生物信息学基础与微生物基因组分析

刘翟

信息中心

2012年6月28日

提纲

- 生物信息学概述
- 微生物基因组学分析
- 系统发生与进化分析

生物信息学

Bioinformatics is the application of computer science and information technology to the field of biology and medicine. Bioinformatics deals with algorithms, databases and information systems, web technologies, artificial intelligence and soft computing, information and computation theory, software engineering, data mining, image processing, modeling and simulation, signal processing, discrete mathematics, control and system theory, circuit theory, and statistics, for generating new knowledge of biology and medicine, and improving & discovering new models of computation. *(from Wikipedia)*

算法和软件

- In mathematics and computer science, an algorithm is an effective method expressed as a finite list of well-defined instructions for calculating a function. Algorithms are used for calculation, data processing, and automated reasoning. In simple words an algorithm is a step-by-step procedure for calculations. *(from Wikipedia)*

- Software is a set of programs, procedures, algorithms and its documentation.

算法很难？！怎么办？

- 对于一般的生物信息学应用需求，不一定需要了解算法的本质
- 软件是算法的表现形式，先接触软件再了解算法
- 和身边的生物信息学家探讨，应该采用什么样的软件
- 目前免费的生物信息学算法（软件）能够解决绝大部分生物信息学的需求
- 当现有的工具无法满足你的需求时：
  - 改良现有算法（如，改变参数，进行微调等）
  - 发明新的算法

数据库与应用平台

数据库和网站是生物信息学最常见的应用形式之一

- 建立和使用针对特定领域的生物信息学数据库和信息平台（网站）是生物信息学的研究工作之一

- 建立数据标准是构建数据库的核心工作之一

- 数据库和信息平台（网站）是为生物学工作者服务的，要充分考虑他们的需求和使用习惯

- 数据库和平台建设要面向特定的生物学需求

Applied bioinformatics

- Applying bioinformatics skills onto biological researches
  - Sequence analysis
  - Phylogenetics
  - Protein domain analysis
  - Data mining
  - Genome analysis
  - ……

数据

算法

应用
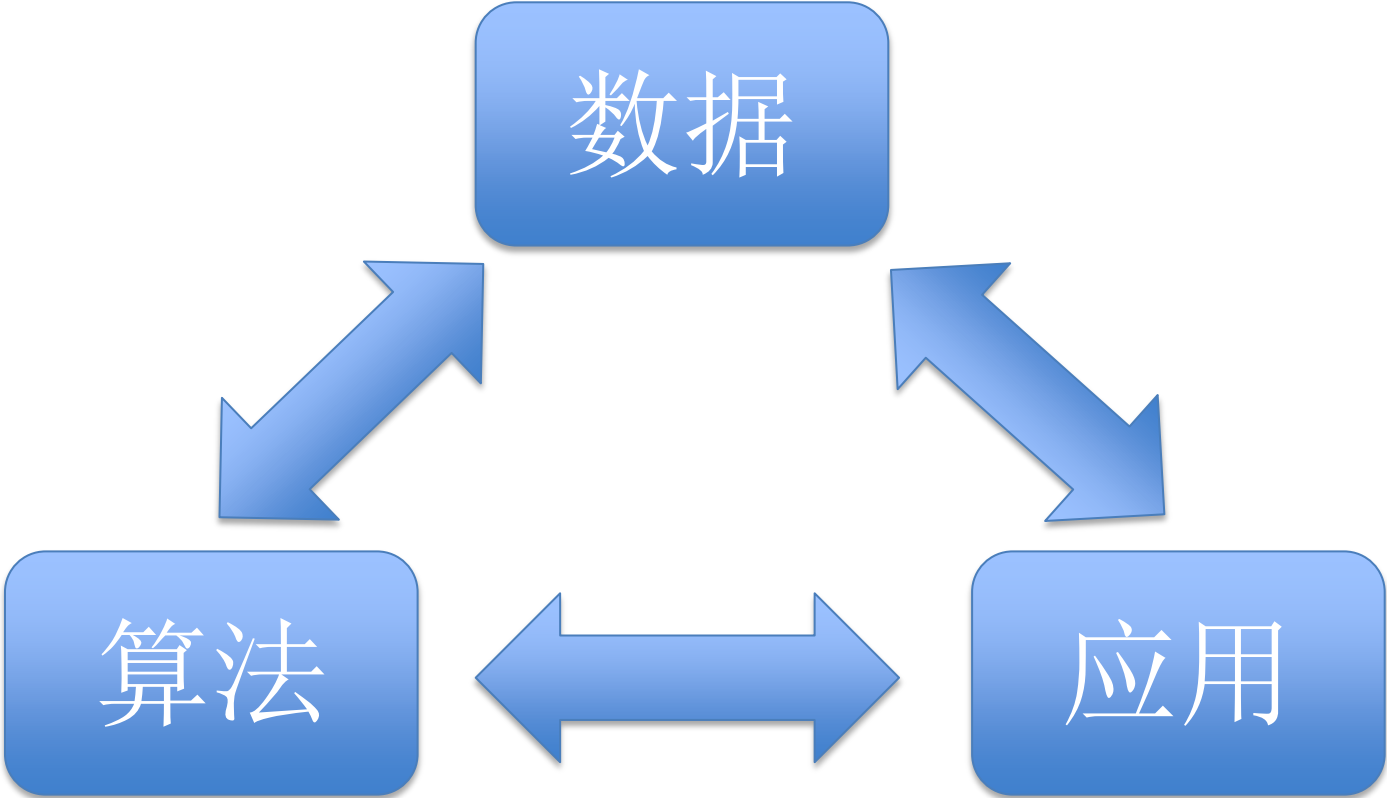
微生物基因组分析

H.hispanica
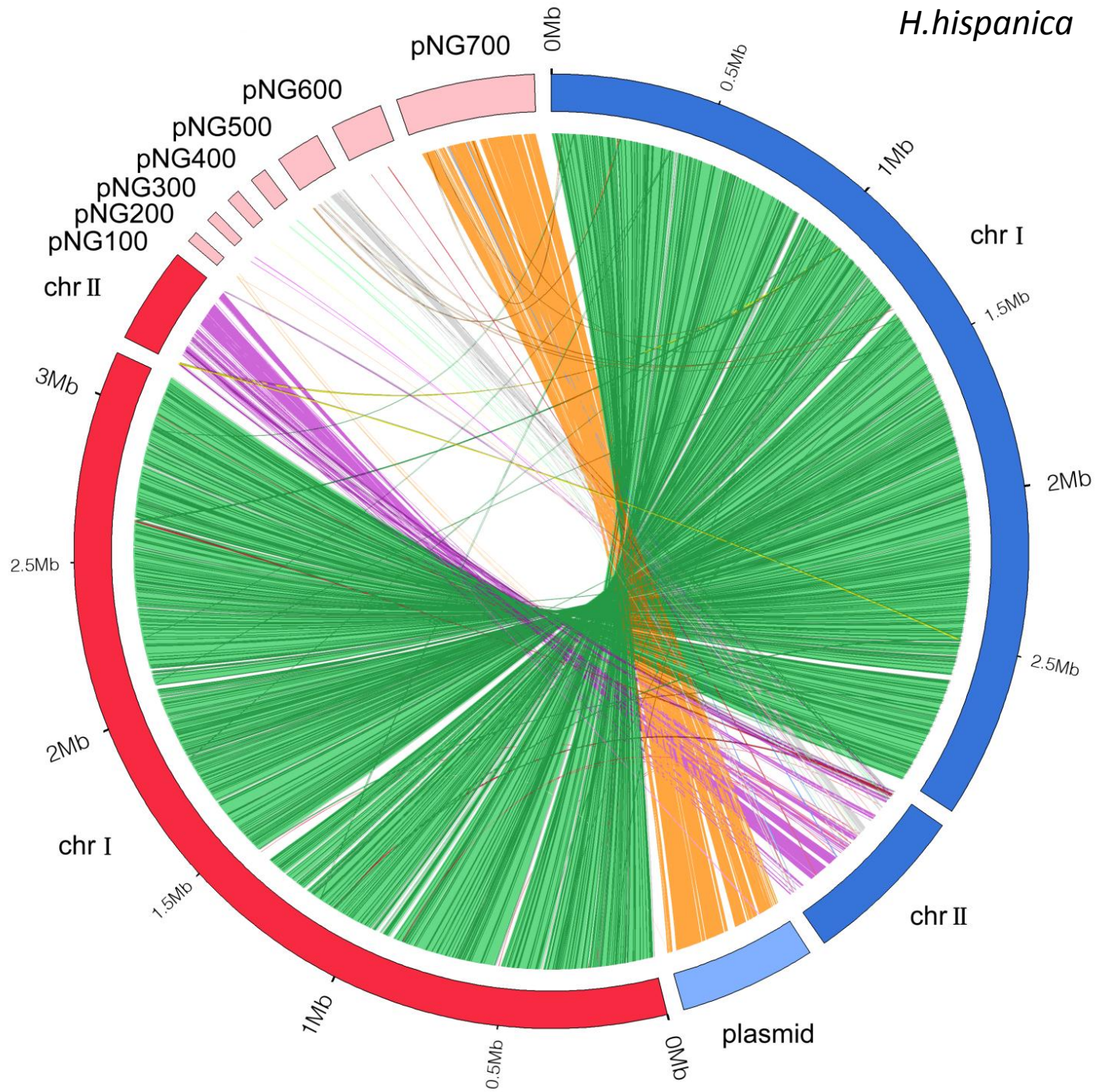
Venn diagram display of homologous proteins between four haloarchaea

Haloarcula hispanica

Haloarcula marismortui ATCC 43049

Halobacterium sp. NRC-1

Haloferax volcanii DS2

Coverage>=60%
Similarity>=60%
E-value<=1e-5

COGs annotation

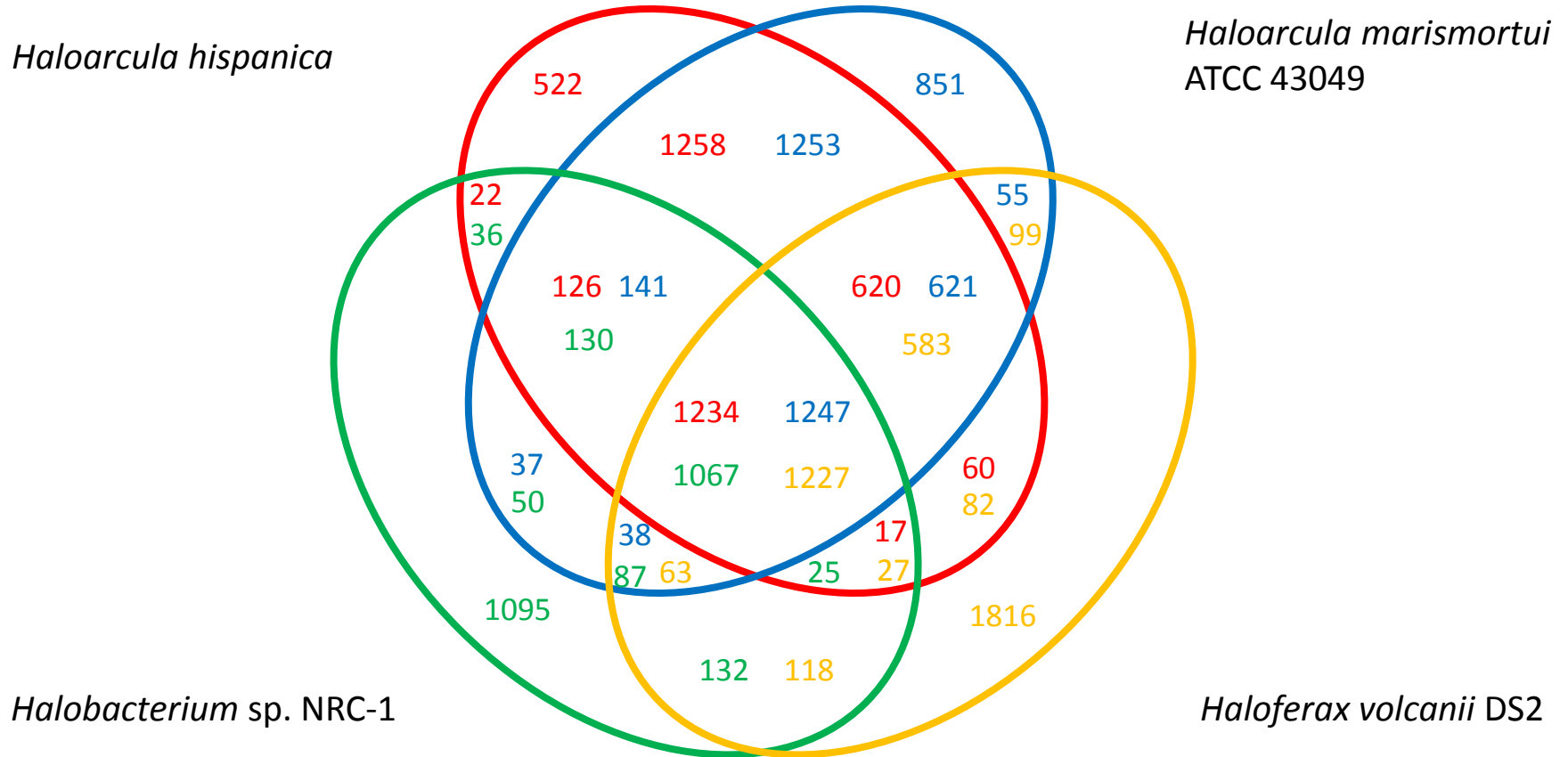| COGs_function	H. hispanica	H. marismortui	H. sp. NRC-1	H. volcanii
[A] RNA processing and modification	0	0	1	1
[B] Chromatin structure and dynamics	3	3	3	3
[C] Energy production and conversion	196	202	269	423
[D] Cell cycle control, cell division, chromosome partitioning	25	29	86	88
[E] Amino acid transport and metabolism	289	280	519	1182
[F] Nucleotide transport and metabolism	67	71	102	120
[G] Carbohydrate transport and metabolism	139	143	164	427
[H] Coenzyme transport and metabolism	128	126	225	328
[I] Lipid transport and metabolism	98	97	90	136
[J] Translation, ribosomal structure and biogenesis	162	161	271	296
[K] Transcription	166	182	239	328
[L] Replication, recombination and repair	150	191	300	365
[M] Cell wall/membrane/envelope biogenesis	108	107	154	246
[N] Cell motility	42	46	64	75
[O] Posttranslational modification, protein turnover, chaperones	125	135	210	254
[P] Inorganic ion transport and metabolism	166	165	487	912
[Q] Secondary metabolites biosynthesis, transport and catabolism	67	62	97	172
[R] General function prediction only	478	494	689	1134
[S] Function unknown	222	212	192	256
[T] Signal transduction mechanisms	145	173	242	477
[U] Intracellular trafficking, secretion, and vesicular transport	32	33	43	72
[V] Defense mechanisms	35	39	108	233

COGs annotation
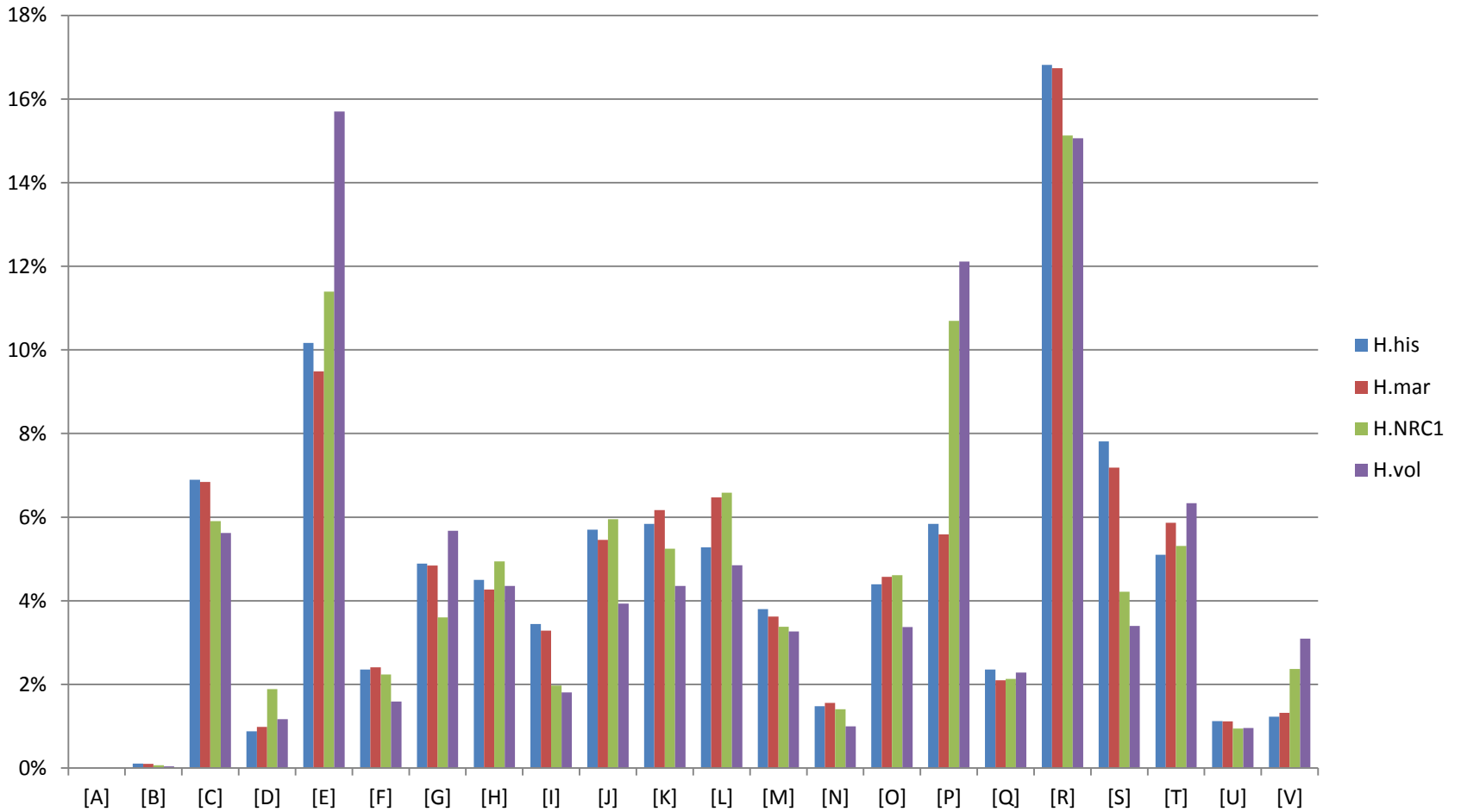
Gene gain and loss(compare with the *H.marismortui* genome) -CELLULAR PROCESSES AND SIGNALING

Gains

Hhis_chr1_orf00123	[M] COG2843 Putative enzyme of poly-gamma-glutamate biosynthesis (capsule formation)
Hhis_chr1_orf00955	[M] COG2335 Secreted and surface protein containing fasciclin-like repeats
Hhis_chr1_orf01616	[M] COG1732 Periplasmic glycine betaine/choline-binding (lipo)protein of an ABC-type transport system (osmoprotectant binding protein)
Hhis_chr1_orf02648	[M] COG0399 Predicted pyridoxal phosphate-dependent enzyme apparently involved in regulation of cell wall biogenesis
Hhis_chr1_orf03002	[O] COG1404 Subtilisin-like serine proteases
Hhis_chr2_orf00071	[V] COG1002 Type II restriction enzyme, methylase subunits
Hhis_chr2_orf00190	[V] COG3440 Predicted restriction endonuclease
Hhis_chr2_orf00287	[T] COG2337 Growth inhibitor
Hhis_chr2_orf00345	[V] COG0286 Type I restriction-modification system methyltransferase subunit
Hhis_plasmid_orf00272	[D] COG2846 Regulator of cell morphogenesis and NO signaling

Losses

Hmar_chr1_YP_136008	[O] COG0755 ABC-type transport system involved in cytochrome c biogenesis, permease component
Hmar_chr1_YP_136002	[O] COG0785 Cytochrome c biogenesis protein
Hmar_chr1_YP_136005	[O] COG1138 Cytochrome c biogenesis factor
Hmar_pNG600_YP_134345	[O] COG2020 Putative protein-S-isoprenylcysteine methyltransferase
Hmar_chr1_YP_136011	[O] COG2332 Cytochrome c-type biogenesis protein CcmE
Hmar_chr1_YP_136004	[O] COG2386 ABC-type transport system involved in cytochrome c biogenesis, permease component
Hmar_chr1_YP_136912	[V] COG1401 GTPase subunit of restriction endonuclease
Hmar_chr1_YP_136477	[V] COG1619 Uncharacterized proteins, homologs of microcin C7 resistance protein MccF
Hmar_pNG600_YP_134389	[V] COG1715 Restriction endonuclease
Hmar_chr1_YP_136547	[V] COG2746 Aminoglycoside N3'-acetyltransferase
Hmar_chr1_YP_136913	[V] COG4268 McrBC 5-methylcytosine restriction system component

Genome Island