

Supporting Online Material

Materials and methods

Genome sequencing

UWE25 was cultivated in *Acanthamoeba* sp. UWC1 (1). Elementary bodies were harvested and purified by step gradient density centrifugation. High molecular weight genomic DNA was isolated from purified elementary bodies by proteinase K/SDS incubation and guanidinium-isothiocyanate/isobutanol precipitation on silica. Whole-genome shotgun libraries were obtained by fragmenting genomic DNA using mechanical shearing and cloning 1.0 - 3.0 kilobase (Kb) fragments into pGEM-T vectors (Promega, Madison, WI, USA). Double-ended plasmid sequencing reactions were carried out using BigDye Terminator chemistry on ABI 3700 capillary sequencers (Applied Biosystems, Foster City, CA). The whole genome sequence of UWE25 was assembled from 27,353 reads with an 8.8-fold redundancy using the Paracel Genome Assembler software (Paracel Inc., Pasadena, CA) and the Staden Package (http://www.mrc-lmb.cam.ac.uk/pubseq/staden_home.html). Gap closure was performed by Vectorette PCR (2), combinatorial PCR, and primer walking. Sequencing or sub-cloning of a single approximately 5 kb PCR fragment, which would have been required to close the genome sequence, was not possible despite of major efforts in two independent laboratories (MWG Biotech and University of Vienna).

Genome analysis

The PEDANT software system was used for genome sequence analysis (3) and annotation (<http://pedant.gsf.de>). Prediction of coding sequences (CDSs) was performed with Orpheus (4). Translated CDSs longer than 180 nucleotides were searched against a non-redundant

Supporting Online Material

protein sequence database derived from PIR (5), SwissProt (6) and trEMBL using blastp (cut-off p-value $\leq 1e-04$) (7) and further analyzed by searching against Pfam (8), Prosite, COGs (9), PIR, and BLOCKS (10). Structural predictions of derived proteins were carried out by the IMPALA software searching against SCOP (11) and PDB, and additional tools implemented in PEDANT. The comprehensive data collected automatically for each CDS by PEDANT were subsequently used for manual annotation, which included additional FastA searches against SwissProt and assignment of CDSs to functional categories according to PEDANT’s functional catalogue FunCat (3). Only CDSs with amino acid sequence homology to characterized proteins (either by knock-out/complementation experiments, protein expression, or by known 3D structure) were annotated as “strongly similar to known proteins” (> 40% amino acid sequence identity to characterized protein) or as “similar to known proteins” (> 20% amino acid sequence identity to characterized protein). CDSs with sequence homology to not yet functionally characterized proteins were classified as “conserved hypothetical proteins” (> 30% amino acid sequence identity to first blast hit) or “hypothetical proteins” (> 20% amino acid sequence identity to first blast hit). CDSs showing no significant homology to proteins in public databases were annotated as “unknown protein”. Biochemical pathway prediction and reconstruction was performed using KEGG (12). tRNA genes were identified with the tRNAscan-SE program (13). Prediction of subcellular location of plant proteins was performed using TargetP (14), iPSORT (15), and ChloroP (16). The origin of replication was determined by cumulative GC-skew analysis using GenSkew (fig. S8; <http://mips.gsf.de/services/analysis/genskew>). Codon usage and codon adaptation index were calculated using the programs cusp and cai

Supporting Online Material

included in the EMBOSS software package (17). Comparative genome sequence analysis with *Chlamydophila pneumoniae* CWL029, *Chlamydophila caviae*, *Chlamydia trachomatis* sv D and *Chlamydia muridarum* were performed using SIMAP, the mips Similarity Matrix of Proteins (<http://mips.gsf.de/services/analysis/simap>). A linear representation of the UWE25 genome is available as fig. S9.

Phylogenetic sequence analysis

The software ARB (18) was used for phylogenetic sequence analysis of selected proteins (<http://www.arb-home.de>). Protein databases were constructed by blastp search (7) against 147 complete microbial and eukaryotic genome sequences. Amino acid sequences were aligned automatically with ClustalW (19) implemented in ARB and the resulting alignment was refined manually. Phylogenetic amino acid sequence trees were constructed by applying neighbor joining (PAM and Kimura correction), PHYLIP distance matrix (Fitch), PHYLIP maximum parsimony methods (20), and maximum likelihood approaches using PROTml 2.3 (JTT amino-acid replacement model) and TREE-PUZZLE (21) implemented in ARB. Maximum parsimony bootstrap values (1000 resamplings) and TREE-PUZZLE support values were calculated to estimate the reliability of obtained tree topologies.

Data deposition

In addition to the genome sequence data deposited at EMBL/GenBank/DDBJ (accession number BX908798), the comprehensive UWE25 genome database is available at <http://mips.gsf.de/services/genomes/uwe25>.

Supporting Online Material

Metabolism of UWE25

Glycolysis

Pathogenic chlamydiae lack a glucokinase and have an incomplete phosphotransferase system (PTS), and therefore rely on import of phosphorylated sugars for glycolysis (22, 23). The UWE25 genome also lacks key components of the PTS and encodes a glucose-6-phosphate transporter (pc0387) orthologous to the respective transporters in pathogenic chlamydiae. In addition, and unlike its pathogenic counterparts, UWE25 possesses a glucokinase (pc0935) and does thus not depend on host-derived phosphorylated sugars. The presence of a pyrophosphate-dependent phosphofructokinase (pc0880) indicates that UWE25 like pathogenic chlamydiae is able to generate four ATP molecules during glycolysis.

Pentose phosphate pathway

The pentose phosphate pathway of UWE25 is complete. Like pathogenic chlamydiae (23), UWE25 is thus able to generate pentose phosphates, erythrose-4-phosphate and to reduce nicotinamide adenine dinucleotide phosphate (NADP) independently from its host.

Oxidative phosphorylation

Both, UWE25 and pathogenic chlamydiae have a respiratory chain most similar to the respiratory chain of *Escherichia coli* expressed under microaerophilic conditions (22, 23). In addition to a Na⁺-dependent NADH oxidoreductase (pc0559-pc0572), a cytochrome bd complex (pc1629, pc1630), and a V-type ATPase (pc1676-pc1680, pc1682) present in

Supporting Online Material

pathogenic chlamydiae, UWE25 also encodes a H⁺-translocating NADH oxidoreductase (pc0559-pc0572), a cytochrome C (pc0909), a cytochrome C oxidase (pc1189-pc1193), and an additional F-type ATPase (pc1667-pc1674).

Amino acids, nucleotides, and cofactors

Amino acid biosynthesis pathways in both, UWE25 and pathogenic chlamydiae, are frequently truncated (23). Like pathogenic chlamydiae, UWE25 is not able to generate threonine, cysteine, valine, leucine, isoleucine, lysine, arginine, phenylalanine, and tyrosine. Surprisingly UWE25 also lacks most key enzymes of the tryptophan biosynthesis pathway, which is complete in some pathogenic chlamydiae (24-29). This clearly contrasts with the generally greater metabolic/biosynthetic capabilities of UWE25. UWE25 probably does not require to synthesize tryptophan because it should not experience tryptophan depletion in its host, which is observed as immune response in mammalian cells upon infection with pathogenic chlamydiae. In contrast to pathogenic chlamydiae, however, UWE25 is able to generate glycine, serine, glutamine, and proline. Purine and pyrimidine biosynthesis pathways are equally incomplete in UWE25 as in pathogenic chlamydiae (23), which thus rely on the import of nucleotides from their host cells. Both, pathogenic chlamydiae and UWE25 lack key enzymes of several cofactor biosynthesis pathways (23) and thus have to acquire cofactors like coenzymeA and nicotinamide adenine dinucleotide (NAD) from their hosts. However, in contrast to pathogenic chlamydiae, UWE25 is able to produce menaquinone, which can be used instead of host-derived ubiquinone in the quinone pool of the respiratory chain.

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)

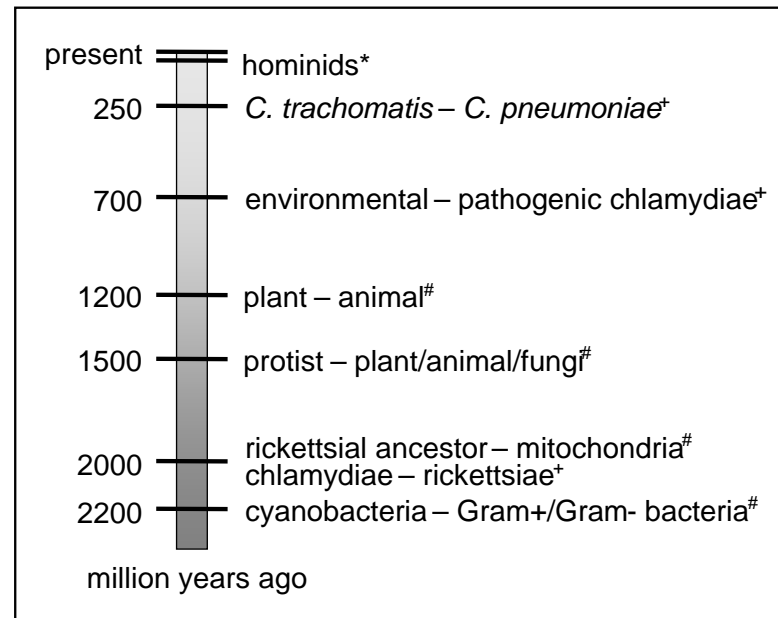
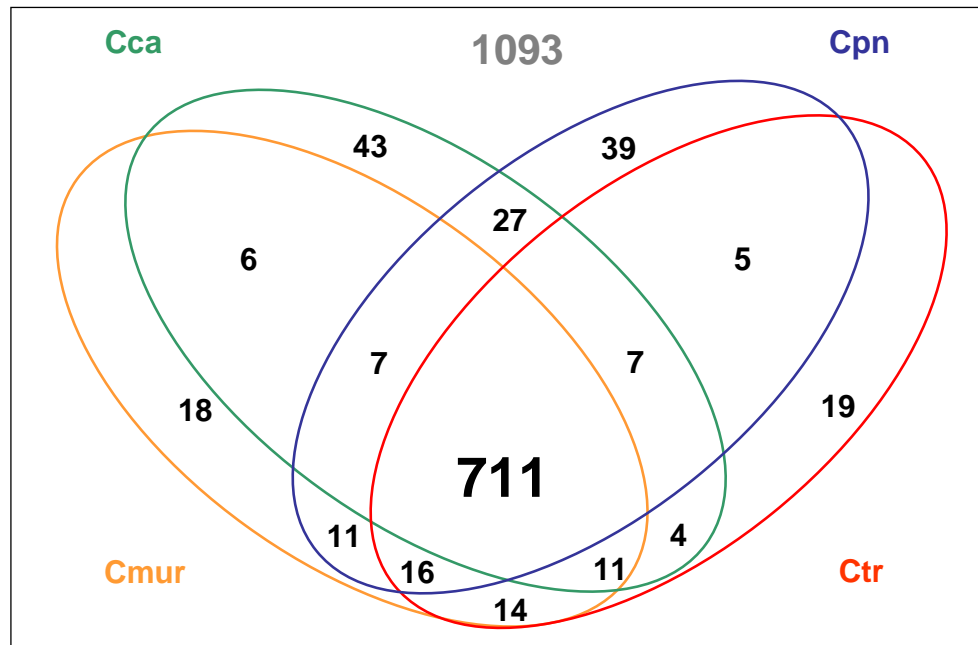


Fig. S1

Evolutionary time scale of chlamydiae. Estimates of major divergence events are based on 16S rRNA dissimilarities and an assumed divergence rate of 1% per 50 million years (30) (+), or taken from Hacia *et al.* (31) (*), and Feng *et al.* (32) (#), respectively.

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)



Chlamydiaceae homologues total: 938

Fig. S2

Number of UWE25 proteins with sequence homology (fastA score/selfscore > 0.1) to proteins of pathogenic chlamydiae. *Chlamydophila pneumoniae* CWL029 (Cpn), *Chlamydophila caviae* (Cca), *Chlamydia trachomatis* sv D (Ctr), *Chlamydia muridarum* (Cmur).

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)

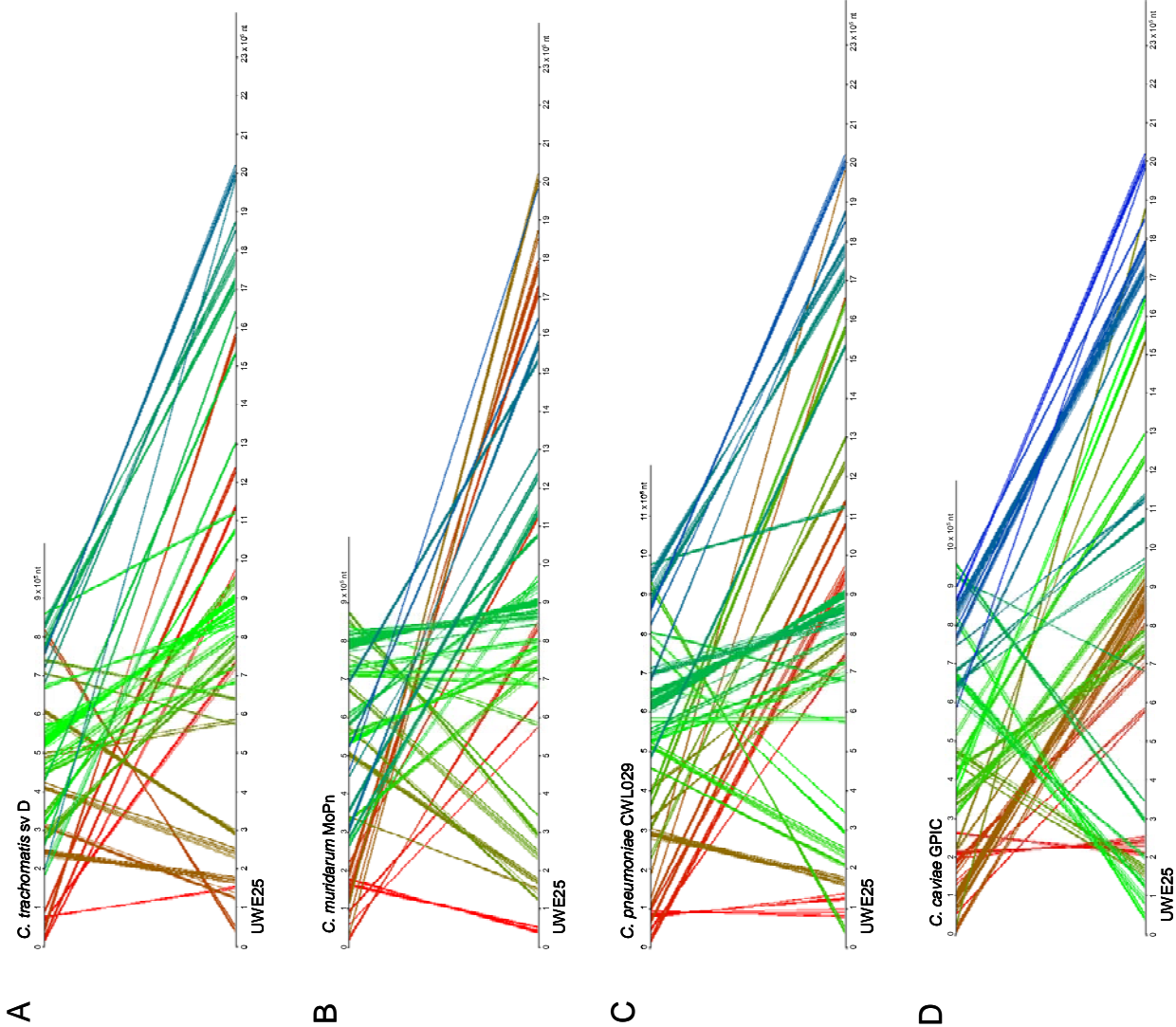


Fig. S3

Comparison of genome synteny between UWE25 and pathogenic chlamydiae based on reciprocal fastA best matches (score/selfscore > 0.2). (A) UWE25 versus *C. trachomatis* sv D; (B) UWE25 versus *C. muridarum* MoPn; (C) UWE25 versus *C. pneumoniae* CWL029; (D) UWE25 versus *C. caviae* GPIC. Genomes have been rotated so that the replication origin is at 0 kb.

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)

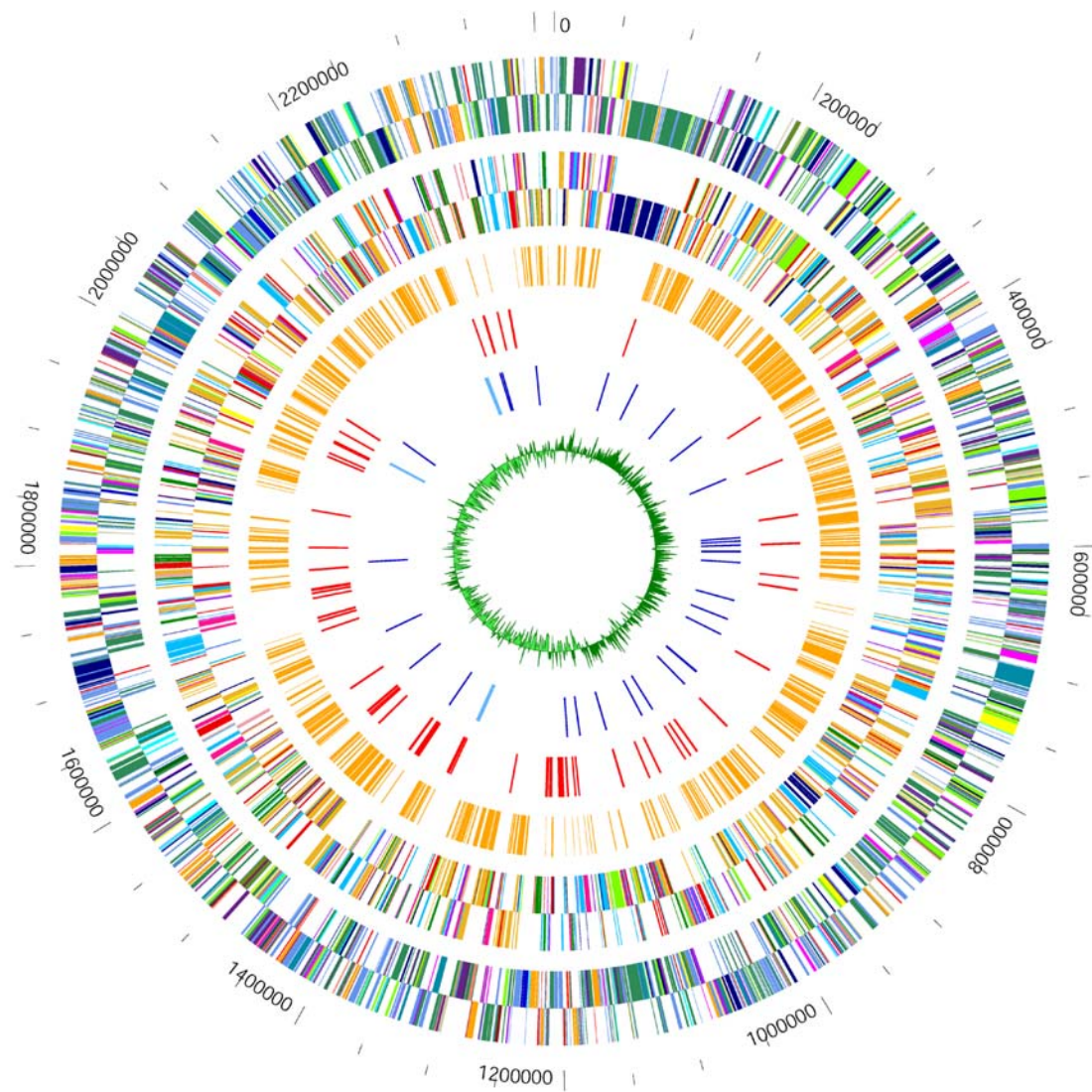


Fig. S4

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)

Fig. S4

Circular representation of the UWE25 genome showing predicted coding regions and other features. Outer two circles, predicted CDSs on the plus and minus strand, respectively, classified by biological role: amino acid metabolism (light orange), cellular processes (cyan), nucleotide metabolism (dark green), lipid/fatty acid metabolism (salmon), metabolism of vitamins/cofactors (purple), energy metabolism (light green), cell cycle/DNA processing (purple), transcription (yellow), protein synthesis and fate (chartreuse), transport (dark blue), detoxification and stress response (magenta), virulence and defense (brown), others (lemon green), cell envelope (olive), mobile elements (red), leu-rich repeat proteins (orange), (conserved) hypothetical proteins (sea green), unknown proteins (light blue), rRNA and tRNA genes (medium blue), pseudogenes (grey). Third and fourth circle, taxonomic classification of first BLAST hit for predicted CDSs on the plus and minus strand, respectively: *Chlamydiaceae* (orange), *Proteobacteria* (light blue), *Firmicutes* (red), *Cyanobacteria* (dark blue), *Actinobacteria* (yellow), *Archaea* (light green), other prokaryotes (purple), fungi (light pink), plants (dark green), mammals (pink), other animals (brown). Fifth circle, CDSs with homology to proteins of pathogenic chlamydiae (reciprocal best hit, fastA score/selfscore > 0.1). Sixth circle, (remnants of) mobile genetic elements. Seventh circle, rRNA (light blue) and tRNA (dark blue) genes. Eighth circle, GC-skew.

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)

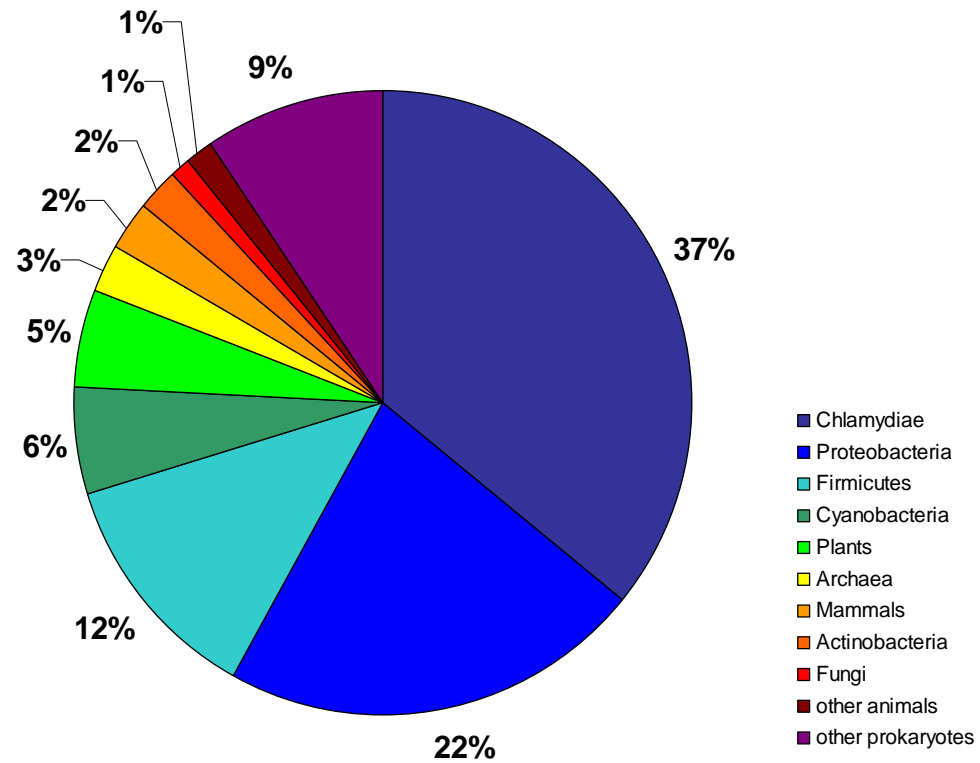


Fig. S5

Distribution of taxonomic classification of first BLAST hits of predicted UWE25 CDSs.

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)

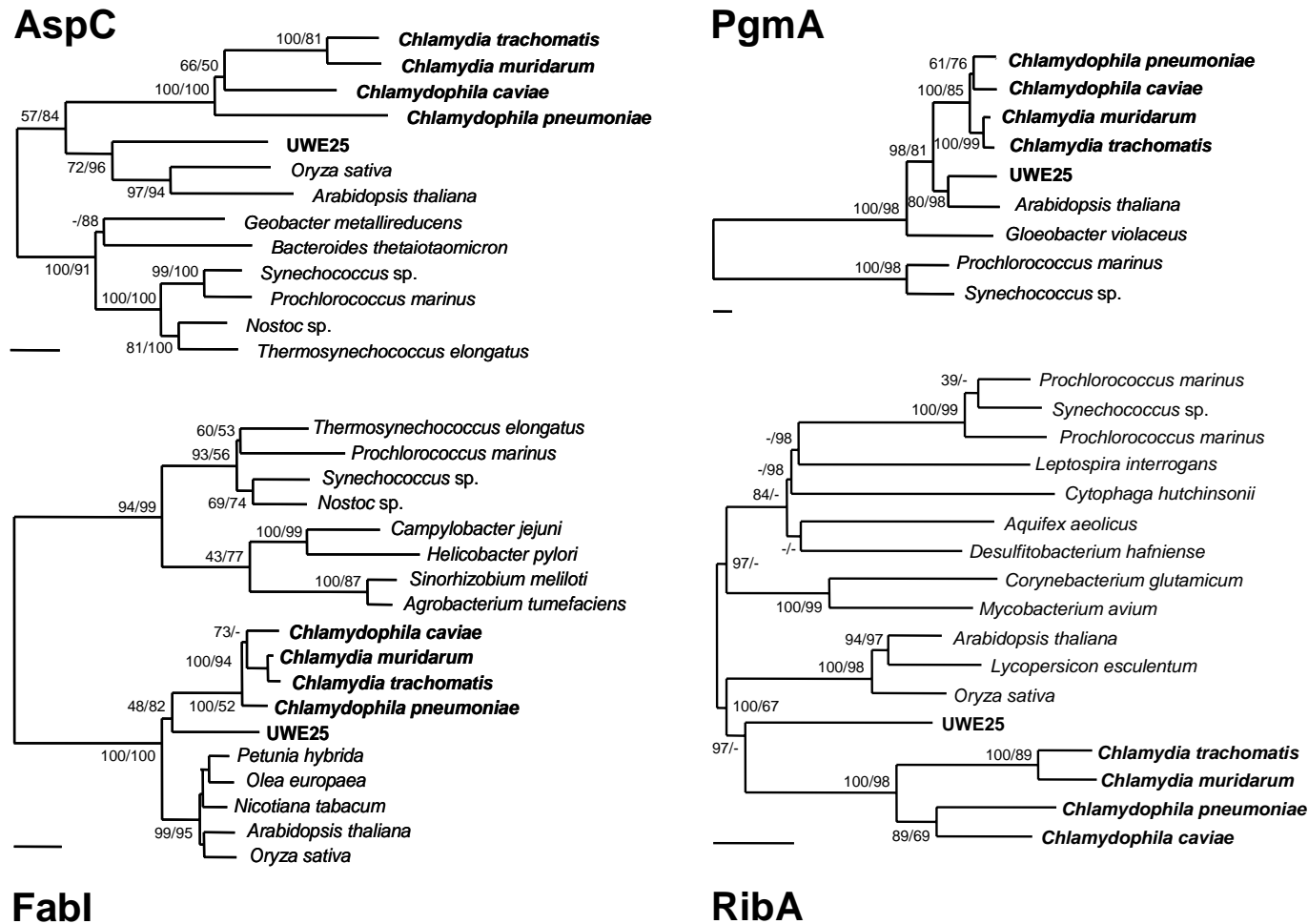


Fig. S6

Phylogeny of four selected UWE25 proteins related to plant proteins. Distance matrix trees (Fitch (20)) are shown. Maximum parsimony bootstrap values and TREE-PUZZLE support values are indicated at each node of the tree. Bar, 10% estimated evolutionary distance.

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)

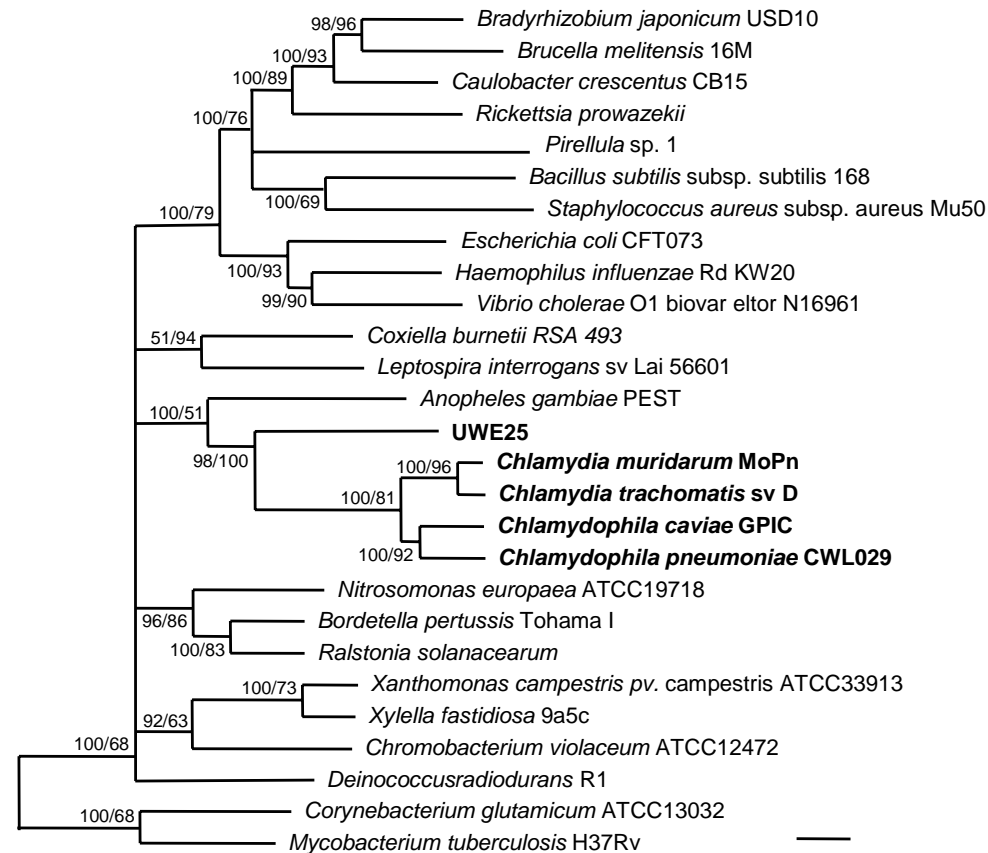


Fig. S7

Phylogeny of the chlamydial tricarboxylic acid (TCA) cycle. Unrooted consensus tree based on a concatenated amino acid sequence alignment comprising FumC (fumarate hydratase), Mdh (malate dehydrogenase), SucB (2-oxoglutarate dehydrogenase) and SucC (succinate-CoA ligase). The monophyletic grouping of pathogenic and environmental chlamydiae (to the exclusion of all other bacteria) demonstrates the origin of these TCA cycle proteins from the last common chlamydial ancestor. Polytomic branching points indicate those parts of the tree for which different treeing methods produced different topologies. Maximum parsimony bootstrap and TREE-PUZZLE support values are indicated at each node of the tree. Bar, 10% estimated evolutionary distance.

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)

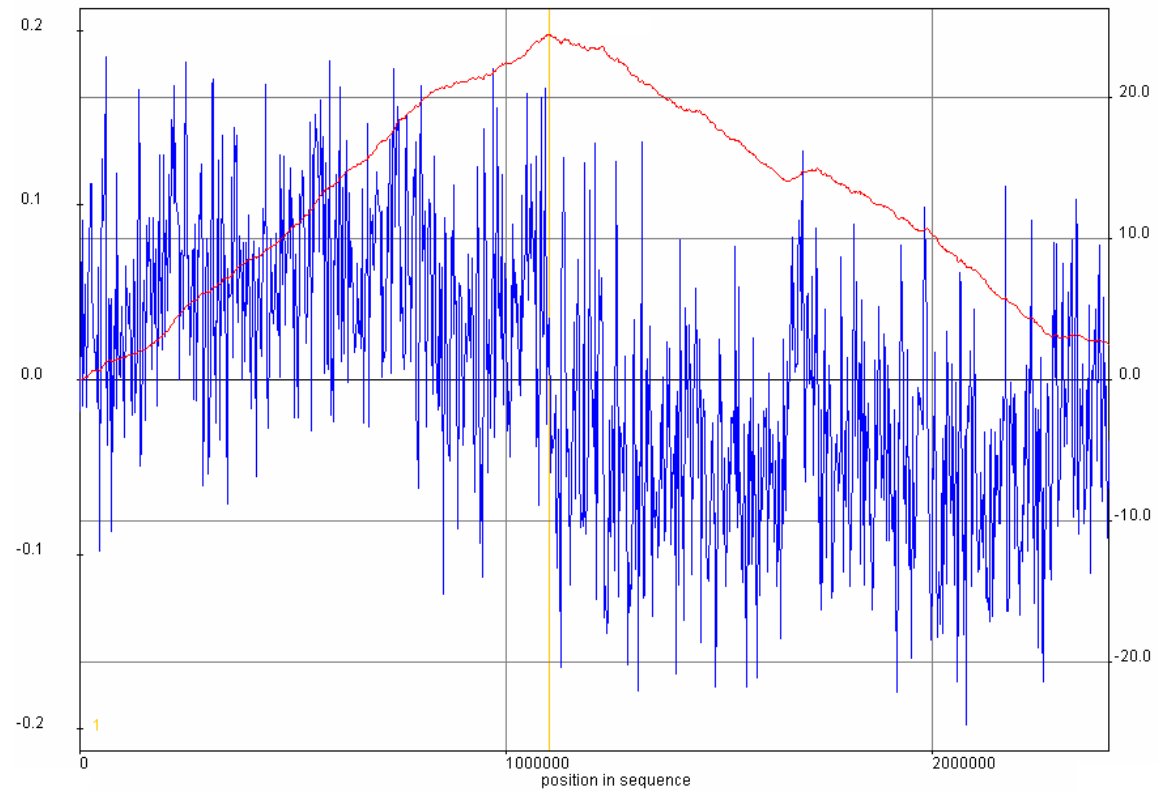
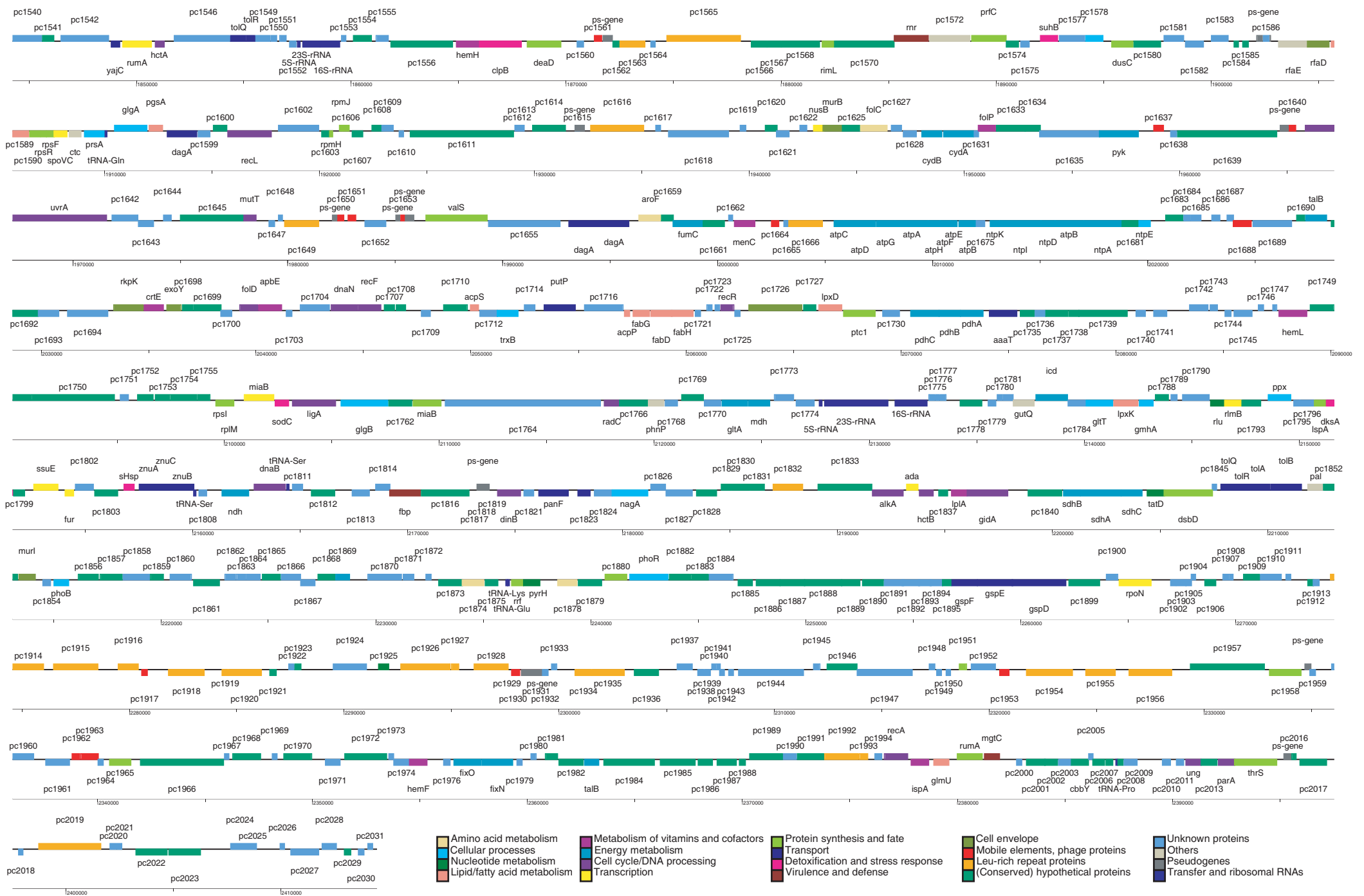


Fig. S8

GC-skew analysis of the UWE25 genome sequence. Blue graph, GC-skew; red graph, cumulative GC-skew; red line, predicted oriC.



Illuminating the evolutionary history of chlamydiae (Horn et al.)

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)

Table S1

Predicted CDSs of UWE25 without sequence homology (fastA score/selfscore cut-off 0.1) to proteins of pathogenic chlamydiae.

| Sequence ID | Gene ID | Description |
|-------------|-------------------|---|
| pc0001 | | unknown protein |
| pc0005 | | conserved hypothetical protein |
| pc0006 | | unknown protein |
| pc0008 | <i>recC</i> | similar to exodeoxyribonuclease V gamma chain |
| pc0011 | | unknown protein |
| pc0012 | | unknown protein |
| pc0013 | <i>ssuA</i> | similar to substrate-binding protein of aliphatic sulfonate ABC transporter |
| pc0017 | <i>rfa, waaF</i> | similar to ADP-heptose--lipopolysaccharide heptosyltransferase II |
| pc0018 | | hypothetical protein |
| pc0025 | | conserved hypothetical protein |
| pc0026 | | conserved hypothetical protein |
| pc0027 | <i>pmA</i> | similar to ribosomal protein L11 methyltransferase |
| pc0029 | | conserved hypothetical protein |
| pc0033 | <i>putA, poaA</i> | similar to bifunctional protein (proline dehydrogenase and delta-1-pyrroline-5-carboxylate dehydrogenase) |
| pc0036 | | unknown protein |
| pc0037 | | similar to metalloprotease |
| pc0038 | | conserved hypothetical protein |
| pc0040 | | hypothetical protein |
| pc0041 | | hypothetical protein |
| pc0042 | | hypothetical protein |
| pc0043 | | unknown protein |
| pc0044 | | hypothetical protein |
| pc0045 | | hypothetical protein |
| pc0046 | | hypothetical protein |
| pc0048 | | hypothetical protein |
| pc0049 | | unknown protein |
| pc0050 | | unknown protein |
| pc0051 | | hypothetical protein |
| pc0052 | | hypothetical protein |
| pc0053 | | hypothetical protein |
| pc0054 | | hypothetical protein |
| pc0055 | | unknown protein |
| pc0056 | | unknown protein |
| pc0057 | | hypothetical protein |
| pc0058 | | hypothetical protein |
| pc0059 | | hypothetical protein |
| pc0060 | | similarity to extracellular metalloproteinase |
| pc0061 | | hypothetical protein |
| pc0062 | | unknown protein |
| pc0063 | | hypothetical protein |
| pc0064 | | unknown protein |
| pc0065 | | hypothetical protein |
| pc0066 | | conserved hypothetical protein |
| pc0068 | | hypothetical protein |
| pc0070 | | unknown protein |
| pc0072 | <i>dprA</i> | similar to protein required for chromosomal DNA transformation |
| pc0074 | | unknown protein |
| pc0075 | | unknown protein |
| pc0076 | | unknown protein |
| pc0079 | | strongly similar to UDP-glucuronat epimerase |
| pc0080 | | unknown protein |
| pc0081 | | unknown protein |
| pc0082 | | conserved hypothetical protein |
| pc0084 | | unknown protein |
| pc0088 | | hypothetical protein |
| pc0090 | | unknown protein |
| pc0093 | | hypothetical protein |

| | | |
|--------|-------------------|---|
| pc0094 | <i>napA</i> | similar to Na(+)/H(+) antiporter |
| pc0100 | | unknown protein |
| pc0104 | | hypothetical protein |
| pc0105 | | unknown protein |
| pc0107 | | unknown protein |
| pc0111 | <i>nrdD</i> | similar to ribonucleoside triphosphate reductase |
| pc0112 | | unknown protein |
| pc0113 | | hypothetical protein |
| pc0114 | <i>batE</i> | similar to batE protein |
| pc0116 | | unknown protein |
| pc0117 | | unknown protein |
| pc0118 | | unknown protein |
| pc0119 | <i>batA</i> | similar to batA protein |
| pc0120 | <i>cicA</i> | similar to protein involved in cell wall biosynthesis, morphogenesis and cell division |
| pc0121 | | conserved hypothetical protein |
| pc0122 | | unknown protein |
| pc0123 | <i>rfbA, rmlA</i> | strongly similar to glucose-1-phosphate thymidyltransferase (dTDP-glucose synthase), rfbA |
| pc0124 | <i>rfbC, rmlC</i> | strongly similar to dTDP-4-dehydrothamnose 3,5-epimerase, rfbC |
| pc0125 | <i>rfbD, rmlD</i> | similar to dTDP-4-keto-L-rhamnose reductase, (TDP-rhamnose synthetase), rfbD |
| pc0126 | <i>rfbB, rmlB</i> | strongly similar to dTDP-glucose 4,6-dehydratase, rfbB |
| pc0127 | <i>emrB</i> | similar to multidrug resistance membrane translocase protein, emrB |
| pc0128 | <i>emrA</i> | strongly similar to multidrug resistance protein, emrA |
| pc0129 | | unknown protein |
| pc0130 | | unknown protein |
| pc0133 | | similar to acylase and diesterase |
| pc0134 | | conserved hypothetical protein |
| pc0138 | | conserved hypothetical protein |
| pc0142 | | hypothetical protein |
| pc0145 | | conserved hypothetical protein |
| pc0147 | | unknown protein |
| pc0148 | <i>asphH</i> | similar to aspartyl/asparaginyl beta-hydroxylase (= peptide-aspartate beta-dioxygenase) |
| pc0149 | <i>arsAB</i> | similar to arsenical pump membrane protein |
| pc0153 | | unknown protein |
| pc0154 | | conserved hypothetical protein |
| pc0157 | | unknown protein |
| pc0158 | | unknown protein |
| pc0159 | | hypothetical protein |
| pc0164 | | unknown protein |
| pc0167 | | unknown protein |
| pc0171 | | similar to O-linked N-acetylglucosamine transferase |
| pc0172 | | similar to O-linked N-acetylglucosamine transferase |
| pc0173 | | similar to O-linked N-acetylglucosamine transferase |
| pc0176 | <i>dinG, yoaA</i> | similar to ATP-dependent DNA helicase dinG |
| pc0178 | | hypothetical protein |
| pc0179 | | hypothetical protein |
| pc0180 | | hypothetical protein |
| pc0181 | <i>wzt</i> | strongly similar to ABC transporter ATP-binding protein wzt |
| pc0182 | <i>wzm</i> | strongly similar to ABC transporter protein wzm |
| pc0183 | | unknown protein |
| pc0185 | | unknown protein |
| pc0186 | <i>wza</i> | similar to polysaccharide export protein wza |
| pc0187 | <i>wzc</i> | similar to Tyrosine-protein kinase |
| pc0188 | | hypothetical protein |
| pc0194 | | unknown protein |
| pc0195 | <i>npr</i> | similar to metalloendopeptidase |
| pc0196 | <i>glpQ, ugpQ</i> | similar to glycerophosphoryl diester phosphodiesterase |
| pc0197 | | unknown protein |
| pc0198 | <i>ppaB</i> | similar to F-box protein |
| pc0199 | <i>ppaA</i> | similar to F-box protein |
| pc0210 | | conserved hypothetical protein |
| pc0211 | | conserved hypothetical protein |
| pc0212 | | unknown protein |
| pc0213 | | unknown protein |
| pc0215 | | unknown protein |

| | | |
|--------|-------------------|---|
| pc0216 | | unknown protein |
| pc0225 | | conserved hypothetical protein |
| pc0228 | | hypothetical protein |
| pc0230 | <i>yodO, kamA</i> | similar to L-lysine 2,3-aminomutase |
| pc0237 | | conserved hypothetical protein |
| pc0238 | <i>hisH</i> | similar to glutamine amidotransferase |
| pc0243 | | hypothetical protein |
| pc0252 | | unknown protein |
| pc0254 | | conserved hypothetical protein |
| pc0255 | | hypothetical protein |
| pc0262 | | hypothetical protein |
| pc0264 | | hypothetical protein |
| pc0265 | | conserved hypothetical protein |
| pc0275 | | unknown protein |
| pc0277 | | unknown protein |
| pc0280 | | unknown protein |
| pc0281 | <i>gcvT</i> | strongly similar to glycine cleavage system T protein |
| pc0283 | <i>gcvP1</i> | strongly similar to glycine dehydrogenase P protein subunit 1 |
| pc0284 | <i>gcvP2</i> | strongly similar to glycine dehydrogenase (decarboxylating) P protein subunit 2 |
| pc0286 | <i>alkB</i> | strongly similar to alkylated DNA repair protein |
| pc0288 | <i>czcA</i> | strongly similar to cation efflux system membrane protein A |
| pc0289 | <i>czcB</i> | similar to cation efflux system membrane protein B |
| pc0290 | <i>czcC</i> | similar to cation efflux system membrane protein C |
| pc0291 | | unknown protein |
| pc0292 | | unknown protein |
| pc0293 | | unknown protein |
| pc0294 | | unknown protein |
| pc0296 | | unknown protein |
| pc0297 | | unknown protein |
| pc0302 | | unknown protein |
| pc0303 | | unknown protein |
| pc0304 | | unknown protein |
| pc0305 | <i>comF</i> | similar to competence-related protein comF |
| pc0306 | <i>corA</i> | similar to divalent cation transport protein |
| pc0308 | | hypothetical protein |
| pc0316 | | hypothetical protein |
| pc0323 | | unknown protein |
| pc0324 | | hypothetical protein |
| pc0330 | | unknown protein |
| pc0331 | | hypothetical protein |
| pc0332 | | hypothetical protein |
| pc0333 | | conserved hypothetical protein |
| pc0334 | <i>ppaA</i> | hypothetical protein |
| pc0335 | <i>ppaA</i> | hypothetical protein |
| pc0336 | <i>ppa</i> | hypothetical protein |
| pc0339 | | hypothetical protein |
| pc0342 | | unknown protein |
| pc0343 | | unknown protein |
| pc0346 | <i>asnS</i> | strongly similar to asparagine-tRNA ligase |
| pc0348 | | conserved hypothetical protein |
| pc0349 | | conserved hypothetical protein |
| pc0352 | | hypothetical protein |
| pc0356 | <i>dtl</i> | strongly similar to D-tyrosyl-tRNA(Tyr) deacylase |
| pc0358 | | unknown protein |
| pc0360 | | unknown protein |
| pc0363 | | unknown protein |
| pc0364 | | unknown protein |
| pc0372 | | unknown protein |
| pc0377 | | unknown protein |
| pc0380 | | unknown protein |
| pc0381 | <i>kefC</i> | similar to glutathione-regulated potassium-efflux system protein |
| pc0386 | | unknown protein |
| pc0388 | | unknown protein |
| pc0398 | <i>ddlA</i> | similar to D-alanine-D-alanine ligase |

| | | |
|--------|-------------------|---|
| pc0399 | <i>zip</i> | similar to eucaryotic myosin heavy chain |
| pc0405 | | hypothetical protein |
| pc0407 | | unknown protein |
| pc0408 | | unknown protein |
| pc0409 | | unknown protein |
| pc0410 | | unknown protein |
| pc0411 | | unknown protein |
| pc0438 | | unknown protein |
| pc0439 | | unknown protein |
| pc0440 | | unknown protein |
| pc0445 | | unknown protein |
| pc0448 | | unknown protein |
| pc0449 | <i>folE</i> | strongly similar to GTP cyclohydrolase I |
| pc0452 | | hypothetical protein |
| pc0456 | | unknown protein |
| pc0457 | | unknown protein |
| pc0460 | <i>glpG</i> | similar to glpG protein |
| pc0461 | <i>rfaE</i> | strongly similar to ADP-heptose synthase |
| pc0463 | | unknown protein |
| pc0464 | | unknown protein |
| pc0465 | <i>dgt</i> | similar to deoxyguanosinetriphosphate triphosphohydrolase (dGTPase) |
| pc0471 | <i>bacA</i> | similar to bacitracin resistance protein (probable undecaprenol kinase) |
| pc0472 | | hypothetical protein |
| pc0474 | <i>hprK, ptsK</i> | strongly similar to HPr(Ser) kinase/phosphatase |
| pc0480 | | conserved hypothetical protein |
| pc0481 | | hypothetical protein |
| pc0485 | <i>ntt_4</i> | similar to ATP/ADP translocase (AATP1) precursor (Arabidopsis thaliana) |
| pc0486 | | hypothetical protein |
| pc0489 | | conserved hypothetical protein |
| pc0490 | | conserved hypothetical protein |
| pc0492 | | unknown protein |
| pc0493 | | unknown protein |
| pc0495 | | conserved hypothetical protein |
| pc0501 | | conserved hypothetical protein |
| pc0505 | | conserved hypothetical protein |
| pc0506 | | unknown protein |
| pc0510 | | unknown protein |
| pc0511 | | unknown protein |
| pc0512 | | similar to antibiotic resistance protein |
| pc0513 | | unknown protein |
| pc0515 | | unknown protein |
| pc0516 | | hypothetical protein |
| pc0517 | | conserved hypothetical protein |
| pc0519 | <i>vatB, satG</i> | strongly similar to streptogramin A acetyltransferase |
| pc0520 | | conserved hypothetical protein |
| pc0521 | | unknown protein |
| pc0522 | | unknown protein |
| pc0523 | | unknown protein |
| pc0524 | | unknown protein |
| pc0525 | | unknown protein |
| pc0528 | | conserved hypothetical protein |
| pc0529 | | unknown protein |
| pc0530 | | unknown protein |
| pc0531 | <i>cpt</i> | similar to chloramphenicol 3-O phosphotransferase |
| pc0532 | | unknown protein |
| pc0533 | | unknown protein |
| pc0535 | | unknown protein |
| pc0536 | | unknown protein |
| pc0537 | | hypothetical protein |
| pc0538 | <i>rtxA</i> | similar to RTX-toxin, partial length |
| pc0539 | | unknown protein |
| pc0542 | | conserved hypothetical protein |
| pc0543 | <i>mocA</i> | strongly similar to oxidoreductase MocA family |
| pc0544 | <i>nasA</i> | similarity to nasA protein |

| | | |
|--------|--------------------|---|
| pc0545 | | unknown protein |
| pc0547 | | unknown protein |
| pc0549 | | conserved hypothetical protein |
| pc0550 | <i>acrA, mexA</i> | similar to multidrug-efflux transport protein <i>acrA</i> |
| pc0551 | <i>acrB, mexB</i> | strongly similar to multidrug-efflux transport protein, <i>acrB</i> |
| pc0552 | <i>oprK</i> | similar to outer membrane protein, componenet of multidrug efflux systems |
| pc0553 | | hypothetical protein |
| pc0554 | | conserved hypothetical protein |
| pc0555 | <i>rfbB, rmlB</i> | similar to dTDP-glucose 4,6-dehydratase, <i>rfbB</i> |
| pc0556 | | unknown protein |
| pc0557 | | unknown protein |
| pc0558 | | unknown protein |
| pc0560 | <i>nuoB, nuo2</i> | strongly similar to NADH-ubiquinone oxidoreductase chain B |
| pc0561 | <i>nuoC, nuoCD</i> | similar to NADH-ubiquinone oxidoreductase chain C/D |
| pc0562 | <i>nuoD, nuoCD</i> | similar to NADH-ubiquinone oxidoreductase chain C/D |
| pc0563 | <i>nuoE, nuo5</i> | similar to NADH-ubiquinone oxidoreductase chain E |
| pc0564 | <i>nuoF, nuo6</i> | strongly similar to NADH-ubiquinone oxidoreductase chain F |
| pc0565 | <i>nuoG, nuo7</i> | similar to NADH-ubiquinone oxidoreductase chain G |
| pc0566 | <i>nuoH, nuo8</i> | strongly similar to NADH-ubiquinone oxidoreductase chain H |
| pc0567 | <i>nuoI, nuo9</i> | strongly similar to NADH-ubiquinone oxidoreductase chain I |
| pc0570 | <i>nuoL, nuo12</i> | similar to NADH-ubiquinone oxidoreductase chain L |
| pc0571 | <i>nuoM, nuo13</i> | similar to NADH-ubiquinone oxidoreductase chain M |
| pc0572 | <i>nuoN, nuo14</i> | similar to NADH-ubiquinone oxidoreductase chain N |
| pc0573 | | conserved hypothetical protein |
| pc0574 | | unknown protein |
| pc0576 | | unknown protein |
| pc0577 | | unknown protein |
| pc0578 | | unknown protein |
| pc0579 | | unknown protein |
| pc0580 | | unknown protein |
| pc0584 | | conserved hypothetical protein |
| pc0585 | | hypothetical protein |
| pc0587 | | hypothetical protein |
| pc0590 | | hypothetical protein |
| pc0591 | <i>dnaK, gnpF</i> | similar to heat shock protein 70 (chaperone protein <i>dnaK</i>) |
| pc0593 | | unknown protein |
| pc0596 | | unknown protein |
| pc0608 | | unknown protein |
| pc0611 | | unknown protein |
| pc0612 | | unknown protein |
| pc0613 | | unknown protein |
| pc0615 | | unknown protein |
| pc0618 | | conserved hypothetical protein |
| pc0623 | <i>nfi</i> | strongly similar to endonuclease V (deoxyinosine 3'endonuclease) |
| pc0626 | | conserved hypothetical protein |
| pc0628 | | unknown protein |
| pc0629 | | unknown protein |
| pc0630 | | similar to Thermostable carboxypeptidase 1 |
| pc0631 | <i>alr</i> | similar to alanine racemase |
| pc0632 | | conserved hypothetical protein |
| pc0633 | <i>acrD</i> | similar to acriflavin resistance protein D |
| pc0634 | | hypothetical protein |
| pc0635 | <i>toiC</i> | similar to outer membrane protein ToIC |
| pc0641 | | hypothetical protein |
| pc0642 | <i>qps</i> | similar to beta-1,4-galactosyltransferase |
| pc0645 | | conserved hypothetical protein |
| pc0647 | | unknown protein |
| pc0650 | | unknown protein |
| pc0659 | | unknown protein |
| pc0660 | | conserved hypothetical protein |
| pc0661 | | unknown protein |
| pc0664 | | unknown protein |
| pc0667 | | unknown protein |
| pc0672 | <i>phrB</i> | similar to photolyase |

| | | |
|--------|-------------------|--|
| pc0673 | | conserved hypothetical protein |
| pc0674 | | hypothetical protein |
| pc0676 | | conserved hypothetical protein |
| pc0677 | <i>recN</i> | similar to DNA repair protein RecN |
| pc0678 | | unknown protein |
| pc0681 | | conserved hypothetical protein |
| pc0691 | | conserved hypothetical protein |
| pc0692 | | unknown protein |
| pc0694 | | unknown protein |
| pc0695 | | conserved hypothetical protein |
| pc0698 | | unknown protein |
| pc0699 | | unknown protein |
| pc0707 | <i>waaC, rfaC</i> | similar to heptosyltransferase I |
| pc0709 | | conserved hypothetical protein |
| pc0710 | | conserved hypothetical protein |
| pc0711 | | conserved hypothetical protein |
| pc0712 | | unknown protein |
| pc0713 | | conserved hypothetical protein |
| pc0714 | | conserved hypothetical protein |
| pc0719 | | similar to serine protease |
| pc0723 | | conserved hypothetical protein |
| pc0726 | | unknown protein |
| pc0727 | | unknown protein |
| pc0729 | | conserved hypothetical protein |
| pc0730 | | unknown protein |
| pc0731 | | unknown protein |
| pc0732 | | unknown protein |
| pc0733 | | unknown protein |
| pc0734 | | hypothetical protein |
| pc0735 | | unknown protein |
| pc0736 | | conserved hypothetical protein |
| pc0737 | | conserved hypothetical protein |
| pc0738 | | unknown protein |
| pc0739 | <i>rhs</i> | similar to rhs core protein with extension |
| pc0741 | | conserved hypothetical protein |
| pc0743 | | conserved hypothetical protein |
| pc0744 | | unknown protein |
| pc0748 | | unknown protein |
| pc0751 | | unknown protein |
| pc0753 | | conserved hypothetical protein |
| pc0754 | | hypothetical protein |
| pc0755 | | unknown protein |
| pc0756 | | unknown protein |
| pc0757 | | unknown protein |
| pc0763 | | unknown protein |
| pc0766 | | unknown protein |
| pc0773 | | unknown protein |
| pc0774 | | unknown protein |
| pc0777 | | unknown protein |
| pc0778 | <i>exoD</i> | similar to exopolysaccharide synthesis protein |
| pc0779 | | hypothetical protein |
| pc0782 | | unknown protein |
| pc0784 | | unknown protein |
| pc0785 | | unknown protein |
| pc0788 | | unknown protein |
| pc0789 | <i>adh3</i> | similar to alcohol dehydrogenase class III |
| pc0790 | | hypothetical protein |
| pc0791 | | hypothetical protein |
| pc0792 | | hypothetical protein |
| pc0796 | | similar to tonoplast intrinsic protein (Aquaporin) |
| pc0797 | | hypothetical protein |
| pc0799 | | unknown protein |
| pc0800 | | hypothetical protein |
| pc0804 | | unknown protein |

| | | |
|--------|------------------|--|
| pc0805 | | unknown protein |
| pc0806 | | unknown protein |
| pc0810 | <i>tdh</i> | strongly similar to threonine 3-dehydrogenase |
| pc0812 | | unknown protein |
| pc0813 | | unknown protein |
| pc0814 | | unknown protein |
| pc0820 | | hypothetical protein |
| pc0822 | | similarity to bumetanide-sensitive Na-K-Cl cotransporter |
| pc0826 | | conserved hypothetical protein |
| pc0828 | | unknown protein |
| pc0829 | | unknown protein |
| pc0830 | | unknown protein |
| pc0831 | | unknown protein |
| pc0833 | | unknown protein |
| pc0834 | | unknown protein |
| pc0835 | | hypothetical protein |
| pc0836 | | unknown protein |
| pc0837 | | unknown protein |
| pc0838 | | unknown protein |
| pc0839 | | unknown protein |
| pc0840 | | unknown protein |
| pc0842 | | unknown protein |
| pc0843 | | hypothetical protein |
| pc0846 | | hypothetical protein |
| pc0847 | | unknown protein |
| pc0848 | | hypothetical protein |
| pc0849 | | unknown protein |
| pc0850 | | hypothetical protein |
| pc0851 | | conserved hypothetical protein |
| pc0852 | | conserved hypothetical protein |
| pc0853 | | conserved hypothetical protein |
| pc0855 | | conserved hypothetical protein |
| pc0856 | <i>pcm</i> | strongly similar to L-isoaspartyl protein carboxyl methyltransferase |
| pc0857 | | conserved hypothetical protein |
| pc0858 | | unknown protein |
| pc0861 | | hypothetical protein |
| pc0862 | | unknown protein |
| pc0863 | | hypothetical protein |
| pc0865 | <i>acnB</i> | strongly similar to aconitate hydratase |
| pc0868 | | conserved hypothetical protein |
| pc0869 | | unknown protein |
| pc0870 | | unknown protein |
| pc0872 | <i>potB</i> | strongly similar to spermidine/putrescine transport system permease, component of ATP-transporter system |
| pc0873 | <i>potC</i> | strongly similar to spermidine/putrescine transport system permease, component of ATP-transporter system |
| pc0874 | <i>potD</i> | similar to spermidine/putrescine-binding protein precursor, component of ABC transporter system |
| pc0876 | | unknown protein |
| pc0878 | <i>sdaB</i> | strongly similar to L-serine ammonia-lyase |
| pc0885 | | unknown protein |
| pc0886 | | unknown protein |
| pc0888 | | conserved hypothetical protein |
| pc0892 | <i>mutM, fpg</i> | similar to formamidopyrimidine-DNA glycosidase |
| pc0893 | <i>rbp</i> | strongly similar to nucleic acid-binding protein |
| pc0894 | | unknown protein |
| pc0895 | | similar to outer membrane protein |
| pc0896 | | hypothetical protein |
| pc0897 | | unknown protein |
| pc0899 | | unknown protein |
| pc0900 | | conserved hypothetical protein |
| pc0901 | | hypothetical protein |
| pc0903 | | conserved hypothetical protein |
| pc0904 | <i>rgpB</i> | strongly similar to rhamnosyltransferase |
| pc0905 | | unknown protein |
| pc0906 | | conserved hypothetical protein |
| pc0907 | | hypothetical protein |

| | | |
|--------|-------------|--|
| pc0908 | | hypothetical protein |
| pc0909 | | conserved hypothetical protein |
| pc0910 | | unknown protein |
| pc0911 | | unknown protein |
| pc0912 | | conserved hypothetical protein |
| pc0913 | | hypothetical protein |
| pc0914 | <i>czcD</i> | similar to cation efflux system protein |
| pc0915 | <i>kefC</i> | similar to glutathione-regulated potassium-efflux system protein |
| pc0917 | | unknown protein |
| pc0918 | | hypothetical protein |
| pc0919 | | conserved hypothetical protein |
| pc0921 | | hypothetical protein |
| pc0922 | | unknown protein |
| pc0923 | | hypothetical protein |
| pc0924 | | unknown protein |
| pc0926 | | unknown protein |
| pc0928 | <i>trpD</i> | similar to anthranilate synthase component II |
| pc0929 | | conserved hypothetical protein |
| pc0930 | <i>cynT</i> | similar to carbonic anhydrase |
| pc0931 | | hypothetical protein |
| pc0932 | | unknown protein |
| pc0934 | | conserved hypothetical protein |
| pc0935 | <i>glk</i> | similar to glucokinase |
| pc0938 | | hypothetical protein |
| pc0940 | | unknown protein |
| pc0941 | | unknown protein |
| pc0942 | | conserved hypothetical protein |
| pc0943 | <i>bic</i> | similar to outer membrane lipoprotein |
| pc0945 | | unknown protein |
| pc0946 | | unknown protein |
| pc0948 | | conserved hypothetical protein |
| pc0949 | | conserved hypothetical protein |
| pc0950 | <i>lig</i> | similar to DNA ligase |
| pc0951 | | conserved hypothetical protein |
| pc0953 | <i>yciF</i> | strongly similar to yciF protein |
| pc0955 | <i>ftn</i> | strongly similar to ferritin |
| pc0956 | | conserved hypothetical protein |
| pc0960 | | unknown protein |
| pc0962 | | unknown protein |
| pc0963 | | conserved hypothetical protein |
| pc0964 | <i>msrA</i> | strongly similar to protein-methionine-S-oxide reductase |
| pc0967 | | unknown protein |
| pc0968 | | unknown protein |
| pc0969 | | conserved hypothetical protein |
| pc0970 | | conserved hypothetical protein |
| pc0973 | | conserved hypothetical protein |
| pc0974 | | conserved hypothetical protein |
| pc0975 | <i>umuC</i> | similar to SOS mutagenesis and repair protein UmuC |
| pc0976 | <i>umuD</i> | strongly similar to SOS mutagenesis and repair protein UmuD |
| pc0977 | | unknown protein |
| pc0979 | | conserved hypothetical protein |
| pc0981 | | hypothetical protein |
| pc0982 | | unknown protein |
| pc0983 | | unknown protein |
| pc0986 | | conserved hypothetical protein |
| pc0990 | | conserved hypothetical protein |
| pc0991 | | unknown protein |
| pc0992 | | conserved hypothetical protein |
| pc0993 | | conserved hypothetical protein |
| pc0996 | | conserved hypothetical protein |
| pc0997 | | unknown protein |
| pc0998 | | unknown protein |
| pc0999 | | conserved hypothetical protein |
| pc1000 | | conserved hypothetical protein |

| | | |
|--------|--------------------|--|
| pc1002 | | unknown protein |
| pc1003 | | hypothetical protein |
| pc1005 | | conserved hypothetical protein |
| pc1006 | | unknown protein |
| pc1007 | | unknown protein |
| pc1008 | | unknown protein |
| pc1009 | | unknown protein |
| pc1010 | | conserved hypothetical protein |
| pc1012 | | unknown protein |
| pc1013 | | unknown protein |
| pc1016 | | unknown protein |
| pc1017 | | conserved hypothetical protein |
| pc1019 | | unknown protein |
| pc1020 | | unknown protein |
| pc1021 | <i>doc</i> | similar to death on curing protein |
| pc1022 | | conserved hypothetical protein |
| pc1023 | | unknown protein |
| pc1024 | | unknown protein |
| pc1025 | | conserved hypothetical protein |
| pc1029 | | conserved hypothetical protein |
| pc1030 | | unknown protein |
| pc1031 | | conserved hypothetical protein |
| pc1032 | | conserved hypothetical protein |
| pc1033 | | hypothetical protein |
| pc1036 | | unknown protein |
| pc1037 | | unknown protein |
| pc1038 | | unknown protein |
| pc1039 | | unknown protein |
| pc1040 | | unknown protein |
| pc1043 | | conserved hypothetical protein |
| pc1044 | | conserved hypothetical protein |
| pc1045 | | unknown protein |
| pc1051 | | unknown protein |
| pc1052 | | unknown protein |
| pc1056 | | conserved hypothetical protein |
| pc1058 | <i>cynT</i> | similar to carbonate dehydratase, cynT |
| pc1061 | | conserved hypothetical protein |
| pc1063 | <i>menA</i> | similar to 1,4-dihydroxy-2-naphthoate octaprenyltransferase, menA |
| pc1064 | <i>menB</i> | strongly similar to naphthoate synthase, menB |
| pc1065 | | conserved hypothetical protein |
| pc1066 | | conserved hypothetical protein |
| pc1067 | | conserved hypothetical protein |
| pc1068 | <i>menD, menCF</i> | similar to menaquinone biosynthesis protein, menD |
| pc1069 | <i>menF</i> | similar to menaquinone-specific isochorismate synthase, menF |
| pc1070 | | hypothetical protein |
| pc1071 | | hypothetical protein |
| pc1077 | | unknown protein |
| pc1079 | <i>wzb</i> | similar to low molecular weight protein-tyrosine-phosphatase |
| pc1085 | | conserved hypothetical protein |
| pc1086 | | hypothetical protein |
| pc1087 | | hypothetical protein |
| pc1091 | | unknown protein |
| pc1092 | | unknown protein |
| pc1093 | <i>ysgA</i> | similar to carboxymethylenebutenolidase |
| pc1094 | | hypothetical protein |
| pc1097 | | unknown protein |
| pc1100 | | hypothetical protein |
| pc1109 | | unknown protein |
| pc1111 | | unknown protein |
| pc1112 | | unknown protein |
| pc1114 | | unknown protein |
| pc1117 | | hypothetical protein |
| pc1119 | | conserved hypothetical protein |
| pc1121 | <i>yedO</i> | similar to 1-aminocyclopropane-1-carboxylate deaminase (ACC deaminase) |

| | | |
|--------|-------------------|--|
| pc1122 | | unknown protein |
| pc1123 | | unknown protein |
| pc1124 | | similar to calcium-dependent protein kinase 9 |
| pc1125 | <i>hlyE</i> | strongly similar to hemolysin III |
| pc1126 | | unknown protein |
| pc1127 | | unknown protein |
| pc1128 | | unknown protein |
| pc1129 | | unknown protein |
| pc1130 | | unknown protein |
| pc1131 | | hypothetical protein |
| pc1132 | <i>proC</i> | similar to pyrroline-5-carboxylate reductase, proC |
| pc1133 | <i>ppiB</i> | strongly similar to peptidylprolyl isomerase II (cyclophilin A) |
| pc1134 | | unknown protein |
| pc1135 | | unknown protein |
| pc1136 | | unknown protein |
| pc1137 | | unknown protein |
| pc1138 | | unknown protein |
| pc1139 | | unknown protein |
| pc1140 | | conserved hypothetical protein |
| pc1141 | | conserved hypothetical protein |
| pc1142 | | conserved hypothetical protein |
| pc1143 | | conserved hypothetical protein |
| pc1144 | | unknown protein |
| pc1145 | | conserved hypothetical protein |
| pc1146 | | conserved hypothetical protein |
| pc1147 | | conserved hypothetical protein |
| pc1148 | | conserved hypothetical protein |
| pc1149 | | conserved hypothetical protein |
| pc1150 | | unknown protein |
| pc1151 | | unknown protein |
| pc1153 | | unknown protein |
| pc1154 | | unknown protein |
| pc1155 | | unknown protein |
| pc1157 | <i>folB, mutT</i> | similar to dGTP pyrophosphohydrolase/dihydroneopterin aldolase (mutT/folB, fusion protein) |
| pc1162 | | conserved hypothetical protein |
| pc1165 | | hypothetical protein |
| pc1167 | | hypothetical protein |
| pc1174 | | unknown protein |
| pc1181 | | hypothetical protein |
| pc1182 | <i>topI</i> | similar to DNA topoisomerase I |
| pc1183 | <i>wbdA</i> | similar to mannosyltransferase |
| pc1184 | | hypothetical protein |
| pc1185 | | unknown protein |
| pc1186 | | unknown protein |
| pc1187 | | unknown protein |
| pc1188 | <i>tspO</i> | similar to outer membrane protein tspO |
| pc1189 | <i>cyoA</i> | strongly similar to cytochrome o ubiquinol oxidase chain II cyoA |
| pc1190 | <i>cyoB</i> | strongly similar to cytochrome o ubiquinol oxidase chain I cyoB |
| pc1191 | <i>cyoC</i> | strongly similar to cytochrome o ubiquinol oxidase chain III cyoC |
| pc1193 | <i>cyoE</i> | strongly similar to heme O synthase (=protoheme IX farnesyltransferase) cyoE |
| pc1194 | | unknown protein |
| pc1195 | | conserved hypothetical protein |
| pc1197 | | hypothetical protein |
| pc1199 | | conserved hypothetical protein |
| pc1201 | | conserved hypothetical protein |
| pc1204 | | unknown protein |
| pc1205 | | unknown protein |
| pc1206 | | unknown protein |
| pc1207 | | unknown protein |
| pc1208 | | unknown protein |
| pc1209 | | conserved hypothetical protein |
| pc1210 | | unknown protein |
| pc1211 | <i>cnp, nprC</i> | similar to metalloprotease |
| pc1212 | | unknown protein |

| | | |
|--------|-------------------|--|
| pc1216 | | hypothetical protein |
| pc1218 | | hypothetical protein |
| pc1219 | <i>oprM</i> | similar to outer membrane protein of AcrAB(MexAB)-OprM multidrug efflux pump |
| pc1220 | <i>DHCR7</i> | similar to 7-dehydrocholesterol reductase |
| pc1222 | | unknown protein |
| pc1223 | | unknown protein |
| pc1224 | | unknown protein |
| pc1227 | | unknown protein |
| pc1230 | | unknown protein |
| pc1231 | | unknown protein |
| pc1233 | | unknown protein |
| pc1234 | | unknown protein |
| pc1240 | <i>glnA</i> | similar to glutamate-ammonia ligase (=glutamine synthetase) type III |
| pc1254 | | unknown protein |
| pc1255 | | similar to eucaryotic stearoyl-CoA 9-desaturase |
| pc1256 | <i>ipgD</i> | similar to virulence protein ipgD (Shigella flexneri plasmid pINV) |
| pc1258 | | unknown protein |
| pc1261 | | conserved hypothetical protein |
| pc1263 | | hypothetical protein |
| pc1264 | <i>neuB</i> | similar to sialic acid synthase |
| pc1265 | <i>deaD</i> | similar to ATP-dependent RNA helicase |
| pc1268 | | hypothetical protein |
| pc1269 | <i>lexA, dinR</i> | similar to SOS response regulator lexA |
| pc1272 | | conserved hypothetical protein |
| pc1273 | | hypothetical protein |
| pc1274 | | conserved hypothetical protein |
| pc1275 | | unknown protein |
| pc1276 | | conserved hypothetical protein |
| pc1278 | | conserved hypothetical protein |
| pc1279 | | conserved hypothetical protein |
| pc1281 | | conserved hypothetical protein |
| pc1282 | | conserved hypothetical protein |
| pc1283 | | unknown protein |
| pc1284 | | unknown protein |
| pc1286 | | unknown protein |
| pc1287 | | unknown protein |
| pc1290 | | unknown protein |
| pc1292 | | unknown protein |
| pc1293 | | conserved hypothetical protein |
| pc1294 | | unknown protein |
| pc1295 | | unknown protein |
| pc1296 | | hypothetical protein |
| pc1300 | | unknown protein |
| pc1301 | | unknown protein |
| pc1302 | <i>galE</i> | similar to UDP-glucose 4-epimerase |
| pc1303 | | unknown protein |
| pc1309 | | unknown protein |
| pc1310 | <i>dpm1</i> | similar to dolichol-phosphate mannosyltransferase |
| pc1313 | | conserved hypothetical protein |
| pc1321 | | conserved hypothetical protein |
| pc1323 | | conserved hypothetical protein |
| pc1324 | <i>proS</i> | similar to prolyl-HRNA synthetase |
| pc1325 | <i>pabC</i> | similar to 4-amino-4-deoxychorismate lyase |
| pc1326 | | unknown protein |
| pc1327 | <i>pabB</i> | similar to para-aminobenzoate synthase component I |
| pc1335 | | hypothetical protein |
| pc1336 | <i>aac</i> | similar to aminoglycoside N6'-acetyltransferase |
| pc1337 | | conserved hypothetical protein |
| pc1338 | <i>tag</i> | strongly similar to 3-methyladenine-DNA glycosylase I |
| pc1339 | | unknown protein |
| pc1340 | | unknown protein |
| pc1341 | | conserved hypothetical protein |
| pc1342 | | unknown protein |
| pc1345 | | unknown protein |

| | | |
|--------|-------------|---|
| pc1346 | | unknown protein |
| pc1347 | | conserved hypothetical protein |
| pc1350 | | unknown protein |
| pc1352 | | unknown protein |
| pc1353 | | conserved hypothetical protein (possible outer surface protein wsp) |
| pc1354 | | unknown protein |
| pc1355 | | unknown protein |
| pc1357 | | unknown protein |
| pc1359 | | hypothetical protein |
| pc1360 | | hypothetical protein |
| pc1361 | | hypothetical protein |
| pc1362 | <i>ackA</i> | strongly similar to acetate kinase |
| pc1363 | | hypothetical protein |
| pc1366 | | unknown protein |
| pc1369 | | similar to glycosyltransferase |
| pc1370 | | hypothetical protein |
| pc1373 | | unknown protein |
| pc1378 | | hypothetical protein |
| pc1379 | | hypothetical protein |
| pc1380 | | unknown protein |
| pc1381 | | unknown protein |
| pc1382 | | unknown protein |
| pc1383 | | unknown protein |
| pc1385 | | unknown protein |
| pc1387 | | unknown protein |
| pc1388 | | unknown protein |
| pc1389 | | unknown protein |
| pc1394 | | similar to UDPgalactose-glucose galactosyltransferase |
| pc1396 | | unknown protein |
| pc1399 | | unknown protein |
| pc1402 | | conserved hypothetical protein |
| pc1404 | | unknown protein |
| pc1408 | | unknown protein |
| pc1410 | | conserved hypothetical protein |
| pc1412 | | unknown protein |
| pc1414 | | unknown protein |
| pc1415 | | unknown protein |
| pc1416 | | unknown protein |
| pc1417 | | unknown protein |
| pc1418 | | unknown protein |
| pc1419 | | hypothetical protein |
| pc1423 | <i>traE</i> | similar to F pilus assembly protein traE |
| pc1424 | | unknown protein |
| pc1425 | <i>traB</i> | similar to F pilus assembly protein traB |
| pc1426 | | conserved hypothetical protein |
| pc1428 | | unknown protein |
| pc1430 | <i>traC</i> | similar to inner-membrane protein traC |
| pc1431 | <i>traF</i> | similar to F pilus assembly protein traF |
| pc1432 | <i>traW</i> | similar to F pilus assembly protein traW |
| pc1433 | <i>trbC</i> | similar to F pilus assembly protein trbC |
| pc1434 | <i>traU</i> | similar to F pilus assembly protein traU |
| pc1435 | | unknown protein |
| pc1436 | | unknown protein |
| pc1437 | <i>traN</i> | similar to conjugative transfer protein traN precursor |
| pc1438 | <i>traF</i> | similar to F pilus assembly protein traF |
| pc1439 | <i>traH</i> | similar to F pilus assembly protein traH |
| pc1440 | <i>traG</i> | similar to conjugative transfer protein traG |
| pc1441 | <i>traD</i> | similar to conjugative transfer protein traD |
| pc1445 | | unknown protein |
| pc1446 | | conserved hypothetical protein |
| pc1448 | | unknown protein |
| pc1449 | | hypothetical protein |
| pc1452 | | unknown protein |
| pc1453 | | conserved hypothetical protein |

| | | |
|--------|-------------------|---|
| pc1454 | | unknown protein |
| pc1455 | | conserved hypothetical protein |
| pc1456 | <i>doc</i> | strongly similar to doc (death on cure) protein of bacteriophage P1 |
| pc1457 | | conserved hypothetical protein |
| pc1458 | | unknown protein |
| pc1459 | | unknown protein |
| pc1460 | | unknown protein |
| pc1462 | | conserved hypothetical protein |
| pc1463 | | unknown protein |
| pc1464 | | unknown protein |
| pc1465 | | hypothetical protein |
| pc1466 | | unknown protein |
| pc1467 | | unknown protein |
| pc1469 | <i>tnpR</i> | strongly similar to resolvase |
| pc1470 | <i>tnpA</i> | strongly similar to transposase, partial length |
| pc1471 | <i>tnpA</i> | strongly similar to transposase, partial length |
| pc1472 | | unknown protein |
| pc1474 | | unknown protein |
| pc1476 | | unknown protein |
| pc1479 | | unknown protein |
| pc1481 | | hypothetical protein |
| pc1482 | | unknown protein |
| pc1485 | <i>dnaK, grpF</i> | similar to heat shock protein 70, dnaK |
| pc1486 | <i>ald</i> | strongly similar to alanine dehydrogenase |
| pc1487 | <i>smc</i> | similar to chromosome segregation SMC protein |
| pc1489 | | unknown protein |
| pc1490 | | similarity to aspartyl aminopeptidase (metalloprotease) |
| pc1491 | | unknown protein |
| pc1492 | | conserved hypothetical protein |
| pc1494 | | unknown protein |
| pc1495 | | hypothetical protein |
| pc1496 | <i>gdhB</i> | similar to eucaryotic NAD-specific glutamate dehydrogenase |
| pc1500 | | unknown protein |
| pc1507 | | hypothetical protein |
| pc1508 | | hypothetical protein |
| pc1509 | | hypothetical protein |
| pc1510 | | hypothetical protein |
| pc1516 | | unknown protein |
| pc1517 | | unknown protein |
| pc1518 | | unknown protein |
| pc1520 | | similar to periplasmic immunogenic protein |
| pc1522 | | unknown protein |
| pc1523 | | unknown protein |
| pc1524 | | hypothetical protein |
| pc1526 | | unknown protein |
| pc1528 | | unknown protein |
| pc1531 | | unknown protein |
| pc1532 | | unknown protein |
| pc1534 | <i>yabJ, yigF</i> | strongly similar to yabJ |
| pc1535 | <i>htpG</i> | similar to heat shock protein HtpG |
| pc1536 | | hypothetical protein |
| pc1537 | | unknown protein |
| pc1540 | | unknown protein |
| pc1542 | | unknown protein |
| pc1546 | | unknown protein |
| pc1549 | | unknown protein |
| pc1551 | | unknown protein |
| pc1553 | | unknown protein |
| pc1554 | | conserved hypothetical protein |
| pc1555 | | unknown protein |
| pc1556 | | conserved hypothetical protein |
| pc1558 | <i>clpB</i> | similar to heat shock protein ClpB |
| pc1559 | <i>deaD</i> | similar to ATP-dependent RNA helicase |
| pc1560 | | unknown protein |

| | | |
|--------|-------------------------|---|
| pc1562 | | conserved hypothetical protein |
| pc1563 | | hypothetical protein |
| pc1564 | | unknown protein |
| pc1565 | | hypothetical protein |
| pc1567 | | hypothetical protein |
| pc1568 | | hypothetical protein |
| pc1569 | <i>rimL</i> | similar to ribosomal-protein-serine acetyltransferase |
| pc1570 | | conserved hypothetical protein |
| pc1574 | | hypothetical protein |
| pc1577 | | unknown protein |
| pc1580 | | hypothetical protein |
| pc1581 | | unknown protein |
| pc1582 | | unknown protein |
| pc1583 | | unknown protein |
| pc1584 | | conserved hypothetical protein |
| pc1586 | | unknown protein |
| pc1587 | <i>hldE, rfaE</i> | similar to bifunctional protein involved in LPS core biosynthesis, hldE |
| pc1588 | <i>hldD, waaD, rfaD</i> | similar to ADP-D-beta-D-heptose epimerase, hldD |
| pc1599 | | unknown protein |
| pc1602 | | unknown protein |
| pc1609 | | unknown protein |
| pc1611 | | conserved hypothetical protein |
| pc1612 | | unknown protein |
| pc1613 | | unknown protein |
| pc1614 | | conserved hypothetical protein |
| pc1615 | | unknown protein |
| pc1616 | | conserved hypothetical protein |
| pc1617 | | unknown protein |
| pc1618 | | unknown protein |
| pc1619 | | unknown protein |
| pc1621 | | unknown protein |
| pc1622 | | unknown protein |
| pc1625 | | conserved hypothetical protein |
| pc1626 | <i>folC</i> | similar to folylpolyglutamate synthase |
| pc1628 | | unknown protein |
| pc1631 | | unknown protein |
| pc1634 | | hypothetical protein |
| pc1635 | | unknown protein |
| pc1637 | | conserved hypothetical protein |
| pc1638 | | unknown protein |
| pc1639 | | conserved hypothetical protein |
| pc1642 | | unknown protein |
| pc1643 | | unknown protein |
| pc1646 | <i>mutT</i> | similar to mutT protein |
| pc1648 | | unknown protein |
| pc1649 | | conserved hypothetical protein |
| pc1650 | | conserved hypothetical protein |
| pc1652 | | unknown protein |
| pc1653 | | conserved hypothetical protein |
| pc1655 | | unknown protein |
| pc1658 | <i>aroF</i> | strongly similar to 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (tyrosine-sensitive), aroF |
| pc1661 | | conserved hypothetical protein |
| pc1662 | | unknown protein |
| pc1663 | <i>menC</i> | similar to o-succinylbenzoate synthase II, menC |
| pc1665 | | unknown protein |
| pc1666 | | hypothetical protein |
| pc1667 | <i>atpC</i> | similar to H ⁺ -transporting two-sector ATPase (epsilon chain, atpC) |
| pc1669 | <i>atpG</i> | similar to H ⁺ -transporting two-sector ATPase (gamma chain, atpG) |
| pc1671 | <i>atpH</i> | similar to H ⁺ -transporting two-sector ATPase (delta chain, atpH) |
| pc1674 | <i>atpB</i> | similar to H ⁺ -transporting two-sector ATPase (chain a, atpB) |
| pc1685 | | unknown protein |
| pc1686 | | unknown protein |
| pc1687 | | unknown protein |
| pc1688 | | hypothetical protein |

| | | |
|--------|-------------------|--|
| pc1689 | | unknown protein |
| pc1692 | | hypothetical protein |
| pc1693 | | unknown protein |
| pc1694 | | unknown protein |
| pc1695 | <i>rkpK</i> | strongly similar to UDPglucose 6-dehydrogenase |
| pc1697 | <i>exoY</i> | strongly similar to exopolysaccharide production protein |
| pc1699 | | conserved hypothetical protein |
| pc1700 | | unknown protein |
| pc1704 | | unknown protein |
| pc1710 | | conserved hypothetical protein |
| pc1712 | | unknown protein |
| pc1714 | | unknown protein |
| pc1715 | <i>putP</i> | strongly similar to sodium/proline symporter |
| pc1716 | | unknown protein |
| pc1721 | | unknown protein |
| pc1722 | | unknown protein |
| pc1723 | | unknown protein |
| pc1725 | | unknown protein |
| pc1729 | <i>ptc1</i> | similar to phosphoprotein phosphatase |
| pc1730 | | unknown protein |
| pc1735 | | conserved hypothetical protein |
| pc1737 | | conserved hypothetical protein |
| pc1738 | | hypothetical protein |
| pc1739 | | hypothetical protein |
| pc1741 | | unknown protein |
| pc1742 | | unknown protein |
| pc1745 | | unknown protein |
| pc1746 | | unknown protein |
| pc1747 | | unknown protein |
| pc1749 | | hypothetical protein |
| pc1750 | | hypothetical protein |
| pc1751 | | unknown protein |
| pc1754 | | hypothetical protein |
| pc1759 | <i>sodC</i> | similar to Superoxide dismutase (Cu-Zn) |
| pc1764 | | unknown protein |
| pc1765 | <i>radC</i> | similar to DNA repair protein radC |
| pc1768 | | unknown protein |
| pc1771 | <i>gltA</i> | similar to citrate (si)-synthase |
| pc1773 | | unknown protein |
| pc1774 | | unknown protein |
| pc1775 | | unknown protein |
| pc1776 | | unknown protein |
| pc1777 | | unknown protein |
| pc1778 | | hypothetical protein |
| pc1781 | | unknown protein |
| pc1783 | <i>icd</i> | strongly similar to isocitrate dehydrogenase (NADP) |
| pc1784 | | unknown protein |
| pc1787 | <i>gmhA, lpcA</i> | strongly similar to phosphoheptose isomerase |
| pc1788 | | hypothetical protein |
| pc1789 | | unknown protein |
| pc1790 | | unknown protein |
| pc1794 | <i>ppx</i> | similar to exopolyphosphatase |
| pc1795 | | unknown protein |
| pc1796 | | unknown protein |
| pc1801 | <i>fur</i> | similar to ferric uptake regulator protein |
| pc1802 | | unknown protein |
| pc1803 | | conserved hypothetical protein |
| pc1804 | <i>sHsp</i> | similar to small heat shock protein |
| pc1808 | | unknown protein |
| pc1809 | <i>ndh</i> | similar to NADH2 dehydrogenase |
| pc1812 | | hypothetical protein |
| pc1813 | | unknown protein |
| pc1814 | | unknown protein |
| pc1815 | <i>fbp</i> | similar to fibronectin/fibrinogen binding protein |

| | | |
|--------|-------------------|---|
| pc1816 | | hypothetical protein |
| pc1817 | | unknown protein |
| pc1818 | | unknown protein |
| pc1819 | | unknown protein |
| pc1820 | <i>dinP, dinB</i> | strongly similar to DNA polymerase IV |
| pc1824 | | unknown protein |
| pc1825 | <i>nagA</i> | similar to beta-N-acetylglucosaminidase |
| pc1826 | | unknown protein |
| pc1827 | | unknown protein |
| pc1828 | | conserved hypothetical protein |
| pc1830 | | conserved hypothetical protein |
| pc1831 | | conserved hypothetical protein |
| pc1832 | | hypothetical protein |
| pc1833 | | conserved hypothetical protein |
| pc1834 | <i>alkA</i> | similar to DNA-3-methyladenine glycosidase II |
| pc1837 | | hypothetical protein |
| pc1840 | | hypothetical protein |
| pc1849 | <i>tolA</i> | similar to tolA protein of Tol-Pal system |
| pc1853 | <i>murI</i> | strongly similar to glutamate racemase |
| pc1858 | | unknown protein |
| pc1860 | | unknown protein |
| pc1861 | | hypothetical protein |
| pc1862 | | unknown protein |
| pc1864 | | unknown protein |
| pc1865 | | conserved hypothetical protein |
| pc1866 | | unknown protein |
| pc1867 | | unknown protein |
| pc1870 | | unknown protein |
| pc1871 | | unknown protein |
| pc1873 | | hypothetical protein |
| pc1878 | | similar to procollagen-lysine 5-dioxygenase |
| pc1879 | | hypothetical protein |
| pc1881 | <i>phoR</i> | similar to two-component sensor histidine kinase phoR |
| pc1882 | | conserved hypothetical protein |
| pc1883 | | conserved hypothetical protein |
| pc1884 | | unknown protein |
| pc1885 | | hypothetical protein |
| pc1886 | | conserved hypothetical protein |
| pc1887 | | conserved hypothetical protein |
| pc1889 | | conserved hypothetical protein |
| pc1890 | | conserved hypothetical protein |
| pc1891 | | unknown protein |
| pc1892 | | unknown protein |
| pc1893 | | unknown protein |
| pc1894 | | unknown protein |
| pc1902 | | unknown protein |
| pc1903 | | unknown protein |
| pc1904 | | unknown protein |
| pc1905 | | unknown protein |
| pc1907 | | unknown protein |
| pc1908 | | unknown protein |
| pc1909 | | hypothetical protein |
| pc1910 | | unknown protein |
| pc1911 | | unknown protein |
| pc1912 | | conserved hypothetical protein |
| pc1913 | | unknown protein |
| pc1914 | | conserved hypothetical protein |
| pc1915 | | conserved hypothetical protein |
| pc1916 | | conserved hypothetical protein |
| pc1918 | | conserved hypothetical protein |
| pc1919 | | conserved hypothetical protein |
| pc1920 | | conserved hypothetical protein |
| pc1921 | | hypothetical protein |
| pc1922 | | unknown protein |

| | | |
|--------|-------------|---|
| pc1924 | | unknown protein |
| pc1925 | | strongly similar to endonuclease G, mitochondrial precursor |
| pc1926 | | conserved hypothetical protein |
| pc1928 | | conserved hypothetical protein |
| pc1932 | | unknown protein |
| pc1934 | | conserved hypothetical protein |
| pc1935 | | conserved hypothetical protein |
| pc1936 | | conserved hypothetical protein |
| pc1937 | | unknown protein |
| pc1938 | | unknown protein |
| pc1940 | | unknown protein |
| pc1941 | | unknown protein |
| pc1942 | | unknown protein |
| pc1943 | | unknown protein |
| pc1944 | | unknown protein |
| pc1946 | | conserved hypothetical protein |
| pc1947 | | unknown protein |
| pc1949 | | unknown protein |
| pc1950 | | unknown protein |
| pc1951 | | strongly similar to metalloproteinase |
| pc1952 | | unknown protein |
| pc1954 | | conserved hypothetical protein |
| pc1955 | | conserved hypothetical protein |
| pc1956 | | conserved hypothetical protein |
| pc1957 | | hypothetical protein |
| pc1958 | | similar to metalloproteinase |
| pc1959 | | unknown protein |
| pc1960 | | unknown protein |
| pc1961 | | unknown protein |
| pc1963 | | conserved hypothetical protein |
| pc1964 | | unknown protein |
| pc1965 | | similar to metalloproteinase |
| pc1966 | | conserved hypothetical protein |
| pc1967 | | unknown protein |
| pc1968 | | conserved hypothetical protein |
| pc1969 | | unknown protein |
| pc1970 | | conserved hypothetical protein |
| pc1971 | | unknown protein |
| pc1972 | | hypothetical protein |
| pc1973 | | unknown protein |
| pc1974 | | unknown protein |
| pc1975 | <i>hemF</i> | strongly similar to coproporphyrinogen oxidase III, aerobic |
| pc1976 | | unknown protein |
| pc1977 | <i>fixO</i> | similar to cytochrome-c oxidase fixO chain |
| pc1978 | <i>fixN</i> | similar to cytochrome-c oxidase fixN chain |
| pc1980 | | unknown protein |
| pc1981 | | conserved hypothetical protein |
| pc1982 | | conserved hypothetical protein |
| pc1984 | | hypothetical protein |
| pc1985 | | conserved hypothetical protein |
| pc1986 | | hypothetical protein |
| pc1989 | | hypothetical protein |
| pc1990 | | unknown protein |
| pc1992 | | conserved hypothetical protein |
| pc1993 | | conserved hypothetical protein |
| pc1999 | <i>mgtC</i> | strongly similar to mgtC protein |
| pc2000 | | unknown protein |
| pc2001 | | conserved hypothetical protein |
| pc2002 | | conserved hypothetical protein |
| pc2003 | | unknown protein |
| pc2005 | | unknown protein |
| pc2009 | | unknown protein |
| pc2011 | | unknown protein |
| pc2016 | | conserved hypothetical protein |

| | | |
|--------|--|--------------------------------|
| pc2017 | | hypothetical protein |
| pc2018 | | unknown protein |
| pc2019 | | hypothetical protein |
| pc2022 | | hypothetical protein |
| pc2023 | | hypothetical protein |
| pc2024 | | unknown protein |
| pc2026 | | unknown protein |
| pc2027 | | unknown protein |
| pc2028 | | unknown protein |
| pc2029 | | conserved hypothetical protein |
| pc2030 | | unknown protein |
| pc2031 | | unknown protein |

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)

Table S2
Predicted CDSs of UWE25 with homology to proteins of plants and cyanobacteria.
Subcellular location of plant homologues were predicted using TargetP, ChloroP, and iPSORT

| Sequence ID | Gene ID | Description | Homology to | Description of homologous protein | Accession nr | Presence in pathogenic chlamydiae | TargetP (14) | | | | ChloroP (16) | | iPSORT (14) | |
|-------------|------------|---|--|---|--------------|-----------------------------------|--------------|-------|-------|----------|--------------|----------|-------------|---|
| | | | | | | | cTP | mTP | SP | Location | cTP | Location | cTP | |
| pc0048 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc0079 | | strongly similar to UDP-glucuronat epimerase | Arabidopsis thaliana | probable nucleotide sugar epimerase | A84889 | - | 0.151 | 0.14 | 0.001 | - | 0.541 | Y | - | - |
| pc0099 | | conserved hypothetical protein | Medicago truncatula | phosphate transporter PHT2-1 | AF533081 | + | 0.639 | 0.507 | 0.005 | C | 0.505 | Y | M | - |
| pc0141 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AK117528 | + | 0.122 | 0.153 | 0.029 | - | 0.483 | - | C | - |
| pc0160 | yibC | similar to ribosomal large chain pseudouridine synthase B | Arabidopsis thaliana | unknown protein | AK117682 | + | 0.038 | 0.503 | 0.182 | M | 0.448 | - | - | - |
| pc0161 | pgm4 | similar to phosphoglycerate mutase | Arabidopsis thaliana | phosphoglycerate mutase homolog | C86354 | + | 0.846 | 0.051 | 0.037 | C | 0.583 | Y | C | - |
| pc0175 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AP001314 | + | 0.209 | 0.024 | 0.264 | - | 0.478 | - | M | - |
| pc0199 | gpaA | similar to F-box protein | Arabidopsis thaliana | hypothetical protein | F86291 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc0324 | | hypothetical protein | Arabidopsis thaliana | hypothetical protein F27K19.220 | T49216 | - | 0.953 | 0.075 | 0.002 | C | 0.593 | Y | C | - |
| pc0325 | | conserved hypothetical protein | Arabidopsis thaliana | hypothetical protein T10H14.190 | T04917 | + | 0.113 | 0.040 | 0.007 | - | 0.440 | - | - | - |
| pc0327 | ispD | similar to 4-diphosphocytidyl-2C-methyl-D-erythritol synthase | Arabidopsis thaliana | unknown protein | AX393036 | + | 0.107 | 0.239 | 0.052 | - | 0.543 | Y | C | - |
| pc0334 | gpaA | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc0346 | asnS | strongly similar to asparagine-tRNA ligase | Oryza sativa (japonica cultivar-group) | putative asparaginyl-tRNA synthetase | AF003849.2 | - | 0.525 | 0.864 | 0.001 | M | 0.541 | Y | M | - |
| pc0395 | ksgA | similar to dimethyladenosine transferase | Arabidopsis thaliana | dimethyladenosine transferase | T51591 | + | 0.967 | 0.028 | 0.024 | C | 0.514 | Y | C | - |
| pc0398 | ddlA | similar to D-alanine-D-alanine ligase | Arabidopsis thaliana | unknown protein | AK118847 | - | 0.378 | 0.075 | 0.028 | - | 0.504 | Y | C | - |
| pc0527 | | conserved hypothetical protein | Oryza sativa (japonica cultivar-group) | Highly conserved hypothetical protein | AAO06953 | + | 0.014 | 0.905 | 0.123 | M | 0.453 | - | M | - |
| pc0643 | prp | strongly similar to polyribonucleotide nucleotidyltransferase | garden pea | polyribonucleotide nucleotidyltransferase | T06540 | + | 0.913 | 0.086 | 0.001 | C | 0.580 | Y | - | - |
| pc0648 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | G96711 | + | 0.114 | 0.680 | 0.001 | M | 0.487 | Y | - | - |
| pc0674 | | hypothetical protein | Arabidopsis thaliana | probable glucose regulated repressor protein | A84649 | - | 0.131 | 0.200 | 0.029 | - | 0.473 | - | - | - |
| pc0685 | aspC | similar to aspartate transaminase | Arabidopsis thaliana | probable aspartate aminotransferase | A84511 | + | 0.152 | 0.207 | 0.075 | - | 0.445 | - | - | - |
| pc0693 | glyS | strongly similar to glycyl-tRNA synthetase | Arabidopsis thaliana | probable aminoacyl-tRNA synthetase | T06672 | + | 0.820 | 0.222 | 0.021 | C | 0.571 | Y | C | - |
| pc0718 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AY065231 | + | 0.934 | 0.211 | 0.017 | C | 0.574 | Y | C | - |
| pc0740 | gcpE | strongly similar to gcpE protein | Catharanthus roseus | GCPE protein | AY184810 | + | 0.496 | 0.098 | 0.013 | C | 0.526 | Y | M | - |
| pc0743 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc0745 | malQ | similar to 4-alpha-glucanotransferase | Arabidopsis thaliana | 4-alpha-glucanotransferase | Y081315 | + | 0.455 | 0.108 | 0.094 | C | 0.502 | Y | M | - |
| pc0754 | | hypothetical protein | Arabidopsis thaliana | probable glucose regulated repressor protein | A84649 | - | 0.131 | 0.200 | 0.029 | - | 0.473 | - | - | - |
| pc0796 | | similar to tonoplast intrinsic protein (Aquaporin) | Nicotiana tabacum | aquaglyceroporin | AJ237751 | - | 0.021 | 0.044 | 0.532 | - | 0.466 | - | - | - |
| pc0798 | | unknown protein | Oryza sativa (japonica cultivar-group) | hypothetical protein | AAO00689 | + | 0.022 | 0.756 | 0.003 | M | 0.473 | - | - | - |
| pc0825 | kdsB | strongly similar to 3-deoxy-manno-octulosonate cytidyltransferase (CMP-KDO synthetase) | Zea mays | CMP-KDO synthetase | AJ242474 | + | 0.480 | 0.211 | 0.016 | C | 0.522 | Y | C | - |
| pc0850 | | hypothetical protein | Arabidopsis thaliana | hypothetical protein | AAG09089 | - | 0.118 | 0.195 | 0.054 | - | 0.430 | - | - | - |
| pc0888 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc0890 | ribA, ribB | strongly similar to 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II | Arabidopsis thaliana | 3,4-dihydroxy-2-butanone 4-phosphate synthase | P47924 | + | 0.953 | 0.184 | 0.007 | C | 0.574 | Y | C | - |
| pc0893 | rbp | strongly similar to nucleic acid-binding protein | Arabidopsis thaliana | glycine-rich RNA binding protein, putative | AY085838 | - | 0.427 | 0.512 | 0.014 | M | 0.555 | Y | M | - |
| pc0992 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |

| | | | | | | | | | | | | | | |
|--------|------------|---|--|--|----------|---|-------|-------|-------|---|-------|---|---|---|
| pc1025 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1031 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1032 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1033 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1106 | | strongly similar to isoamylase | Solanum tuberosum | isoamylase isoform 1 | AY132996 | + | 0.818 | 0.109 | 0.005 | C | 0.534 | Y | C | - |
| pc1131 | | conserved hypothetical protein | Arabidopsis thaliana | hypothetical protein | C85431 | - | 0.982 | 0.084 | 0.014 | C | 0.546 | Y | M | - |
| pc1142 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1152 | fabI | strongly similar to NADH-dependent enoyl-ACP reductase | common tobacco | enoyl-[acyl-carrier-protein] reductase (NADH2) | T03229 | + | 0.980 | 0.027 | 0.007 | C | 0.588 | Y | C | - |
| pc1161 | dnaQ, mutD | similar to DNA polymerase III, epsilon chain, mutD | Arabidopsis thaliana | hypothetical protein A_IJG002P16.22 | T01771 | + | 0.776 | 0.149 | 0.004 | C | 0.496 | - | M | - |
| pc1169 | tyrS | strongly similar to tyrosine-tRNA ligase | Arabidopsis thaliana | putative tyrosyl-tRNA synthetase | AAF32473 | + | 0.860 | 0.191 | 0.004 | C | 0.578 | Y | M | - |
| pc1197 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1238 | fabF | strongly similar to beta-ketoacyl-ACP synthetase | Glycine max | beta-ketoacyl-ACP synthetase I-2 | AF243183 | + | 0.805 | 0.185 | 0.014 | C | 0.575 | Y | C | - |
| pc1272 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1276 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1278 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1282 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1297 | sucD | strongly similar to succinate-CoA ligase (ADP-forming) alpha chain | Lycopersicon esculentum | succinyl-CoA ligase alpha subunit | AY167586 | + | 0.144 | 0.181 | 0.053 | - | 0.478 | - | M | - |
| pc1318 | ats1 | similar to glycerol-3-phosphate acyltransferase | Elaeis guineensis | glycerol-3-phosphate acyltransferase | AF251795 | + | 0.857 | 0.392 | 0.002 | C | 0.590 | Y | M | - |
| pc1343 | ntt_5 | similar to ADP/ATP translocase | Arabidopsis thaliana | putative adenine nucleotide translocase | AY045903 | + | 0.542 | 0.412 | 0.005 | C | 0.502 | Y | M | - |
| pc1462 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1495 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1507 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1508 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1510 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1565 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1589 | | strongly similar to isopentenyl monophosphate kinase (IPK) | Oryza sativa (japonica cultivar-group) | putative 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase | BAB86428 | + | 0.880 | 0.288 | 0.004 | C | 0.557 | Y | C | - |
| pc1596 | glgA | similar to starch synthase, precursor, glgA | Vigna unguiculata | starch synthase, isoform V | AJ006752 | + | 0.055 | 0.077 | 0.187 | - | 0.429 | - | - | - |
| pc1607 | | hypothetical protein | Oryza sativa (japonica cultivar-group) | putative DnaJ domain containing protein | AAL67579 | + | 0.859 | 0.045 | 0.012 | C | 0.553 | Y | M | - |
| pc1616 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1620 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | BT002316 | + | 0.826 | 0.177 | 0.045 | C | 0.548 | Y | C | - |
| pc1652 | | unknown protein | Arabidopsis thaliana | protein T10Q24.19 | B86239 | - | 0.098 | 0.185 | 0.048 | - | 0.467 | - | M | - |
| pc1666 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1673 | atpE | strongly similar to H ⁺ -transporting two-sector ATPase lipid-binding protein (chainC, atpE) | Chaetopharidium globosum chloroplast | subunit III of ATP synthase | AAM96503 | + | 0.111 | 0.038 | 0.866 | S | 0.455 | - | - | - |
| pc1713 | trxB | strongly similar to thioredoxin-disulfide reductase 2 | Arabidopsis thaliana | probable thioredoxin reductase | A84552 | + | 0.315 | 0.051 | 0.303 | C | 0.568 | Y | M | - |
| pc1769 | | hypothetical protein | Arabidopsis thaliana | probable D-amino acid dehydrogenase | B84615 | + | 0.251 | 0.346 | 0.072 | M | 0.468 | - | C | - |
| pc1772 | mdh | strongly similar to NADP-dependent malate dehydrogenase | Arabidopsis thaliana | NADP-dependent malate dehydrogenase | AB019228 | + | 0.951 | 0.131 | 0.006 | C | 0.560 | Y | C | - |
| pc1782 | gutQ | similar to Gut Q protein | Arabidopsis thaliana | sugar-phosphate isomerase-like protein | T47628 | + | 0.150 | 0.070 | 0.026 | - | 0.448 | - | - | - |
| pc1832 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1838 | lplA | similar to lipase-protein ligase | Arabidopsis thaliana | lipase protein ligase-like protein | AB025615 | + | 0.031 | 0.547 | 0.146 | M | 0.440 | - | M | - |
| pc1909 | | hypothetical protein | Chara corallina | myosin | AB007459 | - | 0.059 | 0.178 | 0.191 | - | 0.431 | - | - | - |
| pc1914 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1915 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AY142631 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1918 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AY142631 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1920 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1926 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |
| pc1928 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - | - |

| | | | | | | | | | | | | | |
|--------|------|---|-------------------------------------|--|----------|---|-------|-------|-------|---|-------|---|---|
| pc1935 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - |
| pc1954 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - |
| pc1956 | | conserved hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - |
| pc1981 | | conserved hypothetical protein | Arabidopsis thaliana | hypothetical protein F28B23.16 | G86387 | - | 0.892 | 0.244 | 0.001 | C | 0.587 | Y | C |
| pc1996 | ispA | similar to geranyltransferase | white lupine | farnesyltransferase | T11021 | + | 0.175 | 0.085 | 0.013 | - | 0.425 | - | - |
| pc1998 | rumA | similar to 23S rRNA (Uracil-5-)-methyltransferase | Arabidopsis thaliana | unknown protein | AP001312 | + | 0.854 | 0.077 | 0.070 | C | 0.564 | Y | M |
| pc2012 | ung | strongly similar to uracil-DNA glycosylase | Arabidopsis thaliana | uracil-DNA glycosylase-like protein | AP001303 | + | 0.928 | 0.050 | 0.007 | C | 0.586 | Y | M |
| pc2019 | | hypothetical protein | Arabidopsis thaliana | unknown protein | AF360339 | - | 0.367 | 0.069 | 0.016 | - | 0.464 | - | - |
| pc0005 | | conserved hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S76896 | - | | | | | | | |
| pc0040 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0041 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0042 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0044 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0045 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0046 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0051 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0052 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0053 | | hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein al1820 | AF2033 | - | | | | | | | |
| pc0054 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0057 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0058 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0059 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | |
| pc0061 | | hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein al13194 | AC2205 | - | | | | | | | |
| pc0072 | dprA | similar to protein required for chromosomal DNA transformation | Nostoc sp. (strain PCC 7120) | DNA processing protein | AB1972 | - | | | | | | | |
| pc0106 | glgP | strongly similar to glycogen phosphorylase | Thermosynechococcus elongatus BP-1 | glycogen phosphorylase | BAC08333 | + | | | | | | | |
| pc0108 | tiV | similar to amino acid ABC transporter, periplasmic amino acid-binding protein | Nostoc sp. (strain PCC 7120) | glutamine-binding periplasmic protein of glutamine ABC transporter alr3187 | AD2204 | + | | | | | | | |
| pc0109 | glgC | strongly similar to glucose-1-phosphate adenylyltransferase | Anabaena sp. (strain PCC 7120) | glucose-1-phosphate adenylyltransferase | S24991 | + | | | | | | | |
| pc0115 | | hypothetical protein | Synechocystis sp. | putative protein | AAA85912 | + | | | | | | | |
| pc0181 | wzt | strongly similar to ABC transporter ATP-binding protein wzt | Synechocystis sp. (strain PCC 6803) | ABC-type transport protein sir0982 | S74749 | - | | | | | | | |
| pc0182 | wzm | strongly similar to ABC transporter protein wzm | Synechocystis sp. (strain PCC 6803) | ABC-type transport protein sir0977 | S74745 | - | | | | | | | |
| pc0189 | sufB | strongly similar to ABC transporter protein sufB | Thermosynechococcus elongatus BP-1 | ABC transporter subunit | AP005370 | + | | | | | | | |
| pc0190 | sufC | strongly similar to ABC transporter ATP-binding protein sufC | Nostoc sp. (strain PCC 7120) | ABC transporter ATP-binding protein ycf16 | AF2117 | + | | | | | | | |

| | | | | | | | | | | | | | |
|--------|------------|--|---|---|----------|---|--|--|--|--|--|--|--|
| pc0227 | psf, ygbB | similar to 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MECDP-synthase) | Thermosynechococcus elongatus BP-1 | 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase | AP005376 | + | | | | | | | |
| pc0231 | argS | strong similarity to arginyl-tRNA synthetase (=arginine-tRNA-ligase) | Thermosynechococcus elongatus BP-1 | arginyl-tRNA-synthetase | AP005371 | + | | | | | | | |
| pc0232 | clpB | similar to endopeptidase Clp ATP-binding chain B (heat shock protein) | Nostoc sp. (strain PCC 7120) | endopeptidase Clp ATP-binding chain B | AD2441 | + | | | | | | | |
| pc0311 | mraW, yabC | strongly similar to S-adenosyl-methyltransferase | Synechococcus elongatus | S-adenosyl-methyltransferase mraW | Q8DH87 | + | | | | | | | |
| pc0315 | amiA, cwC | similar to N-acetylmuramoyl-L-alanine amidase | Nostoc sp. (strain PCC 7120) | N-acetylmuramoyl-L-alanine amidase | AD1818 | + | | | | | | | |
| pc0320 | cdsA | similar to phosphatidate cytidyltransferase | Synechocystis sp. (strain PCC 6803) | phosphatidate cytidyltransferase | S77254 | + | | | | | | | |
| pc0321 | uppS | strongly similar to undecaprenyl pyrophosphate synthetase | Nostoc sp. (strain PCC 7120) | undecaprenyl pyrophosphate synthetase | AI2234 | + | | | | | | | |
| pc0371 | fdxC | similar to ferredoxin [2Fe-2S] IV | Thermosynechococcus elongatus BP-1 | probable ferredoxin | BAC09081 | + | | | | | | | |
| pc0390 | | hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein al14382 | AF2353 | + | | | | | | | |
| pc0392 | rsbW | similar to rsbW, negative regulator of sigma-B activity (switch protein/serine kinase) | Synechocystis sp. (strain PCC 6803) | hypothetical protein sir1861 | S77097 | + | | | | | | | |
| pc0443 | clpP | strongly similar to ATP-dependent Clp protease proteolytic subunit | Nostoc sp. (strain PCC 7120) | ATP-dependent Clp proteinase proteolytic chain | AF2350 | + | | | | | | | |
| pc0480 | | conserved hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein sir0904 | S75721 | - | | | | | | | |
| pc0538 | rbxA | similar to RTX-toxin, partial length | Nostoc sp. (strain PCC 7120) plasmid pCC7120delta | hypothetical protein al18511 | AD2564 | - | | | | | | | |
| pc0623 | nfi | strongly similar to endonuclease V (deoxyinosine 3'endonuclease) | Nostoc sp. (strain PCC 7120) | endonuclease V | AF2316 | - | | | | | | | |
| pc0641 | | hypothetical protein | Nostoc sp. (strain PCC 7120) | glycosyl transferase al14431 | AG2359 | - | | | | | | | |
| pc0644 | psO, rs15 | strongly similar to 30S ribosomal protein S15 | Synechocystis sp. (strain PCC6803) | ribosomal protein S15) | S74731 | + | | | | | | | |
| pc0645 | | conserved hypothetical protein | Synechocystis sp. (strain PCC6803) | transposase sir1094 | S75870 | - | | | | | | | |
| pc0729 | | conserved hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein sir0516 | S76649 | - | | | | | | | |
| pc0733 | | unknown protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein sir0516 | S76649 | - | | | | | | | |
| pc0734 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein sir0183 | S76576 | - | | | | | | | |
| pc0736 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein al13114 | AC2195 | - | | | | | | | |
| pc0742 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein sir0024 | AH1809 | + | | | | | | | |
| pc0753 | | conserved hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein sir0064 | S74375 | - | | | | | | | |
| pc0769 | efp | similar to elongation factor P | Thermosynechococcus elongatus BP-1 | translation elongation factor EF-P | AP005373 | + | | | | | | | |
| pc0772 | accC | strongly similar to biotin carboxylase | Thermosynechococcus elongatus BP-1 | biotin carboxylase | AP005375 | + | | | | | | | |
| pc0820 | | hypothetical protein | Synechococcus elongatus PCC 7942 | putative OxPP cycle protein opcA | Q54709 | - | | | | | | | |

| | | | | | | | | | | | | | | | | | | | |
|--------|-------------------|--|-------------------------------------|--|----------|---|--|--|--|--|--|--|--|--|--|--|--|--|--|
| pc0821 | <i>zwf</i> | similarity to glucose-6-phosphate 1-dehydrogenase (G6PD) | Synechocystis sp. (strain PCC 6803) | glucose-6-phosphate 1-dehydrogenase | S77348 | + | | | | | | | | | | | | | |
| pc0868 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | transposase alr0590 | AE1880 | - | | | | | | | | | | | | | |
| pc0903 | | conserved hypothetical protein | Thermosynechococcus elongatus BP-1 | putative transposase | BAC07793 | - | | | | | | | | | | | | | |
| pc0952 | | strongly similar to putative oxidoreductases | Nostoc sp. (strain PCC 7120) | oxidoreductase alr5182 | AF2453 | + | | | | | | | | | | | | | |
| pc0971 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | transposase alr0590 | AE1880 | + | | | | | | | | | | | | | |
| pc0980 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein alr2462 | AG2113 | + | | | | | | | | | | | | | |
| pc0990 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein asl4561 | AI2375 | - | | | | | | | | | | | | | |
| pc0996 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein alr2402 | AC2106 | - | | | | | | | | | | | | | |
| pc1010 | | conserved hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S76678 | - | | | | | | | | | | | | | |
| pc1044 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein alr3204 | AE2206 | - | | | | | | | | | | | | | |
| pc1066 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | transposase alr0590 | AE1880 | - | | | | | | | | | | | | | |
| pc1079 | <i>wzb</i> | similar to low molecular weight protein-tyrosine-phosphatase | Synechococcus elongatus PCC 7942 | putative low molecular weight protein tyrosine phosphatase slr0328 | CAD55613 | - | | | | | | | | | | | | | |
| pc1107 | | hypothetical protein | Synechocystis sp. (strain PCC 6083) | stage II sporulation protein spoIID | S75305 | + | | | | | | | | | | | | | |
| pc1113 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein alr1520 | AB1996 | + | | | | | | | | | | | | | |
| pc1120 | | conserved hypothetical protein | Synechocystis sp. (strain PCC 6803) | hemolysin | S76248 | + | | | | | | | | | | | | | |
| pc1146 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | transposase alr0590 | AE1880 | - | | | | | | | | | | | | | |
| pc1157 | <i>folB, mutT</i> | similar to dGTP pyrophosphohydrolase/dihydroneopterin aldolase (mutT/folB, fusion protein) | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S76176 | - | | | | | | | | | | | | | |
| pc1158 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S76201 | + | | | | | | | | | | | | | |
| pc1162 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein alr0480 | AG1866 | - | | | | | | | | | | | | | |
| pc1195 | | conserved hypothetical protein | Thermosynechococcus elongatus BP-1 | putative transposase | BAC07793 | - | | | | | | | | | | | | | |
| pc1196 | | conserved hypothetical protein | Thermosynechococcus elongatus BP-1 | putative transposase | BAC07792 | + | | | | | | | | | | | | | |
| pc1214 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | transposase alr0590 | AE1880 | + | | | | | | | | | | | | | |
| pc1215 | | conserved hypothetical protein | Thermosynechococcus elongatus BP-1 | putative transposase | BAC07792 | + | | | | | | | | | | | | | |
| pc1226 | | conserved hypothetical protein | Thermosynechococcus elongatus BP-1 | hypothetical protein | BAC09116 | + | | | | | | | | | | | | | |
| pc1266 | | conserved hypothetical protein | Thermosynechococcus elongatus BP-1 | putative dehydrogenase | BAC09443 | + | | | | | | | | | | | | | |
| pc1277 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | transposase alr0590 | AE1880 | + | | | | | | | | | | | | | |
| pc1279 | | conserved hypothetical protein | Synechocystis sp. (strain PCC 6803) | transposase slr1094 | S75870 | - | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | |
|--------|-------------------|---|-------------------------------------|--|----------|---|--|--|--|--|--|--|--|--|--|--|--|--|--|
| pc1362 | <i>ackA</i> | strongly similar to acetate kinase | Nostoc sp. (strain PCC 7120) | acetate kinase | AB2126 | - | | | | | | | | | | | | | |
| pc1369 | | similar to glycosyltransferase | Nostoc sp. (strain PCC 7120) | hypothetical protein alr3062 | AG2188 | - | | | | | | | | | | | | | |
| pc1378 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S76855 | - | | | | | | | | | | | | | |
| pc1379 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S76855 | - | | | | | | | | | | | | | |
| pc1391 | | conserved hypothetical protein | Synechocystis sp. (strain PCC 6803) | adenylate cyclase | S75018 | + | | | | | | | | | | | | | |
| pc1402 | | conserved hypothetical protein | Thermosynechococcus elongatus BP-1 | putative transposase | BAC07793 | - | | | | | | | | | | | | | |
| pc1453 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | HicB protein | AD2465 | - | | | | | | | | | | | | | |
| pc1484 | | conserved hypothetical protein | Thermosynechococcus elongatus BP-1 | putative helicase | BAC09165 | + | | | | | | | | | | | | | |
| pc1534 | <i>yabJ, yigF</i> | strongly similar to yabJ | Synechococcus PCC7002 | hypothetical protein | AAG43443 | - | | | | | | | | | | | | | |
| pc1535 | <i>htpG</i> | similar to heat shock protein HtpG | Thermosynechococcus elongatus BP-1 | heat shock protein | AP005373 | - | | | | | | | | | | | | | |
| pc1554 | | conserved hypothetical protein | Synechocystis sp. (strain PCC 6803) | transforming growth factor-induced protein | S76811 | - | | | | | | | | | | | | | |
| pc1556 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | maltooligosyltrehalose synthase | AG1827 | - | | | | | | | | | | | | | |
| pc1570 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | acetyl-CoA synthetase | AG1902 | - | | | | | | | | | | | | | |
| pc1573 | <i>rfs, prfC</i> | strongly similar to peptide chain release factor 3 | Synechocystis sp. (strain PCC 6803) | translation releasing factor RF-3 | S77410 | + | | | | | | | | | | | | | |
| pc1624 | <i>murB</i> | similar to UDP-N-acetylmuramate dehydrogenase | Thermosynechococcus elongatus BP-1 | UDP-N-acetylenpyruvylglucosamine reductase | BAC07922 | + | | | | | | | | | | | | | |
| pc1637 | | conserved hypothetical protein | Synechocystis sp. (strain PCC 6803) | transposase slr1094 | S75870 | - | | | | | | | | | | | | | |
| pc1640 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | transposase alr0590 | AE1880 | + | | | | | | | | | | | | | |
| pc1696 | <i>crtE</i> | strongly similar to farnesyltransferase | Nostoc sp. (strain PCC 7120) | geranylgeranyl diphosphate synthase | AE1833 | + | | | | | | | | | | | | | |
| pc1697 | <i>exoY</i> | strongly similar to exopolysaccharide production protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S76182 | - | | | | | | | | | | | | | |
| pc1724 | <i>recR</i> | strongly similar to recombination protein RecR | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S76727 | + | | | | | | | | | | | | | |
| pc1728 | <i>lpxD, tirA</i> | similar to UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase | Thermosynechococcus elongatus BP-1 | UDP-3-O-(R-3-hydroxymyristoyl)-glucosamine N-acyltransferase | BAC07585 | + | | | | | | | | | | | | | |
| pc1761 | <i>glgB</i> | strongly similar to 1,4-alpha-glucan branching enzyme (= Glycogen branching enzyme) | Thermosynechococcus elongatus BP-1 | 1,4-alpha-glucan branching enzyme | AP005370 | + | | | | | | | | | | | | | |
| pc1803 | | conserved hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein alr0053 | AE1813 | - | | | | | | | | | | | | | |
| pc1879 | | hypothetical protein | Nostoc sp. (strain PCC 7120) | hypothetical protein alr8078 | AG2560 | - | | | | | | | | | | | | | |
| pc1910 | | unknown protein | Synechococcus sp. PCC 7002 | SMC protein | AJ543651 | - | | | | | | | | | | | | | |
| pc1957 | | hypothetical protein | Synechocystis sp. (strain PCC 6803) | hypothetical protein | S75991 | - | | | | | | | | | | | | | |
| pc1988 | | conserved hypothetical protein | Synechocystis sp. (strain PCC 6803) | conserved hypothetical protein ssr2142 | S74690 | + | | | | | | | | | | | | | |

Illuminating the evolutionary history of chlamydiae (Horn *et al.*)

Table S3

Predicted CDSs present in pathogenic chlamydiae but absent in UWE25 obtained by cluster analysis (fastA score/selfscore cut-off 0.1).

| Description | Gene ID | <i>C. muridarum</i> | <i>C. caviae</i> | <i>C. pneumoniae</i> | <i>C. trachomatis</i> |
|--------------------------------|-------------|--|--|---|----------------------------|
| hypothetical protein | | | | CPn0007, CPn0008, CPn0009, CPn0010, CPn0010.1, CPn0011, CPn0012, CPn0041, CPn0042, CPn0043, CPn0044, CPn0045, CPn0046, CPn0124, CPn0125, CPn0126, CPn1054, CPn1055, CPn1056 | |
| hypothetical protein | | | | CPn0455, CPn0456, CPn0457, CPn0458, CPn0459, CPn0460, CPn0461, CPn0462, CPn0463, CPn0464, CPn0465 | |
| conserved hypothetical protein | | TC0084, TC0085, TC0909, TC0910, TC0911 | CCA00016, CCA00017, CCA00032, CCA00914, CCA00915 | CPn0710, CPn0726, CPn0727, CPn0853 | CT619, CT620, CT621, CT712 |
| conserved hypothetical protein | | TC0328 | CCA00221, CCA00398, CCA00425, CCA00426 | CPn0107, CPn0367, CPn0368, CPn0369, CPn0370, CPn0524 | CT058 |
| hypothetical protein | | TC0496, TC0499, TC0500 | | CPn0186, CPn0210, CPn0212 | CT224, CT228, CT229 |
| conserved hypothetical protein | | TC0650 | CCA00299, CCA00300, CCA00728 | CPn1034, CPn1072, CPn1073 | CT371 |
| Fe-S oxidoreductase | | TC0148, TC0710 | CCA00230 | CPn0513, CPn0911 | CT426, CT767 |
| hypothetical protein | <i>incC</i> | TC0504 | CCA00389, CCA00390, CCA00490 | CPn0292, CPn0404 | CT105, CT233 |
| hypothetical protein | | | CCA00561, CCA00709 | CPn0066, CPn0069, CPn0381 | |
| conserved hypothetical protein | | TC0177 | CCA00824, CCA00825, CCA00826 | CPn0944, CPn0945 | CT795 |
| hypothetical protein | | | | CPn0156, CPn0158, CPn0162, CPn0204 | |
| conserved hypothetical protein | | | CCA00062 | CPn0493, CPn0677, CPn0678 | |
| conserved hypothetical protein | | TC0421 | CCA00525 | CPn0256 | CT144 |
| conserved hypothetical protein | | TC0274 | CCA00291 | CPn0371, CPn0442 | CT006 |
| conserved hypothetical protein | | TC0420 | CCA00524 | CPn0254, CPn0257 | CT143 |
| hypothetical protein | | TC0411 | CCA00537 | CPn0242, CPn0243 | CT134 |
| conserved hypothetical protein | | TC0419 | CCA00523 | CPn0255 | CT142 |
| conserved hypothetical protein | | TC0075 | CCA00924 | CPn0842, CPn0843 | CT702 |
| conserved hypothetical protein | | TC0887 | CCA00974 | CPn0099, CPn0783 | CT598 |
| conserved hypothetical protein | | TC0724 | CCA00156, CCA00576 | CPn0372 | CT440 |
| conserved hypothetical protein | | TC0662 | CCA00263, CCA00708 | CPn0480 | CT383 |
| inclusion membrane protein B | <i>incB</i> | TC0503, TC0644 | CCA00269 | CPn0474 | CT365 |
| conserved hypothetical protein | | TC0607 | CCA00305, CCA00794 | CPn1061 | CT330 |

1/8

| | | | | | |
|--|-------------|--------|------------------------------|------------------|-----------|
| conserved hypothetical protein | | TC0015 | CCA00415, CCA00991 | CPn0766 | CT646 |
| conserved hypothetical protein | | | CCA00538 | CPn0240, CPn0241 | |
| lipid A biosynthesis lauroyl acyltransferase, putative | <i>htrB</i> | TC0278 | CCA00673 | CPn0098 | CT010 |
| conserved hypothetical protein | | TC0234 | CCA00758 | CPn1003 | CT846 |
| conserved hypothetical protein | | TC0020 | CCA01006, CCA01012 | CPn0751 | CT651 |
| MAC/perforin family protein | | TC0431 | | CPn0175, CPn0176 | CT153 |
| conserved hypothetical protein | | TC0024 | CCA00020 | CPn0722 | CT654 |
| conserved hypothetical protein | | TC0028 | CCA00024 | CPn0718 | CT657 |
| conserved hypothetical protein | | TC0027 | CCA00025 | CPn0717 | CT656 |
| conserved hypothetical protein | | TC0039 | CCA00034 | CPn0708 | CT668 |
| major outer membrane protein, porin | <i>ompA</i> | TC0052 | CCA00047 | CPn0695 | CT681 |
| conserved hypothetical protein | | TC0767 | CCA00054 | CPn0688 | CT481 |
| conserved hypothetical protein | | TC0769 | CCA00055 | CPn0687 | CT482 |
| conserved hypothetical protein | | TC0067 | CCA00063 | CPn0676 | CT695 |
| conserved hypothetical protein | | TC0068 | CCA00064 | CPn0675 | G13329150 |
| cytosolic acyl-CoA thioester hydrolase family protein | <i>yclA</i> | TC0822 | CCA00086 | CPn0654 | CT535 |
| conserved hypothetical protein | | TC0816 | CCA00092 | CPn0648 | CT529 |
| conserved hypothetical protein | | TC0777 | CCA00131 | CPn0609 | CT490 |
| conserved hypothetical protein | | TC0756 | CCA00151 | CPn0590 | CT471 |
| conserved hypothetical protein | | TC0741 | CCA00170 | CPn0572 | CT456 |
| conserved hypothetical protein | | TC0729 | CCA00183 | CPn0559 | CT444.1 |
| 15 kDa Cysteine-Rich Protein | <i>crpA</i> | TC0726 | CCA00186 | CPn0556 | CT442 |
| conserved hypothetical protein | | TC0201 | CCA00208 | CPn0537 | CT814.1 |
| conserved hypothetical protein | | TC0711 | CCA00229 | CPn0514 | CT427 |
| conserved hypothetical protein | <i>yprS</i> | TC0671 | CCA00245 | CPn0499 | CT392 |
| amino acid ABC transporter+A20 | <i>artJ</i> | TC0660 | CCA00262 | CPn0482 | CT381 |
| conserved hypothetical protein | | TC0269 | CCA00286 | CPn0001 | CT001 |
| conserved hypothetical protein | | TC0273 | CCA00290 | CPn0443 | CT005 |
| hypothetical protein | | | CCA00296, CCA00297, CCA00298 | CPn1071 | |
| conserved hypothetical protein | | TC0569 | CCA00343 | CPn0055 | CT296 |
| conserved hypothetical protein | | TC0561 | CCA00351 | CPn0065 | CT288 |
| conserved hypothetical protein | | TC0548 | CCA00367 | CPn0425 | CT276 |
| conserved hypothetical protein | | TC0537 | CCA00379 | CPn0415 | CT266 |
| conserved hypothetical protein | | TC0534 | CCA00382 | CPn0412 | CT263 |
| conserved hypothetical protein | | TC0533 | CCA00383 | CPn0411 | CT262 |
| conserved hypothetical protein | | TC0524 | CCA00395 | CPn0399 | CT253 |
| conserved hypothetical protein | | TC0355 | CCA00451 | CPn0331 | CT082 |
| conserved hypothetical protein | | TC0356 | CCA00452 | CPn0330 | CT083 |
| conserved hypothetical protein | | TC0359 | CCA00454 | CPn0328 | CT085 |
| cyclic nucleotide-binding protein, putative | | TC0506 | CCA00488 | CPn0294 | CT235 |
| conserved hypothetical protein | | TC0505 | CCA00489 | CPn0293 | CT234 |
| conserved hypothetical protein | | TC0468 | CCA00494 | CPn0288 | CT195 |
| serine esterase, putative | | TC0413 | CCA00510 | CPn0271 | CT136 |

2/8

| | | | | | |
|--|-------------|------------------------|------------------------------|---------|--------------|
| 4-hydroxybenzoate octaprenyltransferase | <i>ubiA</i> | TC0492 | CCA00516 | CPn0265 | CT219 |
| conserved hypothetical protein | | TC0408 | CCA00547 | CPn0189 | CT131 |
| conserved hypothetical protein | <i>dsbG</i> | TC0449 | CCA00588 | CPn0228 | GI3328581 |
| conserved hypothetical protein | | TC0450 | CCA00590 | CPn0229 | CT178 |
| conserved hypothetical protein | | TC0451 | CCA00591 | CPn0230 | CT179 |
| conserved hypothetical protein | | TC0475 | CCA00606 | CPn0206 | GI3328610 |
| monooxygenase-related protein | <i>mhpA</i> | TC0425 | CCA00615 | CPn0151 | CT148 |
| hypothetical protein | | | CCA00633 | CPn0026 | CT037, CT345 |
| conserved hypothetical protein | | TC0385 | CCA00644 | CPn0133 | CT109 |
| biotin apo-protein ligase-related protein | | TC0305 | CCA00648 | CPn0128 | CT035 |
| conserved hypothetical protein | | TC0286 | CCA00665 | CPn0108 | CT018 |
| conserved hypothetical protein | | TC0598 | CCA00700 | CPn0072 | CT324 |
| hypothetical protein | | TC0258, TC0259, TC0260 | CCA00718 | | CT867, CT868 |
| conserved hypothetical protein | | TC0651 | CCA00729 | CPn1033 | CT372 |
| conserved hypothetical protein | | TC0652 | CCA00730 | CPn1032 | CT373 |
| conserved hypothetical protein | | TC0656 | CCA00735 | CPn1026 | CT377 |
| conserved hypothetical protein | <i>ltuA</i> | TC0253 | CCA00739 | CPn1022 | CT863 |
| conserved hypothetical protein | | TC0251 | CCA00741 | CPn1020 | CT861 |
| Na ⁺ /H ⁺ antiporter, putative | | TC0247 | CCA00746 | CPn1015 | CT857 |
| conserved hypothetical protein | | TC0239 | CCA00753 | CPn1008 | CT850 |
| conserved hypothetical protein | | TC0225 | CCA00767 | CPn0994 | CT837 |
| conserved hypothetical protein | | | CCA00803, CCA00804, CCA00805 | CPn0964 | |
| conserved hypothetical protein | | TC0189 | CCA00813 | CPn0956 | CT805 |
| conserved hypothetical protein | | TC0176 | CCA00827 | CPn0943 | CT794.1 |
| conserved hypothetical protein | | TC0160 | CCA00844 | CPn0925 | CT779 |
| inorganic pyrophosphatase, putative | <i>ppa</i> | TC0153 | CCA00851 | CPn0918 | CT772 |
| icc-related protein | <i>icc</i> | TC0135 | CCA00871 | CPn0897 | CT754 |
| conserved hypothetical protein | | TC0134 | CCA00872 | CPn0896 | CT753 |
| conserved hypothetical protein | | TC0120 | CCA00881 | CPn0887 | CT744 |
| conserved hypothetical protein | | TC0110 | CCA00889 | CPn0878 | CT737 |
| conserved hypothetical protein | <i>ybcL</i> | TC0109 | CCA00890 | CPn0877 | CT736 |
| conserved hypothetical protein | <i>tlhF</i> | TC0092 | CCA00907 | CPn0860 | CT719 |
| conserved hypothetical protein | | TC0091, CT051 | CCA00908 | CPn0859 | CT718 |
| Outer Membrane Protein B | <i>porB</i> | TC0086 | CCA00913 | CPn0854 | CT713 |
| conserved hypothetical protein | | TC0073 | CCA00926 | CPn0840 | CT700 |
| conserved hypothetical protein | | TC0868 | CCA00955 | CPn0808 | CT579 |
| conserved hypothetical protein | | TC0873 | CCA00960 | CPn0803 | CT584 |
| regulatory protein, putative | <i>rbsU</i> | TC0877 | CCA00964 | CPn0793 | CT588 |
| conserved hypothetical protein | | TC0878 | CCA00965 | CPn0792 | CT589 |
| coenzyme PQQ synthesis protein c, putative | | TC0900 | CCA00996 | CPn0761 | CT610 |
| conserved hypothetical protein | | TC0901 | CCA00997 | CPn0760 | CT611 |
| conserved hypothetical protein | | TC0908 | CCA01004 | CPn0753 | CT618 |

3/8

| | | | | | |
|--|---------------|----------------|--------------------|------------------|---------|
| conserved hypothetical protein | | TC0515 | | CPn0303 | CT244 |
| hypothetical protein | | | | CPn0284, CPn0432 | |
| hypothetical protein | | | | CPn0405, CPn1070 | |
| conserved hypothetical protein | | | CCA00015 | CPn0728 | CT622 |
| hypothetical protein | | TC0734 | CCA00177 | CPn0565 | |
| conserved hypothetical protein | | TC0703 | CCA00237 | CPn0507 | CT421.1 |
| conserved hypothetical protein | | | CCA00259 | CPn0485 | CT382.1 |
| phospholipase D family protein | | TC0557 | CCA00357 | CPn0435 | CT284 |
| hypothetical protein | | TC0541 | CCA00396 | CPn0398 | |
| hypothetical protein | | TC0320, TC0321 | CCA00417 | | CT050 |
| conserved hypothetical protein | | TC0345 | CCA00443 | | CT073 |
| phenylacrylic acid decarboxylase | <i>ubiX</i> | TC0493 | CCA00517 | CPn0264 | CT220 |
| conserved hypothetical protein | | | CCA00546 | CPn0190 | CT848 |
| hypothetical protein | | TC0441 | CCA00572 | CPn0170 | |
| inosine-5'-monophosphate dehydrogenase, putative | <i>guaB</i> | TC0443 | CCA00574 | CPn0172 | CT227 |
| hypothetical protein | | | CCA00575, CCA00578 | CPn0177 | |
| disulfide bond formation protein DsbB | <i>dsbB</i> | | CCA00587, CCA00704 | CPn0227 | CT176 |
| Glu/Leu/Phe/Val dehydrogenase family protein | <i>ldh</i> | TC0154 | CCA00850 | CPn0919 | CT773 |
| 5-formyltetrahydrofolate cyclo-ligase, putative | <i>ygfA</i> | TC0018 | CCA00994 | CPn0763 | CT649 |
| conserved hypothetical protein | | TC0412 | | CPn0211 | CT135 |
| hypothetical protein | | TC0495 | | CPn0167 | CT223 |
| conserved hypothetical protein | | | CCA00189 | CPn0553 | |
| conserved hypothetical protein | | | CCA00222 | CPn0523 | |
| DNA-3-methyladenine glycosylase | | | CCA00239 | CPn0505 | |
| conserved hypothetical protein | | | CCA00246 | CPn0498 | |
| hypothetical protein | | | CCA00250 | CPn0494 | |
| conserved hypothetical protein | | | CCA00325 | CPn0034 | |
| hypothetical protein | | | CCA00424 | CPn0376 | |
| inclusion membrane protein B | <i>incB</i> | | CCA00491 | CPn0291 | |
| conserved hypothetical protein | | | CCA00495 | CPn0287 | |
| conserved hypothetical protein | | | CCA00500 | CPn0268 | |
| hypothetical protein | | | CCA00514 | CPn0266 | |
| arginine repressor | <i>argR</i> | | CCA00543 | CPn0193 | |
| 5-methylthioadenosine nucleosidase-related protein | | | CCA00593 | CPn0232 | |
| conserved domain protein | | | CCA00722 | CPn0222 | |
| conserved hypothetical protein | | | CCA00733 | CPn1029 | |
| conserved hypothetical protein | | TC0354 | | CPn0332 | CT081 |
| hypothetical protein | | | | CPn0050, CPn0131 | |
| orotate phosphoribosyltransferase | <i>pyrE</i> | | CCA00132 | CPn0608 | |
| hypothetical protein | | TC0319 | CCA00416 | | CT049 |
| hypothetical protein | | TC0520 | CCA00474 | | CT249 |
| tryptophan synthase, beta subunit | <i>trpB-1</i> | | CCA00559 | | CT170 |
| N-(5'-phosphoribosyl)-anthranilate isomerase, putative | <i>trpF</i> | TC0603 | CCA00565 | | CT327 |
| hypothetical protein | | TC0639 | CCA00710 | | CT360 |

4/8

| | | | | | |
|--------------------------------|---------------|----------------|--------------------|---------|-------|
| conserved hypothetical protein | | | CCA00716 | CPn1046 | |
| conserved hypothetical protein | | TC0306 | | CPn0130 | CT036 |
| hypothetical protein | | TC0574, TC0637 | | | CT300 |
| hypothetical protein | | | | CPn0029 | |
| hypothetical protein | | | | CPn0047 | |
| hypothetical protein | | | | CPn0063 | |
| hypothetical protein | | | | CPn0064 | |
| hypothetical protein | | | | CPn0132 | |
| Hypothetical Protein | <i>yxjG_1</i> | | | CPn0143 | |
| hypothetical protein | | | | CPn0146 | |
| hypothetical protein | | | | CPn0174 | |
| hypothetical protein | | | | CPn0180 | |
| hypothetical protein | | | | CPn0214 | |
| hypothetical protein | | | | CPn0216 | |
| hypothetical protein | | | | CPn0233 | |
| hypothetical protein | | | | CPn0277 | |
| hypothetical protein | | | | CPn0358 | |
| hypothetical protein | | | | CPn0365 | |
| hypothetical protein | | | | CPn0366 | |
| hypothetical protein | | | | CPn0375 | |
| hypothetical protein | | | | CPn0391 | |
| hypothetical protein | | | | CPn0439 | |
| hypothetical protein | | | | CPn0472 | |
| hypothetical protein | | | | CPn0473 | |
| hypothetical protein | | | | CPn0492 | |
| hypothetical protein | | | | CPn0516 | |
| hypothetical protein | | | | CPn0517 | |
| hypothetical protein | | | | CPn0583 | |
| similarity to CHLPS IncA | | | | CPn0585 | |
| hypothetical protein | | | | CPn0656 | |
| hypothetical protein | | | | CPn0685 | |
| hypothetical protein | | | | CPn0686 | |
| hypothetical protein | | | | CPn0731 | |
| hypothetical protein | | | | CPn0745 | |
| hypothetical protein | | | | CPn0882 | |
| hypothetical protein | | | | CPn1027 | |
| hypothetical protein | | | | CPn1040 | |
| dethiobiotin synthetase | <i>bioD</i> | | | CPn1042 | |
| Biotin Synthase | <i>bioB</i> | | | CPn1044 | |
| hypothetical protein | | | | CPn1053 | |
| hypothetical protein | | | | CPn1064 | |
| adherence factor | | TC0437 | CCA00075 | | |
| hypothetical protein | | | CCA00140, CCA00141 | | |
| Inclusion Membrane Protein B | <i>incB</i> | | CCA00318 | | CT232 |

5/8

| | | | | | |
|--------------------------------------|-------------|--------|-----------|---------|---------|
| trp operon repressor | <i>trpR</i> | | CCA00562 | | CT169 |
| tryptophan synthase, alpha subunit | <i>trpA</i> | | CCA00567 | | CT171 |
| conserved hypothetical protein | | | CCA00616 | | CT164 |
| hypothetical protein | | TC0391 | CCA00622 | | |
| conserved domain protein | | | CCA00647 | CPn0129 | |
| conserved hypothetical protein | | TC0751 | CCA00721 | | |
| hypothetical protein | | | CCA00837 | | |
| hypothetical protein | | TC0486 | | | CT214 |
| hypothetical protein | | TC0066 | | | CT694 |
| hypothetical protein | | TC0266 | | | CT873 |
| hypothetical protein | | TC0268 | | | CT875 |
| conserved hypothetical protein | | TC0600 | | | CT326.1 |
| hypothetical protein | | TC0601 | | | CT326.2 |
| hypothetical protein | | TC0766 | | | CT480.1 |
| hypothetical protein | | | | CPn0006 | |
| hypothetical protein | | | | CPn0051 | |
| hypothetical protein | | | | CPn0283 | |
| hypothetical protein | | | | CPn1065 | |
| ABC transporter, permease protein | | TC0698 | | | |
| virulence protein pGP3-D | | | CAA000006 | | |
| hypothetical protein | | | CAA00081 | | |
| hypothetical protein | | | CAA00082 | | |
| hypothetical protein | | | CAA00165 | | |
| conserved hypothetical protein | | | CAA00188 | | |
| thiamine-phosphate pyrophosphorylase | <i>thiE</i> | | CAA00203 | | |
| hydroxyethylthiazole kinase | <i>thiM</i> | | CAA00204 | | |
| hypothetical protein | | | CAA00251 | | |
| hypothetical protein | | | CAA00257 | | |
| hypothetical protein | | | CAA00270 | | |
| hypothetical protein | | | CAA00276 | | |
| hypothetical protein | | | CAA00294 | | |
| hypothetical protein | | | CAA00310 | | |
| hypothetical protein | | | CAA00320 | | |
| hypothetical protein | | | CAA00332 | | |
| conserved hypothetical protein | | | CAA00353 | | |
| conserved hypothetical protein | | | CAA00360 | | |
| hypothetical protein | | | CAA00369 | | |
| hypothetical protein | | | CAA00397 | | |
| hypothetical protein | | | CAA00405 | | |
| hypothetical protein | | | CAA00504 | | |
| hypothetical protein | | | CAA00531 | | |
| hypothetical protein | | | CAA00536 | | |
| inclusion membrane protein A | <i>incA</i> | | CAA00550 | | |
| hypothetical protein | | | CAA00555 | | |

6/8

| | | | | | |
|--|-------------|--------|----------|---------|---------|
| hypothetical protein | | | CCA00557 | | |
| phosphoribosylanthranilate transferase | <i>trpD</i> | | CCA00563 | | |
| indole-3-glycerol phosphate synthase | <i>trpC</i> | | CCA00564 | | |
| hypothetical protein | | | CCA00570 | | |
| hypothetical protein | | | CCA00571 | | |
| hypothetical protein | | | CCA00577 | | |
| hypothetical protein | | | CCA00585 | | |
| hypothetical protein | | | CCA00589 | | |
| hypothetical protein | | | CCA00594 | | |
| hypothetical protein | | | CCA00613 | | |
| hypothetical protein | | | CCA00635 | | |
| hypothetical protein | | | CCA00637 | | |
| hypothetical protein | | | CCA00645 | | |
| hypothetical protein | | | CCA00646 | | |
| hypothetical protein | | | CCA00674 | | |
| hypothetical protein | | | CCA00703 | | |
| hypothetical protein | | | CCA00705 | | |
| hypothetical protein | | | CCA00790 | | |
| hypothetical protein | | | CCA00796 | | |
| hypothetical protein | | | CCA00798 | | |
| hypothetical protein | | | CCA00800 | | |
| hypothetical protein | | | CCA00802 | | |
| conserved hypothetical protein | | | CCA00814 | | |
| hypothetical protein | | | CCA00840 | | |
| hypothetical protein | | | CCA00841 | | |
| hypothetical protein | | | CCA00843 | | |
| C. psittaci hypothetical protein | | | | | CT120 |
| hypothetical protein | | | | | CT159 |
| hypothetical protein | | | | | CT160 |
| hypothetical protein | | | | | CT172 |
| hypothetical protein | | | | | CT196 |
| CHLTR hypothetical protein | | | | | CT222 |
| hypothetical protein (possible 357R?) | | | | | CT357R |
| hypothetical protein | | | | | CT466 |
| hypothetical protein | | | | | CT793 |
| hypothetical protein | | | | | CT813 |
| hypothetical protein | | | | | CT039.1 |
| hypothetical protein | | | | | CT172.1 |
| hypothetical protein | | | | | CT221.1 |
| hypothetical protein | | | | | CT496.1 |
| hypothetical protein | | | | CPn0070 | |
| hypothetical protein | | | | CPn0213 | |
| hypothetical protein | | TC0053 | | | |
| hypothetical protein | | TC0071 | | | |

| | | | | | |
|----------------------------------|------------|--------|--|--|-----------|
| hypothetical protein | | TC0113 | | | |
| hypothetical protein | | | | | GI7190150 |
| hypothetical protein | | TC0115 | | | |
| hypothetical protein | | TC0126 | | | |
| hypothetical protein | | TC0162 | | | |
| hypothetical protein | | TC0179 | | | |
| hypothetical protein | | TC0188 | | | |
| hypothetical protein | | TC0191 | | | |
| hypothetical protein | | TC0198 | | | |
| hypothetical protein | | TC0220 | | | |
| hypothetical protein | | TC0265 | | | |
| hypothetical protein | | TC0287 | | | |
| hypothetical protein | | TC0304 | | | |
| hypothetical protein | | TC0307 | | | |
| hypothetical protein | | TC0337 | | | |
| hypothetical protein | | TC0358 | | | |
| hypothetical protein | | TC0360 | | | |
| hypothetical protein | | TC0377 | | | |
| hypothetical protein | | TC0445 | | | |
| hypothetical protein | | TC0467 | | | |
| hypothetical protein | | TC0469 | | | |
| hypothetical protein | | TC0497 | | | |
| hypothetical protein | | TC0573 | | | |
| helicase, putative | | TC0602 | | | |
| hypothetical protein | | TC0622 | | | |
| conserved hypothetical protein | | TC0624 | | | |
| hypothetical protein | | TC0689 | | | |
| hypothetical protein | | TC0768 | | | |
| uracil phosphoribosyltransferase | <i>upp</i> | TC0833 | | | |
| hypothetical protein | | TC0845 | | | |
| UvrD/REP helicase family protein | | TC0490 | | | |

Supporting Online Material

Supporting reference list

1. R. Gautom, T. R. Fritsche, *J. Euk. Microbiol.* **42**, 452 (1995).
2. C. Arnold, I. J. Hodgson, *PCR Methods Appl.* **1**, 39 (1991).
3. D. Frishman *et al.*, *Bioinformatics* **17**, 44 (2001).
4. D. Frishman, A. Mironov, H. Mewes, M. Gelfand, *Nucleic Acids Res.* **26**, 2941 (1998).
5. P. B. McGarvey *et al.*, *Bioinformatics* **16**, 290 (2000).
6. B. Boeckmann *et al.*, *Nucleic Acids Res.* **31**, 365 (2003).
7. S. F. Altschul, W. Gish, W. Miller, E. W. Meyers, D. J. Lipman, *J. Mol. Biol.* **215**, 403 (1990).
8. A. Bateman *et al.*, *Nucleic Acids Res.* **30**, 276 (2002).
9. R. L. Tatusov *et al.*, *Nucleic Acids Res.* **29**, 22 (2001).
10. J. G. Henikoff, S. Henikoff, *Methods Enzymol.* **266**, 88 (1996).
11. L. Lo Conte *et al.*, *Nucleic Acids Res.* **28**, 257 (2000).
12. J. Wixon, D. Kell, *Yeast* **17**, 48 (2000).
13. T. Lowe, S. Eddy, *Nucleic. Acids Res.* **25**, 955 (1997).
14. O. Emanuelsson, H. Nielsen, S. Brunak, G. von Heijne, *J. Mol. Biol.* **300**, 1005 (2000).
15. H. Bannai, Y. Tamada, O. Maruyama, K. Nakai, S. Miyano, *Bioinformatics* **18**, 298 (2002).
16. O. Emanuelsson, H. Nielsen, G. von Heijne, *Protein Sci.* **8**, 978 (1999).
17. P. Rice, I. Longden, A. Bleasby, *Trends Genet.* **16**, 276 (2000).

Supporting Online Material

18. W. Ludwig *et al.*, *Nucleic Acids Res.* **32**, 1363 (2004).
19. D. G. Higgins, *Methods Mol. Biol.* **25**, 307 (1994).
20. J. Felsenstein, *Cladistics* **5**, 164 (1989).
21. K. Strimmer, A. von Haeseler, *Mol. Biol. Evol.* **13**, 964 (1996).
22. R. S. Stephens, in *Chlamydia* R. S. Stephens, Ed. (American Society for Microbiology, Washington DC, 1999) pp. 9-27.
23. G. McClarty, in *Chlamydia* R. S. Stephens, Ed. (American Society for Microbiology, Washington DC, 1999) pp. 69-100.
24. R. S. Stephens *et al.*, *Science* **282**, 754 (1998).
25. S. Kalman *et al.*, *Nat. Genet.* **21**, 385 (1999).
26. T. D. Read *et al.*, *Nucleic Acids Res.* **28**, 1397 (2000).
27. M. Shirai *et al.*, *Nucleic Acids Res.* **28**, 2311 (2000).
28. T. D. Read *et al.*, *Nucleic Acids Res.* **31**, 2134 (2003).
29. G. Xie, C. Bonner, R. Jensen, *Genome Biology* **3**, research0051.1 (2002).
30. H. Ochman, S. Elwyn, N. A. Moran, *Proc. Natl. Acad. Sci. U S A* **96**, 12638 (1999).
31. J. G. Hacia, *Trends Genet.* **17**, 637 (2001).
32. D. F. Feng, G. Cho, R. F. Doolittle, *Proc. Natl. Acad. Sci. U S A* **94**, 13028 (1997).