

Draft genome sequences of three *Holospora* species (*Holospora obtusa*, *Holospora undulata*, and *Holospora elegans*), endonuclear symbiotic bacteria of the ciliate *Paramecium caudatum*

メタデータ	言語: en 出版者: Wiley 公開日: 2015-09-18 キーワード (Ja): キーワード (En): 作成者: Dohra, Hideo, Tanaka, Kenya, Suzuki, Tomohiro, Fujishima, Masahiro, Suzuki, Haruo メールアドレス: 所属:
URL	http://hdl.handle.net/10297/9122

1 **Title page**

2 **Title**

3 Draft genome sequences of three *Holospora* species (*Holospora obtusa*, *Holospora*
4 *undulata*, and *Holospora elegans*), endonuclear symbiotic bacteria of the ciliate
5 *Paramecium caudatum*

6

7 **Running title**

8 Draft genome sequences of three *Holospora* species

9

10 **Authors**

11 Hideo Dohra^{a,b}, Kenya Tanaka^{c,d}, Tomohiro Suzuki^a, Masahiro Fujishima^{c,d}, and Haruo
12 Suzuki^c

13

14 **Affiliation**

15 ^aInstrumental Research Support Office, Research Institute of Green Science and
16 Technology, ^bDepartment of Biological Science, Graduate School of Science, Shizuoka
17 University, 836 Ohya, Suruga-ku, Shizuoka 422-8529, Japan, ^cDepartment of
18 Environmental Science and Engineering, Graduate School of Science and Engineering,
19 Yamaguchi University, 1677-1 Yoshida, Yamaguchi 753-8512, Japan, ^dNational
20 Bio-Resource Project of the Ministry of Education, Culture, Sports, Science and
21 Technology, Japan.

22

23 **Correspondence:** Haruo Suzuki, Department of Environmental Science and Engineering,
24 Graduate School of Science and Engineering, Yamaguchi University, 1677-1 Yoshida,
25 Yamaguchi 753-8512, Japan. Tel.: +81-83-933-5700; fax: +81-83-933-5273; e-mail:
26 haruo@yamaguchi-u.ac.jp.

27

28 **Keywords:** Endonuclear symbiotic bacteria; *Holospora obtusa*; *Holospora undulata*;
29 *Holospora elegans*; *Paramecium caudatum*.

30 **Abstract.** We present draft genome sequences of three *Holospora* species, hosted by the
31 ciliate *Paramecium caudatum*; i.e. the macronucleus-specific *Holospora obtusa* and the
32 micronucleus-specific *Holospora undulata* and *Holospora elegans*. We investigate
33 functions of orthologous core genes conserved across the three *Holospora* species, which
34 may be essential for the infection and survival in the host nucleus.

35

36 Bacteria of the genus *Holospora* are endonuclear symbionts of the *Paramecium* species
37 (Fokin, *et al.*, 1996). Four *Holospora* species have been identified in the ciliate
38 *Paramecium caudatum*, i.e. a macronuclear specific species *Holospora obtusa* and three
39 micronuclear specific species *Holospora undulata*, *Holospora elegans*, and *Holospora*
40 *recta*. To investigate the *Holospora-Paramecium* symbiosis, the draft genome sequence of
41 *H. undulata* has been determined (Dohra, *et al.*, 2013); however, to date the genome
42 sequences of the other *Holospora* species have not been published (Lang, *et al.*, 2005).
43 Here, we report draft genome sequences of *H. obtusa* strain F1 and *H. elegans* strain E1 as
44 well as updated genome assembly and annotation of *H. undulata* strain HU1.

45 A core genome is a set of all orthologous genes present in all members of a group;
46 e.g. a genus such as *Holospora*, a family such as *Holosporaceae*, an order such as
47 *Rickettsiales*, and a class such as *Alphaproteobacteria* (Charlebois & Doolittle, 2004). Most
48 of the core genes remaining in different species after divergence from a common ancestor
49 should be the essential genes needed by all the species (Liu, *et al.*, 2012). Here, we
50 investigate orthologous core genes conserved across the three *Holospora* species. This
51 information will contribute to the study of important functions characteristic of *Holospora*
52 such as the infection and survival in the host nucleus.

53 Growing *Holospora*-bearing *P. caudatum*, isolating *Holospora* cells from the host
54 nuclei, preparing *Holospora* genomic DNA were performed as described elsewhere (Dohra,
55 *et al.*, 2013) and in Supporting Information, SI Materials and Methods. Genomes were
56 sequenced using the Roche 454 GS FLX (*H. obtusa* and *H. elegans*) and Illumina GAIIx (*H.*
57 *undulata*) technologies. Sequencing statistics for each genome are shown in Supporting
58 information, Table S1. *De novo* genome assembly is a multi-step process, involving contig
59 integration, identification and removal of contaminated sequences with low sequencing
60 depth (details in Supporting Information, Materials and Methods, and Table S2). Among the
61 three *Holospora* genomes, assembly lengths and %G+C contents varied from 1.27 to 1.40
62 Mb, and 35.2 to 36.1 %, respectively.

63 The draft genome sequences of *H. obtusa*, *H. undulata*, and *H. elegans* were
64 automatically annotated using Prokka (<http://vicbioinformatics.com/>) (Seemann, 2014) and
65 then manually curated. The FastOrtho program (<http://enews.patricbrc.org/fastortho/>)
66 grouped a total of 3,553 protein-coding sequences from the three *Holospora* genomes into
67 1,610 ortholog clusters, of which 572 were identified as single-copy core orthologous genes
68 shared by the three genomes (Supporting information, Table S3). COG (Clusters of
69 Orthologous Groups) database defines four major functional categories: “information
70 storage and processing”, “cellular processes and signaling”, “metabolism”, and “poorly
71 characterized”, which are further subdivided into 25 categories (Table 1) (Tatusov, *et al.*,
72 2001). Of the 572 *Holospora* core genes, 488 (85.3%) were assigned to at least one of the
73 COG functional categories (BLAST e-value < 1e-5) (Supporting information, Table S3).
74 The 46 genes were assigned to multiple functional categories; for example, type II secretory
75 pathway proteins were assigned to the COG category N (cell motility) and U (intracellular
76 trafficking, secretion, and vesicular transport).

77 The *Holospora* core genome contained a set of 177 genes for “information storage
78 and processing” (Table 1) which are involved in replication, transcription, and translation
79 (category L, K, and J, respectively). Bacterial core genomes preferentially contain the
80 informational genes (Charlebois & Doolittle, 2004, Segata & Huttenhower, 2011).

81 A set of 138 conserved proteins for “cellular processes and signaling” (Table 1)
82 included a flagellar motor protein in the category N (cell motility). Although flagellar
83 movement in bacteria of the family rickettsiaceae has been reported (Vannini, *et al.*, 2014),
84 flagellar motility is not usually expected in an obligate intracellular organism. The
85 *Holospora* genomes did not contain an entire flagellar apparatus but only a protein
86 annotated as flagellar motor protein (COG1360) or chemotaxis protein MotB (KEGG:
87 K02557), suggesting that this protein could have some other function (Pallen & Matzke,
88 2006).

89 A set of 155 conserved genes for “metabolism” (Table 1) included a pyruvate
90 dehydrogenase complex E1, E2, and E3 component which converts pyruvate into
91 acetyl-CoA, and F0F1 ATP synthase subunits alpha, beta, gamma, a, b, and c in the
92 category C (energy production and conversion). Based on the KEGG database, the pyruvate
93 dehydrogenase complex was involved in glycolysis / gluconeogenesis, citrate cycle (TCA
94 cycle), and pyruvate metabolism. These pathways for energy metabolism have been
95 remained even in the symbiotic strain of *Polynucleobacter necessarius* (a cytoplasmic

96 endosymbiont of the ciliate *Euplotes aediculatus*) although the number of the
97 protein-coding genes of the symbiont's genome was reduced to 61.6% compared with that
98 of the *P. necessarius* free-living strain (Boscaro, *et al.*, 2013). *Holospora* lacked many
99 proteins involved in these pathways. These results suggest that the pyruvate dehydrogenase
100 complex is a possible relic of the ancestral pathway and that *Holospora* strongly rely on the
101 host for energy production.

102 The 84 (14.7%) of the 572 core genes lacking homologs in the COG protein
103 sequence database were not assigned to any of the COG functional categories. Some of
104 these genes are found in clusters of two or three genes (Supporting information, Table S3).
105 The set included periplasmic 5.4 kDa peptide (UniProt: P94818) (Dohra, *et al.*, 1997) and
106 89 kDa periplasmic protein associated with *Holospora* infection (UniProt: Q2ABW8)
107 (Iwatani, *et al.*, 2005), which were previously found in the macronucleus-specific *H. obtusa*.
108 In the present study, their homologs were found in the micronucleus-specific *H. undulata*
109 and *H. elegans*, and were thus included in *Holospora* core genome. These
110 *Holospora*-specific proteins lacking homologs in other bacteria may include proteins that
111 play important roles characteristic of the genus *Holospora*.

112 The Whole Genome Shotgun projects have been deposited at
113 DDBJ/EMBL/GenBank under accessions [AWTR000000000](#) (*Holospora obtusa* F1),
114 [ARPM000000000](#) (*Holospora undulata* HU1), and [BAUP000000000](#) (*Holospora elegans*
115 E1).

116

117 **Acknowledgements**

118 We thank Hans-Dieter Görtz for providing *P. caudatum* clone that bears *H. obtusa* strain F1
119 cells, Sergei I. Fokin for providing *P. caudatum* clone that bears *H. elegans* strain E1 cells,
120 and Torsten Seemann for his advice on Prokka. This work was supported by MEXT
121 TOKUBETSUKEIHI and partly by JSPS KAKENHI Grant Numbers 22370082 and
122 23657157.

123

124 **References**

125 Boscaro V, Felletti M, Vannini C, *et al.* (2013) *Polynucleobacter necessarius*, a model for
126 genome reduction in both free-living and symbiotic bacteria. *Proc Natl Acad Sci U S A* **110**:
127 18590-18595.

128 Charlebois RL & Doolittle WF (2004) Computing prokaryotic gene ubiquity: rescuing the

129 core from extinction. *Genome Res* **14**: 2469-2477.

130 Dale C & Moran NA (2006) Molecular interactions between bacterial symbionts and their
131 hosts. *Cell* **126**: 453-465.

132 Dohra H, Yamamoto K, Fujishima M & Ishikawa H (1997) Cloning and sequencing of
133 gene coding for a periplasmic 5.4 kDa peptide of the macronucleus-specific symbiont
134 *Holospora obtusa* of the ciliate *Paramecium caudatum*. *Zoolog Sci* **14**: 69-75.

135 Dohra H, Suzuki H, Suzuki T, Tanaka K & Fujishima M (2013) Draft Genome Sequence of
136 *Holospora undulata* Strain HU1, a Micronucleus-Specific Symbiont of the Ciliate
137 *Paramecium caudatum*. *Genome Announc* **1**: e00664-00613.

138 Fokin SI, Brigge T, Brenner J & Görtz HD (1996) *Holospora* species infecting the nuclei of
139 *Paramecium* appear to belong into two groups of bacteria. *Eur J Protistol* **32**: 19-24.

140 Iwatani K, Dohra H, Lang BF, Burger G, Hori M & Fujishima M (2005) Translocation of
141 an 89-kDa periplasmic protein is associated with *Holospora* infection. *Biochem Biophys*
142 *Res Commun* **337**: 1198-1205.

143 Lang BF, Brinkman H, Koski L, Fujishima M, Görtz HD & Burger G (2005) On the origin
144 of mitochondria and Rickettsia-related eukaryotic endosymbionts. *Jpn J Protozool* **38**:
145 171-183.

146 Liu W, Fang L, Li M, *et al.* (2012) Comparative genomics of Mycoplasma: analysis of
147 conserved essential genes and diversity of the pan-genome. *PLoS One* **7**: e35698.

148 Pallen MJ & Matzke NJ (2006) From The Origin of Species to the origin of bacterial
149 flagella. *Nat Rev Microbiol* **4**: 784-790.

150 Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*.

151 Segata N & Huttenhower C (2011) Toward an efficient method of identifying core genes
152 for evolutionary and functional microbial phylogenies. *PLoS One* **6**: e24704.

153 Tatusov RL, Natale DA, Garkavtsev IV, *et al.* (2001) The COG database: new
154 developments in phylogenetic classification of proteins from complete genomes. *Nucleic*
155 *Acids Res* **29**: 22-28.

156 Vannini C, Boscaro V, Ferrantini F, *et al.* (2014) Flagellar movement in two bacteria of the
157 family *rickettsiaceae*: a re-evaluation of motility in an evolutionary perspective. *PLoS One*
158 **9**: e87718.

159

160

161 Table 1. Number of genes assigned to each COG functional category among the 572
 162 orthologous genes shared by three *Holospira* genomes.

COG functional category	No. genes
INFORMATION STORAGE AND PROCESSING	
[J] Translation, ribosomal structure and biogenesis	106
[A] RNA processing and modification	0
[K] Transcription	27
[L] Replication, recombination and repair	44
[B] Chromatin structure and dynamics	0
CELLULAR PROCESSES AND SIGNALING	
[D] Cell cycle control, cell division, chromosome partitioning	10
[Y] Nuclear structure	0
[V] Defense mechanisms	8
[T] Signal transduction mechanisms	8
[M] Cell wall/membrane/envelope biogenesis	55
[N] Cell motility	5
[Z] Cytoskeleton	0
[W] Extracellular structures	0
[U] Intracellular trafficking, secretion, and vesicular transport	24
[O] Posttranslational modification, protein turnover, chaperones	28
METABOLISM	
[C] Energy production and conversion	26
[G] Carbohydrate transport and metabolism	22
[E] Amino acid transport and metabolism	33
[F] Nucleotide transport and metabolism	14
[H] Coenzyme transport and metabolism	13
[I] Lipid transport and metabolism	22
[P] Inorganic ion transport and metabolism	18
[Q] Secondary metabolites biosynthesis, transport and catabolism	7
POORLY CHARACTERIZED	
[R] General function prediction only	62

[S] Function unknown

23

163

Table S1. Sequencing statistics for three *Holospira* genomes: *H. obtusa* F1, *H. elegans* E1, and *H. undulata* HU1.

	<i>H. obtusa</i> F1		<i>H. elegans</i> E1		<i>H. undulata</i> HU1	
Platform	454 GS FLX Single-end, No.38,39		454 GS FLX Single-end, No.40		Illumina GA IIX 100bp Paired-end	
Number of reads	38	431,256	187,134	Read 1	30,331,713	
	39	445,222		Read 2	30,331,713	
Total bases of reads	38	145,056,671	64,954,496	Read 1	3,026,102,980	
	39	147,482,825		Read 2	3,050,059,961	
Average of coverage	219		51		4332	
Tools for hybrid assembly	GS De Novo Assembler (Newbler) 2.8 GS Reference Mapper 2.8 SPAdes 2.5.1 CISA 1.3		GS De Novo Assembler (Newbler) 2.8 GS Reference Mapper 2.8 SPAdes 2.5.1 CISA 1.3		ABYSS 1.3.5 Velvet 1.2.08 SPAdes 2.5.1 CISA 1.3 Bowtie2	
Final contigs						
Number of contigs	91		152		203	
Total bases of contigs (bp)	1,334,837		1,268,333		1,402,636	
Min contig length (bp)	233		586		1,065	
Max contig length (bp)	73,978		74,598		59,427	
Average contig length (bp)	14,668.5		8,344.3		6,909.5	
Median contig length (bp)	8,918.0		5,980.5		4,538.0	
N50 contig length (bp)	24,415		12,964		10,884	
G+C content (%)	35.2		36.0		36.1	
Number of CDSs	1117		1211		1224	
Number of rRNAs	3		3		3	
Number of tRNAs	34		31		30	
Number of tmRNAs	1		1		1	

Table S2. Statistics of each assembly step for three *Holospira* genomes: *H. obtusa* F1 (A), *H. elegans* E1 (B), and *H. undulata* HU1 (C).

(A) *H. obtusa* F1

Contig set name	38_Newbler	39_Newbler	SPAdes	CISA	Final contigs
Number of contigs	919	1,447	958	130	91
Total bases of contigs (bp)	2,005,667	3,023,764	1,865,808	1,476,845	1,334,837
Min contig length (bp)	200	200	200	1,095	233
Max contig length (bp)	73,769	73,756	73,973	73,973	73,978
Average contig length (bp)	2,182.5	2,089.7	1,947.6	11,360.4	14,668.5
Median contig length (bp)	803.0	1,126.0	397.5	7,123.5	8,918.0
N50 contig length (bp)	8,743	3,051	8,564	22,124	24,415
G+C content (%)	40.4	45.6	38.3	37.0	35.2

(B) *H. elegans* E1

Contig set name	Newbler	SPAdes 123	SPAdes 127 all	SPAdes 127 remove	CISA	Final contigs
Number of contigs	440	277	550	264	191	152
Total bases of contigs (bp)	1,421,873	1,199,461	1,783,033	1,213,586	1,372,514	1,268,333
Min contig length (bp)	204	201	201	204	586	586
Max contig length (bp)	63,279	41,032	41,043	41,557	74,598	74,598
Average contig length (bp)	3,231.5	4,330.2	3,241.9	4,596.9	7,185.9	8,344.3
Median contig length (bp)	1,056.5	3,336.0	2,101.5	3,556.0	4,209.0	5,980.5
N50 contig length (bp)	9,261	7,204	4,460	7,099	12,630	12,964
G+C content (%)	36.7	36.1	43.0	35.5	36.5	36.0

(C) *H. undulata* HU1

Contig set name	ABYSS	Velvet	SPAdes k55	SPAdes k85	CISA	Final contigs
Number of contigs	452	579	1,084	740	208	203
Total bases of contigs (bp)	1,512,931	1,410,042	1,591,667	1,522,824	1,414,284	1,402,636
Min contig length (bp)	200	200	200	200	1,031	1,065
Max contig length (bp)	33,041	31,335	48,839	32,414	59,427	59,427
Average contig length (bp)	3,347.2	2,435.3	1,468.3	2,057.9	6,799.4	6,909.5
Median contig length (bp)	1,513.5	645.0	315.5	468.0	4,410.5	4,538.0
N50 contig length (bp)	7,206	6,846	6,466	7,599	10,754	10,884
G+C content (%)	36.2	35.9	35.6	36.0	36.1	36.1

Table S3. The 572 core orthologous genes shared by three *Holospora* genomes.

locus_tag	COG			UniProt		
	H.elegans	H.undulata	H.obtusa	H.elegans	H.undulata	H.obtusa
HE1_00004 K737_300134 P618_2000	Predicted DNA i	COG4912L	COG4912L		B3ER71	B3ER71
HE1_00010 K737_300796 P618_2010						
HE1_00011 K737_300795 P618_2010	Asp-tRNAAsn/G COG0154J	COG0154J	COG0154J	Glutamyl-tRNA(D5BRF6	D5BRF6	G5ZWG3
HE1_00012 K737_300794 P618_2010	Asp-tRNAAsn/G COG0721J	COG0721J	COG0721J	Aspartyl/glutar A3EU21	A3EU21	A8FAR9
HE1_00013 K737_300697 P618_2008	Protein chain re COG0216J	COG0216J	COG0216J	Peptide chain r UPI000225B1B	UPI000225B1B	K2EK13
HE1_00039 K737_300329 P618_2002						
HE1_00041 K737_300303 P618_2001	UDP-N-acetylglu COG0766M	COG0766M	COG0766M	UDP-N-acetylglu H6SRD5	H6SRD5	H6SRD5
HE1_00044 K737_301158 P618_2001						
HE1_00045 K737_301157 P618_2001	Integral membr COG0670R	COG0670R	COG0670R	membrane prot UPI000225B1D		H4FAL3
HE1_00049 K737_300566 P618_2003	Uncharacterize COG0728R	COG0728R	COG0728R	Protein MurJ hc K2EGI2	K2EGI2	K2EGI2
HE1_00053 K737_300018 P618_2010	Glucosamine 6- COG0449M	COG0449M	COG0449M	Glutamine--fruc Q2W7U2	Q2W7U2	K2FGB8
HE1_00054 K737_300019 P618_2010	A/G-specific DN COG1194L	COG1194L	COG1194L	A/G-specific ad D7CWZ9	D7CWZ9	F8J950
HE1_00059 K737_300034 P618_2010	Predicted GTPa COG0536R	COG0536R	COG0536R	GTPase Obg J9YZQ3	J9YZQ3	J9YZQ3
HE1_00066 K737_300175 P618_2011	Pyruvate/2-oxo COG1249C	COG1249C	COG1249C	Dihydrolipoyl d G8AQ15	G8AQ15	G8AQ15
HE1_00067 K737_300176 P618_2003	Disulfide bond f		COG1495O	disulphide bonc UPI000225B6D	UPI000225B6D	K2DTR7
HE1_00068 K737_300780 P618_2007						
HE1_00069 K737_300247 P618_2006	Holliday junctio COG2255L	COG2255L	COG2255L	Holliday junctio Q2W2A5	Q2W2A5	Q2W2A5
HE1_00070 K737_300248 P618_2006	Predicted perm COG0628R	COG0628R	COG0628R	Predicted perm H8FNT1	I1XKF5	I1YLG6
HE1_00072 K737_300249 P618_2006	Phosphatidylgly COG0558I	COG0558I	COG0558I	CDP-diacylglyce G0V3U1		K0AZX2
HE1_00082 K737_300429 P618_2000	Asparagine synt COG0367E	COG0367E	COG0367E	Asparagine synt H8FXG3	H8FXG3	B2IYJ4
HE1_00084 K737_301141 P618_2007	Methylase of p COG2890J	COG2890J	COG2890J	Release factor ξ K2HEC1	K2HEC1	A4WRK1
HE1_00085 K737_301140 P618_2009						
HE1_00086 K737_301139 P618_2009						
HE1_00088 K737_300992 P618_2003						
HE1_00098 K737_301020 P618_2001	ATPases involve COG1192D	COG1192D	COG1192D	ATPase Chrom K2EVV3	K2EVV3	G2TCK4
HE1_00099 K737_301021 P618_2001	ABC-type sperr COG1177E	COG1177E	COG1177E	ABC-type trans F2JWX4	F2JWX4	A3YC02
HE1_00100 K737_300430 P618_2008						
HE1_00106 K737_300312 P618_2003	Bacterial cell di COG0772D	COG0772D	COG0772D		K2EHE5	K2EHE5
HE1_00107 K737_300313 P618_2003	UDP-N-acetylm COG0771M	COG0771M	COG0771M	UDP-N-acetylm E4UB27	E4UB27	K2DSU6
HE1_00108 K737_300245 P618_2002	Ribosomal prot COG0261J	COG0261J	COG0261J	50S ribosomal ρ F1Z6W2	F1Z6W2	C9LKW1
HE1_00109 K737_300244 P618_2002	Ribosomal prot COG0211J	COG0211J	COG0211J	50S ribosomal ρ G4E859	F8A8P8	F8A8P8

HE1_00110	K737_300142	P618_2008	Predicted phosphatase	COG0637R	COG0637R	COG0637R	G1X3J9	G2QSG7	G2QSG7	
HE1_00113	K737_300475	P618_2005								
HE1_00114	K737_300474	P618_2001	Response regulator	COG2204T	COG2204T	COG2204T	Nitrogen regulator	B2IB70	B2IB70	E0TBI2
HE1_00115	K737_300473	P618_2001	3-hydroxymyristate	COG0764I	COG0764I	COG0764I	3-hydroxyacyl-CoA synthetase	B6ISU0	B6ISU0	K2JLJ0
HE1_00116	K737_300472	P618_2001	UDP-3-O-acetylglucose	COG1044M	COG1044M	COG1044M	UDP-3-O-acetylglucose	G6F0Q5	G6F0Q5	Q2W4D3
HE1_00117	K737_300471	P618_2001								
HE1_00120	K737_300385	P618_2007								
HE1_00122	K737_300565	P618_2004	CMP-2-keto-3-deoxy-6-phosphogluconate	COG1212M	COG1212M	COG1212M	3-deoxy-mannose-6-phosphate	A0NUL0	A0NUL0	K9HPE2
HE1_00123	K737_300564	P618_2007	Prolyl-tRNA synthetase	COG0442J	COG0442J	COG0442J	prolyl-tRNA synthetase	UPI000225BC54	UPI000225BC54	K2EXC6
HE1_00124	K737_300563	P618_2007	UDP-N-acetylmuramic acid	COG0770M	COG0770M	COG0770M	UDP-N-acetylmuramic acid	I9LJA0	I9LJA0	I9LJA0
HE1_00125	K737_300562	P618_2002	Translation initiation factor	COG0290J	COG0290J	COG0290J	Translation initiation factor	F5RLN3	F5RLN3	F5RLN3
HE1_00126	K737_300561	P618_2002	Threonyl-tRNA synthetase	COG0441J	COG0441J	COG0441J	Threonine--tRNA synthetase	K9GRC1	K9GRC1	H6SRJ6
HE1_00128	K737_300559	P618_2011								
HE1_00129	K737_300558	P618_2011	F ₀ F ₁ -type ATP synthase	COG0055C	COG0055C	COG0055C	ATP synthase stator	UPI000225B705	UPI000225B705	G6A0Q1
HE1_00130	K737_300557	P618_2011	F ₀ F ₁ -type ATP synthase			COG0355C				
HE1_00131	K737_300556	P618_2011	dsRNA-specific endonuclease	COG0571K	COG0571K	COG0571K	Ribonuclease 3	A7HX48	A7HX48	K2FG12
HE1_00132	K737_300555	P618_2011	GTPase	COG1159R	COG1159R	COG1159R	GTPase Era	Q2W512	Q2W512	K2LC06
HE1_00133	K737_300554	P618_2011	1-acyl-sn-glycerol-3-phosphate	COG0204I	COG0204I	COG0204I	Phospholipid/glycerol	G8PIG7		A7IB85
HE1_00135	K737_300836	P618_2006	tRNA-dihydrouridine	COG0042J	COG0042J	COG0042J	Probable tRNA-dihydrouridine	G6EZA8	G6EZA8	Q1RH84
HE1_00136	K737_300835	P618_2006	Signal transduction	COG5000T	COG5000T	COG5000T	Nitrogen regulator	B6IQ43	Q2W4Q3	Q2W4Q3
HE1_00139	K737_300463	P618_2000	Acetyl-CoA carboxylase	COG0777I	COG0777I	COG0777I	Acetyl-coenzyme A carboxylase	K2FGG7	K2FGG7	K2FGG7
HE1_00141	K737_300465	P618_2002	Uncharacterized	COG1968V	COG1968V	COG1968V	Undecaprenyl-capsule	A3W405	D0CSG6	A3W405
HE1_00142	K737_300466	P618_2002	Lipoate-protein	COG0321H	COG0321H	COG0321H	Octanoyltransferase	J1JVW1	K2MNF7	K2EVD4
HE1_00151	K737_300164	P618_2006	DNA-directed RNA polymerase	COG0568K	COG0568K	COG0568K	RNA polymerase	J6J0C4	J6J0C4	B2IDS6
HE1_00152	K737_300165	P618_2011	DNA primase (beta)	COG0358L	COG0358L	COG0358L	DNA primase	K2LHU2	K2LHU2	H8FW02
HE1_00154	K737_300347	P618_2000	Guanylate kinase	COG0194F	COG0194F	COG0194F	Guanylate kinase	D8JYR1	D8JYR1	E1VG03
HE1_00155	K737_300348	P618_2000								
HE1_00156	K737_300349	P618_2000	Aspartate-semialdehyde	COG0136E	COG0136E	COG0136E	Aspartate-semialdehyde	E0XZ82	E0XZ82	I4YZH0
HE1_00157	K737_300350	P618_2008	Parvulin-like protein	COG0760O	COG0760O	COG0760O	Peptidyl-prolyl isomerase			E3EYG2
HE1_00158	K737_300351	P618_2008	Uncharacterized	COG0799S	COG0799S	COG0799S				
HE1_00159	K737_300920	P618_2010	K ⁺ transport system	COG0569P	COG0569P	COG0569P	K ⁺ transport system		B6BPT3	L0IAR6
HE1_00160	K737_300921	P618_2010	Putative Zn-dependent	COG4783R	COG4783R	COG4783R	Peptidase M48, H8FR71	H8FR71	H8FR71	B6ISC4
HE1_00162	K737_300922	P618_2010	Permeases of the			COG0697GER	Transporter, drug	K2FHZ1	K2FHZ1	K2FHZ1
HE1_00163	K737_300923	P618_2010	Ribosomal protein	COG0359J	COG0359J	COG0359J	50S ribosomal protein	K9GY45	K9GY45	G9ZUP8
HE1_00164	K737_300924	P618_2010								

HE1_00165	K737_300925	P618_2010	Ribosomal prot	COG0238J	COG0238J	COG0238J				
HE1_00166	K737_300926	P618_2010	Ribosomal prot	COG0360J	COG0360J	COG0360J	30S ribosomal r	C0DT30	C0DT30	C0DT30
HE1_00167	K737_300927	P618_2001	Predicted trans	COG1475K	COG1475K	COG1475K	Chromosome p	B9NU59	B9NU59	A3TXT6
HE1_00168	K737_300928	P618_2001	Nucleotide-binc	COG0424D	COG0424D	COG0424D	Maf-like proteir	Q989F1	Q989F1	C3X1B2
HE1_00169	K737_300402	P618_2008	Type II secretor	COG1450NU	COG1450NU	COG1450NU		K2EUU3	K2EUU3	K2EUU3
HE1_00177	K737_300069	P618_2006	Permeases of t	COG0477GEPR	COG0477GEPR	COG0477GEPR	Proline/betaine	A8GPJ1	A8GPJ1	A8GPJ1
HE1_00179	K737_300320	P618_2010	4-hydroxybenzc	COG0382H	COG0382H	COG0382H	4-hydroxybenzc	G3ISM0	G3ISM0	K1ZW14
HE1_00181	K737_300987	P618_2010	UDP-N-acetylm	COG0812M	COG0812M	COG0812M	UDP-N-acetyler	A8TR45	A8TR45	A8TR45
HE1_00182	K737_300988	P618_2010	UDP-N-acetylm	COG0773M	COG0773M	COG0773M	UDP-N-acetylm	D2RKRO	D2RKRO	
HE1_00183	K737_300989	P618_2010	UDP-N-acetylgl	COG0707M	COG0707M	COG0707M	UDP-N-acetylgl			B9R1F8
HE1_00186	K737_300804	P618_2000	Uncharacterize			COG5389S		B6JF18	B6JF18	
HE1_00187	K737_300789	P618_2006	Biotin-(acetyl-C	COG0340H	COG0340H	COG0340H	Biotin Syntheta	Q9Z740	Q9Z740	H3ZQU0
HE1_00191	K737_300663	P618_2008	Glutamyl- and g	COG0008J	COG0008J	COG0008J	Glutamate--tRN	Q0AR08	Q0AR08	Q2GDP4
HE1_00204	K737_300431	P618_2009	tRNA delta(2)-is	COG0324J	COG0324J	COG0324J	tRNA dimethyla	Q11HI8	Q11HI8	G5ZWU0
HE1_00205	K737_300778	P618_2003	Ribosome recyc	COG0233J	COG0233J	COG0233J	Ribosome-recyc	D5BR62	G5ZYI0	K2KDB1
HE1_00208	K737_301051	P618_2003								
HE1_00209	K737_301050	P618_2000	Asp-tRNAAsn/G	COG0064J	COG0064J	COG0064J	Aspartyl/glutar	C3PMI9	C3PMI9	D5RQ95
HE1_00221	K737_300178	P618_2008	Pseudouridylat	COG0564J	COG0564J	COG0564J	Pseudouridine s	Q16AB4	D5VG19	E0MJS5
HE1_00225	K737_300871	P618_2011	Uncharacterize	COG3820S	COG3820S	COG3820S		A8GUE2	A8GUE2	G4R9X2
HE1_00226	K737_300872	P618_2009	Tetraacyldisacc	COG1663M	COG1663M	COG1663M	Tetraacyldisacc	K2DWG7	K2DWG7	K2DWG7
HE1_00229	K737_300875	P618_2005	Predicted hydr	COG2945R	COG2945R	COG2945R				
HE1_00230	K737_300876	P618_2007	Malic enzyme	COG0281C	COG0281C	COG0281C	Allosteric NADP	G7UW25	G7UW25	A1WVU3
HE1_00232	K737_300877	P618_2007	Mg/Co/Ni trans	COG2239P	COG2239P	COG2239P	Divalent cation	A4TUW2	K2EJG5	K2EJG5
HE1_00234	K737_300878	P618_2007	Ribosomal prot	COG0291J	COG0291J	COG0291J	50S ribosomal r	K2JRN9	K2JRN9	K2DUG8
HE1_00235	K737_300879	P618_2007	Ribosomal prot	COG0292J	COG0292J	COG0292J	50S ribosomal r	Q3YSX0	Q3YSX0	Q3YSX0
HE1_00236	K737_300880	P618_2007	Uncharacterize	COG3750S	COG3750S	COG3750S		B6ATI7	B6ATI7	B6ATI7
HE1_00237	K737_300881	P618_2007	Predicted ester	COG0400R	COG0400R	COG0400R	Phospholipase/	K9H6P1	K9H6P1	J2GJ61
HE1_00238	K737_300882	P618_2008	Predicted perm	COG0795R	COG0795R		Predicted perm	D3P592	D3P592	D3P592
HE1_00240	K737_300883	P618_2008								
HE1_00242	K737_300260	P618_2009	Pyruvate/2-oxo	COG0022C	COG0022C	COG0022C			K2FF25	K2FF25
HE1_00243	K737_300261	P618_2009	Pyruvate/2-oxo	COG0508C	COG0508C	COG0508C	Dihydrolipoami	Q3YT43	Q3YT43	
HE1_00248	K737_301081	P618_2008	Uncharacterize	COG1619V	COG1619V	COG1619V	Putative carbox	Q92I71	Q92I71	
HE1_00250	K737_301083	P618_2000	Uncharacterize	COG1385S	COG1385S	COG1385S	Ribosomal RNA	G2KQ64	G2KQ64	G2KQ64
HE1_00255	K737_300268	P618_2008								
HE1_00263	K737_300085	P618_2006	Predicted phos	COG0546R	COG0546R	COG0546R	HAD-superfami	K2EHB6	I3TLA1	UPI000225AEF6

HE1_00327 K737_300511 P618_2009								
HE1_00328 K737_300510 P618_2009	Type II secretor	COG1459NU	COG1459NU	COG1459NU	Putative Mannose 6-phosphate 1-phosphotransferase	Q3IFK7	Q3IFK7	C0QQ18
HE1_00329 K737_300509 P618_2009	Type II secretor	COG2804NU	COG2804NU	COG2804NU		K2DV06	K2DV06	K2DV06
HE1_00330 K737_300508 P618_2009								
HE1_00331 K737_300507 P618_2009	Predicted glutamate decarboxylase	COG2071R	COG2071R	COG2071R	Peptidase C26 family	D5RLH5	D5RLH5	H0A8C1
HE1_00335 K737_300809 P618_2006	GTPases - transducin	COG0050J	COG0050J	COG0050J	Elongation factor Tu	H5SH02	H5SH02	H5SH02
HE1_00336 K737_300810 P618_2006	Ribosomal protein L16	COG0051J	COG0051J	COG0051J	30S ribosomal protein L16	Q1D775	Q1D775	Q1D775
HE1_00337 K737_300811 P618_2006	Ribosomal protein L17	COG0087J	COG0087J	COG0087J	50S ribosomal protein L17	Q2N9B1	Q2N9B1	Q2W2J1
HE1_00338 K737_300812 P618_2006	Ribosomal protein L18	COG0088J	COG0088J	COG0088J	50S ribosomal protein L18	K2EI18	K2EI18	K2EI18
HE1_00339 K737_300813 P618_2006	Ribosomal protein L19	COG0089J	COG0089J	COG0089J	50S ribosomal protein L19	E0RQ83	E0RQ83	E3HZS6
HE1_00340 K737_300814 P618_2005	Ribosomal protein L20	COG0090J	COG0090J	COG0090J	50S ribosomal protein L20	A1HQL4	A1HQL4	Q72I07
HE1_00341 K737_300815 P618_2005	Ribosomal protein L21	COG0185J	COG0185J	COG0185J	30S ribosomal protein L21	D5BPY6	D5BPY6	D5BPY6
HE1_00342 K737_300816 P618_2005	Ribosomal protein L22	COG0091J	COG0091J	COG0091J	50S ribosomal protein L22	K2EG79	K2EG79	K2EG79
HE1_00343 K737_300817 P618_2005	Ribosomal protein L23	COG0092J	COG0092J	COG0092J	30S ribosomal protein L23	K9HFT5	K9HFT5	Q2RQW6
HE1_00344 K737_300818 P618_2005	Ribosomal protein L24	COG0197J	COG0197J	COG0197J	50S ribosomal protein L24	F2IUU4	F2IUU4	K2DTD4
HE1_00345 K737_300819 P618_2005								
HE1_00346 K737_300820 P618_2005	Ribosomal protein L25	COG0186J	COG0186J	COG0186J	30S ribosomal protein L25	Q2N9C0	Q2N9C0	Q2N9C0
HE1_00347 K737_300821 P618_2005	Ribosomal protein L26	COG0093J	COG0093J	COG0093J	50S ribosomal protein L26	D3P927	D3P927	Q5X849
HE1_00348 K737_300822 P618_2005	Ribosomal protein L27	COG0198J	COG0198J	COG0198J	50S ribosomal protein L27	D5WS57	D5WS57	D5WS57
HE1_00349 K737_300823 P618_2005	Ribosomal protein L28	COG0094J	COG0094J	COG0094J	50S ribosomal protein L28	Q0AUJ2	Q0AUJ2	Q0AUJ2
HE1_00350 K737_300824 P618_2005	Ribosomal protein L29	COG0199J	COG0199J	COG0199J	30S ribosomal protein L29	B6IRR9	B6IRR9	B6IRR9
HE1_00351 K737_300825 P618_2005	Ribosomal protein L30	COG0096J	COG0096J	COG0096J	30S ribosomal protein L30	H8FN16	H8FN16	D7A967
HE1_00352 K737_300826 P618_2005	Ribosomal protein L31	COG0097J	COG0097J	COG0097J	50S ribosomal protein L31	Q1GPB4	Q1GPB4	Q0BYC9
HE1_00353 K737_300827 P618_2005	Ribosomal protein L32	COG0256J	COG0256J	COG0256J	50S ribosomal protein L32	UPI0001E896B4	UPI0001E896B4	B5YDV9
HE1_00354 K737_300828 P618_2005	Ribosomal protein L33	COG0098J	COG0098J	COG0098J	30S ribosomal protein L33	G5ZX94	G5ZX94	G5ZX94
HE1_00356 K737_300829 P618_2005	Ribosomal protein L34	COG1841J	COG1841J	COG1841J	50S ribosomal protein L34			F4MMF6
HE1_00357 K737_300830 P618_2005	Ribosomal protein L35	COG0200J	COG0200J	COG0200J	50S ribosomal protein L35	B6JEY4	B6JEY4	A5ELK8
HE1_00358 K737_300831 P618_2005	Preprotein translocase	COG0201U	COG0201U	COG0201U	Protein translocase	G5ZX91	G5ZX91	Q1YNG1
HE1_00359 K737_300832 P618_2005	Adenylate kinase	COG0563F	COG0563F	COG0563F	Adenylate kinase	K2EHY5	K2EHY5	K2EHY5
HE1_00363 K737_300162 P618_2005	2-methylthioadenosine synthase	COG0621J	COG0621J	COG0621J	2-methylthioadenosine synthase	A7ICC7	UPI000225BB1F	UPI000225BB1B
HE1_00364 K737_300161 P618_2011	Acetylornithine aminotransferase	COG0624E	COG0624E	COG0624E	Succinyl-diaminopimelate aminotransferase	K2DVT6	K2DVT6	C4GJ26
HE1_00365 K737_300160 P618_2005	Thioredoxin-like protein	COG0694O	COG0694O	COG0694O	NifU domain protein	D5ARE8	D5ARE8	Q9AC07
HE1_00368 K737_300159 P618_2005	Uncharacterized protein	COG2992R	COG2992R	COG2992R	Mannosyl-glycosyltransferase	J9YYF7	J9YYF7	I3TMA6
HE1_00370 K737_300157 P618_2003	Biotin carboxyl carrier protein	COG0511I	COG0511I	COG0511I	Acetyl-CoA carboxylase	E6YVQ6	E6YVQ6	J0Q2Y0
HE1_00373 K737_300155 P618_2011	Translation initiation factor	COG0361J	COG0361J	COG0361J		L8FTM3	L8FTM3	L8FTM3

HE1_00376 K737_300787 P618_2009	Cell division GTI	COG0206D	COG0206D	COG0206D	Cell division prc	UPI000225E77E	UPI000225E77E	D5BPR9
HE1_00377 K737_300786 P618_2009	Actin-like ATPase	COG0849D	COG0849D	COG0849D	Cell division prc	G7Z1Z6	G7Z1Z6	B6IRF6
HE1_00378 K737_300785 P618_2009	Cell division sep	COG1589M	COG1589M	COG1589M	Cell division prc	K2ETE3	K2ETE3	K2ETE3
HE1_00381 K737_300972 P618_2009	DNA segregation	COG1674D	COG1674D	COG1674D		K2EVA1	K2EVA1	
HE1_00382 K737_300973 P618_2009	Ribosomal prot	COG0227J	COG0227J	COG0227J	50S ribosomal r			G6A0J3
HE1_00383 K737_300974 P618_2009	Uridylate kinase	COG0528F	COG0528F	COG0528F	Uridylate kinase	K2EVY2	K2EVY2	K2EVY2
HE1_00384 K737_300975 P618_2009	Translation elong	COG0264J	COG0264J	COG0264J	Elongation factor	Q2GGV3	Q2GGV3	Q5FGZ9
HE1_00385 K737_300976 P618_2009	Ribosomal prot	COG0052J	COG0052J	COG0052J	30S ribosomal r	H8FRB3	H8FRB3	H8FRB3
HE1_00387 K737_300929 P618_2009								
HE1_00388 K737_300930 P618_2009								
HE1_00389 K737_300931 P618_2009	tRNA-(guanine-	COG0336J	COG0336J	COG0336J	tRNA (guanine-	A4ES92	A4ES92	Q1GDC3
HE1_00390 K737_300932 P618_2009	Translation elong	COG0231J	COG0231J	COG0231J	elongation factor	UPI000225A98E	UPI000225A98E	UPI000225A98E
HE1_00391 K737_300933 P618_2010	Predicted ATPase	COG0037D	COG0037D	COG0037D	tRNA(Ile)-lysidir	Q73FR9	Q73FR9	
HE1_00392 K737_300934 P618_2010	dUTPase	COG0756F	COG0756F	COG0756F	Deoxyuridine 5'			I6Z321
HE1_00396 K737_300470 P618_2001	Signal recognition	COG0541U	COG0541U	COG0541U	Signal recognition	Q4UKH4	Q4UKH4	Q4UKH4
HE1_00401 K737_300233 P618_2010	Membrane-fusion		COG0845M	COG0845M	Secretion prote	D2R872	F4QGF8	F4QGF8
HE1_00402 K737_300232 P618_2010	ABC-type unchar	COG4174R	COG4174R	COG4174R	Oligopeptide A	G8AIN0	G8AIN0	
HE1_00403 K737_300231 P618_2010	ABC-type unchar	COG4239R	COG4239R	COG4239R				
HE1_00404 K737_300230 P618_2010	ABC-type unchar	COG4172R	COG4172R	COG4172R	Microcin C ABC	L1LZJ9	L1LZJ9	L1LZJ9
HE1_00413 K737_300053 P618_2006					89-kDa periplas	Q2ABW8	Q2ABW8	Q2ABW8
HE1_00417 K737_300606 P618_2003								
HE1_00418 K737_300605 P618_2004	Exonuclease VII	COG1722L	COG1722L	COG1722L	Exodeoxyribonuc	Q1GR18	Q1GR18	Q1GR18
HE1_00419 K737_300604 P618_2004	Thiol:disulfide is	COG4232OC	COG4232OC	COG4232OC	Thiol:disulfide is	K2ET62	K2ET62	K2ET62
HE1_00420 K737_300727 P618_2008	Pentose-5-phosph	COG0036G	COG0036G	COG0036G	Ribulose-phosph	D0RQR3	D0RQR3	
HE1_00421 K737_300728 P618_2008	Ribose 5-phosph	COG0698G	COG0698G	COG0698G	Ribose 5-phosph	C8X146	C8X146	D1AUS1
HE1_00422 K737_300729 P618_2008	Topoisomerase	COG0550L	COG0550L	COG0550L	DNA topoisomere	G1XWN3	G1XWN3	UPI000225B1E6
HE1_00423 K737_300730 P618_2008	Predicted Rossr	COG0758LU	COG0758LU	COG0758LU	Putative DNA pr	Q4UM41	Q4UM41	Q4UM41
HE1_00425 K737_300732 P618_2008	Ribosomal prot	COG0522J	COG0522J	COG0522J	30S ribosomal r	J7QLY4	J7QLY4	A8I889
HE1_00426 K737_300733 P618_2005	Permeases of th	COG0697GER	COG0697GER	COG0697GER		UPI000225B487	UPI000225B487	UPI000225B487
HE1_00428 K737_300734 P618_2005	Membrane prot	COG0739M	COG0739M		LysM domain p	B6ISM4	K2KE81	G2KRB3
HE1_00429 K737_300735 P618_2005	Seryl-tRNA synt	COG0172J	COG0172J	COG0172J	Serine--tRNA lig	H8FRS5	H8FRS5	H8FRS5
HE1_00435 K737_301144 P618_2004	Adenosylmethic	COG0161H	COG0161H	COG0161H	Ornithine amin	B7P6F6	B7P6F6	B7P6F6
HE1_00436 K737_301145 P618_2010						G2I4E7	G6F398	G6F398
HE1_00437 K737_301146 P618_2010	Tyrosyl-tRNA sy	COG0162J	COG0162J	COG0162J	Tyrosine--tRNA	H6SJN3	H6SJN3	H6SJN3
HE1_00438 K737_301052 P618_2002	Thioredoxin red	COG0492O	COG0492O	COG0492O	Thioredoxin rec			K5ZK49

HE1_00510 K737_300092 P618_2005 SAM-dependen	COG0500QR	COG0500QR	COG0500QR	Ubie_methyltra	C4YZR3	C4YZR3	
HE1_00511 K737_300091 P618_2005 Dethiobiotin syr	COG0132H	COG0132H	COG0132H	ATP-dependent	C4YZR2	C4YZR2	K0PA34
HE1_00513 K737_300979 P618_2011 Ribonucleotide	COG0209F	COG0209F	COG0209F	Ribonucleoside-	B6IX11	B6IX11	Q4FLH6
HE1_00515 K737_300980 P618_2011 Ribonucleotide	COG0208F	COG0208F	COG0208F	Ribonucleoside-	B6IX30	B6IX30	B6IX30
HE1_00518 K737_300982 P618_2002 Valyl-tRNA synt	COG0525J	COG0525J	COG0525J	Valine--tRNA lig	G8ALV7	G8ALV7	G1XXS6
HE1_00520 K737_300520 P618_2007 3-deoxy-D-man	COG1519M	COG1519M	COG1519M	3-deoxy-D-man			I4YV24
HE1_00521 K737_300519 P618_2000 Superoxide disr	COG0605P	COG0605P	COG0605P	Superoxide disr	A3UDL4	A3UDL4	K9SQB8
HE1_00522 K737_300518 P618_2000 Glycosyltransfe	COG0438M	COG0438M	COG0438M	Glycosyl transfe	E2CDJ5	E2CDJ5	A9KFJ3
HE1_00527 K737_300405 P618_2009 ABC-type antim	COG1136V	COG1136V	COG1136V	ABC transporte	F4QGF7	F4QGF7	B0UE61
HE1_00528 K737_300404 P618_2009 ABC-type antim	COG0577V		COG0577V				
HE1_00531 K737_300453 P618_2007 DNA-directed R	COG1758K	COG1758K	COG1758K	DNA-directed R	C0N6P6	C0N6P6	E0MLJ5
HE1_00532 K737_300454 P618_2007 Type IIA topoisc	COG0188L	COG0188L	COG0188L	DNA gyrase sub	UPI000225A8DI	UPI000225A8DI	UPI000225A8DB
HE1_00533 K737_300455 P618_2008 Primosomal prc	COG1198L	COG1198L	COG1198L	primosome ass	UPI000225B45E	UPI000225B45E	UPI000225B458
HE1_00534 K737_300456 P618_2008 Integrase	COG0582L	COG0582L	COG0582L	Tyrosine recom			K2EK11
HE1_00535 K737_300457 P618_2008 Small protein A	COG2913J	COG2913J	COG2913J	Putative outer r	G8AMQ7	G8AMQ7	G2KLU6
HE1_00536 K737_300458 P618_2001 Uncharacterize	COG2859S	COG2859S	COG2859S		Q1H279	Q1H279	
HE1_00539 K737_300461 P618_2001 Predicted S-ade	COG0220R	COG0220R	COG0220R	tRNA (guanine-	H6SMD1	H6SMD1	H8FTP9
HE1_00547 K737_301037 P618_2007 Phospholipid N-	COG3963I	COG3963I	COG3963I		K2FGB1	K2FGB1	K2FGB1
HE1_00548 K737_301036 P618_2007 Glycosyltransfe	COG3306M	COG3306M	COG3306M	Beta-1,4-galact	D0I3Z3	C4WJG7	
HE1_00549 K737_300916 P618_2007					F2ELG7		
HE1_00550 K737_300915 P618_2007							
HE1_00552 K737_300913 P618_2002 Uncharacterize	COG0217S	COG0217S	COG0217S	Probable transc			A7HUZ5
HE1_00553 K737_300912 P618_2002 Lipoproteins	COG0797M	COG0797M	COG0797M	Rare lipoproteir	K2EVS9	K2EVS9	K2EVS9
HE1_00555 K737_300911 P618_2002 D-alanyl-D-alan	COG1686M	COG1686M	COG1686M		K2EHS3	K2EHS3	K2EHS3
HE1_00556 K737_300910 P618_2002 Thymidylate kin	COG0125F	COG0125F	COG0125F	Thymidylate kir	D8JYT6	D8JYT6	D8JYT6
HE1_00557 K737_300909 P618_2002 Holliday junctio	COG0632L	COG0632L	COG0632L	Holliday junctio	I0AGC0	I0AGC0	I0AGC0
HE1_00560 K737_300905 P618_2010					K2B3D9	K2B3D9	K2B3D9
HE1_00561 K737_300904 P618_2010 Preprotein tran			COG0690U				
HE1_00562 K737_300903 P618_2010 Transcription ar	COG0250K	COG0250K	COG0250K	Transcription te	A7HWP9	A7HWP9	A7HWP9
HE1_00563 K737_300902 P618_2010 Ribosomal prot	COG0080J	COG0080J	COG0080J	50S ribosomal ç	A7BRI0	A7BRI0	H8Z309
HE1_00564 K737_300901 P618_2010 Ribosomal prot	COG0081J	COG0081J	COG0081J	50S ribosomal ç	F2J5S1	F2J5S1	H0HV90
HE1_00565 K737_300900 P618_2010 Ribosomal prot	COG0244J	COG0244J	COG0244J	50S ribosomal ç	B6IRP4	B6IRP4	B6IRP4
HE1_00566 K737_300899 P618_2010 Ribosomal prot	COG0222J	COG0222J	COG0222J	50S ribosomal ç	P36247	P36247	K0NJI6
HE1_00567 K737_300898 P618_2010 DNA-directed R	COG0085K	COG0085K	COG0085K	DNA-directed R	B6IRP6	B6IRP6	B6IRP6
HE1_00568 K737_300897 P618_2010 DNA-directed R	COG0086K	COG0086K	COG0086K	DNA-directed R	K2DX89	K2DX89	K2DX89

HE1_00569 K737_300896 P618_2010	Phosphomanno	COG1109G	COG1109G	COG1109G	Phosphoglucosyl	G4E1C6	G4E1C6	
HE1_00570 K737_300895 P618_2010	Phosphopantet	COG0736I	COG0736I	COG0736I				
HE1_00576 K737_300703 P618_2001	DNA uptake lip	COG4105R	COG4105R	COG4105R		K2DJP6	K2DJP6	K2DJP6
HE1_00577 K737_300704 P618_2001	Septum formati			COG2919D				
HE1_00581 K737_301063 P618_2006	Membrane GTP	COG0481M	COG0481M	COG0481M	Elongation fact	K2ELA8	K2ELA8	K2ELA8
HE1_00583 K737_300341 P618_2006	Trypsin-like seri	COG0265O	COG0265O	COG0265O	HtrA protease/c	A8HR77	A8HR77	K9HCB6
HE1_00584 K737_300340 P618_2006	Transcription te	COG0781K	COG0781K	COG0781K	N utilization su	Q89K81	Q89K81	L1Q873
HE1_00585 K737_300339 P618_2002	Dihydrodipicolii	COG0329EM	COG0329EM	COG0329EM	Dihydrodipicolii	F7XVB1	F7XVB1	F7XVB1
HE1_00590 K737_300336 P618_2003	Lipoate synthas	COG0320H	COG0320H	COG0320H	Lipoyl synthase	Q89LR6	Q89LR6	K2K435
HE1_00591 K737_300335 P618_2001	Leucyl aminope	COG0260E	COG0260E	COG0260E	Probable cytos	K2M4V9	K2M4V9	
HE1_00592 K737_300693 P618_2000	Phosphoenolpy	COG0574G	COG0574G	COG0574G	Pyruvate, phos			A4XKV5
HE1_00594 K737_301087 P618_2011	CDP-diglyceride	COG0575I	COG0575I	COG0575I	Phosphatidate	K7ZCY6	Q5FPZ2	C7H142
HE1_00595 K737_301088 P618_2011	Cytosine/adenc	COG0590FJ	COG0590FJ	COG0590FJ	tRNA-specific a	K2DWQ3	K2DWQ3	K2DWQ3
HE1_00596 K737_301089 P618_2011								
HE1_00598 K737_300411 P618_2001	Periplasmic pro	COG0793M	COG0793M	COG0793M	C-terminal proc	F5SZ30	F5SZ30	K2LUV2
HE1_00599 K737_300412 P618_2001	Phosphatidylgly	COG0558I	COG0558I	COG0558I	CDP-diacylglyce	F8E5S1	F8E5S1	F8E5S1
HE1_00604 K737_300193 P618_2000	RimM protein, r	COG0806J	COG0806J	COG0806J	Ribosome matu	B0T563	B0T563	I0FYJ3
HE1_00605 K737_300196 P618_2001	Predicted GTPa	COG0012J	COG0012J	COG0012J	Ribosome-bindi			A5P7D5
HE1_00606 K737_300197 P618_2001	Ribosomal prot	COG0335J	COG0335J	COG0335J	50S ribosomal	K5ZIW1	K5ZIW1	K5ZIW1
HE1_00608 K737_300007 P618_2004	Predicted sugar	COG0794M	COG0794M	COG0794M	Uncharacterize	O67500	O67500	O67500
HE1_00610 K737_300452 P618_2002	Molecular chap	COG0443O	COG0443O	COG0443O	Chaperone prot	Q08UJ1	UPI000225C1FC	UPI000225C1FO
HE1_00612 K737_300451 P618_2002	DnaJ-domain-c	COG1076O	COG1076O	COG1076O	Co-chaperone			A2G406
HE1_00617 K737_300551 P618_2007								
HE1_00618 K737_300550 P618_2007								
HE1_00619 K737_300549 P618_2007	Coproporphyrir	COG0635H	COG0635H	COG0635H	Putative oxyger	A8GV35	A8GV35	
HE1_00620 K737_300548 P618_2007	Methionyl-tRN	COG0143J	COG0143J	COG0143J	Methionine--tR	K2EJU4	K2EJU4	K2EJU4
HE1_00622 K737_300546 P618_2009	NTP pyrophosp	COG0494LR	COG0494LR	COG0494LR	RNA pyrophosp	K2EI56	K2EI56	K2EI56
HE1_00623 K737_300545 P618_2009	Archaeal fructo	COG0483G	COG0483G	COG0483G	inositol monopl	UPI000225A98;	UPI000225A98;	
HE1_00625 K737_300544 P618_2011	ATP-dependent	COG0466O	COG0466O	COG0466O	Lon protease	Q2Y6J2	K9HKG9	
HE1_00626 K737_300543 P618_2011	Uncharacterize	COG1576S	COG1576S	COG1576S	Ribosomal RNA	K2EIP9	K2EIP9	C7MLU1
HE1_00627 K737_300542 P618_2007	ATP/ADP trans	COG3202C	COG3202C	COG3202C	Non-mitochond	Q8KPN2	Q8KPN2	Q8KPN2
HE1_00628 K737_300541 P618_2007	DNA polymeras	COG0847L	COG0847L	COG0847L	DNA polymeras	C6V4W1	C6V4W1	C6V4W1
HE1_00634 K737_300963 P618_2008	UDP-N-acetylm	COG0769M	COG0769M	COG0769M	UDP-N-acetylm	K2EFV8	K2EFV8	K2EFV8
HE1_00635 K737_300964 P618_2008	Stress-induced	COG0271T	COG0271T	COG0271T	BolA-like family	Q0C0U6	Q0C0U6	F4R133
HE1_00636 K737_300965 P618_2008	Phosphoribosyl	COG0152F	COG0152F	COG0152F	Phosphoribosyl	H6SQG2	H6SQG2	A3UDW1

HE1_00637	K737_300966	P618_2006	Uncharacterize	COG1286R	COG1286R	COG1286R	Colicin V produ	Q215U9	Q215U9	Q1QL13
HE1_00638	K737_300967	P618_2003								
HE1_00640	K737_300043	P618_2009	Glutaredoxin-re	COG0278O	COG0278O	COG0278O		I3T6K0	I3T6K0	MOV7Z8
HE1_00641	K737_300042	P618_2009	Site-specific rec	COG4974L	COG4974L	COG4974L	Tyrosine recom	G8PGK9	G8PGK9	G8AI54
HE1_00642	K737_300041	P618_2009	Multimeric flav	COG0655R	COG0655R	COG0655R	FMN reductase	C9NZH2	UPI000248EBOC	
HE1_00645	K737_300667	P618_2006	Uncharacterize	COG2001S	COG2001S	COG2001S	Protein MraZ	K2ETE7	K2ETE7	K2ETE7
HE1_00646	K737_300668	P618_2006	Predicted S-ade	COG0275M	COG0275M	COG0275M	Ribosomal RNA	P60398	P60398	I4YVG0
HE1_00648	K737_300670	P618_2006					Periplasmic pro	A8TR35	A8TR35	
HE1_00649	K737_300671	P618_2006	Glutamate race	COG0796M	COG0796M	COG0796M	Glutamate race	B1XPV8	B1XPV8	K9PZQ5
HE1_00653	K737_300662	P618_2000								
HE1_00654	K737_300109	P618_2003	ATPases with ct	COG0542O	COG0542O	COG0542O	ATP-dependent	G8AT38	G8AT38	K2FGM9
HE1_00658	K737_300443	P618_2003	Kef-type K+ trar	COG0475P	COG0475P	COG0475P		K2DVY3	K2DVY3	K2DVY3
HE1_00659	K737_300444	P618_2003								
HE1_00660	K737_300445	P618_2003	ATPases involve	COG1192D	COG1192D	COG1192D	Chromosome (F	K9HVC8	K9HVC8	K9HVC8
HE1_00661	K737_300446	P618_2005	Signal recogniti	COG0552U	COG0552U	COG0552U	Signal recogniti	K2EH63	K2EH63	K2EH63
HE1_00662	K737_300447	P618_2005	Dimethyladeno	COG0030J	COG0030J	COG0030J	Ribosomal RNA	L9PHV6	A4C996	H8KWR4
HE1_00669	K737_300522	P618_2001	Transaldolase	COG0176G	COG0176G	COG0176G	Probable trans	A9AZ03	A9AZ03	A9AZ03
HE1_00671	K737_300524	P618_2001	Aspartyl-tRNA s	COG0173J	COG0173J	COG0173J	Aspartyl-tRNA s	D9I360	D9I360	D9I360
HE1_00672	K737_300211	P618_2000	Nucleoside dip	COG0105F	COG0105F	COG0105F	Nucleoside dip	A4XY36	E4RSI8	H1GLJ4
HE1_00676	K737_300366	P618_2007	ABC-type mult	COG1132V	COG1132V	COG1132V	ABC-type mult	A8GXJ6	A8GXJ6	A8GXJ6
HE1_00677	K737_300367	P618_2007	Predicted amin	COG0354R	COG0354R	COG0354R				
HE1_00679	K737_300849	P618_2011					5.4 kDa peptide	P94818		P94818
HE1_00680	K737_300848	P618_2011	Pseudouridylat	COG0101J	COG0101J	COG0101J	tRNA pseudouri	B3CMC1	B3CMC1	B3CMC1
HE1_00681	K737_300847	P618_2011	Membrane prot	COG0739M	COG0739M	COG0739M	Peptidase M23	K2KM12	G2KQ43	K2KM12
HE1_00683	K737_300394	P618_2001								
HE1_00684	K737_300322	P618_2002	ABC-type trans	COG0767Q	COG0767Q	COG0767Q	ABC transporte	K7ZD64	F7XV28	F7XV28
HE1_00685	K737_300323	P618_2002	Predicted perip	COG2823R	COG2823R	COG2823R	Transport-assoc	A7HZ52	A7HZ52	
HE1_00689	K737_300503	P618_2009	Outer membrar	COG4775M	COG4775M	COG4775M		K2DSS8	K2DSS8	K2DSS8
HE1_00691	K737_300552	P618_2002	CTP synthase (L	COG0504F	COG0504F	COG0504F	CTP synthase	K2EUJ3	K2EUJ3	K2EUJ3
HE1_00695	K737_300754	P618_2008	Uracil-DNA glyc	COG1573L	COG1573L	COG1573L	DNA polymeras	K2ELY6	K2ELY6	I3TP23
HE1_00696	K737_300755	P618_2008	Ribosomal prot	COG0049J	COG0049J	COG0049J	30S ribosomal	Q2W2I7	Q2W2I7	G1Y3P3
HE1_00697	K737_300756	P618_2008	Ribosomal prot	COG0048J	COG0048J	COG0048J	30S ribosomal	Q2W2I7	A0L5W8	A0L5W8
HE1_00698	K737_300757	P618_2008	Ribosomal prot	COG0103J	COG0103J	COG0103J	30S ribosomal	Q2W2I7	K5XTT4	
HE1_00699	K737_300758	P618_2008	Ribosomal prot	COG0102J	COG0102J	COG0102J	50S ribosomal	Q0FAQ0	J3HV39	A3PK95
HE1_00700	K737_300759	P618_2008	Ribosomal prot	COG0228J	COG0228J	COG0228J	30S ribosomal	Q0FAQ0	A8TYW1	A8TYW1

HE1_00701 K737_300760 P618_2001	Putative transcr	COG1678K	COG1678K	COG1678K	UPF0301 protei	K2LGS8	I3CKN4
HE1_00702 K737_300761 P618_2008	Preprotein tran	COG1862U	COG1862U	COG1862U	Preprotein tran	A9HL42	A0NQN4
HE1_00703 K737_300762 P618_2008	D-alanine-D-ala	COG1181M	COG1181M	COG1181M	D-alanine--D-ala	F8L819	UPI000225C015
HE1_00709 K737_300235 P618_2009	Prolipoprotein (COG0682M	COG0682M	COG0682M	Prolipoprotein (Q9ZE99	Q9ZE99
HE1_00710 K737_300236 P618_2009	Transcription te	COG1158K	COG1158K	COG1158K	Transcription te	G6F1U6	G2I039
HE1_00712 K737_300113 P618_2009	Permeases of t†	COG0697GER	COG0697GER	COG0697GER	S-adenosylmet†	H0U2G9	H0U2G9
HE1_00713 K737_300114 P618_2000							
HE1_00714 K737_300115 P618_2000	Preprotein tran	COG0706U	COG0706U	COG0706U	Membrane prot	K2L193	K2L193
HE1_00715 K737_300116 P618_2000							
HE1_00716 K737_300117 P618_2000	Predicted GTPa	COG0218R	COG0218R	COG0218R	Probable GTP-b	J7QI90	J7QI90
HE1_00718 K737_300118 P618_2000	Permeases of t†		COG0477GEPR	COG0477GEPR	Major facilitato		K2EIL1
HE1_00719 K737_300119 P618_2000	ATP-dependent	COG1074L	COG1074L	COG1074L	ATP-dependent		G4KLD4
HE1_00723 K737_300121 P618_2011	DNA polymeras	COG2927L	COG2927L	COG2927L	DNA polymeras	H8FNZ6	Q2W0Z8
HE1_00724 K737_300122 P618_2011							
HE1_00725 K737_300123 P618_2011	Uncharacterize	COG3494S	COG3494S	COG3494S	Glutamyl-tRNA	A8TUI6	K2DU44
HE1_00726 K737_300124 P618_2011	Apolipoprotein	COG0815M	COG0815M	COG0815M	Apolipoprotein	K2EXI8	G2KPE8
HE1_00727 K737_300125 P618_2011							
HE1_00728 K737_300126 P618_2011	UDP-3-O-acyl-N	COG0774M	COG0774M	COG0774M	UDP-3-O-[3-hyc	E0XYT1	C0QQD8
HE1_00729 K737_300127 P618_2011	NAD-specific gl	COG2902E	COG2902E	COG2902E		A6FJ71	A6FJ71
HE1_00730 K737_300128 P618_2001	F0F1-type ATP (COG0356C	COG0356C	COG0356C	ATP synthase s†	A8TSS1	A8TSS1
HE1_00735 K737_300696 P618_2010	Uncharacterize	COG0779S	COG0779S	COG0779S	Ribosome matu	D3RNU7	K2EIS2
HE1_00736 K737_300695 P618_2010	Transcription el	COG0195K	COG0195K	COG0195K	Transcription te	G8AFH7	D6ZZH9
HE1_00737 K737_300694 P618_2002	Predicted mem	COG0658R	COG0658R	COG0658R	ComEC/Rec2-re	E3I1P8	
HE1_00738 K737_300493 P618_2011	Lauroyl/myristc	COG1560M	COG1560M	COG1560M	Lauroyl/myristc	H8FU01	H8FU01
HE1_00739 K737_301009 P618_2010	Predicted sugar	COG0061G	COG0061G	COG0061G	Probable inorga	C6XLX3	Q1YNC3
HE1_00740 K737_301010 P618_2010	UDP-N-acetyl-D	COG0677M	COG0677M	COG0677M			
HE1_00741 K737_301011 P618_2010	Nucleoside-dip†	COG0451MG	COG0451MG	COG0451MG			
HE1_00746 K737_300884 P618_2010	Cell division prc	COG0768M	COG0768M	COG0768M		K2EHD3	K2EHD3
HE1_00748 K737_300886 P618_2006	Ribosome-bindi	COG0858J	COG0858J				
HE1_00750 K737_300887 P618_2006	Permeases of t†	COG0477GEPR	COG0477GEPR	COG0477GEPR	Major facilitato	K2DX23	D1RK84
HE1_00751 K737_300888 P618_2010	Undecaprenyl p	COG0020I	COG0020I	COG0020I	Isoprenyl transf	F8J9C5	Q98MB9
HE1_00752 K737_300889 P618_2010	DNA polymeras	COG2812L	COG2812L	COG2812L	DNA polymeras	UPI000225C0AI	UPI000225C0AI
HE1_00757 K737_300168 P618_2003	NAD-dependen	COG0272L	COG0272L	COG0272L	DNA ligase	A3UE99	B8ETN6
HE1_00760 K737_300483 P618_2009	Protease subun	COG0740OU	COG0740OU	COG0740OU	ATP-dependent	UPI000225B47E	UPI000225B47E
HE1_00761 K737_300484 P618_2009	ATP-dependent	COG1219O	COG1219O	COG1219O	ATP-dependent	Q11J59	Q11J59

HE1_00762 K737_300485 P618_2007	ABC-type oligoç	COG4166E	COG4166E	COG4166E	ABC transporter		J11J32
HE1_00763 K737_300486 P618_2007	Ribosomal prot	COG0267J	COG0267J	COG0267J	50S ribosomal ç	C3PM21	C3PM21
HE1_00765 K737_300683 P618_2010	Parvulin-like pe	COG07600	COG07600	COG07600	Peptidyl-prolyl	A3UJ90	Q0AQC1
HE1_00766 K737_300682 P618_2010							
HE1_00770 K737_300108 P618_2007	Methionine am	COG0024J	COG0024J	COG0024J	Methionine am	K5Z0G1	Q0BRK1
HE1_00776 K737_300681 P618_2000	Uncharacterizeç	COG4496S	COG4496S	COG4496S		L8XVP6	L8XVP6
HE1_00777 K737_300680 P618_2000	ABC-type uncha	COG2984R	COG2984R	COG2984R		L8Y0D8	L8Y0D8
HE1_00778 K737_300679 P618_2000	ABC-type uncha	COG4120R	COG4120R	COG4120R		L8XYY5	L8XYY5
HE1_00779 K737_300678 P618_2000	ABC-type uncha	COG1101R	COG1101R	COG1101R	Energy-couplinç	L8XYS6	L8XYS6
HE1_00780 K737_300677 P618_2000	Histidyl-tRNA sy	COG0124J	COG0124J	COG0124J	Histidine--tRNA		Q892X7
HE1_00781 K737_300676 P618_2000	Phosphoribosyl	COG0462FE	COG0462FE	COG0462FE			
HE1_00784 K737_300850 P618_2010	Lipid A disaccha	COG0763M	COG0763M	COG0763M	Lipid-A-disaccha	G6EYT3	K7SE98
HE1_00785 K737_300851 P618_2010	NAD/NADP trar	COG3288C	COG3288C	COG3288C	NAD(P) transhy	Q0SD61	COQIR5
HE1_00786 K737_300852 P618_2010	NAD/NADP trar	COG3288C	COG3288C	COG3288C	NAD(P) transhy	H5ST64	H5ST64
HE1_00792 K737_301048 P618_2005	Preprotein tran	COG1952U	COG1952U	COG1952U	Protein-export	H8FTH9	H8FTH9
HE1_00794 K737_301046 P618_2005	1-acyl-sn-glycer	COG0204I	COG0204I	COG0204I	1-acyl-sn-glycer	K2EX71	A9CWH9
HE1_00795 K737_301045 P618_2005	2-methylthioad	COG0621J	COG0621J	COG0621J	(Dimethylallyl)a	K9HGE6	F8JD77
HE1_00796 K737_300745 P618_2005	Predicted hydrç	COG0595R	COG0595R	COG0595R	Metal-depende	D9QI08	B8GWT5
HE1_00797 K737_300744 P618_2010	Response reguli	COG0745TK	COG0745TK	COG0745TK	Two-componen	G1Y4E9	G1Y4E9
HE1_00798 K737_300743 P618_2010						Q3SJT9	D3DIM3
HE1_00799 K737_300742 P618_2010	NAD/NADP trar	COG1282C	COG1282C	COG1282C	NAD(P) transhy	J9Z0R2	A8TJB5
HE1_00800 K737_300741 P618_2010	Trk-type K+ trar	COG0168P	COG0168P	COG0168P	Trk system potç		K7YIU6
HE1_00804 K737_301030 P618_2011	Putative transla	COG0009J	COG0009J	COG0009J	Sua5/YciO/Yrdç	Q5LX83	UPI0002489757
HE1_00806 K737_301028 P618_2003	Transcription el	COG0782K	COG0782K	COG0782K			
HE1_00807 K737_301027 P618_2003	Predicted endoç	COG0816L	COG0816L	COG0816L	Putative Hollida	Q2IXC6	I1B215
HE1_00809 K737_300479 P618_2006							
HE1_00810 K737_300478 P618_2006	Outer membrar	COG1538MU	COG1538MU	COG1538MU	Outer membrar	Q2W526	G6EZ98
HE1_00819 K737_301242 P618_2006	Asparagine synt	COG0367E	COG0367E	COG0367E	Asparagine synt	UPI000225B24E	UPI000225B24E
HE1_00821 K737_300620 P618_2003	Nucleoside pho	COG0775F	COG0775F	COG0775F	AMP nucleosidç	D5QBH0	D5QBH0
HE1_00822 K737_300621 P618_2003	Type IIA topoisc	COG0187L	COG0187L	COG0187L	DNA gyrase sub	A8U128	A8U128
HE1_00824 K737_300622 P618_2003	ABC-type transç	COG1463Q	COG1463Q	COG1463Q	Mce related far	G2KT91	G2KT91
HE1_00825 K737_300024 P618_2011	DNA polymeras	COG0749L	COG0749L	COG0749L	DNA polymeras	K2FGR5	K2FGR5
HE1_00826 K737_300023 P618_2011	Biotin carboxylç	COG0439I	COG0439I	COG0439I	Acetyl-CoA cart	G2ISE2	K9H3E8
HE1_00829 K737_300532 P618_2002	Protein chain rç	COG1186J	COG1186J	COG1186J	Peptide chain rç	K2DTW8	K2DTW8
HE1_00830 K737_300533 P618_2003	Cysteine sulfina	COG1104E	COG1104E	COG1104E	Cysteine desulfç	K2DVY0	K2DVY0

HE1_00831 K737_300534 P618_2003	NifU homolog ir	COG0822C	COG0822C	COG0822C	NifU-like N-tern	L7M1F1	H3E0P7	Q09JH0
HE1_00832 K737_300535 P618_2003	Malate/lactate	COG0039C	COG0039C	COG0039C	Malate dehydr	E8T5Z0	F0S3G2	F0S3G2
HE1_00833 K737_300536 P618_2002	Membrane-fusi	COG0845M	COG0845M	COG0845M	Candidate periç	F5Y0L2	F5Y0L2	H1XX83
HE1_00834 K737_300537 P618_2002	Cation/multidr	COG0841V	COG0841V	COG0841V	Cation/multidr	G8QP91	G7ZAV0	UPI000225B1B4
HE1_00836 K737_300538 P618_2002	Phenylalanyl-tR	COG0016J	COG0016J	COG0016J	Phenylalanine--	K2KL20	K2KL20	K2KL20
HE1_00837 K737_300713 P618_2003	Cysteinyl-tRNA	COG0215J	COG0215J	COG0215J	Cysteine--tRNA	J9DHS7	J9DHS7	B5ZCF1
HE1_00838 K737_300714 P618_2003	Glycosyltransfe	COG3306M	COG3306M	COG3306M	Glycosyltransfe			K2PF12
HE1_00839 K737_300715 P618_2003								
HE1_00841 K737_300718 P618_2003					MotA/TolQ/Ext	K9HM51	K9HM51	K9HM51
HE1_00844 K737_300720 P618_2003	Response reguli	COG0745TK	COG0745TK	COG0745TK	DNA-binding re	K2EJX4	K2EJX4	A5FWT9
HE1_00845 K737_300721 P618_2011								
HE1_00846 K737_300722 P618_2006	Ferredoxin	COG0633C	COG0633C	COG0633C		K8F611	K8F611	K8F611
HE1_00847 K737_300723 P618_2006	Diaminopimelat	COG0253E	COG0253E	COG0253E	diaminopimelat	UPI0002898E7E	UPI0002898E7E	UPI0000E105FD
HE1_00848 K737_300685 P618_2009	2-polyprenyl-3-	COG2227H	COG2227H	COG2227H	3-demethylubic	UPI000262C94/	UPI000262C94/	
HE1_00849 K737_300686 P618_2009	UDP-N-acetylm	COG0472M	COG0472M	COG0472M	Phospho-N-ace	D9QMT0	K9GYY0	I3TVR6
HE1_00851 K737_300502 P618_2001								
HE1_00853 K737_300659 P618_2002	Pseudouridine s	COG0130J	COG0130J	COG0130J	tRNA pseudouri	D9QJ06	D9QJ06	B2IJJ5
HE1_00854 K737_300658 P618_2002	Ribosomal prot	COG0184J	COG0184J	COG0184J	30S ribosomal ç	Q92SW1	Q92SW1	
HE1_00855 K737_300657 P618_2002	Polyribonucleot	COG1185J	COG1185J	COG1185J	Polyribonucleot			G1XXD6
HE1_00856 K737_300656 P618_2002	Ribonucleases (COG1530J	COG1530J	COG1530J	Ribonuclease E	G5ZY04	G5ZY04	G5ZY04
HE1_00859 K737_300959 P618_2001	2-polyprenyl-6-	COG0654HC	COG0654HC	COG0654HC	2-octaprenyl-6-		G9EL60	B0TYS2
HE1_00860 K737_300958 P618_2001								
HE1_00861 K737_300957 P618_2001	Glycerol-3-phos	COG0240C	COG0240C	COG0240C	Glycerol-3-phos	Q2GEH4	F8EU90	F7XX05
HE1_00864 K737_301076 P618_2002	Predicted ATPa	COG0606O	COG0606O	COG0606O		K2FHM7	K2FHM7	K2FHM7
HE1_00865 K737_301077 P618_2002	Preprotein tran	COG0342U	COG0342U	COG0342U	Protein transloc	K2DTQ2	K2DTQ2	K2DTQ2
HE1_00867 K737_301078 P618_2002	Preprotein tran	COG0341U	COG0341U	COG0341U	preprotein tran	UPI000225BAD	UPI000225BAD	UPI000225BADF
HE1_00874 K737_300707 P618_2005								
HE1_00875 K737_300706 P618_2003	Chaperonin Grc	COG0459O	COG0459O	COG0459O	60 kDa chapero	P94820	P94820	P94820
HE1_00876 K737_300705 P618_2003	Co-chaperonin (COG0234O	COG0234O	COG0234O	10 kDa chapero	P94819	P94819	P94819
HE1_00878 K737_300387 P618_2005	Response reguli			COG0745TK				
HE1_00879 K737_300388 P618_2005	Bacterial nuclec	COG0776L	COG0776L	COG0776L	Histone family ç	E8T3H2	E8T3H2	
HE1_00880 K737_300389 P618_2005					Putative PE-PGF	K2FHI8	K2FHI8	K2FHI8
HE1_00884 K737_300302 P618_2003					Internalin-l	L8DSN4	L8DSN4	
HE1_00896 K737_300941 P618_2002	Glycosyltransfe	COG0438M	COG0438M	COG0438M	Glycosyl transfe	K2EKT2	K2EKT2	E1QM80
HE1_00897 K737_300940 P618_2002	Single-stranded	COG0608L	COG0608L	COG0608L	Single-stranded	I3TLZ1	I3TLZ1	I3TLZ1

HE1_00904	K737_300956	P618_2009					K2FG40	K2FG40	K2FG40
HE1_00905	K737_300955	P618_2009	Predicted endo	COG0792L	COG0792L	COG0792L	UPF0102 protei	Q21CJ1	Q21CJ1
HE1_00906	K737_300954	P618_2009							
HE1_00907	K737_300953	P618_2009	Phage-related p	COG4695S	COG4695S	COG4695S	Phage portal pr	UPI000225C13C	UPI000225C13C
HE1_00908	K737_300952	P618_2009	Phage head ma	COG3740R	COG3740R	COG3740R	Phage prohead	J6UB20	J6UB20
HE1_00909	K737_300951	P618_2009	Predicted phag	COG4653R	COG4653R	COG4653R	HK97 family ph	I9CGW5	I9CGW5
HE1_00910	K737_300950	P618_2009							
HE1_00911	K737_300948	P618_2009						K2EK30	K2EK30
HE1_00924	K737_300297	P618_2010	Predicted meta	COG0319R	COG0319R	COG0319R	Endoribonuclea	G7URD7	G7URD7
HE1_00925	K737_300298	P618_2010	Hemolysins and	COG1253R	COG1253R	COG1253R	Putative divaler	G8AG02	G8AG02
HE1_00928	K737_300488	P618_2001	Uncharacterize	COG2121S	COG2121S	COG2121S		K9H5X3	K9H5X3
HE1_00929	K737_300489	P618_2001	Putative Mg2+ ;	COG4536P	COG4536P	COG4536P		K2DVQ8	K2DVQ8
HE1_00930	K737_300490	P618_2001	Shikimate kinas	COG0703E	COG0703E	COG0703E	Shikimate kinas	F3WSN2	F3WSN2
HE1_00931	K737_300491	P618_2001							
HE1_00932	K737_300492	P618_2003	Acyl-[acyl carri	COG1043M	COG1043M	COG1043M	Acyl-(Acyl-carri	B4W7L0	B4W7L0
HE1_00938	K737_300749	P618_2010	Branched-chain	COG0115EH	COG0115EH	COG0115EH	Branched-chain	F7XVU9	F7XVU9
HE1_00939	K737_300750	P618_2010	Uncharacterize	COG0586S	COG0586S	COG0586S		K2DVI7	K2DVI7
HE1_00952	K737_300845	P618_2003	Predicted tRNA	COG0482J	COG0482J	COG0482J	tRNA-specific 2-	A5CFJ6	A5CFJ6
HE1_00953	K737_300844	P618_2003							
HE1_00954	K737_300843	P618_2003	F0F1-type ATP s	COG0711C	COG0711C	COG0711C	ATP synthase s	B6BS78	B6BS78
HE1_00955	K737_300842	P618_2003	F0F1-type ATP s	COG0636C	COG0636C	COG0636C	ATP synthase C		
HE1_00957	K737_300530	P618_2009							
HE1_00959	K737_300354	P618_2001	ABC-type sperr	COG3842E	COG3842E	COG3842E		K2EVU4	K2EVU4
HE1_00963	K737_301060	P618_2000	tmRNA-binding	COG0691O	COG0691O	COG0691O	SsrA-binding pr	P59630	UPI000225B944
HE1_00964	K737_301061	P618_2009	Dioxygenases r	COG2070R	COG2070R	COG2070R		K2EHA9	K2EHA9
HE1_00965	K737_301062	P618_2009						K2FI91	K2FI91
HE1_00967	K737_301085	P618_2000	Permeases of t	COG0477GEPR	COG0477GEPR	COG0477GEPR	Major facilitato	Q5ZXX1	Q5ZXX1
HE1_00979	K737_300607	P618_2006	Predicted ATP-c	COG1066O	COG1066O	COG1066O	DNA repair prot	UPI000225C14C	UPI000225C14C
HE1_00981	K737_300609	P618_2006	Molecular chap	COG0576O	COG0576O	COG0576O	molecular chap	UPI000225A97C	UPI000225A97C
HE1_00982	K737_300610	P618_2006	Transcriptional	COG1420K	COG1420K	COG1420K	Heat-inducible	K2KLR9	K2KLR9
HE1_00983	K737_300611	P618_2000	D-alanyl-D-alan	COG1686M	COG1686M	COG1686M	Penicillin-bindir	Q0FZA7	Q0FZA7
HE1_00984	K737_300612	P618_2009	Recombination;	COG1195L	COG1195L	COG1195L	DNA replication	Q2W9N3	Q2W9N3
HE1_00985	K737_300613	P618_2009	Flagellar motor	COG1360N	COG1360N	COG1360N	OmpA/MotB dc	A5VS68	A5VS68
HE1_00987	K737_300615	P618_2010							
HE1_00989	K737_300617	P618_2002	Membrane cart	COG0744M	COG0744M	COG0744M		K2EXK3	K2EXK3

HE1_00990 K737_300618 P618_2002	Glycyl-tRNA syn	COG0751J	COG0751J	COG0751J	Glycine--tRNA li	K2EI89	G8AFI9	G8AFI9
HE1_00992 K737_300381 P618_2000	Signal peptidas	COG0681U	COG0681U	COG0681U	Signal peptidas	K2EVJ0	K2EVJ0	K2EVJ0
HE1_00993 K737_300380 P618_2000	Nuclease subun	COG0322L	COG0322L	COG0322L	UvrABC system	G8AM02	G8AM02	K2JZL4
HE1_00994 K737_300379 P618_2000	Mg-dependent	COG0084L	COG0084L	COG0084L	Hydrolase, TatC	I2JJV0	I2JJV0	G9ZVR8
HE1_00995 K737_300378 P618_2002	NADH:ubiquino	COG1009CP	COG1009CP	COG1009CP	NADH dehydrog	I2GIR8	I2GIR8	G8US71
HE1_00998 K737_300377 P618_2002	Uncharacterize	COG3002S	COG3002S	COG3002S		D2QHZ1	D2QHZ1	D2QHZ1
HE1_01001 K737_300375 P618_2011	Ribosomal prot	COG0268J	COG0268J	COG0268J	30S ribosomal r	D5BU56	D5BU56	E4TFW0
HE1_01002 K737_300374 P618_2011	ATPase involve	COG0593L	COG0593L	COG0593L	Chromosomal r	H8FMV8	H8FMV8	H8FMV8
HE1_01003 K737_300373 P618_2011	DNA polymeras	COG0592L	COG0592L	COG0592L	DNA polymeras	G6A190	G6A190	Q1GV93
HE1_01004 K737_300372 P618_2009	Phosphatidylse	COG0688I	COG0688I	COG0688I	Phosphatidylse	K2JIH4	K2JIH4	D8JQU8
HE1_01005 K737_300079 P618_2010	Thiol-disulfide i	COG0526OC	COG0526OC	COG0526OC	Thioredoxin	A5CCQ8	A5CCQ8	A5CCQ8
HE1_01006 K737_300078 P618_2010	Uncharacterize	COG1496S	COG1496S	COG1496S	Hypothetical co	B8ENH1	B8ENH1	F2J0F5
HE1_01010 K737_300289 P618_2000	RecA/RadA recc	COG0468L	COG0468L	COG0468L	Protein RecA	C5JAN8	C5JAN8	C5JAN8
HE1_01011 K737_300288 P618_2006	Glutamyl- and g	COG0008J	COG0008J	COG0008J	Glutamate--trN	K2LUA7	K2LUA7	UPI000225BDBE
HE1_01012 K737_300287 P618_2007	ABC-type Mn2+	COG1108P	COG1108P	COG1108P				
HE1_01013 K737_300286 P618_2007	DNA-binding pr	COG1489R	COG1489R	COG1489R	Sugar fermenta			A4F046
HE1_01014 K737_300285 P618_2010	Inactivated sup	COG3893L	COG3893L	COG2887L	Double-strand t	E0TD26	E0TD26	F6EZE7
HE1_01026 K737_300219 P618_2010								
HE1_01027 K737_300220 P618_2010	rRNA methylase	COG0566J	COG0566J	COG0566J	23S rRNA (Guar			F1YTG8
HE1_01028 K737_300221 P618_2010	Ribonuclease H	COG0328L	COG0328L	COG0328L	Ribonuclease H	B8KGL0	B8KGL0	I4C8W0
HE1_01029 K737_300222 P618_2010	Recombination	COG0353L	COG0353L	COG0353L	Recombination	K2JKI4	K2JKI4	D9I3A1
HE1_01030 K737_300223 P618_2010	Uncharacterize	COG0718S	COG0718S	COG0718S	Nucleoid-associ	B3TCV5	B3TCV5	
HE1_01031 K737_300224 P618_2010	ABC-type Mn/Z	COG1121P	COG1121P	COG1121P				
HE1_01032 K737_300225 P618_2010	Ribosomal prot	COG0254J	COG0254J	COG0254J	50S ribosomal r	E3HZ99	E3HZ99	E3HZ99
HE1_01034 K737_300215 P618_2010								
HE1_01036 K737_301022 P618_2010	Signal transduct	COG0642T	COG0642T	COG0642T	Hybrid sensor h	G8ANE6	K2KS51	K2JXJ8
HE1_01038 K737_301023 P618_2010	Type II secretor	COG1450NU	COG1450NU	COG1450NU		K2EK28	K2EK28	K2EK28
HE1_01039 K737_301024 P618_2010	Type II secretor	COG4796U	COG4796U	COG4796U		K2EK28	K2EK28	K2EK28
HE1_01040 K737_300501 P618_2001	Periplasmic con	COG0823U	COG0823U	COG0823U	Protein TolB	K2HCY0	K2HCY0	D5ARB5
HE1_01041 K737_300500 P618_2001	Xaa-Pro amino	COG0006E	COG0006E	COG0006E	Peptidase M24	F7V9N1	F7V9N1	H1XWP0
HE1_01042 K737_300037 P618_2003	Small-conducta	COG0668M	COG0668M	COG0668M	MscS Mechano	C5BLD3	C5BLD3	K2EG92
HE1_01046 K737_300793 P618_2004	Predicted mem	COG0750M	COG0750M	COG0750M				
HE1_01052 K737_301137 P618_2001	Biopolymer trar	COG0848U	COG0848U	COG0848U	Biopolymer trar	A7HV01	A7HV01	Q2W2A8
HE1_01053 K737_301138 P618_2001	Biopolymer trar	COG0811U	COG0811U	COG0811U	Protein TolQ T	H8FNV3	H8FNV3	G9ZVI3
HE1_01057 K737_300711 P618_2001	Permeases of t	COG0477GEPR	COG0477GEPR	COG0477GEPR	MFS transporte	UPI000225A95	UPI000225A95	UPI000225A95

HE1_01058 K737_300710 P618_2001					
HE1_01059 K737_301091 P618_2007	Ribosomal prot	COG0828J	COG0828J	COG0828J	
HE1_01060 K737_301090 P618_2005	Histone acetyltr	COG0454KR	COG0454KR	COG0454KR	Acetyltransfera I7I3Z6
HE1_01061 K737_301175 P618_2000	Fructose-2,6-bi	COG0406G	COG0406G	COG0406G	Phosphoglycer
HE1_01062 K737_301174 P618_2000	Transcriptional	COG1846K	COG1846K		MarR family tra I0QXD2
HE1_01063 K737_301173 P618_2000					A7HIN9
HE1_01064 K737_301172 P618_2005	Transcriptional	COG1846K	COG1846K		Q5WS17
HE1_01076 K737_300207 P618_2007	Isoleucyl-tRNA	COG0060J	COG0060J	COG0060J	Isoleucine--tRN K2FGY8
HE1_01077 K737_300206 P618_2007	FOG: CheY-like	COG0784T	COG0784T	COG2197TK	Response reguli J2LJJ7
HE1_01078 K737_300205 P618_2007	Translation elor	COG0480J	COG0480J	COG0480J	Elongation fact Q2W2I8
HE1_01106 K737_300592 P618_2001	Outer membrar	COG2885M	COG2885M	COG2885M	outer membrar UPI000225AF1/
HE1_01107 K737_300591 P618_2001	Ribosomal prot	COG0099J	COG0099J	COG0099J	30S ribosomal r UPI000225BC8/
HE1_01108 K737_300590 P618_2001	Ribosomal prot	COG0100J	COG0100J	COG0100J	30S ribosomal r B8H4F7
HE1_01109 K737_300589 P618_2001	DNA-directed R	COG0202K	COG0202K	COG0202K	DNA-directed R G2KQT8
HE1_01110 K737_300588 P618_2001	Ribosomal prot	COG0203J	COG0203J	COG0203J	50S ribosomal r B7L0S4
HE1_01112 K737_300587 P618_2001	DnaJ-class mole	COG0484O	COG0484O	COG0484O	Chaperone prot Q9ZDY0
HE1_01113 K737_300586 P618_2001					phosphoglucos UPI000225BC4/
HE1_01114 K737_300585 P618_2001					
HE1_01115 K737_300584 P618_2001	Lipoprotein sigr	COG0597MU	COG0597MU	COG0597MU	Lipoprotein sigr F3L1K1
HE1_01116 K737_300582 P618_2004	ABC-type antim	COG1136V	COG1136V	COG1136V	Lipoprotein rele Q0APW8
HE1_01130 K737_300746 P618_2009	ATPase involve	COG0470L	COG0470L	COG0470L	DNA-directed D G5INT3
HE1_01137 K737_300418 P618_2003	Permeases of tr	COG0477GEPR	COG0477GEPR	COG0477GEPR	Putative transp Q1RHK8
HE1_01138 K737_300419 P618_2003	Methylase invo	COG2226H	COG2226H	COG2226H	Demethylmena I4BAH6
HE1_01139 K737_300420 P618_2002	Multiple antibi	COG2095U	COG2095U	COG2095U	LONFD2
HE1_01140 K737_300421 P618_2002	Dihydrodipicolii	COG0289E	COG0289E	COG0289E	4-hydroxy-tetra C7NV79
HE1_01141 K737_300422 P618_2009	Uncharacterize	COG0316S	COG0316S	COG0316S	Iron-sulfur clust A1TYH3
HE1_01142 K737_300423 P618_2009	Predicted mem	COG0344S	COG0344S	COG0344S	Glycerol-3-phos D9QPY1
HE1_01150 K737_300063 P618_2010	Uncharacterize	COG5323S	COG5323S	COG5323S	K2EVZ3
HE1_01152 K737_300384 P618_2003	Replicative DNA	COG0305L	COG0305L	COG0305L	K2DSK1
HE1_01163 K737_300699 P618_2002	ABC-type multi	COG1132V	COG1132V	COG1132V	ABC-type multi K2KMI5
HE1_01164 K737_300700 P618_2008	Single-stranded	COG0629L	COG0629L	COG0629L	Single-stranded G2KQL9
HE1_01173 K737_300739 P618_2001	Excinuclease AT	COG0178L	COG0178L	COG0178L	Excinuclease AE D8PAT3
HE1_01176 K737_301043 P618_2006	Alanine racema	COG0787M	COG0787M	COG0787M	Alanine racema K2CDY9
HE1_01195 K737_300241 P618_2008	3-oxoacyl-(acyl-	COG0304IQ	COG0304IQ	COG0304IQ	3-oxoacyl-[acyl- K2DSW7
HE1_01196 K737_300240 P618_2008	Acyl carrier pro	COG0236IQ	COG0236IQ	COG0236IQ	

HE1_01203 K737_300648 P618_2004	Predicted amin	COG3173R	COG3173R	COG3173R		G9EPA2	G9EPA2	G9EPA2
HE1_01206 K737_300646 P618_2004	Histone acetyltr	COG0454KR	COG0454KR	COG0454KR		G9EPA3	G9EPA3	G9EPA3
HE1_01207 K737_300645 P618_2004	Histone acetyltr	COG0454KR	COG0454KR	COG0454KR	Acetyltransfera	I2AXK1	I2AXK1	I2AXK1
HE1_01208 K737_300644 P618_2004	Acetyltransfera	COG1670J	COG1670J	COG1670J		G9EPA5	G9EPA5	G9EPA5
HE1_01209 K737_300643 P618_2004	Histone acetyltr	COG0454KR	COG0454KR	COG0454KR	Regulatory prot	I2AXJ9	I2AXJ9	I2AXJ9
HE1_01214 K737_300637 P618_2004	Uncharacterize	COG4994S	COG4994S	COG4994S		C6BGZ5	C6BGZ5	G9EPB2
HE1_01216 K737_300635 P618_2004	Transcriptional	COG1846K	COG1846K	COG1846K				
HE1_01218 K737_300633 P618_2004	Dehydrogenase	COG4221R	COG4221R	COG1028IQR				
HE1_01220 K737_300631 P618_2004						B3ETW6	B3ETW6	K2FHL6
HE1_01223 K737_300629 P618_2004	Uncharacterize	COG2320S	COG2320S	COG2320S		UPI000225B4F8	UPI000225B4F8	B3CRI6
HE1_01224 K737_300628 P618_2004					Acyl-CoA N-acyl	I7I405	I7I405	I7I405
HE1_01226 K737_300626 P618_2004	Predicted meta	COG1878R	COG1878R	COG1878R		G9EPC0	G9EPC0	G9EPC0
HE1_01235 K737_301209 P618_2004	Bacteriophage t	COG5525R	COG5525R	COG5525R	Prophage Lamb		Q73IA7	Q73IA7
HE1_01236 K737_301208 P618_2004								
HE1_01239 K737_300570 P618_2005	Histone acetyltr		COG0454KR	COG0454KR		G9EPG2	G9EPG2	G9EPG2
HE1_01240 K737_301102 P618_2006	Predicted cholir		COG0510M			G9EPG1	G9EPG1	G9EPG1

1 **Supporting Information**

2 **SI Results and Discussion**

3 Bacterial core genomes preferentially contain informational genes, which are
4 involved in replication, transcription, and translation (Charlebois & Doolittle, 2004, Segata
5 & Huttenhower, 2011). In the *Holospira* core genome, a set of 177 genes for “information
6 storage and processing” (Table 1) included those involved in DNA replication (category L),
7 such as replication initiation protein, DNA helicase, DNA gyrase, DNA primase, DNA
8 polymerase III, and DNA ligase. The set included those involved in transcription (category
9 K), such as DNA-directed RNA polymerase subunits and several transcriptional regulators.
10 The set also included those involved in translation (category J), such as aminoacyl-tRNA
11 synthetases, translation initiation and elongation factors, and ribosomal proteins. The
12 informational genes are included in experimentally identified essential genes of
13 *Mycoplasma* genomes (Lin & Zhang, 2011).

14 A set of 138 conserved proteins for “cellular processes and signaling” (Table 1)
15 included chaperones (category O) such as molecular chaperone GrpE (heat shock protein),
16 chaperonin GroEL (HSP60 family) and co-chaperonin GroES (HSP10) (Dohra, *et al.*,
17 1998). The set included actin-like ATPases involved in cell morphogenesis and cell division
18 related Fts (filamentous temperature sensitive) proteins in the category D. The set also
19 included type II secretion system proteins in the category N (cell motility). Protein
20 secretion systems in bacterial symbionts may be involved in the interactions with their
21 hosts (Dale & Moran, 2006).

22 A set of 155 conserved genes for “metabolism” (Table 1) included those involved
23 in transport and metabolism of carbohydrates, amino acids, and inorganic ions (category G,
24 E, and P, respectively). The set also included a pyruvate dehydrogenase complex which
25 converts pyruvate into acetyl-CoA, and F0F1 ATP synthase subunits in the category C
26 (energy production and conversion). Although *Holospira* species rely on the host for their
27 supply of many essential nutrients, they may conduct their own energy production.

28 The 85 conserved genes attributed to the category “poorly characterized” (Table 1)
29 included 23 genes of unknown function (category S), such as those annotated as
30 uncharacterized conserved proteins and predicted membrane proteins. These genes may be
31 further functionally characterized in future experimental studies.

32

33

34 **SI Materials and Methods**

35 **Preparing samples for sequencing *Holospira* genomic DNA.** *Paramecium caudatum*
36 strains used in this study (strain Cd4 bearing *H. obtusa*, strain RB-1s58a2 bearing *H. elegans*,
37 and strain 255 bearing *H. undulata*) were provided by Symbiosis Laboratory, Yamaguchi
38 University with support in part by the National Bio-Resource Project of the Ministry of
39 Education, Culture, Sports, Science and Technology (MEXT). *Holospira*-bearing *P.*
40 *caudatum* clone was grown in a fresh lettuce juice medium with modified Dryl's solution at
41 25 °C (Fujishima, *et al.*, 1990). Non-pathogenic *Klebsiella pneumoniae* cells were
42 inoculated into the medium as food bacteria. Cells of the infectious form of *Holospira*
43 were isolated from homogenates of the host cells at the early stationary phase of growth,
44 using Percoll density gradient centrifugation, and kept at -80 °C until use. The genomic
45 DNA of *H. obtusa* and *H. elegans* was isolated using the Wizard Genomic DNA
46 Purification Kit (Promega), and that of *H. undulata* was isolated using a DNeasy blood and
47 tissue kit (Qiagen).

48

49 **Details on genome assembly.** Reconstructing genome sequences of uncultured bacteria is a
50 challenging and an important task in symbiotic biology. Good draft or even complete
51 genomes can be retrieved from metagenomes (Teeling & Glockner, 2012, Percudani, 2013,
52 Sutcliffe, *et al.*, 2013). The sequencing reads for *H. obtusa* and *H. elegans* were generated
53 by using Roche 454 GS FLX technology (Margulies, *et al.*, 2005), while those for *H.*
54 *undulata* were generated by using Illumina GAIIx technology (Table S1). These raw reads
55 might include contaminating sequences from other organisms with low sequencing depth.
56 Here we describe our computational pipelines to segregate *Holospira* genome sequences
57 from the contaminant sequences.

58 ***H. obtusa* genome assembly.** The draft genome sequence of *H. obtusa* strain F1
59 was generated by a hybrid assembly using the assemblers GS De Novo Assembler
60 (Newbler) version 2.8 and SPAdes version 2.5.1 (Bankevich, *et al.*, 2012), and Contig
61 Integrator for Sequence Assembly (CISA) (Lin & Liao, 2013). Two independently
62 generated 454 GS FLX data-sets, designated as CD38 and CD39, consisted of 431,256
63 reads totaling 145 Mb and 445,222 reads totaling 147 Mb, respectively. The 454 reads were
64 assembled into contigs *de novo* using the Newbler (38_Newbler and 39_Newbler,
65 respectively, in Table S2-A). The reads of CD38 and CD39 were assembled together using
66 the SPAdes with a k-mer size of 123 bp (SPAdes in Table S2-A). The CISA was used to

67 integrate these three assemblies (38_Newbler, 39_Newbler, and SPAdes) into a hybrid
68 contig set (CISA in Table S2-A). The 454 reads were aligned to the 130 hybrid contigs as
69 reference sequences using the software GS Reference Mapper with minimum contig depth
70 of 30 to remove the contigs with very low coverage (Final contigs in Table S2-A). The final
71 genome sequence of *H. obtusa* strain F1 contains 91 contigs consisting of 1,334,837 bp
72 with a G+C content of 35.2%, and an average 219X coverage of the total length of contigs
73 (Table S1).

74 ***H. elegans* genome assembly.** The draft genome sequence of *H. elegans* strain E1
75 was generated by a hybrid assembly using the Newbler, SPAdes, and CISA. 454 GS FLX
76 data, consisted of 187,134 reads totaling 65 Mb, were assembled using the Newbler with
77 screening database option to remove contaminated sequences from *K. pneumoniae* as a
78 component of the culture medium for *P. caudatum* (Newbler in Table S2-B). Sequences
79 used for screening were all available complete chromosomes and plasmids of *K.*
80 *pneumoniae* (accession numbers NC_006625, NC_009648, NC_009649, NC_009650,
81 NC_009651, NC_009652, NC_009653, NC_011281, NC_011282, NC_011283,
82 NC_012731, NC_016838, NC_016839, NC_016840, NC_016841, NC_016845,
83 NC_016846, NC_016847, NC_017540, NC_017541, NC_018522, NC_021231,
84 NC_021232, NC_022078, NC_022082, NC_022083, and NC_022566). Trimmed reads
85 used for assembly with the Newbler were output and assembled using the SPAdes with a
86 k-mer size of 123 bp (SPAdes_123 in Table S2-B). All reads were independently assembled
87 using SPAdes with a k-mer size of 127 bp (SPAdes_127_all in Table S2-B), and aligned to
88 the resulting contigs using the GS Reference Mapper to remove contigs with low coverage
89 (<5) and high G+C content (>50%) (SPAdes_127_remove in Table S2-B). The three
90 assemblies (Newbler, SPAdes_123, and SPAdes_127_remove) were integrated using the
91 CISA (CISA in Table S2-B). The 454 reads were aligned to the resulting 191 contigs using
92 GS Reference Mapper to remove contigs with low sequencing depth; i.e. reads per kilobase
93 per million mapped reads (RPKM) <150 (Final contigs in Table S2-B). The final genome
94 sequence of *H. elegans* strain E1 contains 152 contigs consisting of 1,268,333 bp with a
95 G+C content of 36.0%, and an average 51X coverage of the total length of contigs (Table
96 S1).

97 ***H. undulata* genome assembly.** The draft genome of *H. undulata* strain HU1
98 assembled by ABySS has been reported (Dohra, *et al.*, 2013) (ABySS in Table S2-C). In
99 this study, the draft genome of *H. undulata* was updated by a hybrid assembly using the

100 ABySS, Velvet, SPAdes, and CISA. All the Illumina GA IIx reads were assembled using
101 the Velvet with a k-mer size of 83 bp (Velvet in Table S2-C). A portion (2.5M reads) of the
102 GA IIx reads were assembled using SPAdes with k-mer sizes of 55 and 85 bp (SPAdes_55
103 and SPAdes_85, respectively, in Table S2-C). The four assemblies (ABySS, Velvet,
104 SPAdes_55 and SPAdes_85) were integrated into a hybrid contig set using the CISA (CISA
105 in Table S2-C). All the GA IIx reads were aligned to the hybrid contig set using Bowtie2
106 (Langmead & Salzberg, 2012) to remove contigs with low sequencing depth; i.e. RPKM
107 <40 (Final contigs in Table S2-C). The final genome sequence of *H. undulata* strain HU1
108 contains 203 contigs consisting 1,402,636 bp with a G+C content of 36.1%, and an average
109 4,332X coverage of the total length of contigs (Table S1).

110 Because there is not any finished reference sequence for *Holospora* genomes, we are
111 unable to precisely determine the accuracy of each assembly. Based on statistics for contigs
112 such as the N50 contig size and the number of contigs at least 200 bp long, the updated
113 assembly of HU1 is better than the previously published one. The N50 contig size is larger
114 in the updated assembly of HU1 (10,884 bp) than in the previously published one (7,206).
115 The number of contigs at least 200 bp long is smaller in the updated assembly of HU1
116 (203) than in the previously published one (452).

117

118 **Genome analysis.** We used the G-language Genome Analysis Environment version 1.8.13
119 (<http://www.g-language.org>) (Arakawa, *et al.*, 2003, Arakawa & Tomita, 2006, Arakawa, *et*
120 *al.*, 2008) and R version 3.0.3 for statistical computing (<http://www.R-project.org>)
121 (R_Core_Team, 2014). The draft genome sequences were annotated using the automatic
122 annotation pipeline Prokka 1.8 (<http://vicbioinformatics.com/>) (Seemann, 2014) followed
123 by manual curation of the results. Homologous proteins were identified by the BLAST
124 (Basic Local Alignment Search Tool) program with an E value cutoff of 1e-5, soft filtering
125 and Smith–Waterman final alignment (-F mS -s T options in BLASTP) (Altschul, *et al.*,
126 1997). Each *Holospora* protein was compared by BLAST to the COG (Tatusov, *et al.*,
127 2003) and UniProtKB/Uniref90 (UniProt_Consortium, 2013) protein sequence databases,
128 and was assigned to the functional annotation of the most similar protein sequence in the
129 database. A group of orthologous proteins was built by all-against-all protein sequence
130 comparison using BLAST followed by Markov clustering (MCL) with an inflation factor of
131 1.5 (van Dongen, 2000) using FastOrtho (<http://enews.patricbrc.org/fastortho/>), which is a
132 reimplement of the OrthoMCL program (Li, *et al.*, 2003).

133

134 **References**

- 135 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997)
136 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.
137 *Nucleic Acids Res* **25**: 3389-3402.
- 138 Arakawa K & Tomita M (2006) G-language System as a platform for large-scale analysis
139 of high-throughput omics data. *J Pesticide Sci* **31**: 282-288.
- 140 Arakawa K, Suzuki H & Tomita M (2008) Computational Genome Analysis Using The
141 G-language System. *Genes, Genomes and Genomics* **2**: 1-13.
- 142 Arakawa K, Mori K, Ikeda K, Matsuzaki T, Kobayashi Y & Tomita M (2003) G-language
143 Genome Analysis Environment: a workbench for nucleotide sequence data mining.
144 *Bioinformatics* **19**: 305-306.
- 145 Bankevich A, Nurk S, Antipov D, *et al.* (2012) SPAdes: a new genome assembly algorithm
146 and its applications to single-cell sequencing. *J Comput Biol* **19**: 455-477.
- 147 Charlebois RL & Doolittle WF (2004) Computing prokaryotic gene ubiquity: rescuing the
148 core from extinction. *Genome Res* **14**: 2469-2477.
- 149 Dale C & Moran NA (2006) Molecular interactions between bacterial symbionts and their
150 hosts. *Cell* **126**: 453-465.
- 151 Dohra H, Fujishima M & Ishikawa H (1998) Structure and expression of a
152 GroE-homologous operon of a macronucleus-specific symbiont *Holospora obtusa* of the
153 ciliate *Paramecium caudatum*. *J Eukaryot Microbiol* **45**: 71-79.
- 154 Dohra H, Suzuki H, Suzuki T, Tanaka K & Fujishima M (2013) Draft Genome Sequence of
155 *Holospora undulata* Strain HU1, a Micronucleus-Specific Symbiont of the Ciliate
156 *Paramecium caudatum*. *Genome Announc* **1**: e00664-00613.
- 157 Fujishima M, Nagahara K & Kojima Y (1990) Changes in morphology, buoyant density
158 and protein composition in differentiation from the reproductive short form to the infectious
159 long form of *Holospora obtusa*, a macronucleus-specific symbiont of the ciliate
160 *Paramecium caudatum*. *Zoological Science* **7**: 849-860.
- 161 Langmead B & Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat*
162 *Methods* **9**: 357-359.
- 163 Li L, Stoeckert CJ, Jr. & Roos DS (2003) OrthoMCL: identification of ortholog groups for
164 eukaryotic genomes. *Genome Res* **13**: 2178-2189.
- 165 Lin SH & Liao YC (2013) CISA: contig integrator for sequence assembly of bacterial

166 genomes. *PLoS One* **8**: e60843.

167 Lin Y & Zhang RR (2011) Putative essential and core-essential genes in *Mycoplasma*
168 genomes. *Sci Rep* **1**: 53.

169 Margulies M, Egholm M, Altman WE, *et al.* (2005) Genome sequencing in microfabricated
170 high-density picolitre reactors. *Nature* **437**: 376-380.

171 Percudani R (2013) A Microbial Metagenome (*Leucobacter* sp.) in *Caenorhabditis* Whole
172 Genome Sequences. *Bioinform Biol Insights* **7**: 55-72.

173 R_Core_Team (2014) R: A language and environment for statistical computing. R
174 Foundation for Statistical Computing, Vienna, Austria.

175 Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*.

176 Segata N & Huttenhower C (2011) Toward an efficient method of identifying core genes
177 for evolutionary and functional microbial phylogenies. *PLoS One* **6**: e24704.

178 Sutcliffe B, Midgley DJ, Rosewarne CP, Greenfield P & Li D (2013) Draft Genome
179 Sequence of *Thermotoga maritima* A7A Reconstructed from Metagenomic Sequencing
180 Analysis of a Hydrocarbon Reservoir in the Bass Strait, Australia. *Genome Announc* **1**:
181 e00688-00613.

182 Tatusov RL, Fedorova ND, Jackson JD, *et al.* (2003) The COG database: an updated
183 version includes eukaryotes. *BMC Bioinformatics* **4**: 41.

184 Teeling H & Glockner FO (2012) Current opportunities and challenges in microbial
185 metagenome analysis--a bioinformatic perspective. *Brief Bioinform* **13**: 728-742.

186 UniProt Consortium (2013) Update on activities at the Universal Protein Resource
187 (UniProt) in 2013. *Nucleic Acids Res* **41**: D43-47.

188 van Dongen S (2000) Graph Clustering by Flow Simulation. PhD thesis Thesis, University
189 of Utrecht.

190

191