

**The genome of obligately intracellular *Ehrlichia canis* reveals themes of complex membrane structure, and immune evasion strategies.**

**Running title: *Ehrlichia canis* genome analysis.**

5 **K. Mavromatis<sup>1,\*</sup>, C. Kuyler Doyle<sup>2</sup>, A. Lykidis<sup>1</sup>, N. Ivanova<sup>1</sup>, M. P. Francino<sup>1</sup>, P. Chain<sup>1,3</sup>, M. Shin<sup>1,3</sup>, S. Malfatti<sup>1,3</sup>, F. Larimer<sup>4</sup>, A. Copeland<sup>1</sup>, J. C. Detter<sup>1</sup>, M. Land<sup>4</sup>, P. M. Richardson<sup>1</sup>, X. J. Yu<sup>2</sup>, D. H. Walker<sup>2</sup>, J. W. McBride<sup>2</sup> and N. C. Kyrpides<sup>1</sup>**

<sup>1</sup>Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598; <sup>2</sup>Department of Pathology, Center for Biodefense and Emerging Infectious Diseases and Sealy Center for  
10 Vaccine Development, University of Texas Medical Branch, Galveston, Texas 77555; <sup>3</sup>Lawrence Livermore National Laboratory, Livermore, California 94550; <sup>4</sup>Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

\*Corresponding author: Mavromatis Konstantinos, Department of Energy, Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, Tel +1-925-296-5815, Fax: +1-925-  
15 296-5666, e-mail: [Kmavrommatis@lbl.gov](mailto:Kmavrommatis@lbl.gov)

## Abstract

*Ehrlichia canis*, a small obligately intracellular, tick-transmitted, gram-negative,  $\alpha$ -proteobacterium, is the primary etiologic agent of globally distributed canine monocytic ehrlichiosis. Complete genome sequencing revealed that the *E. canis* genome consists of a single circular chromosome of 1,315,030 bp predicted to encode 925 proteins, 40 stable RNA species, and 17 putative pseudogenes, and a substantial proportion of non-coding sequence (27 %). Interesting genome features include a large set of proteins with transmembrane helices and/or signal sequences, and a unique serine-threonine bias associated with the potential for O-glycosylation that was prominent in proteins associated with pathogen-host interactions. Furthermore, two paralogous protein families associated with immune evasion were identified, one of which contains poly G:C tracts, suggesting that they may play a role in phase variation and facilitation of persistent infections. Proteins associated with pathogen-host interactions were identified including a small group of proteins (12) with tandem repeats and another with eukaryotic-like ankyrin domains (7).

## Introduction

*Ehrlichia*,  $\alpha$ -proteobacteria that belong to the order *Rickettsiales*, cause diseases of veterinary importance and are also responsible for emerging life-threatening anthroponoses. *E. canis* is a small obligately intracellular, gram-negative, dimorphic bacterium transmitted by the brown dog tick, *Rhipicephalus sanguineus*, that resides as a microcolony within a membrane-lined intracellular vacuole (morula), primarily within monocytes and macrophages of mammalian hosts (20, 26, 58). *E. canis* is the primary etiologic agent of canine monocytic ehrlichiosis (CME) and has a complex life cycle involving ticks and is maintained in nature by persistent infection of wild and domestic canids (25). *E. canis* was first described in 1935 in Algeria (16), and is now recognized to have global distribution, including the United States, Europe, South America and Asia. *E. canis* was reported in the United States in 1963 and received more recognition as a pathogen of veterinary importance following outbreaks in British military dogs in Singapore in 1963, and in United States military dogs in Vietnam, that resulted in approximately 200 deaths over a four year period.

Clinically, *E. canis* infections progress in three phases, the acute, subclinical, and chronic (59, 62). With adequate treatment, dogs typically recover from the acute infections, but untreated or inappropriately treated dogs may develop subclinical persistent infections, and thus, can become asymptomatic carriers of the organism for years (20, 31). Dogs that do not eliminate the infection can develop a severe chronic form of the disease, where bone marrow failure with anemia leads to opportunistic infections, poor response to treatment, and death from massive hemorrhage.

Recent molecular characterization of *E. canis* has identified a limited set of major immunoreactive proteins that include glycoproteins and a major outer membrane protein family containing 25 paralogous genes that could be differentially expressed in the tick and mammalian hosts, contributing to persistent infections of the natural hosts (42, 50). Within the group of known major immunoreactive proteins, three glycoproteins have been identified in *E. canis* (gp36, gp140 and gp200) with corresponding orthologs in the human pathogen, *E. chaffeensis* (18, 43, 47). These glycoproteins are among the first such proteins described in pathogenic bacteria, and appear to be important targets of the host immune response, attachment to the host cell, and other potentially significant roles in ehrlichial pathobiology (52). Genes for a type IV secretion system have been identified in *E. canis* (VirB and VirD). Vir proteins play a key role in pathogen – host cell interactions and are associated with protein secretion and inhibition of bacterial inclusion trafficking to the lysosomes; among them VirB9 is a highly antigenic protein expressed both in the mammalian and tick cells (21, 51).

The genus *Ehrlichia* is closely related to the genera *Rickettsia*, *Anaplasma* and *Wolbachia*, whose members have a similar obligately intracellular existence (19). The genomes of three strains of *E. ruminantium* (13), and several related organisms, namely *A. marginale* (6), *R. conorii* (49), *R. typhi* (44), *R. prowazekii* (3), *R. akari*, *R. sibirica*, *Wolbachia* sp. (from *Drosophila melanogaster*) (66) and *Wolbachia* sp. TRS (from *Brugia malayi*) (22), have been sequenced previously. They are all characterized by small gene numbers and low coding density as a result of convergent reductive evolution in response to their intracellular existence.

The pathogenesis of canine and human ehrlichioses as well as the host response to *Ehrlichia* spp. remains poorly understood. The *E. canis* genome was sequenced to advance understanding

of the pathobiology of this obligately intracellular pathogen and its evolutionary relationship with related organisms and to identify potentially immunoprotective antigens which may be critical for development of effective vaccines.

## **Materials and Methods**

5 **Genome Sequencing and Assembly.** The complete genome of *E. canis* was sequenced at the Joint Genome Institute (JGI) using a combination of 3 kb, 8 kb and fosmid (40 kb) libraries. Due to contamination complexities inherent in isolating genomic DNA from obligately intracellular microbes, four libraries had to be constructed, each with varying amounts of host DNA contamination, and >40,000 reads were generated (roughly 20 – fold coverage), though  
10 only 61% of the reads were incorporated into the final *E. canis* genome assembly. Library construction, sequencing, finishing, and automated annotation steps were used as described previously (9). Predicted coding sequences (CDSs) were subjected to manual analysis using JGI's gene models quality assessment pipeline (available at Microbial Genome Analysis Program's web site). The predicted functions were further analyzed using the Integrated  
15 Microbial Genomes (IMG) annotation pipeline (<http://img.jgi.doe.gov>)(38).

**Genome Analysis.** Comparative analysis of *E. canis* and related organisms (*Ehrlichia*, *Anaplasma* and *Rickettsia* spp.) was performed using a set of tools available in IMG. Unique and orthologous *E. canis* genes were identified using BLASTp (cutoff scores of  $E < 10^{-2}$  and 20% identity and reciprocal hits with cutoffs  $E < 10^{-5}$  and 30%, respectively). Signal peptides and  
20 transmembrane helices were predicted using SignalP 3.0 (4) and TMHMM 2.0 (32) set at default values. Nuclear localization signal (NLS) sequences were predicted using PredictNLS server (12), glycosylation sites were predicted using a eukaryotic prediction server, NetOGlyc 3.1 (30), and

repeats were identified using MUMmer 3.0 (33). Ser/Thr clusters were considered to be regions of 15 tandem amino-acids containing at least 5 Ser/Thr residues. In order to identify proteins with an increased Ser/Thr or Cys content, we calculated the average content of all the proteins of the genome in these residues. Proteins that had a percent content of more than the average plus two standard deviations were considered to be rich in these residues. Riboswitch sequences were checked using the RibEx web server (1).

## Results and Discussion

**General Features of the Genome.** The genome of *E. canis* consists of a single circular chromosome spanning 1,315,030 nucleotides and has a G + C content of 28.96 %. The origin of replication was mapped across a region between kilobases 630 and 670, which exhibits characteristics that are frequently associated with the bacterial termination of replication, such as GC-skew and a large number of duplications and rearrangements. *dnaA* is usually associated with origin of replication, but in *E. canis*, *dnaA* is located 250 kb downstream, consistent with its location in other *Ehrlichia* spp. genomes (13). The *E. canis* genome is smaller than other ehrlichiae which are approximately 1.5 Mb, but it has a similar number of predicted genes. Hence, *E. canis* has a higher density of predicted functional genes, but the ratio of coding to noncoding sequence still remains substantially lower than the average bacterium (Table 1). A total of 984 genes were identified in the *E. canis* genome including one copy of each of the ribosomal RNA genes (5S, 16S and 23S). The 5S and 23S rRNA genes form an operon, while the 16S rRNA gene is separated by ~0.8 Mb. This characteristic is unusual for bacterial genomes, which typically have one to multiple copies of rRNAs contained in 16S-23S-5S operons, but is a common feature of rickettsial organisms of genera *Ehrlichia*, *Rickettsia*, *Anaplasma* and

*Wolbachia*. Thirty-six tRNA genes were identified, which include cognates for all amino acids. Probable function was assigned to 654 genes, and 290 were annotated as coding sequences without function assignment (Figure 1).

This genome has undergone a severe loss of metabolic pathway enzymes, as a result of reductive evolution (45). A basic core of biochemical functions shared by other members of the same genus is dictated by their intracellular lifestyle (Figure 3).

Detailed information about the genome properties and genome annotation can be obtained at [http://img.jgi.doe.gov/pub/main.cgi?page=taxonDetail&taxon\\_oid=623500000](http://img.jgi.doe.gov/pub/main.cgi?page=taxonDetail&taxon_oid=623500000).

**Information transfer.** Genes encoding enzymes for DNA replication and repair, RNA synthesis and degradation, ribosomal proteins (except L30 and S22), UV excision nucleases UvrA and D, as well as the RecA, F, O, R, G recombination synaptic, the RuvABC post-synaptic pathways, and a gene for tmRNA, responsible for tagging incomplete proteins for hydrolysis on stalled ribosomes, were identified. Several enzymes present in *Rickettsia* spp. were absent from *E. canis* and other *Ehrlichia* spp. (e.g., DNA polymerase III, chi subunit, recombination protein N [RecN] and Rnase PH).

In addition, *E. canis* does not have a formamidopyrimidine DNA glycosidase, an enzyme that mediates recovery from mutagenesis or lethal cell injury caused by alkylating agents, and removes an oxidatively damaged form of guanine (36). This gene was identified close to the origin of replication in *E. ruminantium* strains, but the high degree of shuffling of this region may be responsible for the loss of this gene in *E. canis*. This gene is present in other rickettsiae, and the loss of this enzyme and potential for an increased A/T mutational pressure and sensitivity to mutagenic agents could be an area of further investigation.

A set of common chaperonins was also identified in the genome (*groEL*, *groES*, *dnaK*, *dnaJ*, *hslU*, *hslV*, *htpG*), and *secA* protein – dependent and *secA* – independent (type I) protein secretion systems for protein translocation were also present in the organism. In addition, genes for twin arginine translocases (*tatA* and *tatC*) were also identified.

- 5 **Transcriptional Regulation.** Members of the order *Rickettsiales* have a small number of transcriptional regulators, a trait also observed in other intracellular pathogens. In *E. canis*, six transcriptional regulators were identified belonging to the COG groups COG1396, COG1475, COG0789, COG0745, COG5606, COG2932. *E. ruminantium* has one additional transcriptional regulator (COG1959), not present in *E. canis*. *Rickettsia*, *Anaplasma* and *Wolbachia* spp. also  
10 have a small number of transcriptional regulators, but not necessarily representatives of the same groups (Supplementary material - Table 1).

It is known that post-transcriptional regulation based on antisense RNA-mRNA interactions also plays a significant role in gene regulation in bacteria and is mediated by the Sm – like protein Hfq (39), or the presence of metabolite-binding domains in certain mRNA molecules  
15 (riboswitches) (37, 65). However, an *hfq* ortholog was not identified in any of the rickettsial genomes indicating that a similar mechanism does not exist, or that post-transcriptional regulation occurs by a different mechanism in these organisms; furthermore, sequences similar to riboswitches were not identified.

The small number of transcription regulators and the absence of alternative regulation methods,  
20 common features among these organisms, appear to be a result of the reductive evolution coupled to the diminished demand for regulation due to the small genome size, as well as the relatively static conditions provided by the intracellular environment of the host cell (7).



**Central Metabolic Pathways and Transporters.** *E. canis* is an aerobic organism that is unable to use glucose or fructose as carbon or energy source, since transport systems and essential enzymes for the utilization of these substrates were not identified. However, amino acid transporters and enzymes for the utilization of aspartate, proline, glutamate, glutamine and arginine were identified, indicating that amino acids constitute the main energy and carbon source for *E. canis*. All enzymes of the tricarboxylic acid pathway (TCA) were present, as well as enzymes that allow the transfer of amino acids to the TCA cycle. A gluconeogenesis pathway exists, which terminates at 6P-fructose and a complete non-oxidative pentose-phosphate pathway. *E. canis* has biosynthetic pathways for proline, glutamate, glutamine, aspartate, lysine and arginine. A complete set of enzymes involved in the biosynthesis of purines and pyrimidines, as well as sets of enzymes involved in the biosynthesis of lipids and phospholipids, enzymes for the metabolism of cofactors (i.e., riboflavin, nicotinate and nicotinamide, biotin, folate, porphyrin) and a partial pathway for ubiquinone biosynthesis, and genes for enzyme complexes typical of aerobic respiration (i.e., ATP synthase complex, NADH dehydrogenase complex, the cytochrome oxidase complex, and a complete succinate dehydrogenase complex) were also present. Unlike *Rickettsia* spp., *E. canis* does not seem to have translocases for ATP or NADH so it appears to rely on its own intracellular ATP production (3, 24). The genome contains several orthologs involved in membrane transport systems that can supply the necessary metabolites for the absent or incomplete pathways.

**Cell Wall Components.** Enzymes relevant to biosynthesis of lipid A and murein sacculus, and metabolism of peptidoglycans and amino sugars are not present. This finding is in agreement with experimental data suggesting that peptidoglycan and lipopolysaccharide are absent from

the outer membranes of *Ehrlichia* spp.. *E. chaffeensis* and *A. phagocytophilum* acquire cholesterol from the host and incorporate it into the outer membrane possibly to compensate for the decreased membrane stability due to the lack of peptidoglycan (34). Enzymes involved in production or biosynthetic modification of cholesterol were not identified in *E. canis* suggesting  
5 that cholesterol is acquired from the host.

A total of 310 proteins (33 % of the proteome) were predicted to be secreted or contain transmembrane helices, indicating that these proteins could interact to form a complex, protein dominated, cell wall structure to compensate for the absence of the peptidoglycan component. A large number of these proteins (121) had unknown function, similar to that observed in other  
10 *Ehrlichia* spp. A large subset (179) were predicted to have a signal peptide at the N-terminus; however, many (39) did not have predicted transmembrane helices, suggesting that they are secreted.

It is likely that ehrlichial outer membranes are primarily dependent on covalent and non-covalent association between outer membrane proteins for structural rigidity (41, 64). Disulfide  
15 bond formation linkages have been demonstrated between major surface proteins (Msp) in *Anaplasma* spp. (64). A group of 36 proteins appears to contain an increased number of cysteine residues, 15 of which are predicted to be either secreted or membrane-associated. Their increased potential to create disulfide bonds suggests involvement in the formation of a protein envelope around the cell (Supplementary material – Table 2) (41, 64).

20 Proteins with tandem repeats play important roles in pathogenicity and pathogen – host cell interactions (10). Twelve proteins containing tandem repeats were identified (Ecaj\_0017, Ecaj\_0060, Ecaj\_0062, Ecaj\_0109, Ecaj\_0220, Ecaj\_0221, Ecaj\_0387, Ecaj\_0472, Ecaj\_0529,

Ecaj\_0530, Ecaj\_07, Ecaj\_0716). Most of these proteins (except Ecaj\_0062, Ecaj\_0220, Ecaj\_0221, Ecaj\_0715) have orthologs in other genomes from *Rickettsiales* which also contain tandem repeats.

Adaptation of the pathogen to the host, especially in persistent infections, dictates that the organism has mechanisms that allow it to evade host defenses. Most likely these are related to reduction in the host innate and adaptive immune responses, for example by alteration of surface architecture and/or expression of different protein variants. *Ehrlichia* spp. lack peptidoglycan and lipopolysaccharide, major pathogen-associated molecular patterns found in cell walls of other Gram - negative bacteria. This condition suggests that ehrlichial cell walls have fundamental structural and composition differences and, thus, may not be recognized by innate pattern recognition receptors such as toll-like receptors 2 and 4 (35).

**Paralogous Protein Families and Repeats.** Compared to other *Ehrlichia* and *Anaplasma* spp., *E. canis* exhibits a smaller number of short sequence repeats, which are distributed evenly within the genome. Conversely, *E. canis* has a similar number of paralogous protein families (15) compared to *E. ruminantium* strains (str. Welgevonden [CIRAD] 21, str. Gardel 19, str. Welgevonden [ARC-OVI] 13) and a similar number of paralogous genes.

Production of different allelic forms of proteins with tandem repeats could account for surface diversity and avoidance of immune recognition. Similar proteins with tandem repeats exist in other rickettsial genomes, suggesting a similar role and possible function in pathogenicity. Two of these proteins, gp140 and gp36 (Ecaj\_0017 and Ecaj\_0109), are glycosylated and predicted to act as adhesins on the basis of the sequence similarity with proteins of *E. chaffeensis* and *E. ruminantium*, respectively, (15, 43, 67), and a similar mode of action for the others cannot be

excluded. Furthermore, Ecaj\_0062 is similar to WASP family proteins which interact with the cytoskeleton facilitating bacterial motility (23, 29). Although motility has not been demonstrated with *Ehrlichia* spp., this protein could mediate a different type of interaction between the pathogen and the host cell.

5 An operon of 22 paralogous membrane proteins was identified. These are major antigenic proteins that have been previously identified and molecularly characterized, and their role in antigenic diversity and immunoprotection has been investigated (42, 50, 54, 55). Notably, *E. canis* contains more total members (25) in this family than other *Ehrlichia* spp (50). This appears to be a result of a recent duplication that created a new locus containing three  
10 additional p28/p30 genes. The large number of major antigenic proteins certainly could facilitate the antigenic variation observed with these organisms; the differential expression of these proteins could allow the organism to avoid host immune surveillance and establish persistent infections (40, 50, 55, 67). Only a few of these proteins appear to be predominantly expressed in cultured tick and mammalian cells (54), and one p28/30 gene transcript was detected in *E.*  
15 *canis* infected ticks; however, multiple transcripts were detected in infected dogs, suggesting that differential expression may not occur in the mammalian host (63), and thus the p28/p30 proteins may not be involved in immune evasion.

**Ankyrin Domain and Actin Polymerization Proteins.** Ankyrin repeats have been found in numerous proteins mediating specific protein – protein interactions (46). Five proteins that  
20 contain ankyrin domains were identified in the genome (Ecaj\_0052, Ecaj\_0221, Ecaj\_0365, Ecaj\_0387, Ecaj\_0627), which have no assigned function. Three of these proteins (Ecaj\_0221, Ecaj\_0365, and Ecaj\_0387) also exhibited high Ser/Thr content.

One of them, gp200 (Ecaj\_0365), is translocated to the nucleus, where it interacts with the chromatin (47), potentially modulating host cell gene expression, an observation similarly reported with an ankyrin domain-containing protein (AnkA) of *A. phagocytophilum* (8). Interestingly the *Ehrlichia* spp. proteins do not have classic nuclear localization signal (NLS) sequences. This absence supports a different mechanism for protein translocation to the host cell nucleus.

It has been reported that some members of *Rickettsiales* utilize host cell actin polymerization to achieve bacterial motility (23, 29). The actin polymerization depends on activation of the host Arp2/3 complex by the WASP family protein RickA. Although motility has not been demonstrated with *Ehrlichia* spp., Ecaj\_0062 exhibits similarity ( $E=5.10^{-5}$ ) to RickA proteins suggesting that a mechanism of interacting with the cytoskeleton exists in this pathogen.

Furthermore, it has been shown in *Pseudomonas aeruginosa* that AnkB, an ankyrin repeat-containing protein, is essential for optimal activity of periplasmic catalase (28). It is therefore reasonable to propose, that these ankyrin – containing proteins may play a significant role in the adaptation of the pathogen to the host cell, mediating a series of interactions or reprogramming of the host cell functions.

**Poly G:C tracts.** Poly G:C tracts which exhibit variation in their length between otherwise identical clones were identified in seven predicted *E. canis* protein-coding genes; conversely, *E. ruminantium* appears to have only one (13). Five of the poly G:C tract genes (Ecaj\_0063, Ecaj\_0065, Ecaj\_0069, Ecaj\_0070 and Ecaj\_0072) appear to be positioned in a cluster of 12 paralogous proteins unique to *Rickettsiales*, while the other two (Ecaj\_0423 and Ecaj\_0496) are in isolated loci unique to *E. canis*. Among these proteins Ecaj\_0063, Ecaj\_0069 and Ecaj\_0070

are predicted to contain transmembrane helices (Supplementary material, table 2). Thus they can act as membrane-bound antigens, and modulation of their expression or production of variants due to the homopolymeric tracts could increase the diversity of the exposed surface. However, to date none of these proteins has ever been identified as such in *E. canis*, *E.*  
5 *chaffeensis*, or *E. ruminantium*, and their role in immune evasion remains to be elucidated.

**Host Interaction and Pathogenicity-Associated Genes.** In the genome of *E. canis* we were able to identify two clusters of Vir homologous proteins. The first cluster contains *virB8/virB9/virB10/virB11/virD4* while the second one contains *virB3/virB4/virB6* and three large *virB6*-related genes. The organization of *virB* genes and the presence of the three large genes  
10 appears to be common among *Ehrlichia* and *Rickettsia* spp., and they have been associated with the secretion of toxins and intracellular existence (21, 51, 57).

**Serine-Threonine Bias and Ehrlichial Glycoproteins.** Glycosylated proteins can play a role in the pathogen – host interaction, as well as in the assembly of the protein envelope, and could be candidates for a vaccine and therapeutic interventions (15, 61). In recent years, protein  
15 glycosylation pathways of bacteria have been increasingly studied (5, 53, 56, 60, 61).

Glycoproteins have been identified in *E. canis*, and these proteins exhibit high serine/threonine content, suggesting that glycans are attached to these residues by O-linkages. Our analysis revealed 180 proteins that exhibit either clustered Ser/Thr residues, or an increased Ser/Thr content. This group contains proteins that have already been identified as glycoproteins  
20 (gp36/Ecaj\_0109, gp140/Ecaj\_0017, gp200/Ecaj\_0365, and p28/Ecaj\_0917). Many (sixty nine) of these are also predicted to be secreted or contain transmembrane helices. These proteins could appear on the surface of the organism, exhibiting different patterns of glycosylation thus

creating a large number of variants. Interestingly, almost all the proteins with tandem repeats (except Ecaj\_0062 and Ecaj\_0472) appear also to be putative glycoproteins. Similarly, proteins with homopolymeric repeats (Ecaj\_0063, Ecaj\_0070 and Ecaj\_0072), as well as VirB components, also exhibit Ser/Thr clusters. Compared to other genera in the *Rickettsiales*,  
5 *Ehrlichia* and *Anaplasma* spp. possess a significantly higher proportion of Ser/Thr rich proteins. This observation, combined with the absence of the peptidoglycan cell wall content, implies a significant role for glycosylated proteins as cell wall constituents, and potential for antigenic variability of the outer cell surface.

Some of the ehrlichial proteins are predicted to have glycosylation sites based on the eukaryotic  
10 glycoprotein database, suggesting that they can be glycosylated by the host machinery when secreted; however glycosylation of recombinant forms of these proteins is also performed by *E. coli* demonstrating that protein glycosylation patterns are also recognized by the bacterial glycosylation machinery (17). Therefore, it is possible that *Ehrlichia* spp. have enzymes that mediate protein glycosylation.

15 Several proteins that participate in protein glycosylation have been identified in *Campylobacter jejuni*, *Neisseria meningitidis* and other species, while a large number of enzymes are still unknown (61). In several cases, ehrlichial proteins which were annotated using a general function prediction exhibit similarity with genes related to protein glycosylation (ABC transporter similar to WlaB and dehydrogenase similar to PtmA). Furthermore, in *C. jejuni* several proteins  
20 belonging to protein glycosylation pathways appear to be positionally clustered with their target genes (60). Although we were not able to conclusively identify proteins that participate in protein glycosylation pathways in the *E. canis* genome, there are three proteins without any

assigned function (Ecaj\_0897, Ecaj\_0900, Ecaj\_0902) that exhibit low similarity to protein glycosylation-related enzymes (Supplementary material -Table 3) and are present in a locus adjacent to Ecaj\_0917, which has been shown to be glycosylated (55). These proteins form candidates for further research.

5 **Comparison with other rickettsial genomes.** The *E. canis* genome exhibits almost complete synteny with *Ehrlichia ruminantium* (Welgevoden and Gardel strains), and to a lower extent with *A. marginale* (St. Maries strain), while no significant synteny is observed when compared with other members of the *Rickettsiales* (Figure 2). Comparison with the completely sequenced rickettsial genomes is illustrated in Figure 3. *E. canis* shares a core of 834 proteins (88% of the  
10 proteome) with *E. ruminantium* strains and 496 proteins (52% of the proteome) with other complete genomes of the *Rickettsiales*. Thirty proteins corresponding to 3.1% of the proteome, are common among *Ehrlichia*, but orthologs for these are not present in the other members of *Rickettsiales*. Among them Ecaj\_0367, Ecaj\_368, Ecaj\_0541 and Ecaj\_0628 have assigned functions and point to acquired functions in this genus or functions lost from the *Rickettsiales*. *E.*  
15 *canis* contains 72 (7.6% of the proteome) genes that are unique to members of the *Rickettsiales*, 22 of which (2.3% of the proteome) do not exhibit any similarity to other proteins of bacterial origin. The function of all these proteins is unknown.

A small number of insertions/deletions as well as local inversions were identified by comparing those genomes. The first insertion in *E. canis* includes genes Ecaj\_0479 – Ecaj\_0485, the second  
20 includes genes Ecaj\_0716 - Ecaj\_0754, which are predicted to be either secreted or membrane proteins. Finally, Ecaj\_0831 - Ecaj\_0833 represents a unique region in *E. canis* and includes three major antigenic proteins from the *map* locus that appear to be duplicated. This duplication



is identical to a region in the primary *p28/p30* locus, suggesting that this is a recent event. Comparison of *E. canis* with other rickettsial genomes revealed 53 proteins of unknown function unique among the *Rickettsiales*, and 2 that were unique to *E. canis* (Figure 1). Most of these proteins are members of two groups of hypothetical proteins (Eca\_479 - Ecaj\_0485, Ecaj\_0716  
5 – Ecaj\_0772).

**Pseudogenes.** Inspection of the genome of *E. canis* revealed a relatively small number of pseudogenes. Seventeen fragments which correspond to 11 genes were identified. These pseudogenes correspond to full length functional genes present in the *E. canis* genome. In the three strains of *E. ruminantium*, the pseudogenes appear to be consistent and homologous. In  
10 contrast, pseudogenes in the *E. canis* genome are found in different relative positions and correspond to fragments from different genes an indication that this genome has reached a minimum number of genes and is not currently losing genes. It appears that duplication events have led to the expansion of the genome, competing with gene loss as a result of reductive evolution, creating paralogous genes and gene fragments identified as pseudogenes (13). The  
15 fact that these fragments display no homology between ehrlichial genomes suggests that duplication events occurred independently in *Ehrlichia* after the separation of the species.

The intracellular lifestyle, which restricts access of this bacterium to new genes, combined with the absence of any apparent transposable elements, results in low numbers of horizontally transferred genes (2). Although there are genes in *E. canis* without orthologs among other  
20 *Rickettsiales* members, but with homologs in other organisms, we were not able to identify any genes that showed higher similarities to counterparts in organisms from other genera as candidates for horizontal transfer events.

**Evolutionary rates of substitution and duplication.** Considerations of the specific associations of *Anaplasma* and *Ehrlichia* spp. with their tick hosts allow us to use the tick fossil record to calibrate the frequency of appearance of duplicated genes and gene fragments within *E. canis*. Given that all species in the genera *Anaplasma* and *Ehrlichia*, as recently redefined (19) on the basis of 16S rRNA phylogeny, are associated with hard tick hosts, we assume that the phylogenetic clade containing these two genera is not older than the earliest appearance of hard ticks in the fossil record, which has been established to be at 120 MYA (14). Moreover, *Anaplasma* spp. are found within tick hosts belonging to the two existing groups of hard ticks, whereas *Ehrlichia* spp. are restricted to Metastricata ticks (19), which appear in the fossil record no earlier than 40 MYA (14). Therefore, we assume that the *Ehrlichia* genus can not be older than 40 MY, and that this is the oldest possible divergence date for *E. canis* and *E. ruminantium*. This assumption can be corroborated by analysis of 16S rRNA divergence of *A. marginale*, *E. ruminantium* and *E. canis*. The distance (uncorrected p or number of nucleotide differences per site) between *E. canis* and *A. marginale* 16S rRNA is 0.077, while the average distance between *E. canis* and the *E. ruminantium* strains is 0.026, as would be expected for a time of divergence 3 times shorter, assuming equal rates of 16S rRNA evolution.

In comparing the *E. canis* genome to those of the 3 sequenced strains of *E. ruminantium* and to *A. marginale*, we have detected 16 gene paralogs that have appeared by gene duplication within the lineage leading to *E. canis* after its divergence from *E. ruminantium*. Given a maximal divergence time of 40 MY for the two *Ehrlichia* species, this represents a minimal rate of generation of duplicate genes of 0.4 new paralogs per MY. This can be compared to the rate of fixation of gene duplicates in *E. coli*, which has been estimated to be around 2 per MY (27), or 5

times higher. Similarly, the rate of 16S rRNA evolution in *Ehrlichia* (0.48 changes per MY or 1% sequence divergence in 300 MY) is around 6 times slower than the estimated rate for *E. coli* (1% 16S rRNA divergence per 50 MY) (48). The similar decrease of 16S rRNA divergence and rate of fixation of duplicates in *Ehrlichia* suggest that both these rates are responding to common factors. An important factor that could affect both rates is generation time, and, remarkably, the number of generations per year has been estimated to be 6 times less in obligately intracellular bacteria than in *E. coli* (11). In addition, we have detected 17 pseudogenic fragments that have originated by gene duplication within the lineage leading to *E. canis*. These fragments total 3459 bp and represent a minimal rate of generation of duplicated base pairs of 86.5 per MY. This rate can be interpreted as a lower boundary to the rate of neutral duplication for coding bases, since it only accounts for bases that have survived deletion pressure after duplication.

**Conclusions.** The genomic sequence of *E. canis* provides the resources necessary for a detailed analysis of this pathogen and for thorough understanding of the molecular basis of host-pathogen interactions. *Ehrlichia* spp. are Gram - negative bacteria, but are deficient in structural components including peptidoglycan and lipopolysaccharide. The lack of these constituents has resulted in the development of complex protein structures in the outer membrane. Proteins that have a potential role in the adaptive strategy of this organism by assisting immune evasion, as well as proteins that mediate host cell – pathogen interactions, were identified, providing subjects for further research for vaccine and drug development.

## **Acknowledgments**

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley  
5 National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. W-7405-ENG-36, Clayton Foundation for Research and the University of Texas Medical Branch Sealy Center for Vaccine Development

## **Accession numbers**

The sequence data have been deposited in Genbank (CP000107)

## References

1. Abreu-Goodger, C., and E. Merino. 2005. RibEx: a web server for locating riboswitches and other conserved bacterial regulatory elements. *Nucleic Acids Res* 33:W690-2.
2. Andersson, S. G., and C. G. Kurland. 1998. Reductive evolution of resident genomes. Trends Microbiol 6:263-8.
3. Andersson, S. G., A. Zomorodipour, J. O. Andersson, T. Sicheritz-Ponten, U. C. Alsmark, R. M. Podowski, A. K. Naslund, A. S. Eriksson, H. H. Winkler, and C. G. Kurland. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396:133-40.
4. Bendtsen, J. D., H. Nielsen, G. von Heijne, and S. Brunak. 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783-95.
5. Benz, I., and M. A. Schmidt. 2002. Never say never again: protein glycosylation in pathogenic bacteria. *Mol Microbiol* 45:267-76.
6. Brayton, K. A., L. S. Kappmeyer, D. R. Herndon, M. J. Dark, D. L. Tibbals, G. H. Palmer, T. C. McGuire, and D. P. Knowles, Jr. 2004. Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proc Natl Acad Sci U S A*.
7. Cases, I., V. de Lorenzo, and C. A. Ouzounis. 2003. Transcription regulation and environmental adaptation in bacteria. *Trends Microbiol* 11:248-53.
8. Caturegli, P., K. M. Asanovich, J. J. Walls, J. S. Bakken, J. E. Madigan, V. L. Popov, and J. S. Dumler. 2000. *ankA*: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infect Immun* 68:5277-83.
9. Chain, P., J. Lamerdin, F. Larimer, W. Regala, V. Lao, M. Land, L. Hauser, A. Hooper, M. Klotz, J. Norton, L. Sayavedra-Soto, D. Arciero, N. Hommes, M. Whittaker, and D. Arp. 2003. Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. *J Bacteriol* 185:2759-73.
10. Citti, C., M. F. Kim, and K. S. Wise. 1997. Elongated versions of Vlp surface lipoproteins protect *Mycoplasma hyorhinis* escape variants from growth-inhibiting host antibodies. *Infect Immun* 65:1773-85.
11. Clark, M. A., N. A. Moran, and P. Baumann. 1999. Sequence evolution in bacterial endosymbionts having extreme base compositions. *Mol Biol Evol* 16:1586-98.
12. Cokol, M., R. Nair, and B. Rost. 2000. Finding nuclear localization signals. *EMBO Rep* 1:411-5.
13. Collins, N. E., J. Liebenberg, E. P. de Villiers, K. A. Brayton, E. Louw, A. Pretorius, F. E. Faber, H. van Heerden, A. Josemans, M. van Kleef, H. C. Steyn, M. F. van Strijp, E. Zwegarth, F. Jongejan, J. C. Maillard, D. Berthier, M. Botha, F. Joubert, C. H. Corton, N. R. Thomson, M. T. Allsopp, and B. A. Allsopp. 2005. The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. *Proc Natl Acad Sci U S A* 102:838-43.
14. de la Fuente, J. 2003. The fossil record and the origin of ticks (Acari: Parasitiformes: Ixodida). *Exp Appl Acarol* 29:331-44.

15. de la Fuente, J., J. C. Garcia-Garcia, A. F. Barbet, E. F. Blouin, and K. M. Kocan. 2004. Adhesion of outer membrane proteins containing tandem repeats of *Anaplasma* and *Ehrlichia* species (*Rickettsiales*: Anaplasmataceae) to tick cells. *Vet Microbiol* 98:313-22.
16. Donatein, A., and F. Lestoquard. 1935. Existence and Algeria d'une rickettsia du chien. *Bull.Soc.Pathol.Exot.* 28:418-419.
- 5 17. Doyle, C. K., K. A. Nethery, V. L. Popov, and J. W. McBride. 2006. Differentially expressed and secreted major immunoreactive protein orthologs of *Ehrlichia canis* and *E. chaffeensis* elicit early antibody responses to epitopes on glycosylated tandem repeats. *Infect Immun* 74:711-20.
- 10 18. Doyle, C. K., X. Zhang, V. L. Popov, and J. W. McBride. 2005. An immunoreactive 38-kilodalton protein of *Ehrlichia canis* shares structural homology and iron-binding capacity with the ferric ion-binding protein family. *Infect Immun* 73:62-9.
19. Dumler, J. S., A. F. Barbet, C. P. Bekker, G. A. Dasch, G. H. Palmer, S. C. Ray, Y. Rikihisa, and F. R. Rurangirwa. 2001. Reorganization of genera in the families Rickettsiaceae and Anaplasmataceae in the order Rickettsiales: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of *Ehrlichia equi* and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*. *Int J Syst Evol Microbiol* 51:2145-65.
- 15 20. Ewing, S. A. 1963. Observations on leukocytic inclusion bodies from dogs infected with *Babesia canis*. *J Am Vet Med Assoc* 143:503-6.
- 20 21. Felek, S., H. Huang, and Y. Rikihisa. 2003. Sequence and expression analysis of virB9 of the type IV secretion system of *Ehrlichia canis* strains in ticks, dogs, and cultured cells. *Infect Immun* 71:6063-7.
22. Foster, J., M. Ganatra, I. Kamal, J. Ware, K. Makarova, N. Ivanova, A. Bhattacharyya, V. Kapatral, S. Kumar, J. Posfai, T. Vincze, J. Ingram, L. Moran, A. Lapidus, M. Omelchenko, N. Kyrpides, E. Ghedin, S. Wang, E. Goltsman, V. Joukov, O. Ostrovskaya, K. Tsukerman, M. Mazur, D. Comb, E. Koonin, and B. Slatko. 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol* 3:e121.
23. Gouin, E., C. Egile, P. Dehoux, V. Villiers, J. Adams, F. Gertler, R. Li, and P. Cossart. 2004. The RickA protein of *Rickettsia conorii* activates the Arp2/3 complex. *Nature* 427:457-61.
- 30 24. Greub, G., and D. Raoult. 2003. History of the ADP/ATP-translocase-encoding gene, a parasitism gene transferred from a Chlamydiales ancestor to plants 1 billion years ago. *Appl Environ Microbiol* 69:5530-5.
- 35 25. Groves, M. G., G. L. Dennis, H. L. Amyx, and D. L. Huxsoll. 1975. Transmission of *Ehrlichia canis* to dogs by ticks (*Rhipicephalus sanguineus*). *Am J Vet Res* 36:937-40.
26. Hildebrandt, P. K., J. D. Conroy, A. E. McKee, M. B. Nyindo, and D. L. Huxsoll. 1973. Ultrastructure of *Ehrlichia canis*. *Infect Immun* 7:265-71.
27. Hooper, S. D., and O. G. Berg. 2003. On the nature of gene innovation: duplication patterns in microbial genomes. *Mol Biol Evol* 20:945-54.
- 40 28. Howell, M. L., E. Alsabbagh, J. F. Ma, U. A. Ochsner, M. G. Klotz, T. J. Beveridge, K. M. Blumenthal, E. C. Niederhoffer, R. E. Morris, D. Needham, G. E. Dean, M. A. Wani, and D. J. Hasset. 2000. AnkB, a periplasmic ankyrin-like protein in *Pseudomonas aeruginosa*, is

- required for optimal catalase B (KatB) activity and resistance to hydrogen peroxide. *J Bacteriol* 182:4545-56.
29. Jeng, R. L., E. D. Goley, J. A. D'Alessio, O. Y. Chaga, T. M. Svitkina, G. G. Borisy, R. A. Heinzen, and M. D. Welch. 2004. A Rickettsia WASP-like protein activates the Arp2/3 complex and mediates actin-based motility. *Cell Microbiol* 6:761-9.
  30. Julenius, K., A. Molgaard, R. Gupta, and S. Brunak. 2005. Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. *Glycobiology* 15:153-64.
  31. Kelch, W. J. 1984. The canine ehrlichiosis (tropical canine pancytopenia) epizootic in Vietnam and its implications for the veterinary care of military working dogs. *Mil Med* 149:327-31.
  32. Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567-80.
  33. Kurtz, S., A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S. L. Salzberg. 2004. Versatile and open software for comparing large genomes. *Genome Biol* 5:R12.
  34. Lin, M., and Y. Rikihisa. 2003. *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* lack genes for lipid A biosynthesis and incorporate cholesterol for their survival. *Infect Immun* 71:5324-31.
  35. Lin, M., and Y. Rikihisa. 2004. *Ehrlichia chaffeensis* downregulates surface Toll-like receptors 2/4, CD14 and transcription factors PU.1 and inhibits lipopolysaccharide activation of NF-kappa B, ERK 1/2 and p38 MAPK in host monocytes. *Cell Microbiol* 6:175-86.
  36. Lu, A. L., X. Li, Y. Gu, P. M. Wright, and D. Y. Chang. 2001. Repair of oxidative DNA damage: mechanisms and functions. *Cell Biochem Biophys* 35:141-70.
  37. Mandal, M., B. Boese, J. E. Barrick, W. C. Winkler, and R. R. Breaker. 2003. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* 113:577-86.
  38. Markowitz, V. M., F. Korzeniewski, K. Palaniappan, E. Szeto, G. Werner, A. Padki, X. Zhao, I. Dubchak, P. Hugenholtz, I. Anderson, A. Lykidis, K. Mavromatis, N. Ivanova, and N. C. Kyrpides. 2006. The integrated microbial genomes (IMG) system. *Nucleic Acids Res* 34:D344-8.
  39. Masse, E., N. Majdalani, and S. Gottesman. 2003. Regulatory roles for small RNAs in bacteria. *Curr Opin Microbiol* 6:120-4.
  40. McBride, J. W., R. E. Corstvet, S. D. Gaunt, C. Boudreaux, T. Guedry, and D. H. Walker. 2003. Kinetics of antibody response to *Ehrlichia canis* immunoreactive proteins. *Infect Immun* 71:2516-24.
  41. McBride, J. W., L. M. Ndip, V. L. Popov, and D. H. Walker. 2002. Identification and functional analysis of an immunoreactive DsbA-like thio-disulfide oxidoreductase of *Ehrlichia* spp. *Infect Immun* 70:2700-3.
  42. McBride, J. W., X. J. Yu, and D. H. Walker. 2000. A conserved, transcriptionally active p28 multigene locus of *Ehrlichia canis*. *Gene* 254:245-52.

43. McBride, J. W., X. J. Yu, and D. H. Walker. 2000. Glycosylation of homologous immunodominant proteins of *Ehrlichia chaffeensis* and *Ehrlichia canis*. *Infect Immun* 68:13-8.
44. McLeod, M. P., X. Qin, S. E. Karpathy, J. Gioia, S. K. Highlander, G. E. Fox, T. Z. McNeill, H. Jiang, D. Muzny, L. S. Jacob, A. C. Hawes, E. Sodergren, R. Gill, J. Hume, M. Morgan, G. Fan, A. G. Amin, R. A. Gibbs, C. Hong, X. J. Yu, D. H. Walker, and G. M. Weinstock. 2004. Complete genome sequence of *Rickettsia typhi* and comparison with sequences of other rickettsiae. *J Bacteriol* 186:5842-55.
45. Moran, N. A. 1996. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc Natl Acad Sci U S A* 93:2873-8.
46. Mosavi, L. K., T. J. Cammett, D. C. Desrosiers, and Z. Y. Peng. 2004. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* 13:1435-48.
47. Nethery, K., C. Doyle, B. Elsom, N. Herzog, V. Popov, and J. McBride. Ankyrin repeat containing immunoreactive 200 kDa glycoprotein (gp200) orthologs of *Ehrlichia chaffeensis* and *E. canis* are translocated to the nuclei of infected monocytes. The 4<sup>th</sup> International Conference on Rickettsiae and Rickettsial Diseases, Logrono, Spain., June 18-21, 2005, Abstract O-60.
48. Ochman, H., and A. C. Wilson. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 26:74-86.
49. Ogata, H., S. Audic, P. Renesto-Audiffren, P. E. Fournier, V. Barbe, D. Samson, V. Roux, P. Cossart, J. Weissenbach, J. M. Claverie, and D. Raoult. 2001. Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. *Science* 293:2093-8.
50. Ohashi, N., Y. Rikihisa, and A. Unver. 2001. Analysis of transcriptionally active gene clusters of major outer membrane protein multigene family in *Ehrlichia canis* and *E. chaffeensis*. *Infect Immun* 69:2083-91.
51. Ohashi, N., N. Zhi, Q. Lin, and Y. Rikihisa. 2002. Characterization and transcriptional analysis of gene clusters for a type IV secretion machinery in human granulocytic and monocytic ehrlichiosis agents. *Infect Immun* 70:2128-38.
52. Popov, V. L., V. C. Han, S. M. Chen, J. S. Dumler, H. M. Feng, T. G. Andreadis, R. B. Tesh, and D. H. Walker. 1998. Ultrastructural differentiation of the genogroups in the genus *Ehrlichia*. *J Med Microbiol* 47:235-51.
53. Power, P. M., and M. P. Jennings. 2003. The genetics of glycosylation in Gram-negative bacteria. *FEMS Microbiol Lett* 218:211-22.
54. Reddy, G. R., and C. P. Streck. 1999. Variability in the 28-kDa surface antigen protein multigene locus of isolates of the emerging disease agent *Ehrlichia chaffeensis* suggests that it plays a role in immune evasion. *Mol Cell Biol Res Commun* 1:167-75.
55. Reddy, G. R., C. R. Sulsona, A. F. Barbet, S. M. Mahan, M. J. Burrige, and A. R. Alleman. 1998. Molecular characterization of a 28 kDa surface antigen gene family of the tribe Ehrlichiae. *Biochem Biophys Res Commun* 247:636-43.
56. Schmidt, M. A., L. W. Riley, and I. Benz. 2003. Sweet new world: glycoproteins in bacterial pathogens. *Trends Microbiol* 11:554-61.
57. Sexton, J. A., and J. P. Vogel. 2002. Type IVB secretion by intracellular pathogens. *Traffic* 3:178-85.

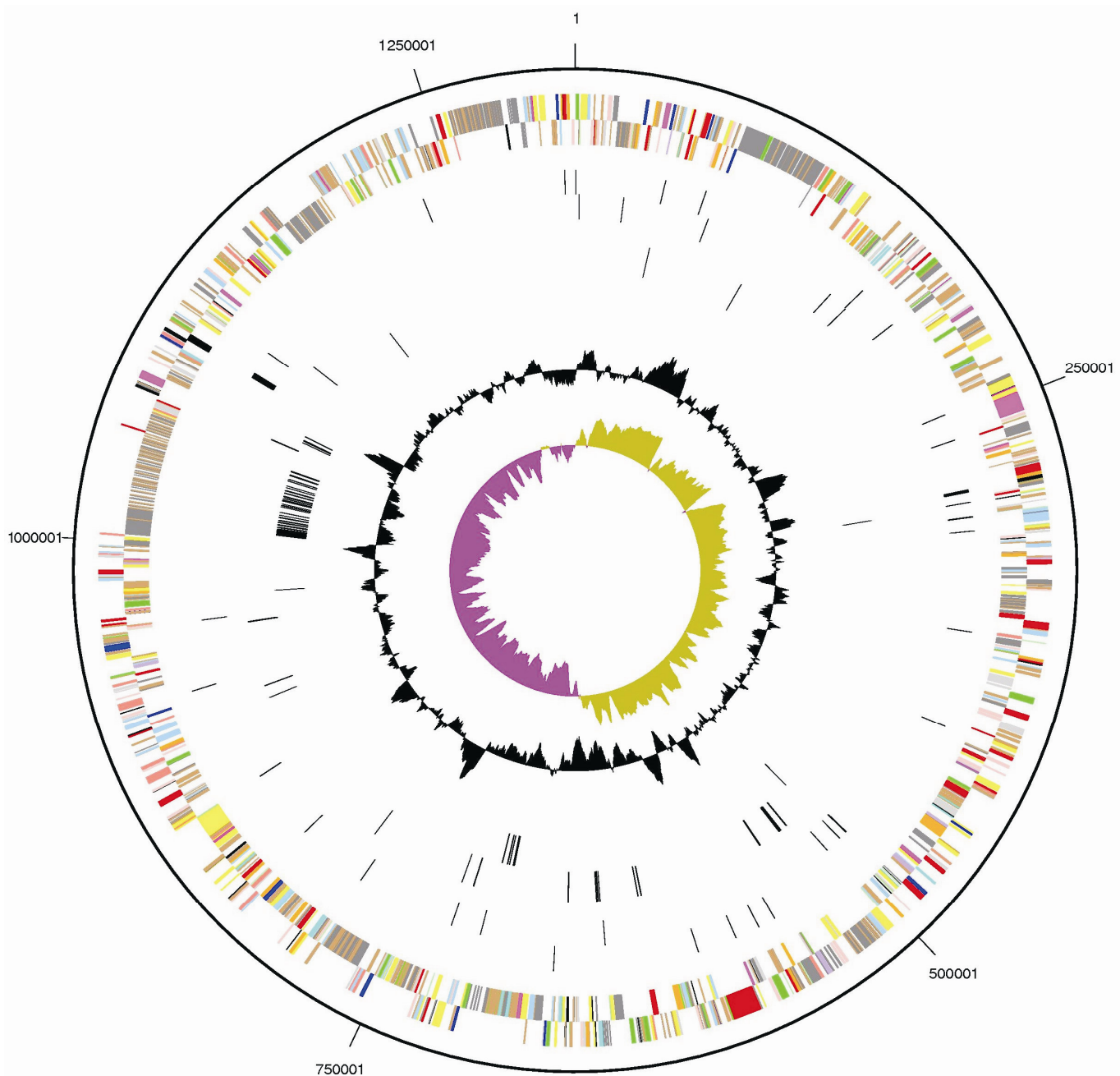


58. Simpson, C. F. 1972. Structure of *Ehrlichia canis* in blood monocytes of a dog. Am J Vet Res 33:2451-4.
59. Skotarczak, B. 2003. Canine ehrlichiosis. Ann Agric Environ Med 10:137-41.
60. Szymanski, C. M., S. M. Logan, D. Linton, and B. W. Wren. 2003. Campylobacter--a tale of two protein glycosylation systems. Trends Microbiol 11:233-8.
- 5 61. Szymanski, C. M., and B. W. Wren. 2005. Protein glycosylation in bacterial mucosal pathogens. Nat Rev Microbiol 3:225-37.
62. Troy, G. C., and S. D. Forrester. Canine Ehrlichiosis, p. 404-418, Viral, Rickettsial, and Mycoplasmal Infections.
- 10 63. Unver, A., N. Ohashi, T. Tajima, R. W. Stich, D. Grover, and Y. Rikihisa. 2001. Transcriptional analysis of p30 major outer membrane multigene family of *Ehrlichia canis* in dogs, ticks, and cell culture at different temperatures. Infect Immun 69:6172-8.
64. Vidotto, M. C., T. C. McGuire, T. F. McElwain, G. H. Palmer, and D. P. Knowles, Jr. 1994. Intermolecular relationships of major surface proteins of *Anaplasma marginale*. Infect Immun 62:2940-6.
- 15 65. Vitreschak, A. G., D. A. Rodionov, A. A. Mironov, and M. S. Gelfand. 2004. Riboswitches: the oldest mechanism for the regulation of gene expression? Trends Genet 20:44-50.
66. Wu, M., L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy, J. C. Brownlie, E. A. McGraw, W. Martin, C. Esser, N. Ahmadinejad, C. Wiegand, R. Madupu, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, A. S. Durkin, J. F. Kolonay, W. C. Nelson, Y. Mohamoud, P. Lee, K. Berry, M. B. Young, T. Utterback, J. Weidman, W. C. Nierman, I. T. Paulsen, K. E. Nelson, H. Tettelin, S. L. O'Neill, and J. A. Eisen. 2004. Phylogenomics of the Reproductive parasite *Wolbachia pipientis* wMel: A streamlined genome overrun by mobile genetic elements. PLoS Biol 2:E69.
- 20 67. Yu, X. J., J. W. McBride, C. M. Diaz, and D. H. Walker. 2000. Molecular cloning and characterization of the 120-kilodalton protein gene of *Ehrlichia canis* and application of the recombinant 120-kilodalton protein for serodiagnosis of canine ehrlichiosis. J Clin Microbiol 38:369-74.

30

Organism Name	Genes	GC Perc	Bases	Coding Bases	Coding/ Total bases
<i>Anaplasma marginale</i> St. Maries	986	49.76%	1197687	1024175	85%
<i>Ehrlichia canis</i> str. Jake	967	28.96%	1315030	955386	73%
<i>Ehrlichia chaffeensis</i> sapulpa	833	29.98%	1005812	731291	73%
<i>Ehrlichia ruminantium</i> str. Welgevonden (CIRAD)	997	27.48%	1512977	967368	64%
<i>Ehrlichia ruminantium</i> str. Gardel	989	27.51%	1499920	968505	64%
<i>Ehrlichia ruminantium</i> str. Welgevonden (ARC-OVI)	959	27.48%	1516355	960154	63%
<i>Rickettsia akari</i> str. Hartford	1217	32.33%	1231060	946961	77%
<i>Rickettsia conorii</i> str. Malish 7	1410	32.44%	1268755	1025607	81%
<i>Rickettsia prowazekii</i> Madrid E	870	29.00%	1111523	845091	76%
<i>Rickettsia sibirica</i> 246	1234	32.47%	1250021	969726	77%
<i>Rickettsia typhi</i> str. Wilmington	874	28.92%	1111496	843854	76%
<i>Wolbachia</i> endosymbiont strain TRS of Brugia malayi	842	34.18%	1080084	730178	68%
<i>Wolbachia pipientis</i> wMel	1232	35.23%	1267782	1021568	81%
Bacteria in IMG (v1.1)	1180331		1333547 111	1121629579	84%

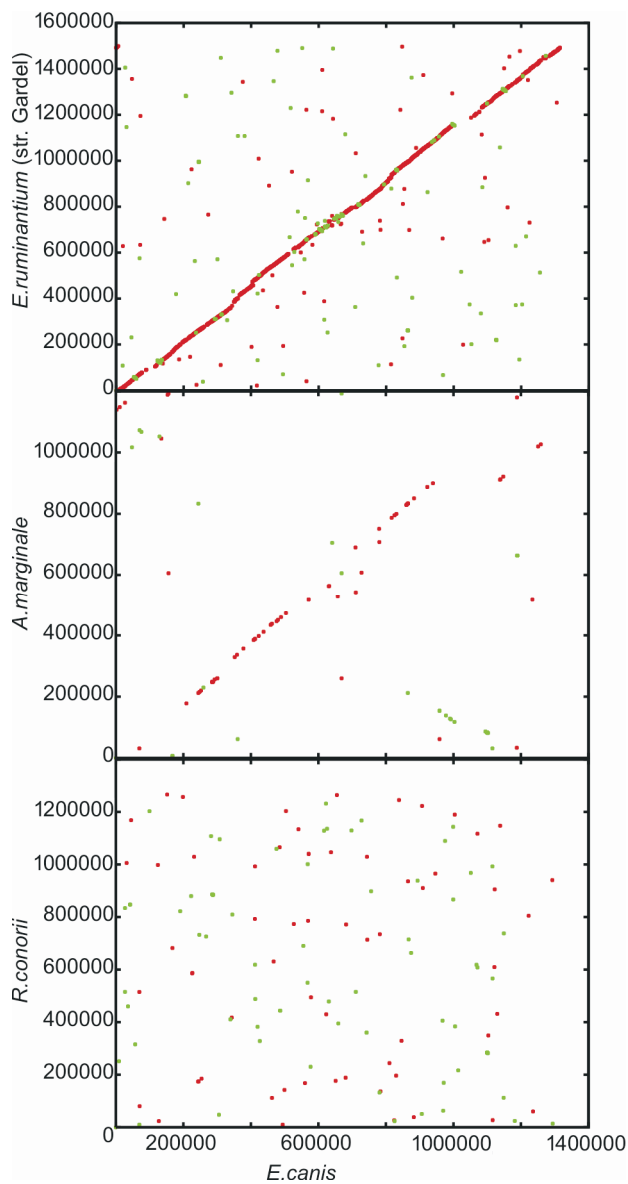
Table 1. Comparison of genome properties of members of Rickettsiales. Data obtained from IMG v1.1 (<http://img.jgi.doe.gov>)



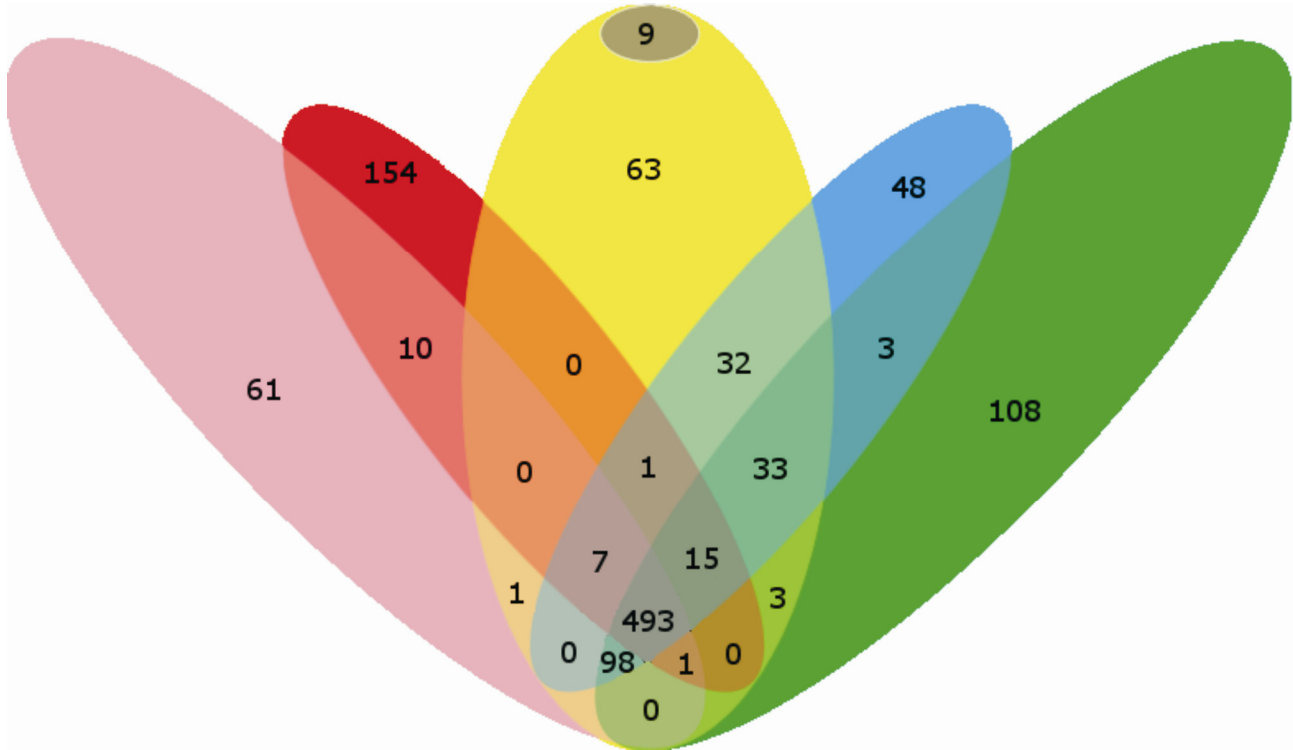
### Figure legends

**Figure 1.** Circular representation of the genome of *E. canis*. From outside to inside, the first two circles represent COG assignment for predicted coding sequences on the plus and minus strands, respectively. Colors indicate: Dark Grey - Hypothetical Proteins, Light Grey - Conserved Hypothetical and Unknown Function, Brown - General function prediction, Red-Replication and

Repair, Green - Energy Metabolism, Blue - Carbon and Carbohydrate Metabolism, Cyan - Lipid Metabolism, Magenta - Transcription, Yellow - Translation, Orange - Amino Acid Metabolism, Pink - Metabolism of Cofactors and Vitamins, Light Red - Purine and Pyrimidine Metabolism, Lavender - Signal Transduction, Sky Blue - Cellular Processes. The third and fourth circle represent unique coding sequences. The fifth and sixth represent the RNA genes. The two innermost circles represent the % GC content and GC skew values, respectively.



**Figure 2.** Synteny plots between *E. canis* (horizontal axis) and *E. ruminantium* (str.Gardel), *A. marginale* and *R. conorii*. Numbers represent base pairs.



**Figure 3.** Venn diagram showing comparison of conserved and unique genes in *Ehrlichia canis* (yellow), *Ehrlichia* spp. (light blue), *Wolbachia* spp.(pink), *Rickettsia* spp.(red), and *Anaplasma* spp (light green). Dark circle represents genes unique in *E. canis*.

	COG0789	COG1327	COG1329	COG1395	COG1396	COG1475	COG1476	COG1521	COG1678	COG1709	COG1959	COG2909	COG2932	COG2944	COG3829
<i>Ehrlichia canis</i> Jake	+	+			+	+		+					+		
<i>Ehrlichia ruminantium</i> str. Gardel	+	+			+	+		+			+		+		
<i>Ehrlichia ruminantium</i> str. Welgevonden (CIRAD)	+	+			+	+		+			+		+		
<i>Ehrlichia ruminantium</i> str. Welgevonden (ARC-OVI)	+	+			+	+		+			+		+		
<i>Anaplasma marginale</i> str. St. Maries	+	+			+	+		+					+		
<i>Rickettsia akari</i> str. Hartford	+		+		+	+	+		+		+				
<i>Rickettsia conorii</i> (strain Malish 7)	+		+		+	+			+		+	+			+
<i>Rickettsia prowazekii</i> (strain Madrid E)	+		+		+	+			+		+				
<i>Rickettsia sibirica</i> 246 rsib_agnrcr	+		+		+	+			+		+				+
<i>Rickettsia typhi</i> (strain Wilmington)			+		+	+			+						

<i>Wolbachia</i> endosymbiont strain TRS of <i>Brugia malayi</i>	+			+						+
<i>Wolbachia pipientis</i> wMel	+					+	+	+		+

Supplementary table 1. Transcriptional regulators distribution in genomes of Rickettsiales. COG assignments were used to identify putative transcriptional regulators.

5

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0001	391						Ubiquinone biosynthesis hydroxylase, UbiH/UbiF/VisC/COQ6
Ecaj_0002	187	+	2				hypothetical
Ecaj_0003	281				2		Glycine--tRNA ligase.
Ecaj_0004	702				2		Glycyl-tRNA synthetase, beta subunit
Ecaj_0005	383				2		DnaJ central domain (CXXCXGXG):Heat shock protein DnaJ, N-terminal:Chaperone DnaJ, C-terminal
Ecaj_0006	276						Nicotinate-nucleotide pyrophosphorylase
Ecaj_0007	192	-	4		2		Protein of unknown function DUF205
Ecaj_0008	157						Crossover junction endodeoxyribonuclease RuvC
Ecaj_0009	275	-	7				Cytochrome c oxidase, subunit III
Ecaj_0010	335						Uroporphyrinogen decarboxylase HemE
Ecaj_0011	296						CBS domain:Transporter associated domain
Ecaj_0012	178						Peptidase U35, phage prohead HK97
Ecaj_0013	394						Phage portal protein, HK97 family
Ecaj_0014	254			C			similar to Methylase involved in ubiquinone/menaquinone biosynthesis
Ecaj_0015	314						Quinolinate synthetase A
Ecaj_0016	126			C			4Fe-4S ferredoxin, iron-sulfur binding domain
Ecaj_0017	689				10		hypothetical
Ecaj_0018	724	-	3				TRAG protein
Ecaj_0019	333				2		Bacterial type II secretion system protein E
Ecaj_0020	446	-	1				Bacterial conjugation TrbI-like protein
Ecaj_0021	275	+	1				Conjugal transfer protein TrbG/VirB9/CagX
Ecaj_0022	235	+	1				VirB8
Ecaj_0023	381						GTP cyclohydrolase II.
Ecaj_0024	357						TPR repeat
Ecaj_0025	195	-	3				hypothetical
Ecaj_0026	268						Diaminopimelate epimerase
Ecaj_0027	144						conserved hypothetical protein
Ecaj_0028	395						Phosphoglycerate kinase.
Ecaj_0029	389						Exonuclease VII, large subunit
Ecaj_0030	222	-	1		2		hypothetical
Ecaj_0031	282						2,3,4,5-tetrahydropyridine-2-carboxylateN-succinyltransferase
Ecaj_0032	442						tRNA modification GTPase TrmE
Ecaj_0033	178						Phosphopantothenoil cysteine decarboxylase.
Ecaj_0034	679						DEAD/DEAH box helicase
Ecaj_0035	321						NADH-ubiquinone oxidoreductase, putative[EC:1.6.5.3 1.6.99.3] KO:K00329 NADH dehydrogenase KO:K00356 NADH dehydrogenase
Ecaj_0036	153						dnaK suppressor protein

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0037	221	-	6				heme exporter protein CcmB KO
Ecaj_0038	81	-	1	C			hypothetical
Ecaj_0039	313	-	6				Cation efflux protein
Ecaj_0040	210	+	0				similar to ABC-type transport system
Ecaj_0041	96						Ribosomal protein S20p
Ecaj_0042	241	-	2				Phospholipid/glycerol acyltransferase
Ecaj_0043	186						Formylmethionine deformylase
Ecaj_0044	262						Phage SPO1 DNA polymerase-related protein
Ecaj_0045	373					2	RecF protein
Ecaj_0046	304						Ornithine carbamoyl transferase
Ecaj_0047	134	+	0			15	conserved hypothetical protein
Ecaj_0048	862	-	1			2	DNA polymerase, family A
Ecaj_0049	223						Ribulose-phosphate 3-epimerase
Ecaj_0050	266	-	6				Protein of unknown function DUF140
Ecaj_0051	239						ABC transporter
Ecaj_0052	614	-	3				Ankyrin
Ecaj_0053	326						Trans-hexaprenyl transtransferase.
Ecaj_0054	491						Glutamine synthetase type I
Ecaj_0055	419						Tyrosyl-tRNA synthetase, class Ib
Ecaj_0056	399						5-aminolevulinic acid synthase
Ecaj_0057	123	-	1				endonuclease distantly related to archaeal Holliday junction resolvase
Ecaj_0058	294	+	6			2	SecD/SecF/SecDF export membrane protein
Ecaj_0059	302						Fructose-bisphosphate aldolase.
Ecaj_0060	3715					9	hypothetical
Ecaj_0061	404					2	Dihydrolipoamide acetyltransferase, long form
Ecaj_0062	958						hypothetical
Ecaj_0063	706	+	1			4	hypothetical
Ecaj_0064	228			C			hypothetical
Ecaj_0065	596						hypothetical
Ecaj_0066	890	+	0			9	hypothetical
Ecaj_0067	696	+	1			2	hypothetical
Ecaj_0068	617					7	hypothetical
Ecaj_0069	824	+	1				hypothetical
Ecaj_0070	79	-	1			3	hypothetical
Ecaj_0071	642					4	hypothetical
Ecaj_0072	917					14	hypothetical
Ecaj_0073	93						hypothetical
Ecaj_0074	527						GMP synthase, C-terminal:GMPsynthase,N-terminal
Ecaj_0075	417						Citrate synthase I
Ecaj_0076	409						glutamate--cysteine ligase- related protein
Ecaj_0077	335						DNA polymerase III delta subunit
Ecaj_0078	411	-	4			2	Protein of unknown function DUF214
Ecaj_0079	263	+	1				conserved hypothetical protein
Ecaj_0080	414	-	2				similar to Uncharacterized membrane-bound protein
Ecaj_0081	85	-	2				hypothetical
Ecaj_0082	184				S		H+-transporting two-sector ATPase, delta(OSCP) subunit
Ecaj_0083	510						ATP synthase F1, alpha subunit

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0084	171						Prokaryotic transcription elongation factor GreA
Ecaj_0085	212						3,4-Dihydroxy-2-butanone 4-phosphate synthase
Ecaj_0086	149						SmpB protein
Ecaj_0087	645	-	15				Cytochrome c assembly protein
Ecaj_0088	803				3		Aminoacyl-tRNA synthetase, class Ia
Ecaj_0089	279						Glycine cleavage T protein(aminomethyl transferase)
Ecaj_0090	463						Amidophosphoribosyl transferase
Ecaj_0091	194						Peptidyl-tRNA hydrolase
Ecaj_0092	210						Ribosomal protein L25
Ecaj_0093	231			C			competence protein F
Ecaj_0094	390						Proteobacterial succinyl- diaminopimelate desuccinylase
Ecaj_0095	570	-	10				TrkA-N
Ecaj_0096	195						conserved hypothetical protein
Ecaj_0097	145	-	4				Conserved hypothetical protein 701
Ecaj_0098	333						Transketolase, central region:Transketolase, C-terminal
Ecaj_0099	631	-	2		16		Hypothetical
Ecaj_0100	478						Peptidase U62, modulator of DNA gyrase
Ecaj_0101	363						Conserved hypothetical protein 92
Ecaj_0102	174						MECDP-synthase
Ecaj_0103	232						4-diphosphocytidyl-2C-methyl-D-erythritolsynthase
Ecaj_0104	426	-	2		5		hypothetical
Ecaj_0105	168						1-(5-Phosphoribosyl)-5-amino-4-imidazole- carboxylate(AIR) carboxylase
Ecaj_0106	149						hypothetical
Ecaj_0107	101						Histone-like bacterial DNA-binding protein
Ecaj_0108	125						Bacterial regulatory protein, MerR
Ecaj_0109	208				3		hypothetical
Ecaj_0110	334						Tryptophanyl-tRNA synthetase, class Ib
Ecaj_0111	200						GrpE protein
Ecaj_0112	362						Riboflavin biosynthesis protein RibD
Ecaj_0113	138			C			hypothetical
Ecaj_0114	539	+	1				CTP synthase
Ecaj_0115	111	-	2		3		Preprotein translocase SecG subunit
Ecaj_0116	310						Phage integrase:Phage integrase, N-terminal SAM- like
Ecaj_0117	227						ABC transporter
Ecaj_0118	764				2		Malate dehydrogenase(oxaloacetate decarboxylating) (NADP+),,
Ecaj_0119	882	-	1		3		hypothetical
Ecaj_0120	489	+	7				Apolipoprotein N-acyltransferase
Ecaj_0121	492	-	14				NADH dehydrogenase (quinone).
Ecaj_0122	104	-	1		S		hypothetical
Ecaj_0123	150	-	1				Mce4/Rv3499c/MTV023.06c protein
Ecaj_0124	100						similar to NADH:ubiquinone oxidoreductase 17.2kD subunit
Ecaj_0125	154						Protein of unknown function DUF193
Ecaj_0126	672				6		hypothetical
Ecaj_0127	1006						surface protein-related protein
Ecaj_0128	348	+	1				Bacterial extracellular solute-binding protein, family 1
Ecaj_0129	88						Ribosomal protein S16



Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0130	442	-	12				General substrate transporter
Ecaj_0131	57						Protein of unknown function DUF343
Ecaj_0132	232						similar to Dehydrogenases with different specificities
Ecaj_0133	346					2	Phenylalanyl-tRNA synthetase, alpha subunit
Ecaj_0134	125						Ribosomal protein L20, bacterial and organelle form
Ecaj_0135	67				S/T		Ribosomal protein L35
Ecaj_0136	214						similar to Uncharacterized protein conserved in bacteria
Ecaj_0137	477						Transcription termination factor Rho
Ecaj_0138	492					2	Dihydrolipoamide dehydrogenase
Ecaj_0139	795	-	1			19	hypothetical
Ecaj_0140	483	-	1				conserved domain protein
Ecaj_0141	211	-	1				conserved hypothetical protein
Ecaj_0142	179						hypothetical
Ecaj_0143	265						Beta-lactamase-like
Ecaj_0144	267					2	Glutamine synthetase, catalytic domain
Ecaj_0145	375	-	7				Binding-protein-dependent transport systems inner membrane component
Ecaj_0146	892						Alanyl-tRNA synthetase, class Iic
Ecaj_0147	296						Succinyl-CoA ligase, alpha subunit
Ecaj_0148	387						Succinyl-CoA synthetase, beta subunit
Ecaj_0149	92						Hypothetical
Ecaj_0150	341	+	0				phosphate ABC transporter, periplasmic phosphate- binding protein,
Ecaj_0151	206	-	1				Electron transport protein SCO1/SenC
Ecaj_0152	346					2	2-nitropropane dioxygenase, NPD
Ecaj_0153	174	-	4				Nickel-dependent hydrogenase b- type cytochrome subunit
Ecaj_0154	540	-	12				Binding-protein-dependent transport systems inner membrane component
Ecaj_0155	506	-	1				Type I secretion membrane fusion protein, HlyD
Ecaj_0156	206						conserved hypothetical protein
Ecaj_0157	545						Beta-lactamase-like
Ecaj_0158	432	+	0				Metallophosphoesterase
Ecaj_0159	124						Ribosomal protein S12, bacterial and chloroplast form
Ecaj_0160	161						Ribosomal protein S7, bacterial and organelle form
Ecaj_0161	691					2	Translation elongation factor G:Small GTP- binding protein domain
Ecaj_0162	396						Translation elongation factor Tu:Small GTP- binding protein domain
Ecaj_0163	66	-	1				SecE component
Ecaj_0164	182						Bacterial transcription antitermination protein NusG
Ecaj_0165	148						Ribosomal protein L11, bacterial
Ecaj_0166	221						Ribosomal protein L1, bacterial and chloroplast form
Ecaj_0167	161						Ribosomal protein L10
Ecaj_0168	133						Ribosomal protein L7/L12
Ecaj_0169	1381						DNA-directed RNA polymerase, beta subunit
Ecaj_0170	1411					2	DNA-directed RNA polymerase.
Ecaj_0171	368						8-amino-7-oxononanoate synthase.
Ecaj_0172	139	-	4				conserved hypothetical protein
Ecaj_0173	210						Ribonuclease H.
Ecaj_0174	1307					11	hypothetical
Ecaj_0175	159	+	2		T		Protein of unknown function DUF68

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0176	199	-	1				hypothetical
Ecaj_0177	265						Bordetella pertussis Bvg accessory factor
Ecaj_0178	351						Dihydroorotate dehydrogenase 2
Ecaj_0179	195						Formylmethionine deformylase
Ecaj_0180	472					3	Argininosuccinate lyase
Ecaj_0181	214						Phospholipase/Carboxylesterase
Ecaj_0182	316			C			Patatin
Ecaj_0183	198						Pyridoxamine-phosphate oxidase.
Ecaj_0184	98						BolA-like protein
Ecaj_0185	272	+	0				unknown protein
Ecaj_0186	286	+	1				hypothetical
Ecaj_0187	1124						DNA polymerase III alpha subunit
Ecaj_0188	429					2	3-phosphoshikimate 1-carboxyvinyltransferase.
Ecaj_0189	137	-	3				Succinate dehydrogenase, cytochrome b subunit
Ecaj_0190	119	-	3				succinate dehydrogenase,
Ecaj_0191	422						hypothetical
Ecaj_0192	270						Phosphomethylpyrimidine kinase
Ecaj_0193	219						chromosomal DNA replication initiator- related protein
Ecaj_0194	203						Ribosomal protein S4, bacterial and organelle form
Ecaj_0196	374	-	8				Protein of unknown function UPF0118
Ecaj_0197	67						Protein of unknown function DUF528
Ecaj_0198	382	+	1				conserved hypothetical protein
Ecaj_0199	262						GCN5-related N-acetyltransferase
Ecaj_0200	169	-	4				Phosphatidylglycerophosphatase A
Ecaj_0201	441					2	Trigger factor
Ecaj_0202	201						Peptidase S14, ClpP
Ecaj_0203	407						ClpX, ATPase regulatory subunit
Ecaj_0204	802						Peptidase S16, ATP-dependent protease La
Ecaj_0205	304						Methionyl-tRNA formyltransferase
Ecaj_0206	275						Rhodanese-like
Ecaj_0207	274						Protein of unknown function DUF299
Ecaj_0208	350						Thiamine monophosphate synthase
Ecaj_0210	429	+	4			3	CBS domain:Protein of unknownfunctionDUF21:Transporter associated domain
Ecaj_0211	95	-	2				Protein of unknown function YGGT
Ecaj_0212	849	+	4				Cell divisionFtsK/SpoIIIE protein
Ecaj_0213	329	-	6			2	hypothetical
Ecaj_0214	392						Acetylornithine and succinylornithine aminotransferase
Ecaj_0215	471	+	2				ATP-binding region, ATPase-like:Histidine kinase A, N-terminal
Ecaj_0216	683					2	DNA mismatch repair protein
Ecaj_0217	375						SMF protein
Ecaj_0218	423	+	0				Beta-ketoacyl synthase:Beta-ketoacyl synthase
Ecaj_0219	92						Phosphopantetheine-binding domain
Ecaj_0220	462					2	Hypothetical
Ecaj_0221	934	-	2			6	Ankyrin
Ecaj_0222	60						Ribosomal protein L33
Ecaj_0223	236	-	6				conserved hypothetical protein

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0224	160	+	4	C			Disulfide bond formation protein DsbB
Ecaj_0225	171						Coq7 family protein
Ecaj_0226	368						tRNA(5-methylaminomethyl-2-thiouridylate)-methyltransferase
Ecaj_0227	337						similar to Tfp pilus assembly protein pilus retraction ATPase PilT
Ecaj_0228	346	-	1				hypothetical
Ecaj_0229	355	-	1		5		hypothetical
Ecaj_0230	487	-	1				hypothetical
Ecaj_0231	329	-	1		3		hypothetical
Ecaj_0232	413				T		hypothetical
Ecaj_0233	640						UvrD/REP helicase
Ecaj_0234	899						DNA gyrase, subunit A
Ecaj_0235	213						Endonuclease III/Nth
Ecaj_0236	201	+	1				conserved hypothetical protein
Ecaj_0237	638						Heat shock protein Hsp90:ATP-binding region, ATPase-like
Ecaj_0238	434						Adenylosuccinate lyase
Ecaj_0239	952				12		hypothetical
Ecaj_0240	248						Biotin--acetyl-CoA-carboxylase ligase
Ecaj_0241	226						Glutathione S-transferase, N-terminal
Ecaj_0242	170	+	0				similar to Uncharacterized protein predicted to be involved in C-type cytochrome biogenesis
Ecaj_0243	341						ABC transporter
Ecaj_0244	57	-	1				Twin-arginine translocation protein TatA/E
Ecaj_0245	196						RecR protein
Ecaj_0248	288				3		Periplasmic solute binding protein
Ecaj_0249	218						ABC transporter
Ecaj_0250	481						2-polyprenylphenol 6-hydroxylase
Ecaj_0251	402	-	12		3		major facilitator family transporter
Ecaj_0252	442						Protein of unknown function UPF0004:tRNA- i(6)A37modification enzyme MiaB
Ecaj_0253	1207				4		conserved hypothetical protein
Ecaj_0254	270						conserved hypothetical protein
Ecaj_0255	913						Dehydrogenase, E1 component
Ecaj_0256	383						Phage major capsid protein, HK97 family
Ecaj_0257	297						Dihydrodipicolinate synthase subfamily
Ecaj_0258	118						Histidine triad (HIT) protein
Ecaj_0259	369				2		hypothetical
Ecaj_0260	805				2		MutS 1 protein
Ecaj_0261	506				3		NAD+ synthase
Ecaj_0262	331						Porphobilinogen synthase.
Ecaj_0263	912						conserved hypothetical protein
Ecaj_0264	416	-	12				General substrate transporter
Ecaj_0265	412	-	12				General substrate transporter
Ecaj_0266	158				3		Single-strand binding protein
Ecaj_0267	459						Malonyl-CoA decarboxylase.
Ecaj_0268	483						Glutamyl-tRNA(Gln) amidotransferase B subunit
Ecaj_0269	281						Short-chain dehydrogenase/reductase SDR
Ecaj_0270	453						Bacterial chromosomal replication initiator protein, DnaA
Ecaj_0271	330	-	1				hypothetical

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0272	195						Cytidyl transferase-related domain
Ecaj_0273	239						Pyridoxal phosphate biosynthetic protein PdxJ
Ecaj_0274	95						Histone-like bacterial DNA-binding protein
Ecaj_0275	297						ATPas
Ecaj_0276	199						Sua5/YciO/YrdC/YwIC
Ecaj_0277	201						Dephospho-CoA kinase
Ecaj_0278	566						Thiamine biosynthesis protein ThiC
Ecaj_0279	205				4		Smr protein/MutS2 C-terminal
Ecaj_0280	139						RNA polymerase, omega subunit
Ecaj_0281	109						hypothetical
Ecaj_0282	830						Leucyl-tRNA synthetase bacterial/mitochondrial, class Ia
Ecaj_0283	261						TatD-related deoxyribonuclease
Ecaj_0284	229						Orotidine 5'-phosphate decarboxylase
Ecaj_0285	257						Survival protein SurE
Ecaj_0286	134	+	1				CcmE/CycJ protein
Ecaj_0287	193						NADH (or F420H2) dehydrogenase, subunit C
Ecaj_0288	186						NADH dehydrogenase (ubiquinone), 20 kDa subunit
Ecaj_0289	124	+	3				NADH-ubiquinone/plastoquinone oxidoreductase, chain 3
Ecaj_0290	951				4		Excinuclease ABC, A subunit
Ecaj_0291	149						hypothetical
Ecaj_0292	149						Riboflavin synthase.
Ecaj_0293	580	-	6				60 kDa inner membrane protein
Ecaj_0294	461	-	11		3		Sodium:alanine symporter
Ecaj_0295	260	-	7				CDP-alcohol phosphatidyltransferase
Ecaj_0296	227	+	1				Phosphatidylserine decarboxylase-related protein
Ecaj_0297	1131				4		conserved hypothetical protein
Ecaj_0301	190						Elongation factor P (EF-P)
Ecaj_0302	256						inositol monophosphatase familyprotein[EC:3.1.3.25] KO: K01092 myo- inositol-1(or4)- monophosphatase
Ecaj_0303	309						Pseudouridine synthase, RluD
Ecaj_0304	459						GGDEF
Ecaj_0305	89						hypothetical
Ecaj_0306	1583						Bacterial NAD-glutamate dehydrogenase
Ecaj_0307	210	-	4				Exopolysaccharide synthesis, ExoD
Ecaj_0308	460						CysteinyI-tRNA synthetase, class Ia
Ecaj_0309	841				7		hypothetical
Ecaj_0310	325	-	1				Ribonucleoside-diphosphate reductase.
Ecaj_0311	90						conserved hypothetical protein
Ecaj_0312	189						hypothetical
Ecaj_0313	336						Protein of unknown function DUF185
Ecaj_0314	605						DNA primase
Ecaj_0315	623						Sigma-70 region 1.1:Sigma-70 region 3:Sigma- 70region 2:Sigma-70 region 4:Sigma-70, non- essentialregion:Sigma-70 region 1.2
Ecaj_0316	318						Dihydrouridine synthase TIM-barrel proteinnifR3
Ecaj_0317	283						4-diphosphocytidyl-2C-methyl-D-erythritolkinase
Ecaj_0318	107				T		CutA1 divalent ion tolerance protein
Ecaj_0319	825	-	1				Response regulator receiver:ATP-binding region, ATPase-like:Histidine kinase A, N-terminal

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0320	220						O-methyltransferase, family 3
Ecaj_0321	100					2	conserved hypothetical protein
Ecaj_0322	467						Protein of unknown function UPF0031:YjeF- related protein, N-terminal
Ecaj_0323	821						Bacterial DNA topoisomerase I
Ecaj_0324	118						hypothetical
Ecaj_0325	158						conserved hypothetical protein
Ecaj_0326	127						Holo-acylcarrierproteinsynthase:Phosphopantethiene-proteintransferasedomain
Ecaj_0327	425						Prolyl-tRNA synthetase, bacterial
Ecaj_0328	322	+	0				conserved hypothetical protein
Ecaj_0329	163						Coenzyme Abiosynthesisprotein:Cytidyltransferase- related domain
Ecaj_0331	322	+	1				Thioredoxin reductase
Ecaj_0332	206						Alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen
Ecaj_0333	399						Aminotransferase, class I and II
Ecaj_0334	631	+	1				PpiC-type peptidyl-prolyl cis-trans isomerase
Ecaj_0335	193				T		Peptidase M22, glycoprotease
Ecaj_0336	302						tRNA pseudouridine synthase B
Ecaj_0337	94						Ribosomal protein S15, bacterial chloroplast and mitochondrial type
Ecaj_0338	797					8	3'exoribonuclease:3'exoribonuclease:3'exoribonuclease:RNA binding S1:KH
Ecaj_0339	264						conserved hypothetical protein
Ecaj_0340	599						Small GTP-binding protein domain:GTP- binding protein LepA
Ecaj_0341	191	-	1				hypothetical
Ecaj_0342	218	+	2			3	hypothetical
Ecaj_0343	182	+	2			2	hypothetical
Ecaj_0344	760	-	2				hypothetical
Ecaj_0345	632	+	2			11	hypothetical
Ecaj_0346	519					24	hypothetical
Ecaj_0347	355	-	2				hypothetical
Ecaj_0348	536	+	2				hypothetical
Ecaj_0349	109						conserved domain protein
Ecaj_0350	336						Peptide chain release factor 2
Ecaj_0351	249						Protein of unknown function DUF28
Ecaj_0352	501						Glutamyl-tRNA(Gln) amidotransferase A subunit
Ecaj_0353	435						Folylpolyglutamate synthetase
Ecaj_0354	299						Porphobilinogen deaminase
Ecaj_0355	209						hypothetical
Ecaj_0356	86			C	S		hypothetical
Ecaj_0357	226					3	hypothetical
Ecaj_0358	247			C			hypothetical
Ecaj_0359	615						Small GTP-binding protein domain:GTP- binding protein TypA
Ecaj_0360	117						hypothetical
Ecaj_0361	168	-	1	C			NADH-quinone oxidoreductase, chain I
Ecaj_0362	237	+	1				Signal peptidase I.
Ecaj_0363	152						HesB/YadR/YfhF
Ecaj_0364	398						Deoxyguanosine triphosphate triphosphohydrolase
Ecaj_0365	1422					5	Ankyrin
Ecaj_0366	276	+	6				hypothetical

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0367	394						Argininosuccinate synthase
Ecaj_0368	227	+	0				hypothetical
Ecaj_0369	231	-	1				hypothetical
Ecaj_0370	424					5	Arginine biosynthesis protein ArgJ
Ecaj_0371	280						AP endonuclease, family1:ExodeoxyribonucleaseIII xth
Ecaj_0372	259	-	6		S		hypothetical
Ecaj_0373	266	-	6				hypothetical
Ecaj_0374	245	+	0		T		Glucose/ribitol dehydrogenase
Ecaj_0375	1050						delta-1-pyrroline-5-carboxylate dehydrogenase 3
Ecaj_0376	429						Adenosylmethionine--8-amino-7-oxononanoateaminotransferase
Ecaj_0377	470					4	Phage uncharacterized protein, C-terminal
Ecaj_0378	210						hypothetical
Ecaj_0379	183						hypothetical
Ecaj_0380	130						conserved hypothetical protein
Ecaj_0381	185						uncharacterized phage protein, putative
Ecaj_0382	134						hypothetical
Ecaj_0383	117						conserved hypothetical protein
Ecaj_0384	43						Ribosomal protein L36
Ecaj_0385	301						Sigma-70 region 2:Sigma-70 region 4
Ecaj_0386	198						Sigma 54 modulation protein/ribosomalproteinS30EA
Ecaj_0387	4246					17	Ankyrin
Ecaj_0388	282	-	1				H+-transporting two-sector ATPase, gamma subunit
Ecaj_0389	191						GTP cyclohydrolase I.
Ecaj_0390	445						Peptidase U62, modulator of DNA gyrase
Ecaj_0391	339						FAD-dependent pyridine nucleotide- disulphideoxidoreductase
Ecaj_0392	263						Ribosomal RNA adenine dimethylase
Ecaj_0393	256	+	1				Triosephosphate isomerase
Ecaj_0394	327	+	2			5	putative immunogenic protein
Ecaj_0395	351						Peptidase M22, glycoprotease
Ecaj_0396	197	-	3				RDD
Ecaj_0397	119						Dihydroneopterin aldolase
Ecaj_0398	314	-	1				L-lactate dehydrogenase
Ecaj_0399	320	-	1				hypothetical
Ecaj_0400	76						hypothetical
Ecaj_0401	150	-	1				Ribose/galactose isomerase
Ecaj_0402	240						Ubiquinone biosynthesis O-methyltransferase
Ecaj_0403	133	+	2				cytochrome c-type biogenesis protein CcmH, putative KO: K03702 excinuclease ABC subunit B
Ecaj_0404	358	+	1			3	hypothetical
Ecaj_0405	117			C			Ferredoxin
Ecaj_0406	619						Heat shock protein Hsp70
Ecaj_0407	146						Heat shock protein DnaJ, N-terminal
Ecaj_0408	142						HesB/YadR/YfhF
Ecaj_0409	139						Nitrogen-fixing NifU-like, N-terminal
Ecaj_0410	423					2	Aminotransferase, class V
Ecaj_0411	534					2	Protein of unknown function UPF0074
Ecaj_0412	242						conserved hypothetical protein

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0413	512						Lysyl-tRNA synthetase (archaeal), class 1c
Ecaj_0414	146	-	1				hypothetical
Ecaj_0415	201	-	4	C			hypothetical
Ecaj_0416	175	-	4	C			hypothetical
Ecaj_0417	799						DNA gyrase, subunit B
Ecaj_0418	306					2	Aspartate carbamoyltransferase
Ecaj_0419	246						tRNA pseudouridine synthase
Ecaj_0420	73						conserved hypothetical protein
Ecaj_0421	684						NADH-quinone oxidoreductase, chain G
Ecaj_0422	370	-	8				Respiratory-chain NADH dehydrogenase, subunit 1
Ecaj_0423	258			C			hypothetical
Ecaj_0424	290						hypothetical
Ecaj_0425	80			C			hypothetical
Ecaj_0426	394						NADH dehydrogenase I, D subunit
Ecaj_0427	181						NADH dehydrogenase (ubiquinone), 24 kDa subunit
Ecaj_0428	291					5	hypothetical
Ecaj_0429	252					5	hypothetical
Ecaj_0430	184						hypothetical
Ecaj_0431	437					10	hypothetical
Ecaj_0432	470						Glutamyl-tRNA synthetase bacterial/mitochondrial
Ecaj_0433	127						hypothetical
Ecaj_0434	224			C			hypothetical
Ecaj_0435	196						hypothetical
Ecaj_0436	196						hypothetical
Ecaj_0437	510						Carboxyl transferase
Ecaj_0438	201					2 U	hypothetical
Ecaj_0440	173						Peptide chain release factor 1 (short gene)
Ecaj_0442	360						Peptide chain release factor 1
Ecaj_0443	201						GTP-binding domain
Ecaj_0444	324						Acetylglutamate kinase
Ecaj_0445	383	+	0			2	conserved hypothetical protein
Ecaj_0446	283						Coproporphyrinogen oxidase.
Ecaj_0447	438	-	1				Protein of unknown function DUF195
Ecaj_0448	228	-	1				type IV secretion system protein VirB8, putative
Ecaj_0449	427	-	9			3	Sodium:dicarboxylate symporter
Ecaj_0450	331					2	hypothetical
Ecaj_0451	51					U	hypothetical
Ecaj_0452	66					U	hypothetical
Ecaj_0453	109						hypothetical
Ecaj_0454	424						Seryl-tRNA synthetase, class IIa
Ecaj_0455	408						similar to 3-hydroxyacyl-CoA dehydrogenase
Ecaj_0456	154						conserved hypothetical protein
Ecaj_0457	301						tRNA isopentenyltransferase
Ecaj_0458	457	-	5				Divalent cation transporter
Ecaj_0459	508					2	ATP synthase F1, beta subunit
Ecaj_0460	129						H <sup>+</sup> -transporting two-sector ATPase, delta/epsilon subunit
Ecaj_0461	221						Transaldolase.

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0462	94						conserved hypothetical protein
Ecaj_0463	1329	-	2			3	hypothetical
Ecaj_0464	760						AAA ATPase, central region:Clp, N terminal
Ecaj_0465	228	-	7				Integral membrane protein TerC family
Ecaj_0466	116						Ribosome-binding factor A
Ecaj_0467	849						Initiation factor 2:Small GTP- binding proteindomain
Ecaj_0468	518						Transcription terminationfactorNusA:Transcription termination factor NusA, C-terminalduplication
Ecaj_0469	1039	+	12				Acriflavin resistance protein
Ecaj_0470	251	-	5				Twin-arginine translocating C- subunit
Ecaj_0471	423						IspG protein
Ecaj_0472	642	+	0				hypothetical
Ecaj_0473	388						1-deoxy-D-xylulose 5-phosphate reductoisomerase
Ecaj_0474	475	-	13				Proton-translocating NADH-quinone oxidoreductase, chain N
Ecaj_0475	486	+	14				Proton-translocating NADH-quinone oxidoreductase, chain M
Ecaj_0476	622	-	14				NADH-plastoquinone oxidoreductase, chain 5
Ecaj_0477	109	-	3				NADH-ubiquinone oxidoreductase, chain 4L
Ecaj_0478	201	+	5				NADH-ubiquinone/plastoquinone oxidoreductase, chain 6
Ecaj_0479	194						hypothetical
Ecaj_0480	206					2	hypothetical
Ecaj_0481	170						hypothetical
Ecaj_0482	93						hypothetical
Ecaj_0483	90						hypothetical
Ecaj_0484	199						hypothetical
Ecaj_0485	201						hypothetical
Ecaj_0486	428						Respiratory-chain NADH dehydrogenase, 51kDa subunit
Ecaj_0487	89						Ribosomal protein L27
Ecaj_0488	103						Ribosomal protein L21
Ecaj_0489	422						Enolase
Ecaj_0490	341						GTP1/OBG sub-domain
Ecaj_0491	298						Bacterial methyltransferase
Ecaj_0492	174						5-formyltetrahydrofolate cyclo-ligase
Ecaj_0493	240						Conserved hypothetical protein 46
Ecaj_0494	148						Alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen
Ecaj_0495	1121						Isoleucyl-tRNA synthetase, class Ia
Ecaj_0496	98						Hypothetical
Ecaj_0497	194						Hypothetical
Ecaj_0498	577						Arginyl-tRNA synthetase, class Ic
Ecaj_0499	249						recO-related protein KO: K03584 DNA repair protein RecO (recombination protein O)
Ecaj_0500	186						hypothetical
Ecaj_0501	172						Ribosomal RNA methyltransferase RrmJ/FtsJ
Ecaj_0502	507					2	Mg chelatase-related protein
Ecaj_0503	111	+	2		T		hypothetical
Ecaj_0504	280	-	5				Ribonuclease BN
Ecaj_0505	315						Thiamine-monophosphate kinase
Ecaj_0506	240						DNA polymerase III, epsilon subunit :DNA polymerase 3, epsilon subunit



Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0507	491	+	0		S		conserved hypothetical protein
Ecaj_0508	652	+	1			4	hypothetical
Ecaj_0509	539	+	0			2	hypothetical
Ecaj_0510	252	+	1				Cytochrome c1
Ecaj_0511	409	-	9				Cytochrome b/b6, N-terminal:Cytochrome b/b6,C- terminal
Ecaj_0512	188	-	1				Ubiquinol-cytochrome c reductase, iron- sulfur subunit
Ecaj_0513	291	-	5				hypothetical
Ecaj_0514	268	-	8				ABC transporter, family 3
Ecaj_0515	121						Protein of unknown function DUF598
Ecaj_0516	289						Elongation factor Ts
Ecaj_0517	292						Ribosomal protein S2, bacterial and organelle form
Ecaj_0518	193						Maf-like protein
Ecaj_0519	83						Translation initiation factor IF-1
Ecaj_0520	357	+	1				Secretion protein HlyD
Ecaj_0521	467						Dihydrolipoamide dehydrogenase
Ecaj_0522	388	+	1			3	Bacterial quinoprotein
Ecaj_0523	502						Phosphoglycerate mutase,2,3-bisphosphoglycerate- independent
Ecaj_0524	440						Small GTP-binding protein domain:GTP- binding domain
Ecaj_0525	372						Carbamoyl-phosphate synthase, small subunit
Ecaj_0526	320						LytB protein
Ecaj_0527	154						DeoxyUTP pyrophosphatase, subfamily 1
Ecaj_0528	200					2	conserved hypothetical protein
Ecaj_0529	2031	-	6			27	similar to Type IV secretory pathwayVirB6components
Ecaj_0530	1445	-	6			23	TrbL/VirB6 plasmid conjugal transfer protein
Ecaj_0531	936	+	7				TrbL/VirB6 plasmid conjugal transfer protein
Ecaj_0532	840	-	9			2	TrbL/VirB6 plasmid conjugal transfer protein
Ecaj_0533	801						CagE, TrbE, VirB family component of typeIVtransporter system
Ecaj_0534	102	-	2				Type IV secretory pathway, VirB3-like
Ecaj_0535	206						Superoxide dismutase.
Ecaj_0536	422	+	6			2	Binding-protein-dependent transport systemsinnermembrane component
Ecaj_0537	300						Lipoate synthase
Ecaj_0538	443	-	1				hypothetical
Ecaj_0539	657						Biotin/lipoyl attachment:Carbamoyl- phosphatesynthase L chain, ATP-binding:Carbamoyl- phosphatesynthetase large chain, N-terminal:Biotincarboxylase,C-terminal
Ecaj_0540	324						Pseudouridine synthase, RluD
Ecaj_0541	422					2	Diaminopimelate decarboxylase
Ecaj_0542	101						Ribosomal protein L28
Ecaj_0543	48					U	Hypothetical protein
Ecaj_0544	659						Primosomal protein n
Ecaj_0545	325	+	1				hypothetical
Ecaj_0546	241						Uroporphyrinogen III synthase HEM4
Ecaj_0547	575						Peptidase M24
Ecaj_0548	157						conserved hypothetical protein
Ecaj_0549	276						conserved hypothetical protein
Ecaj_0550	297						Small GTP-binding protein domain:GTP- binding protein Era
Ecaj_0551	450						Signal recognition particle protein

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0552	532	-	14			3	NADH/Ubiquinone/plastoquinone (complex I)
Ecaj_0553	155						Protein of unknown function UPF0083
Ecaj_0554	635					3	Heat shock protein Hsp70
Ecaj_0555	609						Ribonuclease E and G
Ecaj_0556	112	+	3				NADH-ubiquinone oxidoreductase, chain 4L
Ecaj_0557	140	-	4				Na <sup>+</sup> /H <sup>+</sup> antiporter MnhB subunit -related protein
Ecaj_0558	182	-	5				conserved hypothetical protein
Ecaj_0559	94	-	3				Na <sup>+</sup> /H <sup>+</sup> antiporter subunit
Ecaj_0560	106	+	3				Multiple resistance and pH regulation protein F
Ecaj_0561	662						Bacterial transketolase
Ecaj_0562	492						thermostable carboxypeptidase 1, putative[EC:3.4.17.19]
Ecaj_0563	163						OmpA/MotB domain
Ecaj_0564	432						Adenylosuccinate synthetase
Ecaj_0565	157						Conserved hypothetical protein 250
Ecaj_0566	623						Ribonucleoside-diphosphate reductase.
Ecaj_0567	284						Polyprenyl synthetase
Ecaj_0568	304	+	1				conserved domain protein
Ecaj_0570	354						FAD dependent oxidoreductase
Ecaj_0571	489					2	DnaB-like helicase,N-terminal:DnaB-like helicase, C-terminal
Ecaj_0572	142	-	1				conserved hypothetical protein
Ecaj_0573	140						Cytidine/deoxycytidylate deaminase, zinc- binding region
Ecaj_0574	320					2	Beta-ketoacyl-acyl carrier protein synthase III(FabH)
Ecaj_0575	338						Fatty acid synthesis plsX protein
Ecaj_0576	62						Ribosomal protein S32, bacterial and organelle form
Ecaj_0577	397						Queuine/other tRNA- ribosyltransferase:Queuine tRNA-ribosyltransferase
Ecaj_0578	254						ABC transporter
Ecaj_0579	268						Dihydrodipicolinate reductase
Ecaj_0580	164					3	Flavin reductase-like
Ecaj_0581	293	-	8				4-hydroxybenzoate polyprenyl transferase, proteobacterial
Ecaj_0582	296	-	7	C			4-hydroxybenzoate octaprenyltransferase
Ecaj_0583	45						Ribosomal protein L34
Ecaj_0584	120						hypothetical
Ecaj_0585	437	-	11			2	Sodium:alanine symporter
Ecaj_0586	673	-	13				ComEC/Rec2-related protein
Ecaj_0587	786						Bacterial phenylalanyl-tRNA synthetase, beta subunit
Ecaj_0588	133						Ribosomal protein L17
Ecaj_0589	373						DNA-directed RNA polymerase.
Ecaj_0590	127						Ribosomal protein S11
Ecaj_0591	124	-	1				Ribosomal protein S13
Ecaj_0592	222			C		2	Adenylate kinase
Ecaj_0593	437	-	10				SecY protein
Ecaj_0594	157						Ribosomal protein L15, bacterial form
Ecaj_0595	175						Ribosomal protein S5, bacterial and organelle form
Ecaj_0596	122						Ribosomal protein L18P/L5E
Ecaj_0597	179						Ribosomal protein L6
Ecaj_0598	133				S		Ribosomal protein S8

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0599	102				S		Ribosomal protein S14
Ecaj_0600	178						Ribosomal protein L5:Ribosomal protein L5
Ecaj_0601	110						Ribosomal protein L24
Ecaj_0602	120						Ribosomal protein L14, bacterial and organelle form
Ecaj_0603	84						Ribosomal protein S17
Ecaj_0604	68						Ribosomal protein L29
Ecaj_0605	137						Ribosomal protein L16
Ecaj_0606	212						Bacterial ribosomal protein S3
Ecaj_0607	115						Ribosomal protein L22, bacterial and organelle form
Ecaj_0608	94						Ribosomal protein S19, bacterial and organelle form
Ecaj_0609	277						Ribosomal protein L2, bacterial and organelle form
Ecaj_0610	97						Ribosomal L23 protein
Ecaj_0611	206						Ribosomal protein L4/L1e
Ecaj_0612	232						Ribosomal protein L3
Ecaj_0613	109						Ribosomal protein S10, bacterial form
Ecaj_0614	396						Translation elongation factor Tu:Small GTP- binding protein domain
Ecaj_0615	253						RNA methyltransferase TrmH, group 3
Ecaj_0616	215						Cytidylate kinase.
Ecaj_0617	568						RNA binding S1
Ecaj_0618	289	-	1				Peptidase S49, SppA
Ecaj_0619	88						Histone-like bacterial DNA-binding protein
Ecaj_0620	98						hypothetical
Ecaj_0621	123						conserved hypothetical protein
Ecaj_0622	306	-	3				hypothetical
Ecaj_0623	338						Ferrochelataase
Ecaj_0624	352						AFG1-like ATPase
Ecaj_0625	113						hypothetical
Ecaj_0626	286	-	3				Band 7 protein
Ecaj_0627	152						Ankyrin
Ecaj_0628	119						conserved hypothetical protein
Ecaj_0629	855					2	UvrD/REP helicase
Ecaj_0630	325					3	Zinc-containing alcohol dehydrogenase superfamily
Ecaj_0631	593	-	5			5	Type I secretion system ATPase, PrtD
Ecaj_0632	76	-	2			2	hypothetical protein
Ecaj_0633	281						Dihydropteroate synthase
Ecaj_0634	360	-	2				Hypothetical
Ecaj_0635	1077					6	Carbamoyl-phosphate synthase, large subunit , glutamine-dependent
Ecaj_0636	99					2	Hypothetical
Ecaj_0637	462						Fumarate hydratase, class II
Ecaj_0638	215						DNA/pantothenate metabolism flavoprotein,C- terminal
Ecaj_0639	448						Dihydroorotase multifunctional complex type
Ecaj_0640	215						Lipoate-protein ligase B
Ecaj_0641	209						Phosphoribosylglycinamide formyltransferase
Ecaj_0642	501						Leucyl aminopeptidase.
Ecaj_0643	225	+	0				Intradiol ring-cleavage dioxygenase
Ecaj_0644	81				T		hypothetical
Ecaj_0645	858						AAA ATPase, central region:Clp, N terminal

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0646	242						Short-chain dehydrogenase/reductase SDR
Ecaj_0647	553						Chaperonin Cpn60/TCP-1
Ecaj_0648	95						Chaperonin Cpn10
Ecaj_0649	231						DNA repair protein RadC
Ecaj_0651	74					U	hypothetical protein
Ecaj_0652	265						Phosphoribosyl formylglycinamide synthase I, putative [EC:6.3.5.3] KO:K01952phosphoribosylformylglycinamide synthase
Ecaj_0653	626				2		Glucose-inhibited division protein A subfamily
Ecaj_0654	310						GlpX
Ecaj_0655	202	+	0		2		Peptidase M23B
Ecaj_0656	152						conserved hypothetical protein
Ecaj_0657	325						Biotin synthase
Ecaj_0658	1005				2		Phosphoribosyl formylglycinamide synthase.
Ecaj_0659	173						7,8-Dihydro-6-hydroxymethylpterin-pyrophosphokinase,HPPK
Ecaj_0660	198			C			hypothetical
Ecaj_0661	436						Peptidase M48, Ste24p
Ecaj_0662	119				S		Iojax-related protein
Ecaj_0663	294						conserved hypothetical protein
Ecaj_0664	211						hypothetical
Ecaj_0665	343						Phosphoribosyl formylglycinamide cyclo-ligase
Ecaj_0666	311						Phage integrase:Phage integrase, N-terminalSAM- like
Ecaj_0667	325	+	1		2		Glycerol-3-phosphate dehydrogenase (NAD(P)+).
Ecaj_0668	472						Helix-turn-helix, Fis-type
Ecaj_0669	276	-	1				Cell division protein FtsQ
Ecaj_0670	314						Glutathione synthase.
Ecaj_0671	184						Transport-associated domain
Ecaj_0672	591				2		Aspartyl-tRNA synthetase
Ecaj_0673	212				T		HAD-superfamily hydrolase, subfamily IA,variant 1
Ecaj_0674	164	-	3				RDD
Ecaj_0675	875						Pyruvate,phosphate dikinase
Ecaj_0676	230						hypothetical
Ecaj_0677	497	-	13				NADH dehydrogenase (quinone).
Ecaj_0678	174	-	1				Cytochrome c, class I
Ecaj_0679	258						Alpha/beta hydrolase fold
Ecaj_0680	301				4		Methenyltetrahydrofolate cyclohydrolase.
Ecaj_0681	210						Guanylate kinase.
Ecaj_0682	235	-	6				Cytochrome c-type biogenesis protein CcmC
Ecaj_0683	191						Holliday junction DNA helicase RuvA KO:K03550holliday junction DNA helicase RuvA
Ecaj_0684	330						Holliday junction DNA helicase RuvB
Ecaj_0685	387	-	11				Drug resistance transporter Bcr/CflA subfamily
Ecaj_0686	294						Thymidylate synthase (FAD).
Ecaj_0687	259			C			Succinate dehydrogenase/fumarate reductase iron- sulfur protein
Ecaj_0688	599						Succinate dehydrogenase or fumarate reductase, flavoprotein subunit
Ecaj_0689	88	-	1				hypothetical
Ecaj_0690	582	+	5		3		ABC transporter, transmembrane region:ABCtransporter
Ecaj_0691	110				S		CyaY protein
Ecaj_0692	422						Glycine hydroxymethyltransferase.

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0693	209						Ribosomal protein L9:Ribosomal protein L9
Ecaj_0694	96						Ribosomal protein S18
Ecaj_0695	110						Ribosomal protein S6
Ecaj_0696	201	-	4				Colicin V production protein
Ecaj_0697	884	-	20				TRAP C4-dicarboxylate transport system permease DctM subunit
Ecaj_0698	451				3		DNA repair protein RadA
Ecaj_0699	166	+	1				hypothetical
Ecaj_0700	77						BolA-like protein
Ecaj_0701	111						Glutaredoxin-related protein
Ecaj_0702	678				2		NAD-dependent DNA ligase
Ecaj_0703	387						3'-5' exonuclease
Ecaj_0704	714	+	5				ATP-binding region, ATPase-like:Histidine kinase, HAMP region:Histidine kinase A, N-terminal
Ecaj_0705	88						Cold-shock DNA-binding domain
Ecaj_0706	274						conserved hypothetical protein
Ecaj_0707	188						dCTP deaminase.
Ecaj_0708	211						ABC transporter
Ecaj_0709	348	-	8				Cytochrome oxidase assembly
Ecaj_0710	247	+	1				DSBA oxidoreductase
Ecaj_0711	1108	-	2		3		hypothetical
Ecaj_0712	414						Histidyl-tRNA synthetase, class IIa
Ecaj_0713	259						SAICAR synthetase
Ecaj_0714	519						hypothetical
Ecaj_0715	1919	+	1		28		hypothetical
Ecaj_0716	1602	-	1		15		hypothetical
Ecaj_0717	227	-	1				hypothetical
Ecaj_0718	257	-	2				hypothetical
Ecaj_0719	209			C		U	hypothetical
Ecaj_0720	222	-	1	C			hypothetical
Ecaj_0721	224	-	1	C			hypothetical
Ecaj_0722	71				S		hypothetical
Ecaj_0723	249	+	1				hypothetical
Ecaj_0724	137	+	1				hypothetical
Ecaj_0725	191	+	1				hypothetical
Ecaj_0726	215	-	1				hypothetical
Ecaj_0727	98			C			hypothetical
Ecaj_0728	229	+	1				hypothetical
Ecaj_0729	103						hypothetical
Ecaj_0730	166			C			hypothetical
Ecaj_0731	221	-	1				hypothetical
Ecaj_0732	114	-	1				hypothetical
Ecaj_0733	160						hypothetical
Ecaj_0734	133						hypothetical
Ecaj_0735	101	-	1	C			hypothetical
Ecaj_0736	189	+	1				hypothetical
Ecaj_0737	183	+	1				hypothetical
Ecaj_0738	142	+	1				hypothetical

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0739	221	+	1				hypothetical
Ecaj_0740	195	-	1				hypothetical
Ecaj_0741	115						hypothetical
Ecaj_0742	69						hypothetical
Ecaj_0743	85						hypothetical
Ecaj_0744	128	+	1	C			hypothetical
Ecaj_0745	99	-	3				hypothetical
Ecaj_0746	244	+	1			2	hypothetical
Ecaj_0747	141	+	1				hypothetical
Ecaj_0748	122	-	1	C			hypothetical
Ecaj_0749	144	-	1				hypothetical
Ecaj_0750	190						Methylpurine-DNA glycosylase (MPG)
Ecaj_0751	243	-	1				hypothetical
Ecaj_0752	274	-	1				hypothetical
Ecaj_0753	272	-	1	C			hypothetical
Ecaj_0754	270	+	2				hypothetical
Ecaj_0755	243	-	1				hypothetical
Ecaj_0756	333						hypothetical
Ecaj_0757	217	-	6				Phosphatidate cytidyltransferase
Ecaj_0758	186						Ribosome recycling factor
Ecaj_0759	245						Aspartate/glutamate/uridylyltransferase
Ecaj_0760	997	+	1				conserved hypothetical protein
Ecaj_0761	147						RNase H
Ecaj_0762	353	+	1		S		hypothetical
Ecaj_0763	53				S/T		hypothetical
Ecaj_0764	191						Hypothetical
Ecaj_0765	1129					4	Transcription-repair coupling factor
Ecaj_0766	65	-	1	C		U	hypothetical
Ecaj_0767	92	+	1	C	S		hypothetical
Ecaj_0768	46					U	hypothetical
Ecaj_0769	64	-	1			U	hypothetical
Ecaj_0770	107	-	1				hypothetical
Ecaj_0771	246					7	hypothetical
Ecaj_0772	164	-	1				hypothetical
Ecaj_0773	213				T		Lumazine-binding protein
Ecaj_0774	468						conserved hypothetical protein
Ecaj_0775	422						Insulinase-like:Peptidase M16, C-terminal
Ecaj_0776	240	+	0				conserved hypothetical protein
Ecaj_0777	178						Bacterial protein export chaperone SecB
Ecaj_0778	213						Mitochondrial import inner membrane translocase, subunit Tim44
Ecaj_0779	412	-	12				permease, putative
Ecaj_0780	203						Thymidylate kinase
Ecaj_0781	327						Malonyl CoA-acyl carrier protein transacylase
Ecaj_0782	103						hypothetical
Ecaj_0783	264						NAD(+) kinase.
Ecaj_0784	486						IMP dehydrogenase
Ecaj_0785	284						Alpha/beta hydrolase fold

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0786	328					2	Pyruvate dehydrogenase (lipoamide).
Ecaj_0787	271	+	1				Conjugal transfer protein TrbG/VirB9/CagX
Ecaj_0788	111						Thioredoxin
Ecaj_0789	412						Protein of unknown function UPF0004:MiaB- like tRNA modifying enzyme
Ecaj_0790	63						exodeoxyribonuclease small subunit
Ecaj_0791	98						ETC complex I subunit conserved region
Ecaj_0792	306	-	10				Auxin Efflux Carrier
Ecaj_0793	183						Conserved hypothetical protein 95
Ecaj_0794	400	+	1				conserved hypothetical protein
Ecaj_0795	444						Glutamyl-tRNA synthetase bacterial/mitochondrial
Ecaj_0796	121	-	3				hypothetical
Ecaj_0797	262						Thiazole biosynthesis
Ecaj_0798	59						similar to ThiS sulfur transfer protein
Ecaj_0799	475					2	hypothetical
Ecaj_0800	187						Nitrogen-fixing NifU, C-terminal
Ecaj_0801	141	-	1			2	Heat shock protein DnaJ, N-terminal
Ecaj_0802	102				T		hypothetical
Ecaj_0803	470						Transcription termination factor Rho
Ecaj_0804	190						20S proteasome, A and B subunit s
Ecaj_0805	493					6	Heat shock protein HslU
Ecaj_0806	232						Ubiquinone/menaquinone biosynthesis methyltransferase
Ecaj_0807	507						Methionyl-tRNA synthetase, class Ia
Ecaj_0808	418						Aspartate kinase
Ecaj_0809	248	-	2				Cytochrome-c oxidase.
Ecaj_0810	519	-	12				Cytochrome-c oxidase.
Ecaj_0811	296	-	9				Protoheme IX farnesyltransferase
Ecaj_0812	255	+	0				Protein of unknown function UPF0169
Ecaj_0813	422						Phosphoribosylglycinamide synthetase
Ecaj_0814	149	-	1		S		YajC
Ecaj_0815	376						DNA polymerase III, beta chain
Ecaj_0816	217						3'-5' exonuclease
Ecaj_0817	267						Response regulator receiver:Transcriptional regulatory protein, C terminal
Ecaj_0818	212						Helix-turn-helix motif
Ecaj_0819	173						Inorganic diphosphatase.
Ecaj_0820	347	-	1			2	N-acetyl-gamma-glutamyl-phosphate reductase
Ecaj_0821	150						Ribosomal protein S9
Ecaj_0822	157						Ribosomal protein L13, bacterial and organelle form
Ecaj_0823	433					3	Outer membrane efflux protein
Ecaj_0824	243					4	hypothetical
Ecaj_0825	352						Protein of unknown function DUF59
Ecaj_0826	319					5	Ribose-phosphate pyrophosphokinase
Ecaj_0827	115					2	Glutamyl-tRNA(Gln) amidotransferase C subunit
Ecaj_0828	884					3	Aconitate hydratase 1
Ecaj_0829	135						Protein of unknown function DUF525
Ecaj_0830	364						Phosphoribosylaminoimidazole carboxylase, ATPase subunit
Ecaj_0831	281	+	0				Surface antigen msp4

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0832	277	+	0				Surface antigen msp4
Ecaj_0833	288	+	1				Surface antigen msp4
Ecaj_0834	783	-	1				hypothetical
Ecaj_0835	1268			C	6		hypothetical
Ecaj_0836	87	-	1	C			hypothetical
Ecaj_0837	346	-	1				hypothetical
Ecaj_0838	1511	+	1				hypothetical
Ecaj_0839	792						CagE, TrbE, VirB family component of type IV transporter system
Ecaj_0840	121	+	3				conserved hypothetical protein
Ecaj_0841	122	+	3				conserved hypothetical protein
Ecaj_0842	126	+	3		2		conserved hypothetical protein
Ecaj_0843	356	-	1				HflK
Ecaj_0844	291	-	1				HflC
Ecaj_0845	472	+	0				Peptidase S1, chymotrypsin:PDZ/DHR/GLGF domain
Ecaj_0846	194	+	1				Hypothetical
Ecaj_0847	172						Ribonuclease III.
Ecaj_0848	175	+	1				Cytochrome c oxidase assembly proteinCtaG/Cox11
Ecaj_0849	439	+	0				Insulinase-like:Peptidase M16, C-terminal
Ecaj_0850	452	+	0		2		Insulinase-like:Peptidase M16, C-terminal
Ecaj_0851	226	+	4		2		Hypothetical
Ecaj_0852	154	-	4				Peptidase A8, signal peptidase II
Ecaj_0853	306						Riboflavin kinase / FAD synthetase
Ecaj_0854	82						Glutaredoxin
Ecaj_0855	281						Modification methylase HemK
Ecaj_0856	267						Peptidase M24A, methionine aminopeptidase, subfamily 1
Ecaj_0857	401				3		Dihydrolipoamide succinyltransferase
Ecaj_0858	173						Bacterial transferase hexapeptide repeat
Ecaj_0859	384	+	1				Serine-type D-Ala-D-Ala carboxypeptidase.
Ecaj_0860	410	-	8				Cytochrome c biogenesis protein, transmembrane region
Ecaj_0861	94						Protein of unknown function DUF339
Ecaj_0862	381	-	4		3		Peptidase M50, putative membrane- associated zinc metallopeptidase
Ecaj_0863	770	+	0		5		Bacterial surface antigen (D15):Surface antigen variable number
Ecaj_0864	183	+	0				conserved hypothetical protein
Ecaj_0865	149						Beta-hydroxyacyl-(acyl-carrier- protein)dehydratase FabZ
Ecaj_0866	505						AICARFT/IMPCHase bienzyme
Ecaj_0867	186	-	4				CDP-diacylglycerol--glycerol-3-phosphate3-phosphatidyltransferase
Ecaj_0868	437	-	12				Cytochrome c oxidase, subunit I
Ecaj_0869	209	+	2				conserved hypothetical protein
Ecaj_0870	636	-	1		8		Pentapeptide repeat
Ecaj_0871	420						Glycerol-1-phosphate dehydrogenase [NAD(P)].
Ecaj_0872	244	-	7				H+-transporting two-sector ATPase, A subunit
Ecaj_0873	74	+	2				H+-transporting two-sector ATPase, C subunit
Ecaj_0874	168	-	1				H+-transporting two-sector ATPase, B/B' subunit
Ecaj_0875	162	-	1				hypothetical
Ecaj_0876	412						Cell division protein FtsA
Ecaj_0877	484	+	12		3		Cation transporter
Ecaj_0878	434						PP-loop



Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0879	612	+	2				Peptidase M41, FtsH
Ecaj_0880	262	-	7				Prolipoprotein diacylglyceryl transferase
Ecaj_0881	306					4	hypothetical
Ecaj_0882	528	-	5				hypothetical
Ecaj_0883	509	-	6				SecD/SecF/SecDF export membrane protein
Ecaj_0884	53						Hypothetical
Ecaj_0885	261	-	1				UBA/THIF-type NAD/FAD binding fold:MoeZ/MoeB
Ecaj_0886	199						Orotate phosphoribosyltransferase, Thermus type
Ecaj_0887	358						RecA bacterial DNA recombination protein
Ecaj_0888	232						Dethiobiotin synthase
Ecaj_0889	311						Cell division transporter substrate- binding protein FtsY
Ecaj_0890	483						Isocitrate dehydrogenase (NAD+).
Ecaj_0891	585						Bacterial RecJ exonuclease
Ecaj_0892	397						Polynucleotide adenyltransferase
Ecaj_0893	146						Nucleoside-diphosphate kinase.
Ecaj_0894	216						Helix-turn-helix motif
Ecaj_0895	297	+	1			2	Surface antigen msp4
Ecaj_0896	276	+	1				Surface antigen msp4
Ecaj_0897	149						hypothetical
Ecaj_0898	317	+	1				Surface antigen msp4
Ecaj_0899	284	+	0				Surface antigen msp4
Ecaj_0900	265	-	2				Surface antigen msp4
Ecaj_0901	290	+	1				Surface antigen msp4
Ecaj_0902	279	+	0				Surface antigen msp4
Ecaj_0903	186						hypothetical
Ecaj_0904	290	+	1				Surface antigen msp4
Ecaj_0905	280	+	0				Surface antigen msp4
Ecaj_0906	282	+	1				Surface antigen msp4
Ecaj_0907	300	+	0				Surface antigen msp4
Ecaj_0908	293	+	0				Surface antigen msp4
Ecaj_0909	295	+	1				Surface antigen msp4
Ecaj_0910	294	+	0				Surface antigen msp4
Ecaj_0911	281	+	0				Surface antigen msp4
Ecaj_0912	277	+	0				Surface antigen msp4
Ecaj_0913	288	+	1				Surface antigen msp4
Ecaj_0914	284	+	0				Surface antigen msp4
Ecaj_0915	281	+	0			3	Surface antigen msp4
Ecaj_0916	279	+	0				Surface antigen msp4
Ecaj_0917	289	+	0			2	Surface antigen msp4
Ecaj_0918	276	+	0			3	major outer membrane protein P30- 20;bacterialsurface antigen
Ecaj_0919	121						hypothetical
Ecaj_0920	183						hypothetical
Ecaj_0921	863						SecA protein
Ecaj_0922	134			C			hypothetical
Ecaj_0923	185						hypothetical
Ecaj_0924	145						hypothetical
Ecaj_0925	421						Cell division protein FtsZ

Locus tag	Size	SP	TH	Cys rich	Ser/Thr rich	Unique	annotation
Ecaj_0926	136						conserved hypothetical protein
Ecaj_0927	155						Protein of unknown function UPF0054
Ecaj_0928	256				2		ATPase, ParA type
Ecaj_0929	284						ParB-like partition protein
Ecaj_0930	174						RimM protein:PRC-barrel
Ecaj_0931	236						tRNA (guanine-N1-)-methyltransferase
Ecaj_0932	126						Ribosomal protein L19
Ecaj_0933	161			C			CinA, C-terminal
Ecaj_0934	634				2		Threonyl-tRNA synthetase, class IIa
Ecaj_0935	160						Initiation factor 3:Initiation factor 3
Ecaj_0936	156						Protein of unknown function UPF0079
Ecaj_0937	235	-	6				Protein of unknown function UPF0005
Ecaj_0938	326	-	6				hypothetical
Ecaj_0939	334				2		Glyceraldehyde-3-phosphate dehydrogenase, typeI
Ecaj_0940	220						ThiJ/PfpI
Ecaj_0941	269				2		Pyrroline-5-carboxylate reductase.
Ecaj_0942	492						DNA-directed DNA polymerase.
Ecaj_0943	102				S		conserved hypothetical protein TIGR00103
Ecaj_0944	337						Aspartate-semialdehyde dehydrogenase, USG- 1related
Ecaj_0945	394						S-adenosylmethionine synthetase

5 Supplementary material - Table 2. List of proteins properties in *E.canis*. Locus tag and length of protein is presented in the first two columns. Prediction of presence of signal peptide is presented in third column. Number of predicted transmembrane helices is presented in forth column. C in fifth column indicates protein with an increased Cys content. Number of Ser/Thr rich clusters of proteins is presented in sixth column, in case a protein do not contain such clusters but exhibits a large number of Ser or Thr residues proteins are marked with S or T respectively. U indicates that this protein is unique in *E.canis*.

Protein associated with glycosylation in <i>C.jenuni</i> and <i>N.meningitidis</i>	<i>E.canis</i> gene	Gene annotation in <i>E. canis</i>	%id	alignment length	e-value	bit score
acpP2	Ecaj_0062	hypothetical	33.33	39	4.3	20.8
acpP2	Ecaj_0071	hypothetical	32.14	56	0.14	25.8
acpP2	Ecaj_0113	hypothetical	23.68	38	4.3	20.8
acpP2	Ecaj_0127	surface protein-related protein	40	15	4.3	20.8
acpP2	Ecaj_0177	Bvg accessory factor	24.14	58	5.7	20.4
acpP2	Ecaj_0219	Phosphopantetheine-binding domain	29.82	57	0.012	29.3
acpP2	Ecaj_0297	conserved hypothetical protein	41.67	24	5.7	20.4
acpP2	Ecaj_0387	Ankyrin	29.41	51	1.5	22.3
acpP2	Ecaj_0493	Conserved hypothetical protein 46	36.84	19	2.5	21.6
acpP2	Ecaj_0730	hypothetical	50	18	2.5	21.6
acpP2	Ecaj_0863	Bacterial surface antigen (D15):Surface antigen variable number	31.82	44	1.1	22.7
acpP3	Ecaj_0080	similar to Uncharacterized membrane-bound protein	52.94	17	3.4	21.2
acpP3	Ecaj_0578	ABC transporter	30.43	23	4.4	20.8
acpP3	Ecaj_0728	hypothetical	70	10	9.8	19.6
acpP3	Ecaj_0875	hypothetical	46.15	26	3.4	21.2

Protein associated with glycosylation in <i>C.jejuni</i> and <i>N.meningitidis</i>	<i>E.canis</i> gene	Gene annotation in <i>E. canis</i>	%id	alignment length	e-value	bit score
Cj1293	Ecaj_0239	hypothetical	40	20	7.1	23.1
Cj1293	Ecaj_0760	conserved hypothetical protein	45	20	7.1	23.1
Cj1293	Ecaj_0928	ATPase, ParA type	29.73	37	9.2	22.7
Cj1294	Ecaj_0062	hypothetical	21.51	186	0.087	29.6
Cj1294	Ecaj_0253	conserved hypothetical protein	23.01	113	1.3	25.8
Cj1294	Ecaj_0321	conserved hypothetical protein	25	60	6.3	23.5
Cj1294	Ecaj_0410	Aminotransferase, class V	20.59	68	2.1	25
Cj1294	Ecaj_0920	hypothetical	23	100	1.3	25.8
Cj1296	Ecaj_0320	O-methyltransferase, family 3	45	20	3.5	21.9
Cj1296	Ecaj_0785	Alpha/beta hydrolase fold	27.78	126	0.029	28.9
Cj1296	Ecaj_0892	Polynucleotide adenyllyltransferase	37.5	48	3.5	21.9
Cj1297	Ecaj_0250	2-polyprenylphenol 6-hydroxylase	26.83	41	5	21.9
Cj1297	Ecaj_0253	conserved hypothetical protein	42.86	21	3.8	22.3
Cj1297	Ecaj_0271	hypothetical	32.14	28	1	24.3
Cj1297	Ecaj_0313	Protein of unknown function DUF185	40.74	27	8.5	21.2
Cj1297	Ecaj_0387	Ankyrin	28.57	42	1.3	23.9
Cj1297	Ecaj_0472	hypothetical	43.48	23	8.5	21.2
Cj1297	Ecaj_0753	hypothetical	26.79	56	6.5	21.6
Cj1303	Ecaj_0253	conserved hypothetical protein	22.5	120	5.8	23.5
Cj1303	Ecaj_0574	Beta-ketoacyl-acyl carrier protein synthaseIII(FabH)	27.22	338	2.00E-30	124
Cj1303	Ecaj_0711	hypothetical	24.72	89	0.9	26.2
Cj1303	Ecaj_0760	conserved hypothetical protein	24.64	69	7.6	23.1
Cj1312	Ecaj_0062	hypothetical	29.82	57	4.2	23.5
Cj1312	Ecaj_0198	conserved hypothetical protein	40	25	7.2	22.7
Cj1312	Ecaj_0210	Protein of unknownfunction DUF21	45	20	9.4	22.3
Cj1312	Ecaj_0508	hypothetical	28.26	92	0.38	26.9
Cj1312	Ecaj_0566	Ribonucleoside-diphosphate reductase.	22.67	75	5.5	23.1
Cj1312	Ecaj_0711	hypothetical	46.67	15	7.2	22.7
Cj1312	Ecaj_0844	HflC	38.46	26	4.2	23.5
Cj1312	Ecaj_0902	Surface antigen msp4	28.26	46	9.4	22.3
Cj1312	Ecaj_0923	hypothetical	40	25	7.2	22.7
Cj1313	Ecaj_0231	hypothetical	23.3	103	2.5	23.1
Cj1313	Ecaj_0344	hypothetical	29.73	37	3.3	22.7
Cj1313	Ecaj_0387	Ankyrin	39.47	38	9.6	21.2
Cj1313	Ecaj_0628	conserved hypothetical protein	77.78	9	7.4	21.6
Cj1313	Ecaj_0666	Phage integrase	23.4	47	4.3	22.3
Cj1319	Ecaj_0255	Dehydrogenase, E1 component	54.55	22	0.81	26.2
Cj1319	Ecaj_0322	Protein of unknown function UPF0031 :YjeF- relatedprotein, N-terminal	26.98	63	3.1	24.3
Cj1319	Ecaj_0570	FAD dependent oxidoreductase	37.5	40	4	23.9
Cj1319	Ecaj_0646	Short-chain dehydrogenase/reductase SDR	42.86	21	5.2	23.5
Cj1319	Ecaj_0823	Outer membrane efflux protein similar to Uncharacterized membrane-bound protein	23.81	63	5.2	23.5
Cj1320	Ecaj_0080	similar to Uncharacterized membrane-bound protein	29.55	44	3.8	24.3
Cj1320	Ecaj_0201	Trigger factor	21.43	112	3.8	24.3
Cj1320	Ecaj_0316	Dihydrouridine synthase TIM-barrel protein nifR3	28.12	32	6.4	23.5
Cj1320	Ecaj_0333	Aminotransferase, class I and II	26.76	71	0.069	30
Cj1320	Ecaj_0387	Ankyrin	32.35	34	6.4	23.5
Cj1320	Ecaj_0410	Aminotransferase, class V	19.28	83	1.3	25.8
Cj1320	Ecaj_0411	Protein of unknown function UPF0074	28.29	152	0.014	32.3
Cj1320	Ecaj_0508	hypothetical	38.18	55	6.4	23.5
Cj1320	Ecaj_0711	hypothetical	22.58	155	3.8	24.3
Cj1320	Ecaj_0835	hypothetical	32.43	37	4.9	23.9
Cj1320	Ecaj_0870	Pentapeptide repeat	22.45	343	2.2	25
Cj1320	Ecaj_0885	UBA/THIF-type NAD/FAD binding fold:MoeZ/MoeB	26.47	68	1.3	25.8
Cj1321	Ecaj_0119	hypothetical	31.11	45	6.9	21.9
Cj1321	Ecaj_0157	Beta-lactamase-like	20.17	119	2.4	23.5
Cj1321	Ecaj_0387	Ankyrin	47.62	21	1.8	23.9
Cj1321	Ecaj_0415	hypothetical	29.55	44	9	21.6
Cj1324	Ecaj_0060	hypothetical	30.43	46	2.8	24.6

Protein associated with glycosylation in <i>C.jejuni</i> and <i>N.meningitidis</i>	<i>E.canis</i> gene	Gene annotation in <i>E. canis</i>	%id	alignment length	e-value	bit score
Cj1324	Ecaj_0174	hypothetical	34.48	29	8.1	23.1
Cj1324	Ecaj_0367	Argininosuccinate synthase	27.78	54	0.43	27.3
Cj1324	Ecaj_0387	Ankyrin	37.14	35	3.6	24.3
Cj1324	Ecaj_0740	hypothetical	28	50	8.1	23.1
Cj1324	Ecaj_0878	PP-loop	30.88	68	2.1	25
Cj1329	Ecaj_0013	Phage portal protein, HK97 family	33.96	53	1.9	25
Cj1329	Ecaj_0044	Phage SPO1 DNA polymerase-related protein	48	25	9.5	22.7
Cj1329	Ecaj_0067	hypothetical	21.84	87	7.3	23.1
Cj1329	Ecaj_0322	Protein of unknown function similar to 3-hydroxyacyl-CoA dehydrogenase	32.43	37	3.3	24.3
Cj1329	Ecaj_0455	similar to 3-hydroxyacyl-CoA dehydrogenase	22.81	57	7.3	23.1
Cj1329	Ecaj_0919	Hypothetical	31.25	48	1.1	25.8
galE	Ecaj_0139	Hypothetical	33.33	21	9.1	22.7
galE	Ecaj_0260	MutS 1 protein	30.65	62	5.3	23.5
galE	Ecaj_0271	Hypothetical	25.49	51	7	23.1
galE	Ecaj_0365	Ankyrin	33.33	45	9.1	22.7
galE	Ecaj_0834	Hypothetical	26.53	49	5.3	23.5
Glycosylated	Ecaj_0017	Hypothetical				
Glycosylated	Ecaj_0109	gp36				
Glycosylated	Ecaj_0113	Hypothetical				
neuA2	Ecaj_0052	Ankyrin	34.29	35	5.8	22.7
neuA2	Ecaj_0116	Phage integrase:Phage integrase, N-terminalSAM- like	41.38	29	0.53	26.2
neuA2	Ecaj_0248	Periplasmic solute binding protein	30	30	9.9	21.9
neuA2	Ecaj_0297	conserved hypothetical protein	31.58	38	4.5	23.1
neuA2	Ecaj_0469	Acriflavin resistance protein	28.57	28	9.9	21.9
neuA2	Ecaj_0856	Peptidase M24A, methionine aminopeptidase, subfamily 1	37.84	37	1.2	25
neuB2	Ecaj_0071	Hypothetical	26	50	5.4	23.5
neuB2	Ecaj_0719	Hypothetical	20.24	84	1.9	25
neuB2	Ecaj_0824	Hypothetical	24.19	62	2.4	24.6
neuB3	Ecaj_0095	TrkA-N	23.19	69	4.3	23.9
neuC2	Ecaj_0378	Hypothetical	44.44	27	4.9	23.9
neuC2	Ecaj_0411	Protein of unknown function UPF0074	23.64	110	0.99	26.2
neuC2	Ecaj_0414	Hypothetical	30.19	53	4.9	23.9
neuC2	Ecaj_0450	Hypothetical	29.69	64	0.2	28.5
neuC2	Ecaj_0505	Thiamine-monophosphate kinase	39.29	28	6.4	23.5
neuC2	Ecaj_052	Small GTP-binding protein domain	24.32	111	2.2	25
neuC2	Ecaj_0739	Hypothetical	29.09	55	8.4	23.1
neuC2	Ecaj_0875	Hypothetical	24.32	74	4.9	23.9
pgIA	Ecaj_0485	Hypothetical	23.91	46	4.8	23.9
pgIA	Ecaj_0514	ABC transporter, family 3	33.33	39	3.7	24.3
pgIA	Ecaj_0838	Hypothetical	33.33	33	2.1	25
pgIA	Ecaj_0844	HflC	39.39	33	6.3	23.5
pgIA	Ecaj_0870	Pentapeptide repeat	26.83	41	8.2	23.1
pgIB	Ecaj_0154	Binding-protein-dependent transport systems inner membrane component	23.12	199	1.8	26.2
pgIB	Ecaj_0475	Proton-translocating NADH-quinoneoxidoreductase, chain M	25.95	131	0.37	28.5
pgIB	Ecaj_0531	TrbL/VirB6 plasmid conjugal transfer protein	26.23	122	1.8	26.2
pgIB	Ecaj_0765	Transcription-repair coupling factor	50	18	1.1	26.9
pgIB	Ecaj_0897	Hypothetical	25	84	3.2	25.4
pgIB	Ecaj_0900	Surface antigen msp4	28.89	45	2.4	25.8
pgIC	Ecaj_0018	TRAG protein	26.56	64	2.8	23.5
pgIC	Ecaj_0104	Hypothetical	29.41	34	6.2	22.3
pgIC	Ecaj_0324	Hypothetical	27.78	54	2.1	23.9
pgIC	Ecaj_0416	Hypothetical	58.33	12	8.1	21.9
pgIC	Ecaj_0631	Type I secretion system ATPase, PrtD	39.47	38	2.8	23.5
pgIC	Ecaj_0770	Hypothetical	39.13	23	4.7	22.7
pgID	Ecaj_0198	conserved hypothetical protein	39.29	28	3.7	23.1
pgID	Ecaj_0293	60 kDa inner membrane protein	29.03	62	1.3	24.6
pgID	Ecaj_0631	Type I secretion system ATPase, PrtD	34.78	46	0.97	25
pgID	Ecaj_0902	Surface antigen msp4	50	20	3.7	23.1

Protein associated with glycosylation in <i>C.jenuni</i> and <i>N.meningitidis</i>	<i>E.canis</i> gene	Gene annotation in <i>E. canis</i>	%id	alignment length	e-value	bit score
pgIE	Ecaj_0018	TRAG protein	28.36	67	1.7	25.4
pgIE	Ecaj_0060	Hypothetical	28.85	52	4.9	23.9
pgIE	Ecaj_0333	Aminotransferase, class I and II	45.45	33	0.12	29.3
pgIE	Ecaj_0410	Aminotransferase, class V	21.59	176	0.006	33.5
pgIE	Ecaj_0444	Acetylglutamate kinase	23.85	109	0.069	30
pgIE	Ecaj_0630	Zinc-containing alcohol dehydrogenase superfamily	26.83	41	6.5	23.5
pgIE	Ecaj_0706	conserved hypothetical protein	31.58	38	8.4	23.1
pgIE	Ecaj_0830	Phosphoribosyl aminoimidazole carboxylase, ATPase subunit	27.85	79	0.2	28.5
pgIE	Ecaj_0834	Hypothetical	32.65	49	6.5	23.5
pgIF	Ecaj_0119	Hypothetical	42.11	19	8.1	23.9
pgIF	Ecaj_0181	Phospholipase/Carboxylesterase	38.71	31	2.1	25.8
pgIF	Ecaj_0416	Hypothetical	21.78	101	0.95	26.9
pgIF	Ecaj_0489	Enolase	24.74	97	2.8	25.4
pgIF	Ecaj_0529	similar to Type IV secretory pathwayVirB6components	22.81	57	8.1	23.9
pgIF	Ecaj_0570	FAD dependent oxidoreductase	32.14	28	6.2	24.3
pgIF	Ecaj_0646	Short-chain dehydrogenase/reductase SDR	29.58	71	0.003	35
pgIG	Ecaj_0328	conserved hypothetical protein	25.23	111	0.15	28.5
pgIG	Ecaj_0380	conserved hypothetical protein	29.27	41	6.1	23.1
pgIG	Ecaj_0463	Hypothetical	24.19	62	2.1	24.6
pgIG	Ecaj_0474	Proton-translocating NADH-quinoneoxidoreductase, chain N	30.61	49	2.1	24.6
pgIG	Ecaj_0733	Hypothetical	37.5	24	6.1	23.1
pgIG	Ecaj_0830	Phosphoribosyl aminoimidazole carboxylase, ATPase subunit	25.49	51	4.7	23.5
pgIH	Ecaj_0208	Thiamine monophosphate synthase	20.83	72	0.41	27.3
pgIH	Ecaj_0760	conserved hypothetical protein	43.75	32	4.5	23.9
pgIH	Ecaj_0778	Mitochondrial import inner membrane translocase, subunit Tim44	24.44	45	7.7	23.1
pgIH	Ecaj_0789	Protein of unknown function UPF0004	34.38	32	1.6	25.4
pgII	Ecaj_0018	TRAG protein	36.73	49	2.2	24.6
pgII	Ecaj_0018	TRAG protein	25.71	70	4.9	23.5
pgII	Ecaj_0063	Hypothetical	25.21	119	0.34	27.3
pgII	Ecaj_0127	surface protein-related protein	24.27	206	2.2	24.6
pgII	Ecaj_0365	Ankyrin	35.48	62	0.008	32.7
pgII	Ecaj_0387	Ankyrin	20.53	151	0.58	26.6
pgII	Ecaj_0740	Hypothetical	20	110	2.2	24.6
pgII	Ecaj_0778	Mitochondrial import inner membrane translocase, subunit Tim44	23.73	59	0.34	27.3
pgII	Ecaj_0838	Hypothetical	37.93	29	8.4	22.7
pgIJ	Ecaj_0229	Hypothetical	33.33	36	7.9	23.1
pgIJ	Ecaj_0711	Hypothetical	27.63	76	0.11	29.3
pgIJ	Ecaj_0714	Hypothetical	17.5	40	7.9	23.1
pgIJ	Ecaj_0823	Outer membrane efflux protein	29.41	51	0.42	27.3
pgIJ	Ecaj_0835	Hypothetical	25	56	2.7	24.6
pseA	Ecaj_0017	Hypothetical	32.35	34	8.2	23.1
pseA	Ecaj_0093	competence protein F	27.78	54	3.7	24.3
pseA	Ecaj_0116	Phage integrase:Phage integrase, N-terminalSAM- like	31.25	64	0.97	26.2
pseA	Ecaj_0126	Hypothetical	22.22	126	2.8	24.6
pseA	Ecaj_0142	Hypothetical	29.27	41	4.8	23.9
pseA	Ecaj_0676	Hypothetical	24.03	129	0.018	32
pseA	Ecaj_0713	SAICAR synthetase	24.68	77	0.97	26.2
ptmA	Ecaj_0269	Short-chain dehydrogenase/reductase SDR	25.75	233	4.00E-12	63.2
ptmA	Ecaj_0297	conserved hypothetical protein	28	50	6.6	22.7
ptmA	Ecaj_0374	Glucose/ribitol dehydrogenase	23.4	265	6.00E-14	69.3
ptmA	Ecaj_0387	Ankyrin	29.55	44	6.6	22.7
ptmA	Ecaj_0394	putative immunogenic protein	24.39	41	2.3	24.3
ptmA	Ecaj_0431	Hypothetical	34.62	52	1	25.4
ptmA	Ecaj_0464	AAA ATPase, central region:Clp, N terminal	22.08	77	0.35	26.9

Protein associated with glycosylation in <i>C.jenuni</i> and <i>N.meningitidis</i>	<i>E.canis</i> gene	Gene annotation in <i>E. canis</i>	%id	alignment length	e-value	bit score
ptmA	Ecaj_0813	Phosphoribosylglycinamide synthetase	25.93	108	1.7	24.6
ptmA	Ecaj_0902	Surface antigen msp4	25.86	58	0.46	26.6
ptmB	Ecaj_0297	conserved hypothetical protein	23.88	134	1.2	25
ptmB	Ecaj_0491	Bacterial methyltransferase	25.86	58	4.5	23.1
ptmB	Ecaj_0687	Succinate dehydrogenase/fumarate reductase iron- sulfur protein	40.62	32	7.8	22.3
ptmB	Ecaj_0702	NAD-dependent DNA ligase	28.87	97	0.001	35
ptmB	Ecaj_0863	Bacterial surface antigen (D15)	28.95	38	5.9	22.7
wlaB	Ecaj_0051	ABC transporter	30.74	231	1.00E-20	92.8
wlaB	Ecaj_0117	ABC transporter	32.28	189	8.00E-18	83.6
wlaB	Ecaj_0243	ABC transporter	29.86	221	8.00E-21	93.6
wlaB	Ecaj_0249	ABC transporter	27.14	199	6.00E-13	67.4
wlaB	Ecaj_0297	60 kDa inner membrane protein	25.32	79	4.5	24.6
wlaB	Ecaj_0508	hypothetical	27.5	80	7.7	23.9
wlaB	Ecaj_0520	Secretion protein HlyD	21.74	92	7.7	23.9
wlaB	Ecaj_0533	CagE, TrbE, VirB family component of typeIV transporter system	41.03	39	2	25.8
wlaB	Ecaj_0550	Small GTP-binding protein domain:GTP-binding protein Era	40	25	1.2	26.6
wlaB	Ecaj_0551	Signal recognition particle protein	24.59	61	0.9	26.9
wlaB	Ecaj_0568	conserved domain protein	30	90	0.9	26.9
wlaB	Ecaj_0578	ABC transporter	31.25	224	4.00E-19	87.8
wlaB	Ecaj_0671	Transport-associated domain	27.69	65	2.6	25.4
wlaB	Ecaj_0690	ABC transporter	26.08	464	2.00E-37	148
wlaB	Ecaj_0696	Colicin V production protein	52.63	19	7.7	23.9
wlaB	Ecaj_0703	3'-5' exonuclease	25.42	59	7.7	23.9
wlaB	Ecaj_0708	ABC transporter	27.93	179	2.00E-09	55.5
wlaB	Ecaj_0715	hypothetical	23.66	131	5.9	24.3
wlaB	Ecaj_0751	hypothetical	27.54	69	7.7	23.9
wlaB	Ecaj_0765	Transcription-repair coupling factor	35	40	5.9	24.3
wlaB	Ecaj_0844	HflC	24.44	45	7.7	23.9
wlaB	Ecaj_0861	Protein of unknown function DUF339	31.37	51	4.5	24.6
wlaB	Ecaj_0872	H <sup>+</sup> -transporting two-sector ATPase, A subunit	38.24	34	7.7	23.9
wlaB	Ecaj_0936	Protein of unknown function UPF0079	35.29	34	0.037	31.6
wlaJ	Ecaj_0025	hypothetical	23.53	85	1.1	25
wlaJ	Ecaj_0174	hypothetical	29.63	54	2.4	23.9
wlaJ	Ecaj_0465	Integral membrane protein TerC family	25	56	1.4	24.6
wlaJ	Ecaj_0547	Peptidase M24	30.23	43	4	23.1
wlaJ	Ecaj_0718	hypothetical	35.42	48	2.4	23.9
wlaJ	Ecaj_0852	Peptidase A8, signal peptidase II	46.15	13	9	21.9
wlaJ	Ecaj_0856	Peptidase M24A, methionine aminopeptidase, subfamily 1	34.29	35	1.4	24.6
wlaJ	Ecaj_0936	Protein of unknown function UPF0079	45.45	22	5.3	22.7

Supplementary material – Table 3. Proteins or *E.canis* that exhibit similarity with proteins involved in protein glycosylation in *C.jenuni* and *N.meningitidis*