Genome analysis of *Elusimicrobium minutum*, the first cultivated representative of the *Elusimicrobia* phylum (formerly Termite Group 1)

D. P. R. Herlemann[1], O. Geissinger[1], W. Ikeda-Ohtsubo[1], V. Kunin[2], H. Sun[2], A. Lapidus[2], P. Hugenholtz[2] and A. Brune[1]*

[1] *Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse, 35043 Marburg, Germany*

[2] *US DOE Joint Genome Institute, 2800 Mitchell Drive B100, Walnut Creek, CA 94598-1698, USA*

Running title:

Genome analysis of *Elusimicrobium minutum*

* Corresponding author: Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse, 35043 Marburg, Germany. Phone: +49-6421-178701, Fax: +49-6421-178709. E-mail: brune@mpi-marburg.mpg.de

Abstract

The candidate phylum Termite group 1 (TG1), is regularly encountered in termite hindguts but is present also in many other habitats. Here we report the complete genome sequence (1.64 Mbp) of *Elusimicrobium minutum* strain Pei191^{T}, the first cultured representative of the TG1 phylum. We reconstructed the metabolism of this strictly anaerobic bacterium isolated from a beetle larva gut and discuss the findings in light of physiological data. *E. minutum* has all genes required for uptake and fermentation of sugars via the Embden-Meyerhof pathway, including several hydrogenases, and an unusual peptide degradation pathway comprising transamination reactions and leading to the formation of alanine, which is excreted in substantial amounts. The presence of genes encoding lipopolysaccharide biosynthesis and the presence of a pathway for peptidoglycan formation are consistent with ultrastructural evidence of a Gram-negative cell envelope. Even though electron micrographs showed no cell appendages, the genome encodes many genes putatively involved in pilus assembly. We assigned some to a type II secretion system, but the function of 60 *pilE*-like genes remains unknown. Numerous genes with hypothetical functions, e.g., polyketide synthesis, non-ribosomal peptide synthesis, antibiotic transport, and oxygen stress protection, indicate the presence of hitherto undiscovered physiological traits. Comparative analysis of 22 concatenated single-copy marker genes corroborated the status of *Elusimicrobia* (formerly TG1) as a separate phylum in the bacterial domain, which was so far based only on 16S rRNA sequence analysis.

Introduction

At least half of the phylum-level lineages within the domain *Bacteria* do not comprise pure cultures, but are rather represented only by 16S rRNA gene sequences of environmental origin (43). The number of such "candidate phyla" is still growing, and the biology of the members of these phyla is usually completely obscure. The first sequences of the candidate phylum Termite Group 1 (TG1; 23) were obtained from the hindgut of the termite *Reticulitermes speratus*, where they represent a substantial portion of the gut microbiota (21, 41). Meanwhile, numerous sequences affiliated with this phylum have been retrieved also from habitats other than termite guts. They form several deep-branching lineages comprising sequences derived not only from intestinal tracts but also from soils, sediments, and contaminated aquifers (14, 20).

Recently, we were able to isolate strain Pei191^T, the first pure-culture representative of the TG1 phylum, from the gut of a humivorous scarab beetle larva, *Pachnoda ephippiata* (14). Based on the 16S rRNA gene sequence, strain Pei191^T is a member of the "Intestinal cluster", which consists of sequences derived from invertebrate guts and cow rumen (20) and is only distantly related to the Endomicrobia, a lineage of TG1 bacteria comprising endosymbionts of termite gut protozoa (24, 42, 54). It is an obligately anaerobic ultramicrobacterium that grows heterotrophically on glucose and produces acetate, hydrogen, ethanol, and alanine as major products (14). The species description of *Elusimicrobium minutum*, with strain Pei191^T as the type strain, and the proposal of *Elusimicrobia* as the new phylum name are published in a companion paper (14).

Here, we report the complete genome sequence of *E. minutum*, focusing on a reconstruction of the metabolism of this strictly anaerobic bacterium. The implications of these findings are

discussed in light of physiological data, and potential functions indicated by the genome annotation are compared to requirements imposed by the intestinal environment. Using the concatenated sequences of 22 single-copy marker genes of *E. minutum* and of the uncultivated "*Endomicrobium*" strain Rs-D17, an endosymbiont of termite gut flagellates (22), we also investigated the phylogenetic position of *Elusimicrobia* relative to other bacterial phyla.

Material and Methods

DNA preparation. A 400-ml culture of *Elusimicrobium minutum* strain Pei191^T grown on glucose (14) was harvested by centrifugation. Cells were resuspended in 500 µl TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), and 30 µl of 10% SDS and 3 µl of proteinase K (20 mg/ml) were added. The mixture was incubated at 37 °C for 1 h. The lysate was extracted three times with an equal volume of phenol-chloroform-isoamyl alcohol (49:49:1, by vol) using Phase Lock Gel tubes (Eppendorf). The supernatant was transferred to a fresh tube, and the DNA was precipitated with 0.6 volumes of isopropanol, washed with ice-cold 80% (vol/vol) ethanol, and air-dried. Quality and quantity were checked by agarose gel electrophoresis.

Genome sequencing, assembly, and gap closure. The genome of *E. minutum* was sequenced at the Joint Genome Institute (JGI) using a combination of 8-kb and 40-kb Sanger libraries and 454 pyrosequencing. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/. 454 pyrosequencing reads were assembled using the Newbler assembler (Roche). Large Newbler contigs were chopped into 1871 overlapping fragments of 1000 bp and entered into the assembly as pseudo-reads. The

sequences were assigned quality scores based on Newbler consensus q-scores with modifications to account for overlap redundancy and adjust inflated q-scores. A hybrid assembly of 454 and Sanger reads was performed using the PGA assembler. Possible mis-assemblies were corrected and gaps between contigs were closed by custom primer walks from sub-clones or PCR products. The error rate of the completed genome sequence of *E. minutum* is less than 1 in 50,000. The complete nucleotide sequence and annotation of *E. minutum* has been deposited at GenBank under accession number CP001055.

Annotation. Sequences were automatically annotated at the Oak Ridge National Laboratory (ORNL) according to the genome analysis pipeline described in Hauser et al. (18). All automatic annotations with functional prediction were also checked manually with the annotation platform provided by Integrated Microbial Genomes (IMG) (37). For each gene, the specific functional assignments suggested by the matches with the NCBI non-redundant database were compared to the domain-based assignments supplied by the COG/PFAM/TIGRFAM/INTERPRO databases, and if necessary corrected accordingly. When it was not possible to infer function or COG domain membership (RPS BLAST against COG PSSM with e-value > 10^{-2}), genes were annotated as predicted to be novel. For all the genes, the subcellular location of their potential gene products was determined based on the presence of transmembrane helices and signal peptides. Putative transport proteins were compared to those in the Transport Classification Database (http://www.tcdb.org). Genes were viewed graphically with Integrated Microbial Genomes. Metabolic pathways were reconstructed using MetaCyc as a reference data set (7). Detailed information about the automatic genome annotation can be obtained from the JGI IMG website (http://img.jgi.doe.gov/w/doc/about_index.html). Insertion sequences were detected with IS Finder (http://www-is.biotoul.fr/).

Phylogenetic analyses. A concatenated gene tree was created using a set of 22 conserved single-copy phylogenetic marker genes derived from the set used by Ciccarelli et al. (9). The marker genes were extracted from *E. minutum* and 279 microbial reference genomes (including "*Endomicrobium*" strain Rs-D17) in the IMG database ver. 2.50 (38), concatenated, and aligned with MUSCLE (11). The alignment and sequence-associated data (e.g., organism name) were then imported into ARB (33) and manually refined. A mask was created using the base frequency filter tool (20% minimal identity) to remove regions of ambiguous positional homology, yielding a masked alignment of 3982 amino acids, which is available on request from the authors. Several combinations of outgroups to the TG1 taxa (*E. minutum* and "*Endomicrobium*" strain Rs-D17) were selected for phylogenetic inference to establish the monophyly of the TG1 phylum and to identify any specific associations with other phyla that may exist (10). Maximum-likelihood trees were constructed from the masked datasets using RAxML ver. HPC-2.2.3 (53).

The phylogenetic relationships of the [NiFe] hydrogenase were determined using the ARB program suite (33). The sequences of *E. minutum* and *Thermoanaerobacter tengcongensis* were aligned with the sequences of the large subunit given in Vignais et al. (57). Highly variable positions (< 20% sequence similarity) were filtered from the data set, resulting in 560 unambiguously aligned amino acids, and phylogenetic distances were calculated using the Protein maximum-likelihood algorithm provided in the ARB package.

Clustered, regularly interspaced short palindromic repeats (CRISPR) arrays were identified using PILER-CR (12). Prophages or other elements targeted by CRISPRs were identified by pair-wise comparison of spacers to the rest of the genome using BLASTN (2).

Results and Discussion

Genome structure. *E. minutum* has a relatively small circular chromosome of 1,643,562 bp (Fig. 1), with an average G+C content of 39.0 mol%. No plasmids were found. The genome contains 1597 predicted genes, of which 1529 (95.7%) code for proteins, 48 (3.1%) code for RNA genes, and 20 (1.3%) are pseudogenes. Of the protein-coding genes, 1141 (74.6%) were assigned to specific domains in the COG database, and 388 (25.4%) are predicted to be novel (Table 1). The genome contains only a single rRNA operon, which is in agreement with the long doubling time of the organism (11–20 h; 14). The G+C content of the rRNA genes deviates from that of the rest of the genome, which is typical for mesophilic bacteria (40). There are 45 genes encoding tRNAs for the 20 standard amino acids; tRNA genes with anticodons for unusual amino acids were not present. The substantial asymmetry in gene density on the two DNA strands on both sides of the origin indicates the switching between leading and lagging strands typical of bacteria with a bifurcating replication mechanism (28).

The genome contains one array of clustered, regularly interspaced short palindromic repeats (CRISPR) comprising 13 repeat/spacer units, flanked by an operon containing CRISPR-associated genes; this region is characterized by a lower G+C content (Fig. 1). CRISPR elements are widespread in the genomes of almost all archaea and many bacteria and are considered one of the most ancient antiviral defense systems in the microbial world (37, 52). One of the *E. minutum* spacers had an identical match within the genome, highlighting the location of an intact 34-kb prophage. The detailed annotation of all protein-coding genes and their COG assignments is presented in the supplementary material (Table S1). We detected 63 putative insertion sequences (IS) in the genome, but most of them had only low similarities to sequences from known IS families (Table S2).

Phylogeny and taxonomy. As expected for the first cultivated representative of a candidate phylum, many genes from the *E. minutum* genome are only distantly related to homologs identified in genomes from other bacterial phyla. The recent publication of a composite genome of "*Endomicrobium*" strain Rs-D17, recovered from a homogeneous population of endosymbionts isolated from a single protist cell in a termite hindgut (22), provides a phylogenetic reference point for analysis. A comparative analysis of 22 concatenated single-copy marker genes confirmed a highly reproducible relationship between *E. minutum* and "*Endomicrobium*" strain Rs-D17 (Fig. 2), as predicted already by 16S rRNA-based phylogeny (20). The analysis also reinforced the phylum-level status proposed for the *Elusimicrobia* lineage (formerly TG1; 23) since no robust associations to other bacterial phyla were identified.

Energy metabolism. Pure cultures of *E. minutum* convert sugars to H_2, CO_2, ethanol, and acetate as major fermentation products (14). A full reconstruction of the energy metabolism by manual genome annotation (Table S1) revealed that *E. minutum* uses a set of pathways typical of many strictly fermentative organisms (Fig. 3, blue box). Hexoses are imported via several phosphotransferase systems (PTS) or permeases. PTS systems for fructose, glucose, and *N*-acetylglucosamine, three of the five substrates supporting growth of *Elusimicrobium minutum* (14), were present. The resulting sugar phosphates are converted to fructose 6-phosphate and degraded to pyruvate via the classical Embden-Meyerhof pathway (EMP); 2-dehydro-3-deoxy-phosphogluconate aldolase, the key enzyme of the Entner-Doudoroff pathway, is absent.

Pyruvate is further oxidized to acetyl-CoA by pyruvate:ferredoxin oxidoreductase (PFOR). The acetyl-CoA is converted to acetate by phosphotransacetylase and acetate kinase. There are

two enzymes potentially involved in hydrogen formation: a membrane-bound [NiFe] hydrogenase and a soluble [FeFe] hydrogenase. The [NiFe] hydrogenase operon comprises the genes encoding the typical subunits; the large subunit contains the two conserved CxxC motifs found in complex-I-related [NiFe] hydrogenases, and the small subunit has the typical $CxxCx_nGxCxxxGx_mGCPP$ (*E. minutum*: $n = 61$, $m = 24$) motif (1). There is also an operon of five genes with high similarity to maturation proteins required for the synthesis of the catalytic metallocluster of [NiFe] hydrogenases (25). Comparative analysis of the genes coding for the large subunit (*echD*) revealed that the enzyme belongs to group IV [NiFe] hydrogenases (Fig. 4). Hydrogenases of this group function as redox-driven ion pumps, coupling the reduction of protons by ferredoxin with the generation of a proton-motive force (44, 50), suggesting that this type of energy conservation may be present also in *E. minutum* (Fig. 3).

The second hydrogenase shows the typical structure and sequence motifs of a cytosolic NADH-dependent [FeFe] hydrogenase (Fig. 5; 51), including the typical H-cluster motif (57). Since the reduction of NADH to hydrogen is thermodynamically favorable only at low hydrogen partial pressure (46), this enzyme is probably not involved in hydrogen formation in batch culture, where hydrogen accumulates to substantial concentrations (14). Here, the stoichiometry of less than 2 H_2 per glucose indicates that H_2 is formed only via the ferredoxin-driven [NiFe] hydrogenase; the NADH formed during glycolysis is regenerated by the reduction of acetyl-CoA to ethanol (Fig. 3).

Although it remains to be shown whether *E. minutum* shifts from ethanol to H_2 formation at low hydrogen partial pressures to increase its energy yield, the presence of the second hydrogenase may be an adaptation to the low hydrogen partial pressures in its habitat. Hydrogen concentrations in the hindgut of *Pachnoda ephippiata* were typically below the

176 detection limit of the hydrogen microsensor (60–70 Pa) (30), which is close to the threshold
177 concentration (< 10 Pa) permitting H_2 formation from NADH (46).

178 **Anabolism.** Although the presence of fructose 1,6-bisphosphatase indicates the possibility for
179 gluconeogenesis via the EMP, *E. minutum* requires a hexose for growth (14). The absence of
180 genes coding for 2-oxoglutarate dehydrogenase, succinate dehydrogenase, and succinyl-CoA
181 synthetase is typical for strict anaerobes and documents that *E. minutum* does not possess a
182 complete tricarboxylic acid (TCA) cycle. The reductive branch of the incomplete TCA cycle
183 is initiated by phosphoenol pyruvate (PEP) carboxykinase and allows the interconversion of
184 oxaloacetate, malate, and fumarate. The oxidative branch of the pathway starts with citrate
185 synthase and allows the formation of 2-oxoglutarate. Typical for anaerobic microorganisms,
186 the citrate synthase of *E. minutum* belongs to the *Re*-type (32). The products of the incomplete
187 TCA cycle are precursors of several amino acids. The biosynthetic pathways for the formation
188 of glutamate, glutamine, proline, aspartate, lysine, threonine, and cystathione are present. Also
189 the pathways for the formation of alanine, cysteine, glycine, histidine, and serine, starting with
190 intermediates of the EMP, are almost fully represented by the corresponding genes (Table S1,
191 Fig. S1). However, the genes for the synthesis of other proteinogenic amino acids (arginine,
192 asparagine, isoleucine, leucine, methionine, phenylalanine, tyrosine, tryptophan, and valine)
193 are lacking, which would explain why *E. minutum* requires small amounts of yeast extract in
194 the medium (14).

195 The genome of *E. minutum* does not possess an oxidative pentose phosphate pathway, which
196 is typically involved in the regeneration of NADPH. This important coenzyme is probably
197 regenerated by the alternative route of pyruvate formation from PEP (formation of
198 oxaloacetate by PEP carboxykinase, NADH-dependent reduction of oxaloacetate by malate

dehydrogenase, and $NADP^+$-dependent oxidative decarboxylation of malate by malic enzyme; Fig. 3, green box), as proposed for *Corynebacterium glutamicum* (45). $NADP^+$ is required for the *de novo* biosynthesis of nucleic acids. The presence of the genes required for the non-oxidative pentose phosphate pathway (transaldolase and transketolase) allows the reconstruction of the pathways for purine and pyrimidine nucleotide biosynthesis almost completely (Table S1) and also explains the catabolism of ribose via the EMP (14).

Also the genes coding for the synthesis of lipopolysaccharides and peptidoglycan are well represented (Table S1). This is in agreement with the results of electron microscopy, which showed that *E. minutum* possesses the typical cell envelope architecture of gram-negative bacteria (14). The pathways for vitamin synthesis are absent or at most rudimentary (Table S1), which would be another reason why the bacterium requires small amounts of yeast extract in the growth medium (14).

A large open reading frame (3008 amino acids) was assigned to the polyketide synthase gene family. Interestingly, the polyketide synthase gene shows a relatively high G+C content (46%; Fig. 1), suggesting an origin from horizontal gene transfer. The presence of a polyketide synthase and a putative non-ribosomal peptide synthetase (1284 amino acids) is rather unusual for anaerobic bacteria (48). The function of the two enzymes remains to be investigated.

Peptide degradation. *E. minutum* has a particular pathway for catabolic utilization of amino acids, which may lead to additional energy conservation (Fig. 3, yellow box). The pathway comprises the transfer of amino groups from peptide-derived amino acids to pyruvate via a homolog of a non-specific aminotransferase (58), resulting in alanine formation. The 2-oxoacids produced by the transamination can be oxidatively decarboxylated to the corresponding acyl-CoA esters, probably by the gene products annotated as 2-

222 oxoacid:ferredoxin oxidoreductases. Substrate-level phosphorylation is accomplished via an
223 acyl-CoA synthetase (ADP-forming), resulting in the formation of ATP and the corresponding
224 fatty acid. The genome also encodes proton-dependent oligopeptide transporters, ABC-type
225 transport systems for peptides, and numerous proteolytic and peptolytic enzymes, some of
226 which have typical signal peptides, indicating extracellular proteinase activity (Table S1).

227 A comparable peptide utilization pathway is also present in *Pyrococcus furiosus* (34, 36, 19).
228 Besides the PFOR, a homodimer that typically oxidizes only pyruvate and a few other
229 oxoacids, e.g., 2-oxoglutarate (39), *E. minutum* also possesses a homologue of a
230 heterotetrameric 2-oxoisovalerate:ferredoxin oxidoreductase (VFOR) with a broad substrate
231 specificity, especially for branched-chain 2-oxoacids (19). In addition, a putative two-subunit
232 indolepyruvate:ferredoxin oxidoreductase (IFOR) is present. The large number of different
233 acyl-CoA esters resulting from the oxidative decarboxylation of various amino acids seem to
234 be converted to their corresponding acids by a single ADP-dependent acetyl-CoA synthetase;
235 the homolog in *P. furiosus* is reportedly rather unspecific and processes also branched-chain
236 derivatives (35).

237 The operation of this peptide utilization pathway in *E. minutum* is supported by the
238 observation that most proteinogenic (and even some non-proteinogenic) amino acids are
239 converted to their corresponding oxidative decarboxylation products during growth on
240 glucose. Further evidence was provided by ^{13}C-labeling, which demonstrated that the carbon
241 skeleton of the putative transamination product, alanine, is derived from glucose (14). In
242 principle, *E. minutum* also possesses the capacity for the net amination of pyruvate to alanine
243 (Fig. 3, green box), which has been proposed to function as an additional electron sink in *P.*
244 *furiosus* (26).

A combination of glucose fermentation with the oxidative decarboxylation of an amino acid can increase the free-energy change of the metabolism, as exemplified by the case of valine ($\Delta G^{\circ\prime}$ values calculated according to 56; data for isobutyrate from 60).

$$\text{Glucose} + H_2O \rightarrow \text{Ethanol} + \text{Acetate}^- + H^+ + 2\ H_2 + 2\ CO_2$$

$$\Delta G^{\circ\prime} = -225\ \text{kJ mol}^{-1}$$

$$\text{Glucose} + \text{Valine} + H_2O \rightarrow \text{Ethanol} + \text{Alanine} + \text{Isobutyrate}^- + H^+ + 2\ H_2 + 2\ CO_2$$

$$\Delta G^{\circ\prime} = -245\ \text{kJ mol}^{-1}$$

However, since substrate-level phosphorylation in the peptide utilization pathway occurs at the expense of ATP generation from carbohydrates (i.e., pyruvate oxidation), the co-fermentation of amino acids becomes energetically productive only if this opens up the possibility for additional energy conservation. Interestingly, *E. minutum* possesses a Na^+/alanine symporter, which could couple export of the accumulating alanine with the generation of an electrochemical sodium gradient. Together with the H^+/Na^+ antiporter encoded in the genome, the sodium gradient can be converted into a proton-motive force, which would either drive the generation of additional ATP via ATP synthase or avoid the hydrolysis of ATP necessitated by the dissipation of the proton motive force in other transport processes (27), such as the proton-dependent import of amino acids or oligopeptides (Fig. 3).

Secretion. A large number of proteins (40%) encoded in the genome of *E. minutum* contain a signal peptide, indicating their export from the cell (Table S1). These putatively exported proteins comprise almost all of the proteins in COG category U (intracellular trafficking, secretion, and vesicular transport) and more than half of the predicted novel proteins.

266 The results of the manual annotation revealed that *E. minutum* possesses a variant of the
267 general secretion pathway (GSP). The Sec translocon (encoded by *secADFYEG*) lacks a SecB
268 subunit; SecB is probably replaced by one of the more general chaperones (DnaJ or DnaK)
269 (59). There are numerous genes encoding the typical type II secretion system (T2SS), but
270 several essential components of the machinery are missing in the annotation (Table 2). Most
271 of these components are poorly conserved (encoded by *gspABCNS*; 8) and might have simply
272 escaped detection. Some of the missing elements might have been annotated as elements of
273 type IV pili (T4P), which are related structures with numerous similar components (55). T4P
274 are probably absent in *E. minutum* because the PilMNOP components, which are essential for
275 functional pili (6, 5), are lacking and no pilus-like structures are seen in ultra-thin sections of
276 *E. minutum* (14). The absence of *gspL* and *gspM* in *E. minutum* is more critical because the
277 encoded proteins have no homologs in T4P and are usually indicative of a T2SS. However,
278 also the T2SS of *Acinetobacter calcoaceticus* and *Bdellovibrio bacteriovorus* lack the GspLM
279 components (8), and the pathogen *Francisella tularensis* ssp. *novicidia* uses a T2SS even
280 lacking the GspLMC components to export chitinases, proteinases, and β-glucosidases (16).
281 The presence of two ATPases in *E. minutum*, which are typical for T4P, does not necessarily
282 argue against a T2SS; the T2SS of *Aeromonas hydrophila* also has two ATPases, and they are
283 thought to increase the efficiency of the secretory process (47).

284 The number of *pilE*-like genes in the genome of *E. minutum* is much higher than the number
285 of all other components of the T2SS. Sixty *pilE*-like genes (members of COG4968) are spread
286 over the genome (Table S1). It has been shown that variable gene copies of *pilE* play a role in
287 immune evasion because they lead to antigenic variations in the pilins of the *Neisseria*
288 *gonorrhoeae* T4P (17). Although the pilins of T2SS reach through the periplasm and the outer
289 membrane, their importance as an antigen is not clear. It is also not clear whether antigenic

290 variation is important for the colonization of the insect gut. Although insects lack an adaptive
291 immune system with antigen-specific antibodies, it has been reported that a response to an
292 immune challenge can be enhanced by previous exposure (29).

293 Comparative analysis revealed that only the encoded N-terminal methylase domain is
294 conserved between the *E. minutum pilE*-like genes and *pilE* genes from other organisms. This
295 effectively reduces the comparable region to only ~50 amino acids and compromises
296 phylogenetic inference. However, it appears that most of the *E. minutum* copies (57/60) form a
297 monophyletic group, which suggests a large lineage-specific expansion of this gene family, or
298 at least an expansion of the gene domain (data not shown). Indeed, the numerous copies of the
299 *pilE*-like genes of the *E. minutum* genome alone increase the size of the COG4968 family in
300 the IMG database by almost 10% because there are only 682 representatives present in 1087
301 other microbial genomes (38). Since *E. minutum* lacks observable pili and many of the *pilE*-
302 like genes appear in operons of diverse function, we speculate that this gene family is
303 involved in some other aspect(s) of endogenous regulation, perhaps not related to pili or
304 secretion at all, and have undergone a lineage-specific expansion in response to environmental
305 selection.

306 In addition to the type-II-like secretion system, the genome contains numerous ABC
307 transporters (Table S1). Together with outer membrane efflux proteins (OMP, MFP), they
308 may constitute type I secretion systems with various functions.

309 **Oxygen stress**. In agreement with the obligately anaerobic nature of *E. minutum*, the genome
310 contains no cytochrome genes and no pathways for the biosynthesis of quinones,
311 corroborating the absence of any respiratory electron transport chains. However, *E. minutum*
312 has a six-gene "oxygen stress protection" cluster consisting of ruberythrin (*rbr*), superoxide

313 reductase (*sor*), rubredoxin:oxygen oxidoreductase (*roo*), and rubredoxin (*rub*) (Fig. 6). The
314 *roo* gene of *E. minutum* has similarity to the corresponding genes of *Desulfovibrio gigas* and
315 *Moorella thermoacetica*, which have been shown to reduce molecular oxygen by reduced
316 rubredoxin (15, 49). The presence of an oxygen-reducing system may explain the ability of *E.*
317 *minutum* to retard the diffusive influx of oxygen into deep-agar tubes (14) and may play an
318 important role in survival in the intestinal tract of insects, a habitat constantly exposed to the
319 influx of oxygen (4, 31).

320 **Ecological considerations.** The genome of *E. minutum* revealed several adaptations of the
321 bacterium to its environment. As a member of the "Intestinal Cluster", *E. minutum* is probably
322 a resident inhabitant of the gut of *P. ephippiata*, which is thought to assist in digestion (31). *P.*
323 *ephippiata* feeds on a humus-rich diet, and its gut contains high concentrations of glucose,
324 peptides, and amino acids (3). With its putative capacity for proteinase secretion, the potential
325 to maximize ATP yield in a coupled fermentation of sugars and amino acids, and the ability to
326 cope with the exposure to molecular oxygen and reactive oxygen species, *E. minutum* appears
327 to be well adapted to this habitat. As with other intestinal bacteria, it requires complex
328 nutritive supplements and lacks pathways for the synthesis of most vitamins and certain amino
329 acids. Although the genome of *E. minutum* is relatively small, there are no indications for an
330 obligate association with its host. Genes encoding glycosyl hydrolases involved in the
331 degradation of polysaccharides (other than glycogen) were not identified, indicating that *E.*
332 *minutum* does not participate in the digestion of plant fibers.

Acknowledgements

We thank members of the JGI production sequencing, quality assurance and genome biology programs and the IMG team for their assistance in genome sequencing, assembly, annotation, and loading of the genome into IMG. These activities were supported by the 2007 Community Sequencing Program. D.H. and W.I.-O. were supported by stipends of the International Max Planck Research School for Molecular, Cellular, and Environmental Microbiology and the Deutscher Akademischer Austauschdienst. We thank Henning Seedorf, Reiner Hedderich, and Rolf Thauer (Marburg) for helpful advice. This work was financed in part by a grant of the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center Transregio 1 (SFB-TR1) and by the Max Planck Society. Other parts of this work were performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

References

1. **Albracht, S. P. J.** 1994. Nickel hydrogenases: in search of the active site. *Biochim. Biophys. Acta* **1188:** 167–204.

2. **Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

3. **Andert, J., O. Geissinger, and A. Brune.** 2008. Peptidic soil components are a major dietary resource for the humivorous larvae of *Pachnoda* spp. (Coleoptera: Scarabaeidae). *J. Ins. Physiol.* **54**: 105–113.

4. **Brune, A., P. Frenzel, and H. Cypionka.** 2000. Life at the oxic-anoxic interface: microbial activities and adaptations. *FEMS Microbiol. Rev.* **24:** 691–710.

5. **Carbonnelle, E., S. Hélaine, L. Prouvensier, X. Nassif, and V. Pelicic.** 2005. Type IV pilus biogenesis in *Neisseria meningitidis*: PilW is involved in a step occurring after pilus assembly, essential for fibre stability and function. *Mol. Microbiol*. **55:** 54–64.

6. **Carbonnelle, E., S. Hélaine, X. Nassif, and V. Pelicic.** 2006. A systematic genetic analysis in *Neisseria meningitidis* defines the Pil proteins required for assembly, functionality, stabilization and export of type IV pili. *Mol. Microbiol.* **61:** 1510–1522.

7. **Caspi, R., H. Foerster, C. A. Fulcher, R. Hopkinson, J. Ingraham, P. Kaipa, M. Krummenacker, S. Paley, J. Pick, S. Y. Rhee, C. Tissier, P. Zhang, and P. D. Karp.** 2006. MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* **34:** D511–516.

8. **Cianciotto, N. P.** 2005. Type II secretion: a protein secretion system for all seasons. *Trends Microbiol.* **13:** 281–288.

9. **Ciccarelli, F. D., T. Doerks, C. von Mering, C. J. Creevey, B. Snel, and P. Bork.** 2006. Toward automatic reconstruction of a highly resolved tree of life. *Science* **311:** 1283–1287.

10. **Dalevi, D., P. Hugenholtz, and L. L. Blackall.** 2001. A multiple-outgroup approach to resolving division-level phylogenetic relationships using 16S rDNA data. *Int. J. Syst. Evol. Microbiol.* **51:** 385–391.

11. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32:** 1792–1797.

12. **Edgar, R. C.** 2007. PILER-CR: Fast and accurate identification of CRISPR repeats. *Bioinformatics* **8:** 18

13. **Filloux, A.** 2004. The underlying mechanism of type II protein secretion. Biochim. *Biophys. Acta* **1694:** 163–179.

14. **Geissinger, O., D. P. R. Herlemann, U. G. Maier, E. Mörschel, and A. Brune.** *Elusimicrobium minutum* gen. nov., sp. nov., the first isolate of the Termite Group I phylum (submitted).

15. **Gomes, C. M., G. Silva, S. Oliveira, J. LeGall, M.-Y. Liu, A. V. Xavier, C. Rodrigues-Pousada, and M. Teixeira.** 1997. Studies on the redox centers of the terminal oxidase from *Desulfovibrio gigas* and evidence for its interaction with rubredoxin. *J. Biol. Chem.* **272:** 22502–22508.

16. **Hager, A. J., D. L. Bolton, M. R. Pelletier, M. J. Brittnacher, L. A. Gallagher, R. Kaul, J. S. Skerrett, I. S. Miller, and T. Gulna.** 2006. Type IV pili-mediated secretion modulates *Francisella* virulence. *Mol. Microbiol.* **62:** 227–237.

17. **Häggblom, P., E. Segal, E. Billyard, and M. So.** 1985. Intragenic recombination leads to pilus antigenic variation in *Neisseria gonorrhoeae*. *Nature* **315:** 156–158.

18. **Hauser, L., F. Larimer, M. Land, M. Shah, and E. Uberbacher.** 2004. Analysis and annotation of microbial genome sequences, p. 225–238. *In* Setlow J. K. (ed), *Genetic Engineering*, Vol. 26. Kluwer Academic, New York, NY.

19. **Heider, J., X. Mai, and M. W. Adams.** 1996. Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A-dependent enzyme involved in peptide fermentation by hyperthermophilic archaea. *J. Bacteriol.* **178:** 780–787.

20. **Herlemann, D. P. R., O. Geissinger, and A. Brune.** 2007. The Termite Group I phylum is highly diverse and widespread in the environment. *Appl. Environ. Microbiol.* **73**: 6682–6685.

21. **Hongoh, Y., M. Ohkuma, and T. Kudo.** 2003. Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera, Rhinotermitidae). *FEMS Microbiol Ecol.* **44**: 231–242.

22. **Hongoh, Y., V. K. Sharma, T. Prakash, S. Noda, T. D. Taylor, T. Kudo, Y. Sakaki, A. Toyoda, M. Hattori, and M. Ohkuma.** 2008. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc. Natl. Acad. Sci. U.S.A.* **105:** 5555–5560.

410 **23. Hugenholtz, P., B. M. Goebel, and N. R. Pace.** 1998. Impact of culture-independent
411 studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol.* **180**:
412 4765–4774.

413 **24. Ikeda-Ohtsubo, W., M. Desai, U. Stingl, and A. Brune.** 2007. Phylogenetic diversity
414 of "Endomicrobia" and their specific affiliation with termite gut flagellates.
415 *Microbiology* **153:** 3458–3465.

416 **25. Jacobi A., R. Rossmann, and A. Böck.** 1992. The *hyp* operon gene products are
417 required for the maturation of catalytically active hydrogenase isoenzymes in
418 *Escherichia coli*. *Arch. Microbiol.* **158**: 444–451.

419 **26. Kengen, S. W. M., and A. J. M. Stams.** 1994. Formation of l-alanine as a reduced end
420 product in carbohydrate fermentation by the hyperthermophilic archaeon *Pyrococcus*
421 *furiosus*. *Arch. Microbiol.* **161:** 168–175.

422 **27. Konings W.N.** 2006. Microbial transport: adaptions to natural environments. *Antonie v.*
423 *Leeuwenhoek* **90:** 325–42.

424 **28. Koonin E.V. and Y.I. Wolf.** 2008. Genomics of bacteria and archaea: the emerging
425 dynamic view of the prokaryotic world. Nucleic Acids Res. **36**: 6688–719.

426 **29. Kurtz, J.** 2004. Memory in the innate and adaptive immune systems. *Microbes and*
427 *Infection* **6:** 1410–1417.

428 **30. Lemke T., T. van Alen, J. H. P. Hackstein, and A. Brune.** 2001. Cross-epithelial
429 hydrogen transfer from the midgut compartment drives methanogenesis in the hindgut
430 of cockroaches. *Appl. Environ. Microbiol.* **67**: 4657–4661.Tables

31. Lemke, T., U. Stingl, M. Egert, M. W. Friedrich, and A. Brune. (2003) Physicochemical conditions and microbial activities in the highly alkaline gut of the humus-feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Appl. Environ. Microbiol.* **69**: 6650–6658.

32. Li, F., C. H. Hagemeier, H. Seedorf, G. Gottschalk, and R. K. Thauer. 2007. *Re*-citrate synthase from *Clostridium kluyveri* is phylogenetically related to homocitrate synthase and isopropylmalate synthase rather than to *Si*-citrate synthase. *J. Bacteriol.* **189:** 4299–304.

33. Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüßmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K.-H. Schleifer. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**: 1363–1371.

34. Mai, X., and M. W. W. Adams. 1994. Indolepyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus*. A new enzyme involved in peptide fermentation. *J. Biol. Chem.* **269:** 16726–32.

35. Mai, X., and M. W. W. Adams. 1996a. Characterization of a fourth type of 2-keto acid-oxidizing enzyme from a hyperthermophilic archaeon: 2-ketoglutarate ferredoxin oxidoreductase from *Thermococcus litoralis*. *J. Bacteriol.* **178**: 5890–5896.

451 **36. Mai, X., and M. W. W. Adams.** 1996b. Purification and characterization of two
452 reversible and ADP-dependent acetyl coenzyme A synthetases from the
453 hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Bacteriol.* **178:** 5897–5903.

454 **37. Makarova, K. S., N. V. Grishin, S. A. Shabalina, J. I. Wolf, and V. K. Eugene.**
455 2006. A putative RNA-interference-based immune system in prokaryotes:
456 computational analysis of the predicted enzymatic machinery, functional analogies with
457 eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct.* **1**: **7**.

458 **38. Markowitz, V. M., E. Szeto, K. Palaniappan, Y. Grechkin, K. Chu, I. M. Chen, I.**
459 **Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, N. N. Ivanova, and N. C.**
460 **Kyrpides.** 2008. The integrated microbial genomes (IMG) system in 2007: data content
461 and analysis tool extensions. *Nucleic Acids Res.* **36:** D528–D533.

462 **39. Moulis, J. M., V. Davasse, J. Meyer, and J. Gaillard.** 1996. Molecular mechanism of
463 pyruvate-ferredoxin oxidoreductases based on data obtained with the *Clostridium*
464 *pasteurianum* enzyme. *FEBS Lett.* **380:** 287–290.

465 **40. Muto, A., and S. Osawa.** 1987. The guanine and cytosine content of genomic DNA and
466 bacterial evolution. *Proc. Natl. Acad. Sci. U.S.A.* **84:** 166–9.

467 **41. Ohkuma, M., and T. Kudo.** 1996. Phylogenetic diversity of the intestinal bacterial
468 community in the termite *Reticulitermes speratus*. *Appl Environ Microbiol.* **62**: 461–
469 468.

470 **42. Ohkuma, M., T. Sato, S. Noda, S. Ui, T. Kudo, and Y. Hongoh.** 2007. The candidate
471 phylum 'Termite Group 1' of bacteria: phylogenetic diversity, distribution, and

endosymbiont members of various gut flagellated protists. *FEMS Microbiol. Ecol.* **60:** 467–476.

43. Rappe, M. S., and S. J. Giovannoni. 2003. The uncultured microbial majority. *Annu. Rev. Microbiol.* **57:** 369–394.

44. Sapra, R., K. Bagramyan, and M. W. Adams. 2003. A simple energy-conserving system: proton reduction coupled to proton translocation. *Proc. Natl. Acad. Sci. U.S.A.* **100:** 7545–50.

45. Sauer, U., and B. J. Eikmanns. 2005. The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol. Rev.* **29:** 765–794.

46. Schink, B. 1997. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol. *Mol. Biol. Rev.* **61:** 262–280.

47. Schoenhofen, I. C., G. Li, T. G Strozen, and S. P. Howard. 2005. Purification and characterization of the N-terminal domain of ExeA: a novel ATPase involved in the type II secretion pathway of *Aeromonas hydrophila*. *J. Bacteriol.* **187:** 6370–63708.

48. Seedorf, H., W. F. Fricke, B. Veith, H. Brüggemann, H. Liesegang, A. Strittmatter, M. Miethke, , Buckel, W., J. Hinderberger, F. Li, C. Hagemeier, R. K. Thauer, and G. Gottschalk. 2008. The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. *Proc. Natl. Acad. Sci. U.S.A.* **105:** 2128–2133.

49. Silaghi-Dumitrescu, R., E. D. Coulter,A. Das, L. G. Ljungdahl, G. N. Jameson, B. H. Huynh, and D.M. Kurtz, Jr.. 2003. A flavodiiron protein and high molecular

493 weight rubredoxin from *Moorella thermoacetic*a with nitric oxide reductase activity.
494 *Biochemistry* **42**: 2806–2815.

495 **50. Soboh, B., D. Linder, and R. Hedderich.** 2002. Purification and catalytic properties of
496 a CO-oxidizing:H_2-evolving enzyme complex from *Carboxydothermus*
497 *hydrogenoformans*. *Eur. J. Biochem.* **269:** 5712–21.

498 **51. Soboh, B., D. Linder, and R. Hedderich.** 2004. A multisubunit membrane-bound
499 [NiFe] hydrogenase and an NADH-dependent Fe-only hydrogenase in the fermenting
500 bacterium *Thermoanaerobacter tengcongensis*. *Microbiology* **150:** 2451–2463.

501 **52. Sorek, R., Y. Zhu, C. J. Creevey, P. M. Francino, P. Bork, and E. M. Rubin.** 2007.
502 Genome-wide experimental determination of barriers to horizontal gene transfer.
503 *Science* **318**: 1449–1452.

504 **53. Stamatakis, A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic
505 analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–90.

506 **54. Stingl, U., R. Radek, H. Yang, and A. Brune.** 2005. '*Endomicrobia*': Cytoplasmic
507 symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl.*
508 *Environ. Microbiol.* **71**: 1473–1479.

509 **55. Strom, M. S., D. Nunn, and S. Lory.** 1991. Multiple roles of the pilus biogenesis
510 protein PilD: Involvement of the PilD in excretion of enzymes from *Pseudomonas*
511 *aeruginosa*. *J. Bacteriol.* **173**: 1175–1180.

512 **56. Thauer, R. K., K. Jungermann, and K. Decker.** 1977. Energy conservation in
513 chemotrophic anaerobic bacteria. *Bacteriol. Rev.* **41:** 100–180.

514 **57. Vignais, P. M., B. Billoud, and J. Meyer.** 2001. Classification and phylogeny of
515 hydrogenases. *FEMS Microbiol. Rev.* **25:** 455–501.

516 **58. Ward, D. E., S. W. Kengen, J. van der Oost, and W. M. de Vos.** 2000. Purification
517 and characterization of the alanine aminotransferase from the hyperthermophilic
518 archaeon *Pyrococcus furiosus* and its role in alanine production. *J. Bacteriol.* **182:**
519 2559–2566.

520 **59. Wild, J., E. Altman, T. Yura, and C. A. Gross.** 1992. DnaK and DnaJ heat shock
521 proteins participate in protein export in *Escherichia coli*. *Genes Dev.* **6:** 1165–1172.

522 **60. Wu, W.-M., M. Jain, K. Hickey, and J. G. Zeikus.** 1996. Perturbation of syntrophic
523 isobutyrate and butyrate degradation with formate and hydrogen. *Biotechnol. Bioeng.*
524 **52**: 404–411.

Tables

525 TABLE 1. Summary of the functional assignment, according to COG domain, of the 1529
526 protein-coding genes in the *Elusimicrobium minutum* genome. Details are shown in the
527 supplementary material (Table S1).

COG group	Number of genes [a]	Gene frequency (%)	COG function definition
C	67	4	Energy production and conversion
D	19	1	Cell cycle control, cell division, chromosome partitioning
E	83	5	Amino acid transport and metabolism
F	53	3	Nucleotide transport and metabolism
G	73	5	Carbohydrate transport and metabolism
H	42	3	Coenzyme transport and metabolism
I	37	2	Lipid transport and metabolism
J	117	8	Translation, ribosomal structure, and biogenesis
K	46	3	Transcription
L	76	5	Replication, recombination, and repair
M	109	7	Cell wall/membrane/envelope biogenesis
N	84	5	Cell motility
O	45	3	Posttranslational modification, protein turnover, chaperones
P	24	2	Inorganic ion transport and metabolism
Q	10	1	Secondary metabolites biosynthesis, transport, and catabolism
R	123	8	General function prediction only
S	72	5	Function unknown
T	32	2	Signal transduction mechanisms
U	114	7	Intracellular trafficking, secretion, and vesicular transport
V	19	1	Defense mechanisms
–	388	25	Unassigned (predicted to be novel)

528 [a] A number of genes belong to more than one category

TABLE 2. Comparison of the components of the type II secretion system (*gsp* genes) and type IV pili (*pil* genes) present in *Aeromonas hydrophila* and *Francisella tularensis* ssp. *novicida* with those of *Elusimicrobium minutum*. The information is based on COG assignment and was collected from the IMG platform. Homologous structures present in both systems are given in the same row. Bold letters indicate typical components of the respective system; nomenclature follows that of Filloux (13).

COG	Function	*gsp*	*pil*	*Aeromonas hydrophila* [a,b]	*Francisella tularensis* [b,c]	*Elusimicrobium minutum*
4796	Secretin	**D**	Q	+	+	+
2804	Fimbrial assembly	**E**	B	+	+	+
1459	Fimbrial assembly	**F**	C, Y1	+	+	+
4969	Pilin	**G**	A	+	+	+
1989	Prepilin kinase	**O**	**D**	+	+	+
3168	Stabilizing lipoprotein	**S**	**P**	+	+	–
4726	Pilin-like	**K**	X	+	–	–
2165	Minor pilin	**H, I, J**		+	–	+
3149	Membrane location	**M**		+	–	–
3031	Unknown	**C**		+	–	–
3297	Function unknown	**L**		+	–	–
3267	Unknown	A		+	–	–
3063	Fimbrial assembly		**F**	+	–	+
4972	Fimbrial biogenesis		**M**	+	–	+

3156	Fimbrial assembly	**N**	+	–	–
3176	Fimbrial assembly	O	+	–	–
2805	Twitching motility	T	+	+	+
4968	Pilin-like	E	+	+	+
4966	Pilin-like	W	+	+	–
4970	Pilin-like	U	+	+	–
4967	Pilin-like	V	+	–	–
5008	Twitching motility	U	+	–	–
642	Two-component system	S	+	+	+
745	Chemosensory	H, G	+	+	+
835	Chemosensory	I	+	–	–
840	Chemosensory	J	+	–	+

[a] Organism possesses type IV pili (13)

[b] Organism possesses a type II secretion system (13, 16)

[c] The type II secretion system is incomplete and pili-like fibers were not detected (16).

Figure legends

FIG. 1. Genomic organization of the *Elusimicrobium minutum* chromosome. The two outermost rings show the genes encoded on the forward and reverse strand (scale in mega base pairs). The third ring depicts the location of tRNA genes. The fourth ring shows the G+C content and the innermost ring the GC skew. The polyketide synthase (PKS) and rRNA operons have a relatively high G+C content; a prophage and several predicted novel genes have a relatively low G+C content. GC skew was used to identify the origin of replication (Ori).

FIG. 2. An unrooted maximum-likelihood tree of 280 bacterial genomes, including the two sequenced representatives of the phylum *Elusimicrobia*, representing the regions of the bacterial domain currently mapped by genome sequences. The tree is based on a concatenated alignment of 22 single-copy genes. Reproducibly monophyletic groups of taxa (>98% bootstrap values, except for the *Deltaproteobacteria*; 82%) are grouped into wedges for clarity. The apparent relationship between *Elusimicrobia* and the *Synergistetes* is not stable.

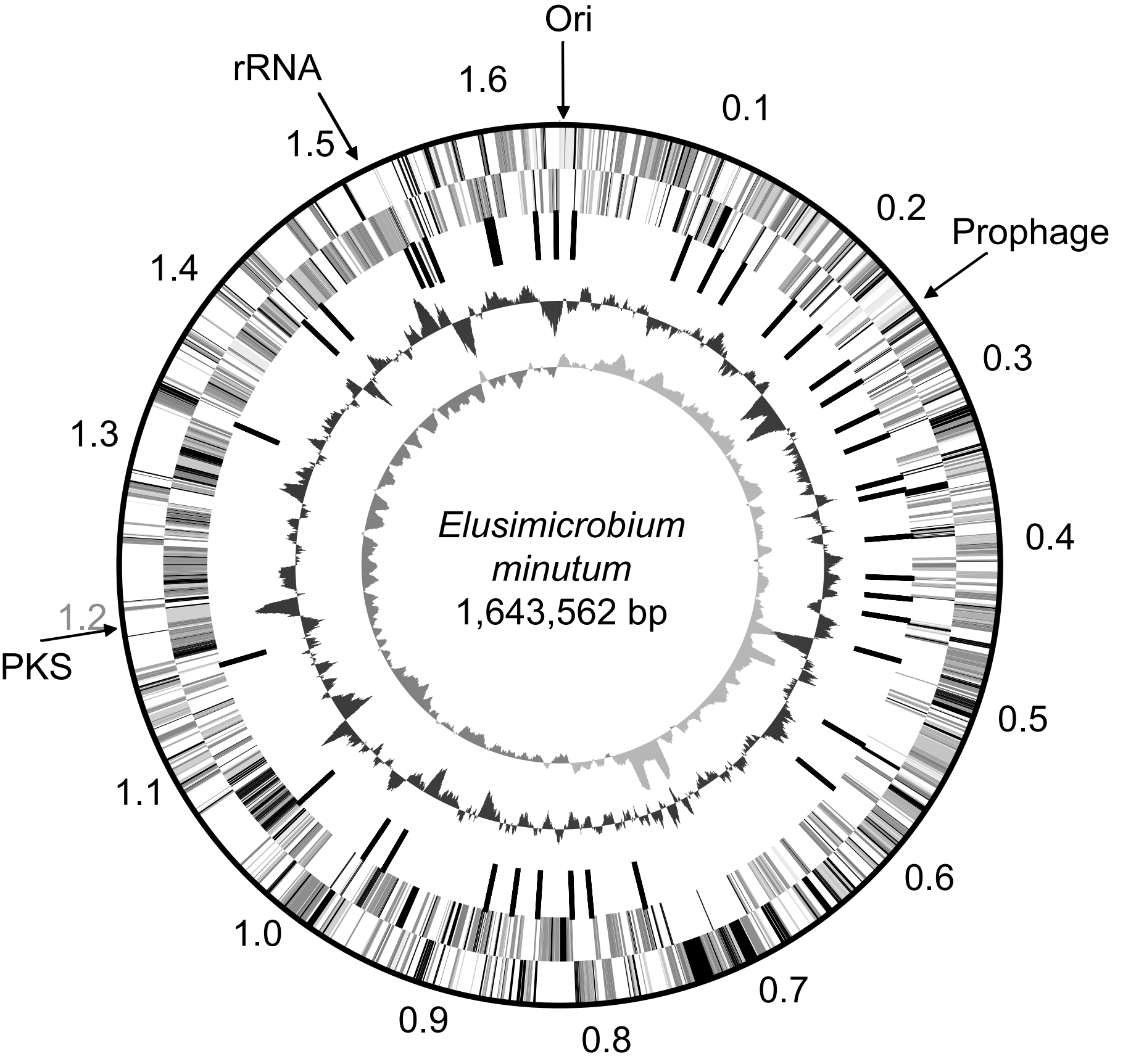
FIG. 3. Schematic overview of the energy metabolism in *Elusimicrobium minutum*. Sugars are degraded via the Embden-Meyerhof pathway and pyruvate-ferredoxin oxidoreductase (PFOR) (blue box). NADH is recycled by reduction of acetyl-CoA to ethanol or, at low hydrogen partial pressure, by the cytoplasmic [FeFe] hydrogenase. Reduced ferredoxin is regenerated by the membrane-bound [NiFe] hydrogenase. Amino acids are metabolized by transamination with pyruvate and subsequently oxidatively decarboxylated to the corresponding acids by several homologs of PFOR (yellow box). Alanine can be generated not only by transamination but also by reductive amination of

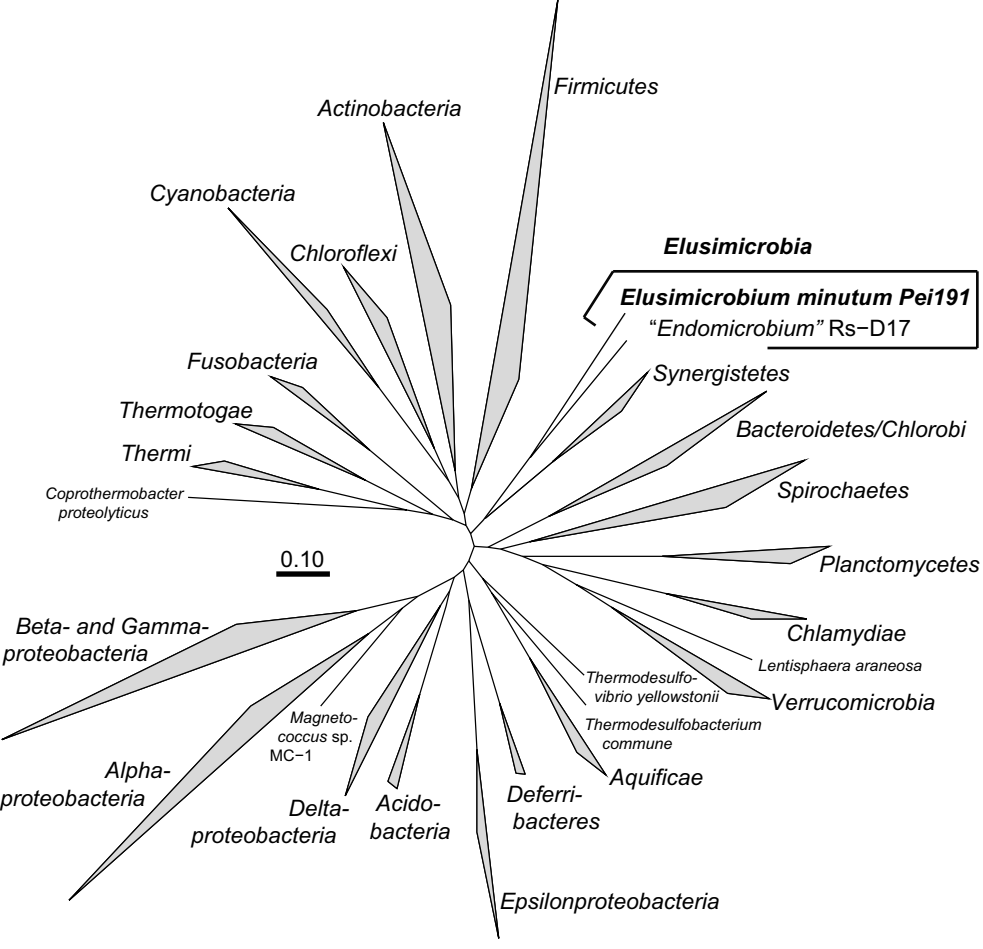
pyruvate (green box). The export of alanine generates a sodium-motive force, which is coupled to the proton-motive force, the synthesis/hydrolysis of ATP via ATP synthase, and the proton-dependent uptake of amino acids or oligopeptides. Pathways were reconstructed based on the manually annotated genome and results from batch culture experiments (16).

FIG. 4. Maximum-likelihood tree of [NiFe] hydrogenases, based deduced amino acid sequences of the large subunit. The sequences of *Elusimicrobium minutum* and *Thermoanaerobacter tengcongensis* fall within the radiation of the sequences assigned to group IV [NiFe] hydrogenases by (54). The topology of the tree was tested separately by neighbor-joining and RAxML, with bootstrapping provided in the ARB package (31).

FIG. 5. Organization of the genes encoding the subunits of the [FeFe] hydrogenase of *T. tengcongensis* (48) and their predicted homologs in *Elusimicrobium minutum*. The displayed length is proportional to the size of the corresponding ORF. *hydA*, *hydB*, and *hydC* have deduced amino acid sequence identities of 46, 56, and 40%, respectively *hydD* is not present in *E. minutum.* White symbols: hypothetical function.

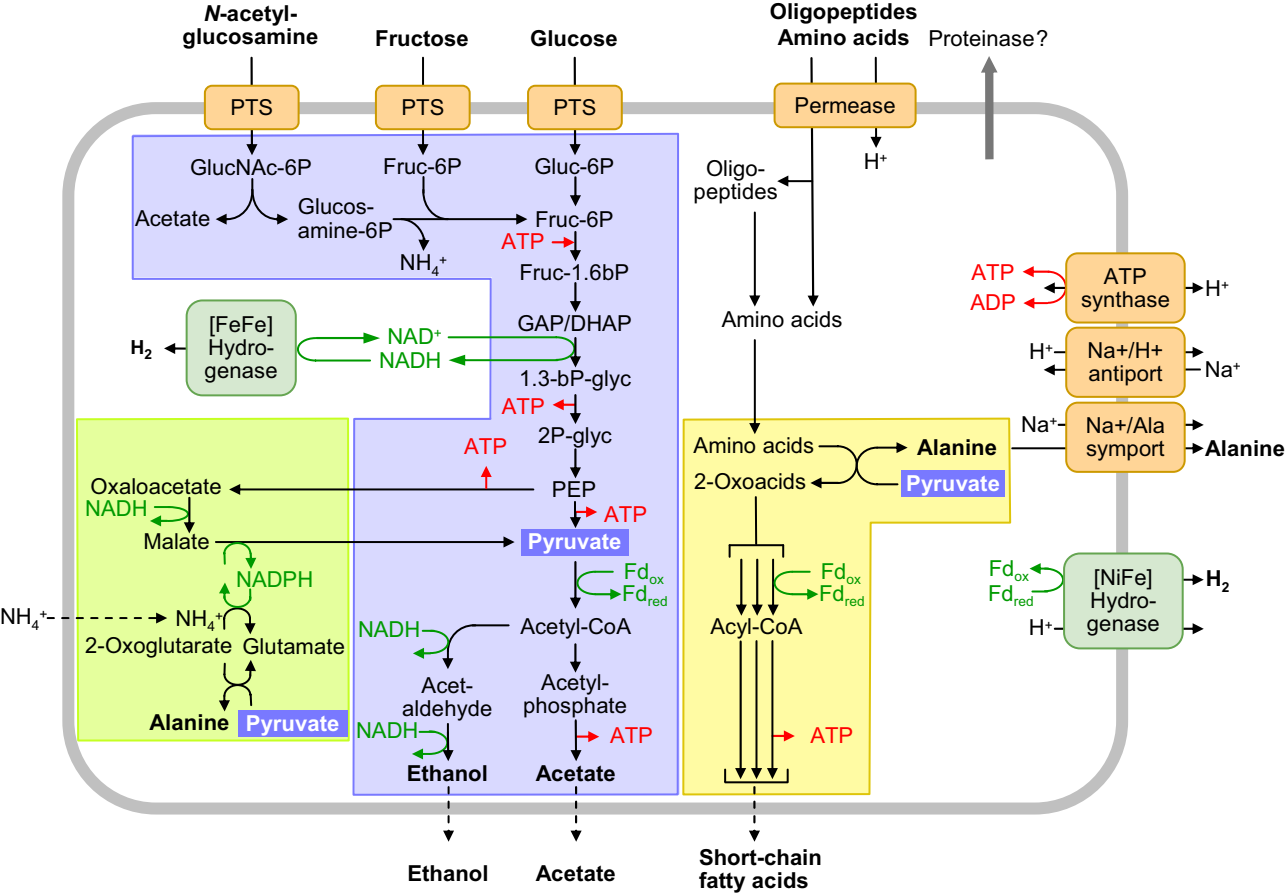
FIG. 6. Organization of the genes encoding the "oxidative stress protection" cluster in *Moorella thermoacetica, Desulfovibrio gigas*, and their predicted homologs in *Elusimicrobium minutum.* The displayed length is proportional to the size of the corresponding ORF. The genes for ruberythrin (*rbr*), superoxide reductase (*sor*), rubredoxine:oxygen oxidoreductase (*roo*), rubredoxin (*rub*) and rubredoxin-like (*rbl*) in *E. minutum* have high sequence similarities to their homologs in *Desulfovibrio* spp. and other *Deltaproteobacteria*. White symbol: hypothetical function.

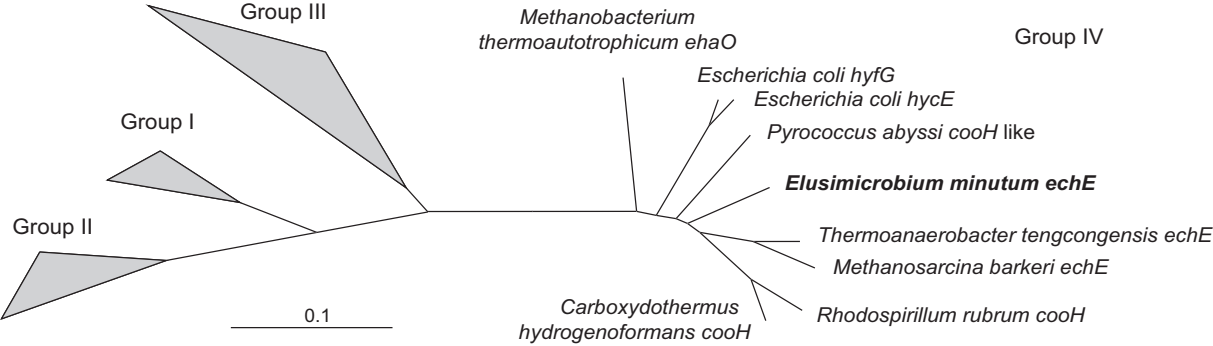

Ori
rRNA
1.6
0.1
1.5
0.2
Prophage
1.4
0.3
1.3
Elusimicrobium
minutum
1,643,562 bp
0.4
1.2
PKS
0.5
1.1
0.6
1.0
0.7
0.9
0.8

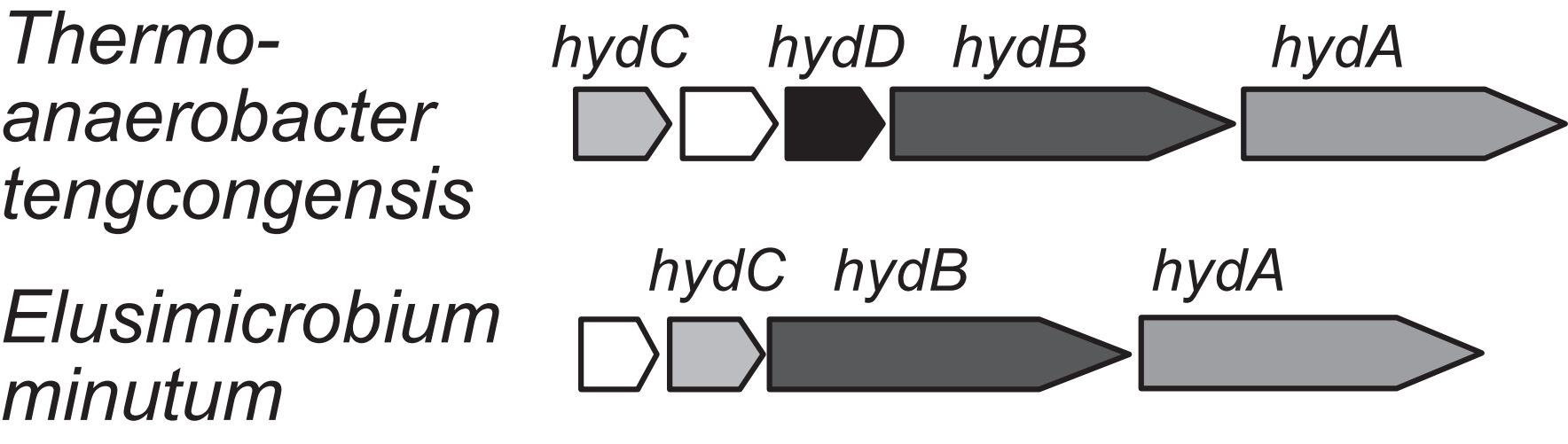


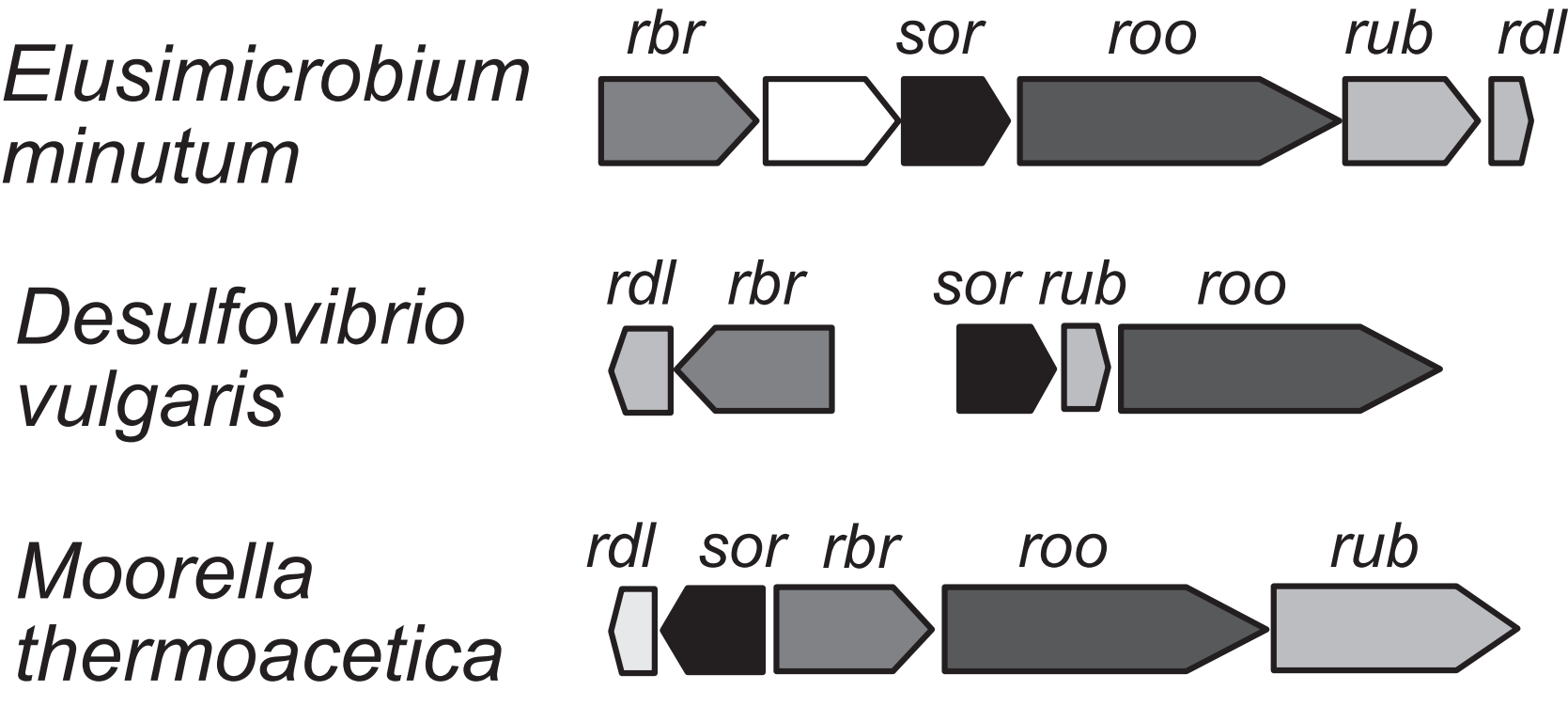
Herlemann et al.

Figure 2


N-acetyl-glucosamine
Fructose
Glucose
Oligopeptides
Amino acids
Proteinase?
PTS
PTS
PTS
Permease
H+
GlucNAc-6P
Fruc-6P
Gluc-6P
Oligo-peptides
Acetate
Glucos-amine-6P
Fruc-6P
ATP
NH4+
Fruc-1.6bP
ATP
ADP
ATP synthase
H+
[FeFe] Hydro-genase
H2
NAD+
NADH
GAP/DHAP
Amino acids
H+
Na+/H+ antiport
Na+
1.3-bP-glyc
ATP
2P-glyc
Amino acids
Alanine
Na+
Na+/Ala symport
Alanine
ATP
2-Oxoacids
Pyruvate
Oxaloacetate
NADH
PEP
ATP
Malate
Pyruvate
Fdox
Fdred
[NiFe] Hydro-genase
H2
NADPH
NH4+
NH4+
Fdox
Fdred
NADH
Acetyl-CoA
Acyl-CoA
2-Oxoglutarate
Glutamate
Acet-aldehyde
Acetyl-phosphate
NADH
ATP
ATP
Alanine
Pyruvate
Ethanol
Acetate
Ethanol
Acetate
Short-chain fatty acids


Group III
Methanobacterium
thermoautotrophicum ehaO
Group IV
Escherichia coli hyfG
Escherichia coli hycE
Group I
Pyrococcus abyssi cooH like
Elusimicrobium minutum echE
Group II
Thermoanaerobacter tengcongensis echE
Methanosarcina barkeri echE
0.1
Carboxydothermus
hydrogenoformans cooH
Rhodospirillum rubrum cooH


Thermo-
anaerobacter
tengcongensis
hydC
hydD
hydB
hydA
Elusimicrobium
minutum
hydC
hydB
hydA


Elusimicrobium minutum
rbr
sor
roo
rub
rdl
Desulfovibrio vulgaris
rdl
rbr
sor
rub
roo
Moorella thermoacetica
rdl
sor
rbr
roo
rub