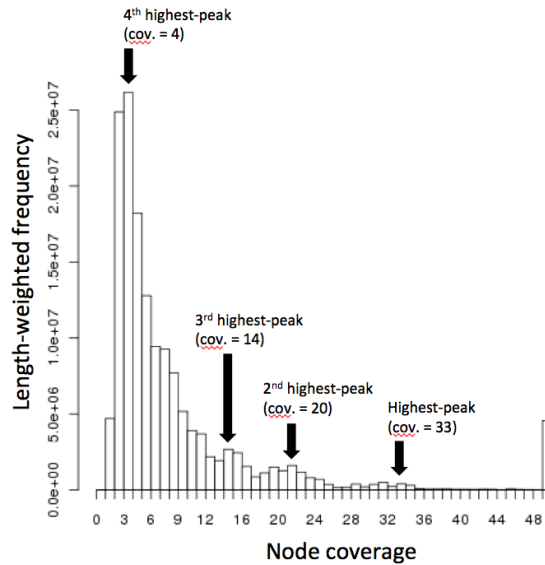


Figure S1: Node coverage histogram of the MH0006 dataset.

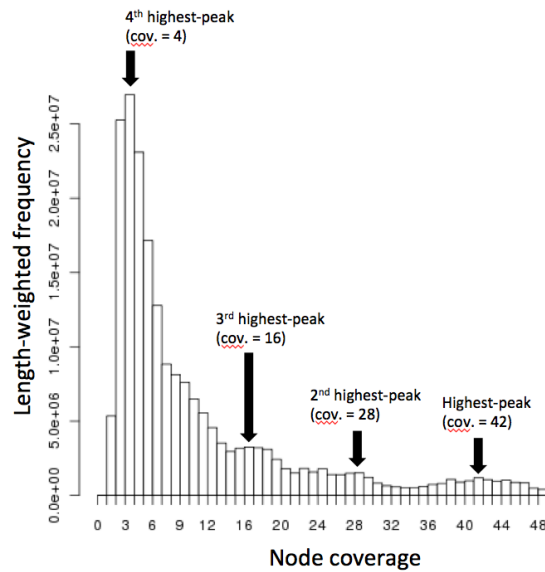


Figure S2: Node coverage histogram of the MH0012 dataset.

\*These authors equally contributed as joint first authors.

†These authors equally contributed as joint first authors.

‡to whom correspondence should be addressed

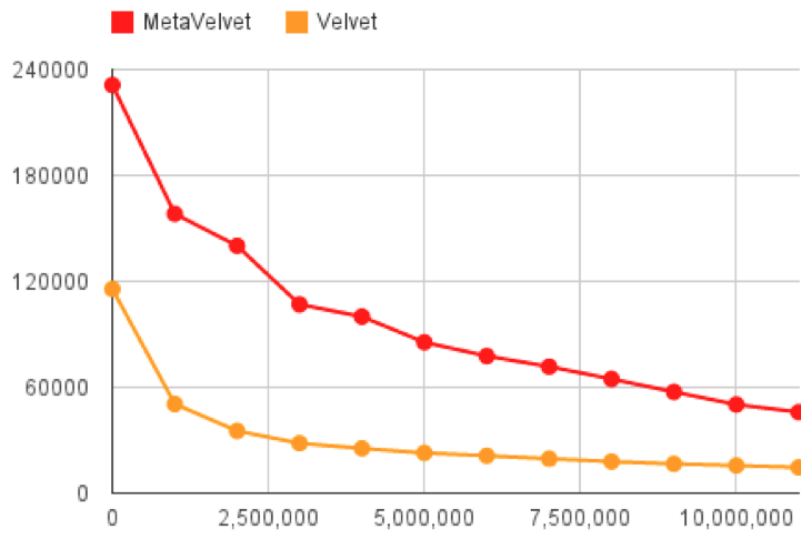


Figure S3: N-len( $x$ ) plot of the MH0006 dataset.

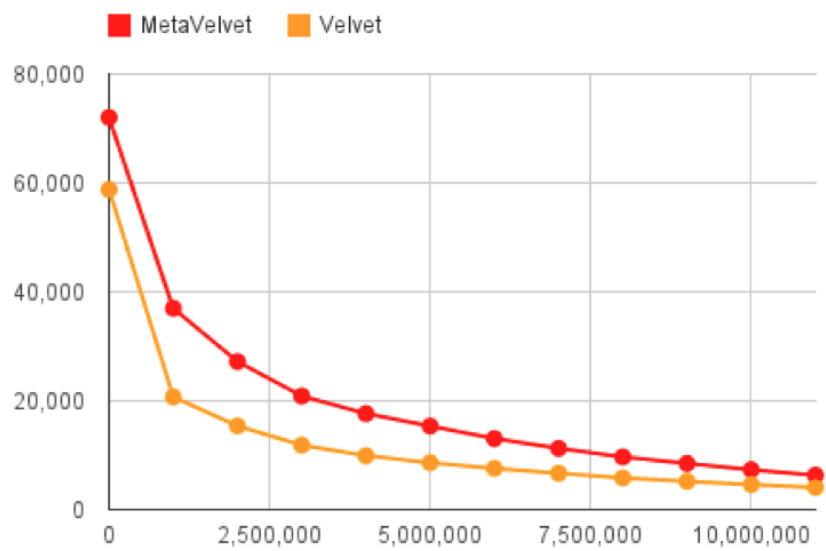


Figure S4: N-len( $x$ ) plot of the MH0047 dataset.

Table S1: Taxonomic profile of the order-level simulation dataset

Domain	Phylum	Class	Order	Family	Genus	Species	Strain	Length	coverage	#. read pairs
Bacteria	Proteobacteria	Alphaproteobacteria	<b>Caulobacterales</b>	Caulobacteraceae	Caulobacter	<i>caescentius</i>	NA1000	4,042,929	90.08	2,291,222
Bacteria	Proteobacteria	Alphaproteobacteria	<b>Rhizobiales</b>	Rhizobiaceae	Rhizobium	<i>etli</i>	CFN 42	4,381,608	30.00	821,552
Bacteria	Proteobacteria	Alphaproteobacteria	<b>Rhodobacteriales</b>	Rhodobacteriaceae	Dimoroseobacter	<i>shibae</i>	DFL 12	3,789,584	224.47	5,316,521
Bacteria	Proteobacteria	Alphaproteobacteria	<b>Parvularculales</b>	Parvularculaceae	Parvularcula	<i>bermudensis</i>	HTCC2503	2,902,643	57.69	1,046,594
Bacteria	Proteobacteria	Alphaproteobacteria	<b>Rhodospirillales</b>	Rhodospirillaceae	Azospirillum	<i>sp.</i>	B510	3,311,395	48.86	1,011,126
Bacteria	Proteobacteria	Alphaproteobacteria	<b>Sphingomonadales</b>	Erythrobacteraceae	Erythrobacter	<i>litoralis</i>	HTCC2594	3,052,398	30.00	572,325
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Acidithiobacillales</b>	Acidithiobacillaceae	Acidithiobacillus	<i>ferrooxidans</i>	ATCC 23270	2,982,397	356.34	6,642,138
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Aeromonadales</b>	Aeromonadaceae	Aeromonas	<i>hydrophila</i>	hydrophila ATCC 7966	4,744,448	30.00	889,584
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Alteromonadales</b>	Alteromonadaceae	Alteromonas	<i>macleodii</i>	Deep ecotype	4,412,282	54.95	1,498,817
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Cardiobacteriales</b>	Cardiobacteriaceae	Dichelo bacter	<i>nodosus</i>	VCSI703A	1,389,350	45.97	399,169
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Chromatiales</b>	Ectothiorhodospiraceae	Alkalimnicola	<i>ehritchei</i>	MLHE-1	3,275,944	30.00	614,240
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Enterobacteriales</b>	Enterobacteriaceae	Escherichia	<i>coli</i>	K-12 substr. MG1655	4,639,675	144.19	4,181,236
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Legionellales</b>	Legionellaceae	Legionella	<i>pneumophila</i>	str. Paris	3,503,610	30.00	656,927
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Methylococcales</b>	Methylococcaceae	Methylococcus	<i>capsulatus</i>	str. Bath	3,304,561	134.78	2,783,604
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Oceanospirillales</b>	Halomonadaceae	Chromohalobacter.	<i>salzigens</i>	DSM 3043	3,696,649	30.00	693,122
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Pasteurellales</b>	Pasteurellaceae	Pasteurella.	<i>multocida</i>	multocida str. Pm70	2,257,487	100.76	1,421,665
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Pseudomonadales</b>	Pseudomonadaceae	Pseudomonas	<i>putida</i>	W619	5,774,330	87.97	3,174,887
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Thiotrichales</b>	Piscirickettsiaceae	Thiomicrospira	<i>crunigena</i>	XCL-2	2,427,734	99.93	1,516,337
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Vibrionales</b>	Vibrionaceae	Vibrio	<i>cholerae</i>	O1 biovar eltor str. N16961	2,961,149	174.74	3,233,874
Bacteria	Proteobacteria	Gammaproteobacteria	<b>Xanthomonadales</b>	Xanthomonadaceae	Xanthomonas	<i>campestris</i>	pv. campestris	5,079,002	275.25	8,737,628

Table S2: Taxonomic profile of the family-level simulation dataset

Domain	Phylum	Class	Order	Family	Genus	Species	Strain	Length	coverage	#. read pairs
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bartoniellaceae	<i>Bartonella</i>	<i>bacilliformis</i>	KC583	1,445,021	124.84	1,127,480
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	<i>Methylocella</i>	<i>silvestris</i>	BL2	4,305,430	30.00	807,268
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	<i>Nitrobacter</i>	<i>hamburgensis</i>	X14	4,406,967	30.00	826,306
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Brucellaceae	<i>Brucella</i>	<i>suus</i>	1330	2,107,794	30.00	395,211
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Hypnomicrobacteriaceae	<i>Hypnomicrobium</i>	<i>denitrificans</i>	ATCC 51888	3,638,969	30.17	686,272
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methyllobacteriaceae	<i>Methyllobacterium</i>	<i>chloromethanicum</i>	GM4	5,777,908	484.74	17,504,858
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Phyllobacteriaceae	<i>Mesorhizobium</i>	<i>loti</i>	MAFF303099	7,036,071	147.34	6,479,319
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	<i>Rhizobium</i>	<i>etli</i>	CFN 42	4,381,608	31.19	854,268
Bacteria	Proteobacteria	Alphaproteobacteria	Rhizobiales	Xanthobacteraceae	<i>Azorhizobium</i>	<i>caulinodans</i>	ORS 571	5,369,772	101.13	3,394,174
Bacteria	Proteobacteria	Alteromonadales	Alteromonadales	Alteromonadaceae	<i>Alteromonas</i>	<i>macleodii</i>	Deep ecotype	4,412,282	30.00	827,303
Bacteria	Proteobacteria	Gammmaproteobacteria	Alteromonadales	Alteromonadaceae	<i>Colwellia</i>	<i>psychrerythraea</i>	34H	5,373,180	413.22	13,876,916
Bacteria	Proteobacteria	Gammmaproteobacteria	Alteromonadales	Ferrimonadaceae	<i>Ferrimonas</i>	<i>balearica</i>	DSM 9799	4,279,159	120.27	3,216,658
Bacteria	Proteobacteria	Gammmaproteobacteria	Alteromonadales	Idiomarinaceae	<i>Idiomarina</i>	<i>lohiensis</i>	L2TR	2,839,318	58.56	1,039,213
Bacteria	Proteobacteria	Gammmaproteobacteria	Alteromonadales	Pseudoalteromonadaceae	<i>Pseudoalteromonas</i>	<i>haloplanktis</i>	TAC125	3,214,944	40.59	815,636
Bacteria	Proteobacteria	Gammmaproteobacteria	Alteromonadales	Psychromonadaceae	<i>Psychromonas</i>	<i>ingrahamii</i>	37	4,559,598	141.39	4,029,249
Bacteria	Proteobacteria	Gammmaproteobacteria	Alteromonadales	Shewanellaceae	<i>Shewanella</i>	<i>oneidensis</i>	MR-1	4,969,803	48.15	1,495,658
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	<i>Bacillus</i>	<i>subtilis</i>	subsp. subtilis str. 168	4,214,630	56.21	1,480,619
Bacteria	Firmicutes	Bacilli	Bacillales	Listeriaceae	<i>Listeria</i>	<i>monocytogenes</i>	str. 4b F2365	2,905,187	38.08	691,432
Bacteria	Firmicutes	Bacilli	Bacillales	Paenibacillaceae	<i>Brevibacillus</i>	<i>brevis</i>	NBRC 100599	6,296,436	55.18	2,171,406
Bacteria	Firmicutes	Bacilli	Bacillales	Alicyclobacillaceae	<i>Alicyclobacillus</i>	<i>actidocaldarius</i>	actidocaldarius DSM 446	3,018,755	34.91	658,708

Table S3: Taxonomic profile of the genus-level simulation dataset

Domain	Phylum	Class	Order	Family	Genus	Species	Strain	Length	coverage	#. read pairs
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Yersinia</b>	<i>pestis</i>	Angola	4,504,254	390.28	10,986,994
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Candidatus Moranella</b>	<i>endobia</i>	PCIT	538,294	243.89	820,513
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Sodalis</b>	<i>glossiniidius</i>	morsitans	4,171,146	92.53	2,412,319
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Citrobacter</b>	<i>koseri</i>	ATCC BAA-895	4,720,462	30.00	885,087
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Cronobacter</b>	<i>sakazakii</i>	ATCC BAA-894	4,368,373	30.00	819,070
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Dickeya</b>	<i>dadantii</i>	Ech703	4,679,450	166.48	4,868,965
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Edwardsiella</b>	<i>ictaluri</i>	93-146	3,812,315	81.42	1,939,877
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Enterobacter</b>	<i>asburiae</i>	LF7a	4,812,833	72.73	2,187,857
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Erwinia</b>	<i>billingiae</i>	Eb661	5,100,167	30.00	956,281
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Escherichia</b>	<i>coli</i>	str. K-12 substr. MG1655	4,639,675	81.31	2,357,927
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Klebsiella</b>	<i>varuicola</i>	At-22	5,458,505	30.00	1,023,470
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Pantoea</b>	<i>vagans</i>	C9-1	4,024,986	241.81	6,082,961
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Pectobacterium</b>	<i>atrosepticum</i>	SCRI1043	5,064,019	30.00	949,504
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Photobhabdus</b>	<i>asymbiotica</i>	subsp. asymbiotica ATCC 43949	5,064,808	63.75	2,018,159
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Proteus</b>	<i>mirabilis</i>	H14320	4,063,606	83.65	2,124,570
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Rahnella</b>	<i>sp.</i>	Y9602	4,864,217	57.65	1,752,746
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Salmonella</b>	<i>bongori</i>	NCTC-12419	4,460,105	30.00	836,270
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Serratia</b>	<i>proteamaculans</i>	568	5,448,853	96.20	3,276,158
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Shigella</b>	<i>flemeni</i>	5 str. 8401	4,574,284	80.21	2,293,062
Bacteria	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	<b>Xenorhabdus</b>	<i>bovienii</i>	SS-2004	4,225,498	182.67	4,824,299

Table S4: Taxonomic profile of the species-level simulation dataset

Domain	Phylum	Class	Order	Family	Genus	Species	Strain	Length	coverage	#. read pairs
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>amyloliquefaciens</i>	DSM 7	3,980,199	92.04	2,289,500
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>anthracis</i>	Ames	5,227,293	30.04	981,581
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>atrophaceus</i>	1942	4,168,266	270.68	7,051,549
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>cellulosilyticus</i>	DSM 2522	4,681,672	31.46	920,597
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>cereus</i>	Q1	5,214,195	100.35	3,270,220
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>clausii</i>	KSM K16	4,303,871	114.47	3,079,139
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>coagulans</i>	36D1	3,552,226	215.27	4,779,398
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>cytotoxicus</i>	NVH 391 98	4,087,024	93.45	2,386,983
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>halodurans</i>	C 125	4,202,352	47.38	1,244,520
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>licheniformis</i>	ATCC 14580	4,222,597	30.00	791,737
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>megaterium</i>	DSM319	5,097,447	123.48	3,934,058
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>pseudofirmus</i>	OF4	3,858,997	30.00	723,562
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>pumilus</i>	SAFR 032	3,704,465	303.41	7,024,887
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>selenitireducens</i>	MLS10	3,592,487	161.57	3,627,841
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>subtilis</i>	subsp. subtilis str. 168	4,214,630	30.00	790,243
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>thurmgienensis</i>	AI Hakan	5,257,091	240.41	7,899,114
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>tusciae</i>	DSM 2912	3,384,766	30.00	634,644
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	<i>weihenstephanensis</i>	KBAB4	5,262,775	30.00	986,770
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Geobacillus	<i>kaustophilus</i>	HTA426	3,544,776	30.00	664,646
Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae	Geobacillus	<i>thermoglucosidasius</i>	C56 YS93	3,893,306	42.95	1,045,116