

1 **Additional File 5.**

2 Putative functions and taxonomic assignments of predicted ORFs of 2 fosmid inserts:

3 Scaffold\_56 (GenBank: JQ335997) and Scaffold\_90 (GenBank: JQ335998).

ORF	Putative function (COG category if available)	Closest relative (Organism / Phylum)	Size (aa)	GC (%)
<b>Scaffold_56</b>				
1	Uncharacterized vancomycin resistance protein (V)	<i>Clostridium ramosum</i> / Firmicutes	417	63
2	Sigma 54 modulation protein / SSU ribosomal protein S30P (J)	<i>Ruminococcaceae bacterium</i> / Firmicutes	48	63
3	Hypothetical protein	No significant similarity found	76	63
4	ATP-dependent metalloprotease FtsH (O)	<i>Ruminococcaceae bacterium</i> / Firmicutes	732	65
5	Tyrosine-tRNA ligase (J)	<i>Subdoligranulum variabile</i> / Firmicutes	409	64
6	Integrase family protein (L)	<i>Ruminococcaceae bacterium</i> / Firmicutes	134	58
7	Beta-glucosidase-related glycosidases (G)	<i>Roseburia intestinalis</i> / Firmicutes	810	60
8	Transcriptional regulator, TetR family (K)	<i>Clostridium acetobutylicum</i> / Firmicutes	191	50
9	Na <sup>+</sup> /melibiose symporter and related transporters (G)	<i>Marvinbryantia formatexigens</i> / Firmicutes	195	54
10	Na <sup>+</sup> /melibiose symporter and related transporters (G)	<i>Marvinbryantia formatexigens</i> / Firmicutes	365	59
11	Beta-glucosidase-related glycosidases (G)	<i>Ruminococcus torques</i> / Firmicutes	890	60
12	Beta-galactosidase / beta-glucuronidase (G)	<i>Marvinbryantia formatexigens</i> / Firmicutes	771	60
13	L-arabinose isomerase (G)	<i>Subdoligranulum variabile</i> / Firmicutes	111	53
14	Sugar (pentulose and hexulose) kinases (G)	<i>Roseburia intestinalis</i> / Firmicutes	512	64
15	Toxin-antitoxin system, toxin component, PIN family (R)	<i>Eubacterium saepenum</i> / Firmicutes	137	47
16	Addiction module antitoxin, RelB/DinJ family (L)	<i>Butyrivibrio fibrisolvens</i> / Firmicutes	94	48
17	Transposase, IS4 family (L)	<i>Ruminococcus albus</i> / Firmicutes	472	53
18	Transposase, IS4 family (L)	<i>Ruminococcus albus</i> / Firmicutes	430	53
19	SH3 domain protein (T)	<i>Clostridium cellulolyticum</i> / Firmicutes	264	65
20	Hypothetical protein	No significant similarity found	77	65
21	Conserved hypothetical protein	<i>Ruminococcaceae bacterium</i> / Firmicutes	78	61
22	Hypothetical protein	No significant similarity found	32	60
23	IS186 transposase (L)	<i>Escherichia coli</i> / Proteobacteria	371	52
24	Hypothetical protein	No significant similarity found	62	62
25	Hypothetical protein	No significant similarity found	72	66
26	Conserved hypothetical protein (S)	<i>Eubacterium dolichum</i> / Firmicutes	992	59
27	Prophage protein (R)	<i>Coprobacillus sp.</i> / Firmicutes	364	65
28	Prophage protein (S)	<i>Coprobacillus sp.</i> / Firmicutes	213	63
<b>Scaffold_90</b>				
1	Branched-chain amino acid transporter AzIC (E)	<i>Eubacterium yurii</i> / Firmicutes	58	47
2	NAD-dependent protein deacetylases, SIR2 family (K)	<i>Ruminococcus albus</i> / Firmicutes	286	47
3	Conserved hypothetical protein (R)	<i>Jonquetella anthropi</i> / Synergistetes	98	43
4	Nitroreductase (C)	<i>Ruminococcus sp.</i> / Firmicutes	195	51
5	NADPH:quinone reductase / Zn-dependent oxidoreductases (C)	<i>Synergistetes bacterium</i> / Synergistetes	334	51
6	Predicted Fe-S oxidoreductases (R)	<i>Bacteroides pectinophilus</i> / Bacteroidetes	328	50
7	Hypothetical protein	No significant similarity found	48	59
8	Transcriptional regulator, HxIR family (K)	<i>Butyrivibrio proteoclasticus</i> / Firmicutes	111	44
9	Beta-glucosidase-related glycosidases (G)	<i>Cellulosilyticum ruminicola</i> / Firmicutes	835	51
10	Na <sup>+</sup> /melibiose symporter and related transporters (G)	<i>Marvinbryantia formatexigens</i> / Firmicutes	469	47
11	Transcriptional regulator, AraC family (K)	<i>Marvinbryantia formatexigens</i> / Firmicutes	412	44
12	Conserved hypothetical protein (S)	<i>Bacteroides capillosus</i> / Bacteroidetes	207	48
13	Putative chlorophyll synthesis pathway, BchC (E)	<i>Clostridium hathewayi</i> / Firmicutes	338	52
14	Transcriptional regulator, TetR family (K)	<i>Clostridium hathewayi</i> / Firmicutes	197	48
15	Hypothetical protein	No significant similarity found	75	54
16	Hypothetical protein	No significant similarity found	103	44
17	Conserved hypothetical protein (S)	<i>Butyrivibrio proteoclasticus</i> / Firmicutes	488	46
18	Conserved hypothetical protein (S)	<i>Roseburia inulinivorans</i> / Firmicutes	176	47
19	Conserved hypothetical protein (S)	<i>Holdemanella filiformis</i> / Firmicutes	399	48
20	Conserved hypothetical protein (S)	<i>Clostridiales bacterium</i> / Firmicutes	366	42
21	NADH-dependent butanol dehydrogenase A, Fe-dependent (C)	<i>Clostridium methylpentosum</i> / Firmicutes	396	56
22	Integrase family protein (L)	<i>Clostridium scindens</i> / Firmicutes	411	48
23	Integrase family protein (L)	<i>Desulfotomaculum acetoxidans</i> / Firmicutes	324	50
24	Integrase family protein (L)	<i>Desulfotomaculum acetoxidans</i> / Firmicutes	348	47
25	4-carboxymuconolactone decarboxylase (S)	<i>Marvinbryantia formatexigens</i> / Firmicutes	264	50
26	Transcriptional regulator, MerR family (K)	<i>Clostridium methylpentosum</i> / Firmicutes	133	51
27	Conserved hypothetical protein	<i>Biifidobacterium breve</i> / Actinobacteria	203	48
28	Beta-glucosidase-related glycosidases (G)	<i>Ruminococcus gnavus</i> / Firmicutes	619	56
29	Beta-glucosidase-related glycosidases (G)	<i>Marvinbryantia formatexigens</i> / Firmicutes	221	55
30	Beta-glucosidase-related glycosidases (G)	<i>Marvinbryantia formatexigens</i> / Firmicutes	1076	54
31	Conserved hypothetical protein	<i>Marvinbryantia formatexigens</i> / Firmicutes	436	45
32	Response regulator of the LysR/AlgR family (K)	<i>Marvinbryantia formatexigens</i> / Firmicutes	97	46

4

5