

Supplemental Table 4. Gene statistics for both EN 588 and EN 604 metatranscriptomes. Combined statistics are shown in the first column. Individual EN 588 and EN 604 statistics are shown in the right two columns. Statistics were found using JGI Institute Integrated Microbial Genomes (IMG) web interface.

Gene Statistics	2016/2017 combined		2016		2017	
	Number	% of Total	Number	% of Total	Number	% of Total
DNA, Total number of bases	2352243881	100.00%	519221020	100.00%	1312434210	100.00%
DNA coding number of bases	N/A	0.00%		0.00%		0.00%
DNA G+C number of bases	1195912772	0.00% ¹	250480305	0.00% ¹	679061557	0.00% ¹
DNA Scaffolds	7215668	100.00%	1761728	100.00%	3914884	100.00%
Genes Total number	7492116	100.00%	1761728	100.00%	4101045	100.00%
Protein coding genes	7437098	99.27%	1745243	99.06%	4074241	99.35%
RNA genes	55018	0.73%	16485	0.94%	26804	0.65%
rRNA Genes	19277	0.26%	5561	0.32%	9403	0.23%
5S rRNA	1333	0.02%	280	0.02%	698	0.02%
16S rRNA	5980	0.08%	1817	0.10%	2927	0.07%
23S rRNA	11964	0.16%	3464	0.20%	5778	0.14%
tRNA genes	14678	0.20%	4413	0.25%	7463	0.18%
Other RNA genes	21063	0.28%	6511	0.37%	9938	0.24%
Protein Coding Genes with predicted function	2366065	31.58%	649400	36.86%	1226598	29.91%
Without predicted function	5071033	67.68%	1095843	62.20%	2847643	69.44%
Protein coding genes with enzymes	841131	11.23%	235757	13.38%	435577	10.62%
Protein coding genes connected to KEGG pathways	1166320	15.57%	321472	18.25%	604766	14.75%
Not connected to KEGG pathways	3866350	51.61%	396875	22.53%	3469475	84.60%
Protein coding genes connected to KEGG Orthology (KO)	1627603	21.72%	451175	25.61%	843986	20.58%
Not connected to KEGG Orthology (KO)	3589050	47.90%	3587795	20.37%	3230255	78.77%
Protein coding genes connected to MetaCyc pathways	549561	7.34%	155404	8.82%	283067	6.90%
Not connected to MetaCyc pathways	4240301	56.60%	449127	25.49%	3791174	92.44%
Protein coding genes with COGs	2085127	27.83%	583695	33.13%	1070761	26.11%
With TIGRfam	221945	2.96%	56107	3.18%	118560	2.89%