The Candidate phylum "Termite Group 1"

Diversity, distribution, metabolism and evolution of representatives of an unexplored bacterial phylum

Dissertation

Zur Erlangung des Doktorgrades der
Naturwissenschaften im Fachbereich Biologie der
Philipps-Universität Marburg

Vorgelegt von

Daniel Philipp Ralf Herlemann

aus Offenburg

Marburg/Lahn 2009

Die Untersuchungen zur folgenden Arbeit wurden von Juni 2006 bis Juli 2009 am Max-Planck-Institut für terrestrische Mikrobiologie in Marburg unter Leitung von Prof. Dr. Andreas Brune durchgeführt.

Vom Fachbereich Biologie der Philipps-Universität Marburg als Dissertation angenommen am:

Erstgutachter: Prof. Dr. Andreas Brune
Zweitgutachter: Prof. Dr. Uwe Maier

The following publications are integrated in this thesis:

Herlemann, D. P. R., Geissinger, O., and A. Brune. 2007. The Termite Group 1 phylum is highly diverse and widespread in the environment. *Appl. Environ. Microbiol.* **73**, 6682-6685.

Geissinger O., Herlemann, D. P. R., Maier U., and A. Brune. 2009. *Elusimicrobium minutum* gen. nov., the first isolate of the Termite Group 1 phylum. *Appl. Environ. Microbiol.* **75**, 2831-2840.

Herlemann, D. P. R., Geissinger, O., Ikeda-Ohtsubo, W., Kunin, V., Sun, H, Lapidus A., Hugenholtz P., and A. Brune. 2009. Genome analysis of *Elusimicrobium minutum*, the first cultivated representative of the Elusimicrobia phylum (formerly Termite Group 1). *Appl. Environ. Microbiol.* **75**, 2841-2849.

Erklärung

Ich versichere, dass ich meine Dissertation

„The Candidate phylum “Termite Group 1” – Diversity, distribution, metabolism and evolution of representatives of an unexplored bacterial phylum”

selbständig und ohne unerlaubte Hilfe angefertigt habe und mich keiner als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe. Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, Juli 2009

Danksagung

An erster Stelle möchte ich mich bei meinem Doktorvater Prof. Dr. Andreas Brune für die Überlassung des Themas sowie die stets offene Tür bei Fragen und Diskussionen als auch den Freiraum für die Entfaltung eigener Ideen bedanken.

Herrn Prof. Dr. Uwe Maier danke ich für die Übernahme des Zweitgutachtens und Prof. Dr. Ralf Conrad danke ich für die Möglichkeit in seiner Abteilung zu arbeiten. Ebenfalls möchte ich dem IMPRS „Thesis advisory comitee“ bestehend aus Prof. Dr. Michael Friedrich, Dr. Werner Liesack und Prof. Dr. Rolf Thauer für umfangreiche Diskussionen und Hinweise im Laufe meiner Doktorarbeit danken.

Weiterer Dank gilt allen aktuellen und ehemaligen Mitgliedern der „Termitengruppe“ für die gute und fröhliche Arbeitsatmosphäre. Im Besonderen danke ich unserer technischen Assistentin Katja Meuser für die große Hilfe bei vielen kleinen Dingen ohne die ein Labor nicht funktioniert.

Mein letzter und größter Dank gilt meiner Familie und Inga Krämer für die Unterstützung und Standhaftigkeit in schwierigen Situationen.

Table of Contents

1. Introduction

Termites (Insecta, Isoptera) are terrestrial arthropods abundant in tropical habitats and also present in temperate zones. Termite colonies can exceed 6000 individuals per m^2, and they have a high impact on the dynamics of carbon and nitrogen in soil (3, 9). The majority of the termite species are able to digest lignocellulosic compounds like wood, grass, and plant litter. This ability has recently made them the subject of biofuel-production studies and has attracted much attention to the digestive processes within the termite intestinal tracts (28).

Termites are divided into seven families: Mastotermitidae, Kalotermitidae, Hodotermitidae, Termopsidae, Rhinotermitidae, Serritermitidae, and Termitidae (1). The Termitidae represent the evolutionarily higher termites, characterized by a nutritionally diverse diet and the lack of cellulolytic gut flagellate symbionts. All six families of lower termites feed exclusively on wood and have relatively simple intestinal tracts consisting of three compartments (foregut, midgut, and hindgut). The midgut contains primarily enzymes provided by the host, whereas the digestion in the hindgut depends on special single-celled eukaryotes responsible for cellulose degradation. The microbial activity yields high concentrations of fermentation products that are absorbed by the termite and further oxidized (6). A major part of this rich reservoir of microbial diversity in the hindgut has remained elusive for classical cultivation techniques. Owing to the difficulty in cultivating most of the members from the termite gut microbiota and the complexity of the community, our understanding of the major bacterial groups in the termite gut is still poor.

Microbial diversity in the gut of lower termites

Interdisciplinary research, combining classical microbiological techniques with genetic and biochemical approaches, became one of the most successful strategies for the characterization of diversity in the termite gut. A standard method for the identification and classification of bacteria is performed by targeting the small subunit ribosomal RNA (SSU rRNA). The SSU rRNA is an ideal phylogenetic marker because it is present in all living organisms and has highly conserved and variable regions (32). Comparative analysis of the SSU rRNA sequence not only resulted in the separation of the three domains Bacteria, Archaea, and Eukarya (32), but also transformed microbial taxonomy from an identification system to an evolutionary-based

systematic framework (26, 33). The diversity of 16S rRNA sequences present in a sample is therefore used to describe the microbial diversity.

Anaerobic flagellate protists are essential for the depolymerization of cellulose and hemicellulose in lower termites (15, 4). Molecular studies showed that these flagellates belong to the phyla Parabasalia and Preaxostyla (18, 5) and occur exclusively in the gut of termites and wood-feeding cockroaches. Other than *Trichomitopsis termopsidis* (23, 34), termite gut flagellates have not been cultivated in axenic culture. Flagellates in the hindgut of lower termites occupy the majority of the volume (19). Their surface is usually colonized by many prokaryotic ectosymbionts (2). In addition, the cytoplasm provides an important habitat for many endosymbiotic prokaryotes (7).

The adaptation of termite gut prokaryotes to gut-specific niches is reflected in the phylogeny of the 16S rRNA sequences. Sequences derived from termites often form unique monophyletic lineages, indicating that many bacterial species are autochthonous members of the gut microbiota (35, 10, 24). However, the bacterial community structure in termite guts based on 16S rRNA gene libraries differs considerably from that of cultivation-based studies (6). Ohkuma (24) estimated that the gut of *Reticulitermes speratus* harbors 700 bacterial phylotypes from more than 15 bacterial phyla in *Reticulitermes speratus*, among which Spirochaetes, Bacterioidetes, and the candidate phylum "Termite Group 1" (TG1) are the most dominant representatives. In contrast, most cultivable heterotrophic bacteria in the hindgut of the closely related termite *Reticulitermes flavipes* are *Streptococcus*, *Enterococcus*, and *Bacteroides* spp. (29).

The formal description of genus and species is not possible for bacteria detected by 16S rRNA analysis alone, because the current rules of bacterial nomenclature require that the organism is available as an axenic culture. However, if sufficient genetic, phenotypic, and ecological information of an organism can be retrieved, it is possible to propose a provisional *Candidatus* taxon (20, 21). Such candidate taxa are increasingly used for classification of bacteria even up to the level of candidate phyla (14).

Candidate phylum "Termite Group 1"

An example of a deep-branching bacterial lineage from termite intestinal tracts with no cultivated representative is the candidate phylum "Termite Group 1" (TG1; 25, 14). It had been shown that members of this phylum comprise up to 30% of the gut microbiota in *Reticulitermes santonensis* (Figure 1A; 35). Fluorescence *in situ* hybridization experiments demonstrated that TG1 bacteria are intracellular symbionts of flagellate protists (Figure 1B; 31). These so-called endomicrobia are a monophyletic lineage of bacteria found exclusively in the hindgut of lower termites and wood-feeding cockroaches (31). The investigation of the gut wall from *Reticulitermes speratus* also revealed sequences that are affiliated with TG1 but distinct from the endomicrobia lineage (22). Moreover, public databases contain phylogenetically unassigned sequences from various habitats that seem to be affiliated with the phylum, suggesting that members of TG1 are present not only in termite protists but also found in other habitats. These sequences indicate that the description of TG1 only as endomicrobia needs to be revised and the presence of TG1 bacteria in habitats other than termite flagellates needs to be investigated.

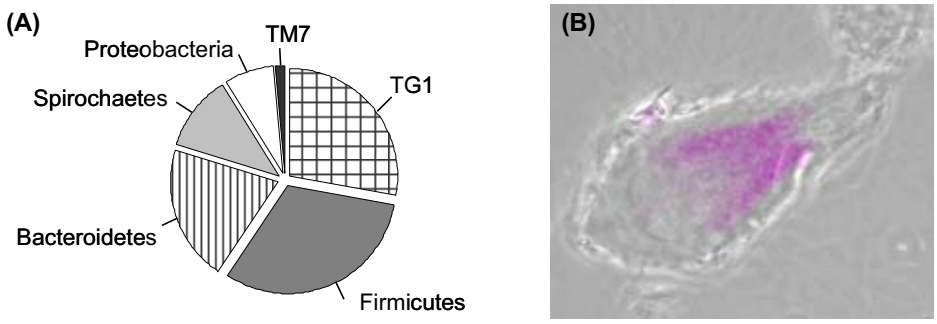


Figure 1. (A) Distribution among clones derived from *Reticulitermes santonensis* hindgut fluid (35). (B) Localization of endomicrobia (purple) in a *Trichynmpha* flagellate by fluorescence *in situ* hybridization (modified according to 31).

Functional characterization of uncultivated microorganisms

Describing the microbial diversity of a habitat is only the first step, yet an important one in advancing the understanding about the community. The SSU rRNA information cannot be extrapolated to physiology and metabolism of an organism (27). Even though members of TG1 constitute a large fraction of bacteria in lower termites, none have been grown in culture and little is understood about their function in the gut system. A detailed study of a single representative, including both physiological and molecular investigations, increases the

knowledge about the bacteria from a phylum enormously, because identified genes can be extrapolated to orthologous genes in related bacteria. Most genomes available today do not cover many habitats, and most bacterial phyla apart from the Proteobacteria and the Firmicutes are poorly or not at all represented by cultures whose genome sequence is available (Figure 2; 13, 12). The identification of functions and phylogenetic assignment of a putative open reading frame relies on the existence of related reference genomes. The lack of well-described reference species in many phyla often leads to a prediction of function below a reliable threshold; therefore, it is important to extend the sequencing of genomes to other phyla accordingly.

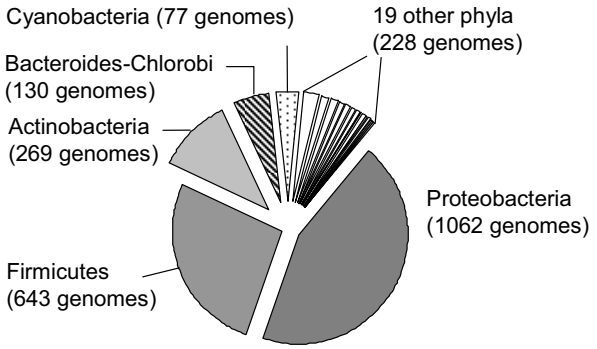


Figure 2. Phylogenetic distribution of the currently 2416 publicly available genomes according to the genomes online database (14). In addition to the phyla illustrated in the pie chart, there are 19 phyla with only a few genome sequences: Spirochaetes (52 genomes), Chlamydiae (29 genomes), Fusobacteria (28 genomes), Chloroflexi (21 genomes), Thermotogae (15 genomes), Deinococcus-Thermus (14 genomes), Synergistetes (14 genomes), Aquificae (11 genomes), Planctomycetes (10 genomes), Verrucomicrobia (9 genomes), Acidobacteria (5 genomes), Deferribacteres (5 genomes), Nitrospira (5 genomes), Dictyoglomi (2 genomes), Gemmatimonadetes (2 genomes), Lentisphaerae (2 genomes), Thermodesulfobacteria (2 genomes), Chrisogenetes (1 genome) and Fibrobacteres (1 genome).

Endomicrobia in termite guts

Symbiosis has various characteristics that range as a gradient of interaction from parasitism to commensalism to mutualism (8). The symbiosis between prokaryotes and protozoa can be quite specific, and recently it has been shown that members of CET cospeciate with their hosts (17), suggesting an obligate symbiotic relationship.

Obligate endosymbionts may never be available in pure culture because of their strong dependence on their hosts. Culture-independent approaches allow the recovery of genomes of obligate symbionts, providing an opportunity for comparative genome analysis to reconstruct the putative metabolism. Genomes of compartmented endosymbionts were successfully

recovered from *Buchnera aphidicola* by the application of metagenomics (30). Metagenomic approaches are based on randomly sequencing genome fragments of a sample. This includes usually all members in a microbial community and therefore a physical enrichment of the target population improves the yield of target genomic fragments enormously. This strategy was also applied to *Candidatus* "Endomicrobium trichonymphae" (CET) by filtering CET from *Trichonympha* flagellates (Figure 3; 16). This metagenome project should yield large genome fragments, which contain enough information to investigate the function of CET in the gut of termites.

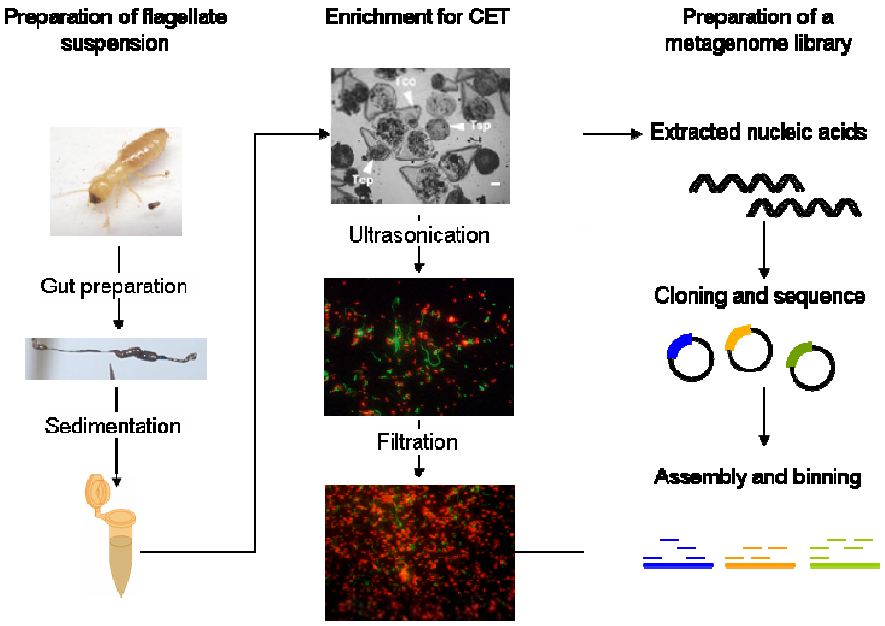


Figure 3. Schematic representation of the physical enrichment of *Candidatus* "Endomicrobium trichonymphae" (CET) from *Zootermopsis nevadensis* and the general procedure for the construction of a metagenome library (16).

Chapter outline

The aims of these studies were to increase the current knowledge about the candidate phylum "Termite Group 1". In the first part of this thesis, (Chapter 2), I report my investigations of the environmental distribution of representatives of the "Termite Group 1" phylum outside the gut of lower termites by screening various habitats with phylum-specific primers.

The majority of this thesis concerns *Elusimicrobium minutum*, the first and currently only isolate from the TG1 phylum (Chapters 3 and 4). The isolate allowed the investigation of the physiology of a member of this phylum. Based on the complete genome sequence of this bacterium, the metabolism was reconstructed and unique genomic characteristics were identified. The genome was sequenced in collaboration with the Joint Genome Institute (Walnut Creek, Calif., USA).

During the course of these studies, the complete genome sequence of *Candidatus* “Endomicrobium trichonymphae” strain Rs-D17 from *Trichonympha agilis* was published (11). The genome sequence answered many of the questions about the function of the abundant *Candidatus* “Endomicrobium trichonymphae” in termites. In the last part of the thesis (Chapter 5), the genome from *Elusimicrobium minutum* was compared with that of strain Rs-D17 and metagenomic data of *Candidatus* "Endomicrobium trichonymphae" derived from physically enriched *Trichonympha* spp. of *Zootermopsis nevadensis*. This comparison gave first insights into the mechanisms of genome evolution in CET.

References

[1] **Abe, T., D. E. Bignell, and M. Higashi (eds.).** 2000. Termites: Evolution, Sociality, Symbiosis, Ecology. Kluwer Academic Publishers, Dordrecht.

[2] **Berchtold, M., A. Chatzinotas, W. Schönhuber, A. Brune, R. Amann, D. Hahn, and H. König.** 1999. Differential enumeration and *in situ* localization of micro-organisms in the hindgut of the lower termite *Mastotermes darwiniensis* by hybridization with rRNA-targeted probes. *Arch. Microbiol.* **172:**407–416.

[3] **Bignell, D. E., and P. Eggleton.** 2000. Termites in ecosystems. *In* Abe, T., D. E Bignell . and M. Higashi (ed.), Termites: Evolution, Sociality, Symbioses, Ecology. Kluwer Academic Publisher, Dordrecht, pp. 363–387.

[4] **Breznak, J. A., and A. Brune.** 1994. Role of microorganisms in the digestion of lignocellulose by termites. *Annu. Rev. Entomol.* **39:**453–487.

[5] **Brugerolle, G., and R. Radek.** 2006. Symbiotic protozoa of termites. *In* König, H. and A. Varma (ed.), Intestinal Microorganisms of Termites and Other Invertebrates. Springer Verlag, Berlin, pp. 243–269.

[6] **Brune, A.** 2006. Symbiotic associations between termites and prokaryotes. *In* Dworkin, M., S. Falkow, E. Rosenberg, K-H. Schleifer, and E. Stackebrandt, (ed.), The Prokaryotes, 3rd ed.: Symbiotic associations, Biotechnology, Applied Microbiology. Springer, New York, NY, pp. 439–474.

[7] **Brune, A., and U. Stingl.** 2005. Prokaryotic symbionts of termite gut flagellates: phylogenetic and metabolic implications of a tripartite symbiosis. *In* Overmann, J. (ed.), Molecular Basis of Symbiosis. Springer, Berlin, pp. 39–60.

[8] **deBary, A.** 1878. Ueber Symbiose. *In* Ber. Vers. Deut. Naturf. Aerzte: Cassel 1878, Versammlung Deutscher Naturforscher und Aerzte, Cassel, pp. 121–126.

[9] **Eggleton, P., and I. Tayasu.** 2001. Feeding groups, lifetypes and the global ecology of termites. *Ecol. Res.* **16:**941–960.

[10] **Hongoh, Y., P. Deevong, T. Inoue, S. Moriya, S. Trakulnaleamsai, M. Ohkuma, C. Vongkaluang, N. Noparatnaraporn, and T. Kudo.** 2005. Intra- and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host. *Appl. Environ. Microbiol.* **71:**6590–6599.

[11] **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.

[12] **Raymond, J.** 2008. Coloring in the tree of life. *Trends Microbiol.* **16:**41–3.

[13] **Hugenholtz P.** 2002. Exploring prokaryotic diversity in the genomic era. *Genome Biol.* **3:**0003.1–8.

[14] **Liolios, K., N. Tavernarakis, P. Hugenholtz, and N. C. Kyrpides.** 2006. The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide. *Nucleic Acids Res.* **1:**D332–4.

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

[15] **Hungate, R. E.** 1955. Mutualistic intestinal protozoa. *In* Hutner, SH; Lwoff, A (ed.), Biochemistry and Physiology of Protozoa. Academic Press, New York, NY, pp. 159–199.

[16] **Ikeda-Ohtsubo, W.** 2007. Endomicrobia in termite guts: symbionts within symbiont. Doctoral thesis, Philipps-Universität Marburg.

[17] **Ikeda-Ohtsubo, W., and A. Brune.** 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and *Candidatus* "Endomicrobium trichonymphae". *Mol. Ecol.* **18:**332–342.

[18] **Inoue, T., O. Kitade, T. Yoshimura, and I. Yamaoka.** 2000. Symbiotic associations with protists. *In* Abe, T.; Bignell, D.E.; Higashi, M. (ed.), Termites: Evolution, Sociality, Symbiosis, Ecology. Kluwer Academic Publishers, Dordrecht, pp. 275–288.

[19] **Katzin, L. I., and H. Kirby.** 1939. The relative weights of termites and their protozoa. *J. Parasitol.* **25:**444–445.

[20] **Murray, R. G., and K. H. Schleifer.** 1994. Taxonomic notes: a proposal for recording the properties of putative taxa of prokaryotes. *Int. J. Syst. Bacteriol.* **44:**174–176.

[21] **Murray, R. G. E., and E. Stackebrandt.** 1995. Taxonomic note: Implementation of the provisional status *Candidatus* for incompletely described prokaryotes. *Int. J. Syst. Bacteriol.* **45:**186–187.

[22] **Nakajima, H., Y. Hongoh, R. Usamib, T. Kudo, and M. Ohkuma.** 2005. Spatial distribution of bacterial phylotypes in the gut of the termite *Reticulitermes speratus* and the bacterial community colonizing the gut ephitelium. *FEMS Microbiol. Ecol.* **54:**247–255.

[23] **Odelson, D. A., and J. A. Breznak.** 1985. Nutrition and growth characteristics of *Trichomitopsis termopsidis*, a cellulolytic protozoan from termites. *Appl. Environ. Microbiol.* **49:**614–621.

[24] **Ohkuma, M.** 2008. Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends Microbiol.* **16:**345–352.

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

[26] **Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace, and D. A. Stahl.** 1986. Microbial Ecology and Evolution: A Ribosomal RNA Approach. *Annu. Rev. Microbiol.* **40:**337–365.

[27] **Rosselló-Mora, R., and R. Amann.** 2001. The species concept for prokaryotes. *FEMS Microbiol.* Reviews **25:**39–67.

[28] **Schubert, C.** 2006. Can biofuels finally take center stage? *Nat. Biotechnol.* **24:**777–784.

[29] **Schultz, J. E., and J. A. Breznak.** 1978. Heterotrophic bacteria present in hindguts of wood-eating termites *Reticulitermes flavipes* (Kollar). *Appl. Environ. Microbiol.* **35:**930–936.

[30] **Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa.** 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **407:**81–6.

[31] **Stingl, U., R. Radek, H. Yang., and A. Brune.** 2005. '*Endomicrobia*': Cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* **71:**1473−1479.

[32] **Woese, C. R., and G. E. Fox.** 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* **74:**5088–90.

[33] **Woese, C.** 1998. The universal ancestor. *Proc. Natl. Acad. Sci. U.S.A.* **95:**6854–6859.

[34] **Yamin, M. A.** 1978. Axenic cultivation of the cellulolytic flagellate *Trichomitopsis termopsidis* (Cleveland) from the termite *Zootermopsis*. *J. Protozool.* **25:**535–538.

[35] **Yang, H., D. Schmitt-Wagner, U. Stingl, and A. Brune.** 2005. Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ. Microbiol.* **7:**916–932.

2. The Termite Group 1 Phylum is highly diverse and widespread in the environment

Daniel P. R. Herlemann, Oliver Geissinger, and Andreas Brune

Published in Applied and Environmental Microbiology, Oct. 2007 Vol. 73, No. 20

Summary

The bacterial candidate phylum "Termite Group 1" (TG1) presently consists mostly of endomicrobia, which are endosymbionts of flagellate protists occurring exclusively in the hindgut of termites and wood-feeding cockroaches. Here, we show that public databases contain many, mostly undocumented, 16S rRNA gene sequences from other habitats that are affiliated with the TG1 phylum but only distantly related to endomicrobia. Phylogenetic analysis of the expanded dataset revealed several diverse and deep-branching lineages comprising clones from many different habitats. In addition, we designed specific primers to explore the diversity and environmental distribution of bacteria in the TG1 phylum.

Authors' contribution: Molecular, phylogenetic and bioinformatic analysis was performed and planed by D. H., who also prepared the manuscript draft together with A. B. O. G. supplied some 16S rRNA sequences and was involved in the revision and discussion of the manuscript.

Introduction

The Termite Group 1 (TG1) represents a deep branch in the tree of bacterial 16S rRNA gene sequences (18) and has been recognized as a candidate phylum (10). TG1 comprises a large number of the bacteria in the hindgut of *Reticulitermes* spp. (6, 25), where they occur as intracellular symbionts of flagellate protists (22). These symbionts, for which the name endomircrobia has been proposed, form a monophyletic lineage occurring exclusively in the hindgut of termites and wood-feeding cockroaches (11, 22).

However, a few sequences only distantly related to the endomircrobia but clearly affiliated with the TG1 phylum have been reported to occur also in habitats other than termite guts (3, 19, 20, 24). Moreover, Nakajima et al. (17) obtained two sequences from the gut of *Reticulitermes speratus* that fall outside the endomircrobia lineage. Presently, public databases contain a growing number of sequences from various habitats that are phylogenetically unassigned but seem to be affiliated with the TG1 phylum.

In this study, we screened public databases for hitherto unrecognized TG1 sequences and conducted a comprehensive phylogenetic analysis of the expanded dataset. In addition, we designed specific PCR primers to investigate the diversity and environmental distribution of major lineages of TG1 bacteria in soils, sediments, and intestinal tracts.

Data mining

Sequences affiliated with the TG1 phylum were retrieved from GenBank (http://www.ncbi.nlm.nih.gov/) using various characteristic oligonucleotide signatures deduced from the originally available 16S rRNA gene sequences, and added to the database of the ARB program suite (15). By continuously adapting the signatures to the growing dataset (details not shown), we obtained approximately 50 previously unassigned phylotypes that fell into the radiation of the TG1 phylum. Most were from large-scale diversity studies of various environments, including soils, sediments, and intestinal tracts. Sequences were aligned with the ARB Fast Aligner tool. The alignment was manually corrected, and highly variable regions and ambiguous positions were excluded from the analysis. Rigorous chimera checking with Bellerophon (9) and fractional treeing (14) identified only one sequence as a putative chimera (DQ830579), which was removed from the dataset. All shorter sequences (500–1300 bp) were added to the core tree using the parsimony tool implemented in ARB.

Primer design and PCR

Phylogenetic analysis of the near-full-length sequences (>1300 bp) fully supported that the TG1 phylum forms a separate line of descent in the bacterial tree, consisting of several diverse and deep-branching lineages (Figure 1; sequences in bold). The design of a single phylum-specific primer that excluded all representatives of other phyla proved to be impossible. A primer set for the specific amplification of endomircrobia (table 1, primer set 1) had been designed already in a previous study (22). Based on the expanded dataset, we designed three additional primer sets (primer sets 2–4) that covered most of the other lineages in the TG1 phylum. When applicable, information from shorter sequences was included to improve primer design.

The primers were used for PCR-based screening of various habitats for the presence of TG1 bacteria. DNA was extracted by bead beating (16), and humic substances were removed by passing the aqueous extracts over an Autoseq G-50 column (Amersham Bioscience). PCR amplifications used a standard protocol optimized for the respective primer pairs (Table 1). The products of two identical reactions were combined, cleaned with the MinElute PCR purification kit (Qiagen), and cloned with the pGEM-T Easy Vector kit (Promega). Positive clones were amplified with M13 vector primers and checked for inserts on a 1% agarose gel. Clones with correct insert lengths were sorted by restriction fragment length polymorphism analysis as previously described (21). Inserts were sequenced on both strands, and sequences were submitted to the EMBL database (accession numbers: AM491071–491086, AM491098, AM491123, AM491125; http://www.ebi.ac.uk/).

Phylogenetic analysis

Phylogenetic analysis of the resulting clone libraries documented that all primer pairs were highly specific and amplified only 16S rRNA genes of bacteria in the TG1 phylum.

Primer set 1 gave a PCR product only with the termite hindgut samples (table 2). This corroborates the specificity of this primer set for its target group, the endomircrobia (Figure 1, lineage I), which seem to be restricted to termites and wood-feeding cockroaches harboring gut flagellates (1, 22). endomircrobia sequences from termite guts were not further investigated since they are the subject of a separate study (11).

Table 1. Primer sets used for the amplification of major lineages in the TG1 phylum (Figure 1), and annealing temperatures and $MgCl_2$ concentrations used in the respective PCR assay (see Table 2).

Primer set	Target group [a]	Primer name	*E. coli* position	Sequence (5' to 3')	T (°C)	$MgCl_2$ (mM)
1	I	TG1-209f (Ref. 22)	210–237	AAT GCG TTT TGA GAT GGT CCT G	54.0	2
		TG1-1325r (Ref. 22)	1325–1343	GAT TCC TAC TTC ATG TGG		
2	II	EluD540f	539–562	AGG TGG CAA GCG TTA CTC GGA AT	63.9	1.5
		EluD1300r	1286–306	TCT GAA CTG GGG CCG GCT TTT		
3	III	EluB22f	22–40	GCT CAG AGT TAA CGC TGG C	61.7	1.5
		EluB998r	981–998	GTC GTT CGA GCC CAG GTA A		
4	IV	EluC104f	104–126	GGC AGA CGA GTG AGT ARC ACG TA	57.0	2
		EluC1173r	1153–1173	ACG TTA TCC GCG GCA GTC TCC		
T	II–IV	27f (Ref.12)	8–27	AGAGTTTGATCCTGGCTCAG	55.0	2
		Elu1058r	1042–1058	CCATGCAGCACCTCGGC		

[a] With the exception of primer set T, all primer sets had at least 2–3 mismatches to sequences outside the TG-1 phylum present in the greengenes database (2; http://greengenes.lbl.gov/).

Primer set 2 matched the sequences of three deep-branching lineages (IIa–c) comprising clones from soil or sediment samples. An amplification product of the expected size was obtained only with Italian rice soil. The resulting clone library contained several phylotypes falling into lineage IIa and a single phylotype falling into lineage IIb (Figure 1). Generally, the large number of sequences from soils and sediments retrieved from public databases suggests that bacteria in lineage II are widely distributed in these habitats.

Primer set 3 was designed to target the sequences in lineage III, which consists of sequences from intestinal habitats and also includes strain Pei191, the first isolate from the TG1 phylum recently obtained from the gut of a beetle larva (O. Geissinger and A. Brune, in preparation). Several new phylotypes were obtained also from the hindgut of *R. santonensis* and *Zootermopsis nevadensis* (Table 2). All were distantly related to the clones previously retrieved from a *R. speratus* gut wall sample (17), underlining that termite guts harbor a second lineage of TG1 bacteria besides the endomircrobia. Amplification products of cow rumen yielded a diverse but monophyletic group of sequences clustering with a single TG1 sequence (AB034017) contained in a bacterial clone library from this habitat and erroneously assigned to the Proteobacteria (23).

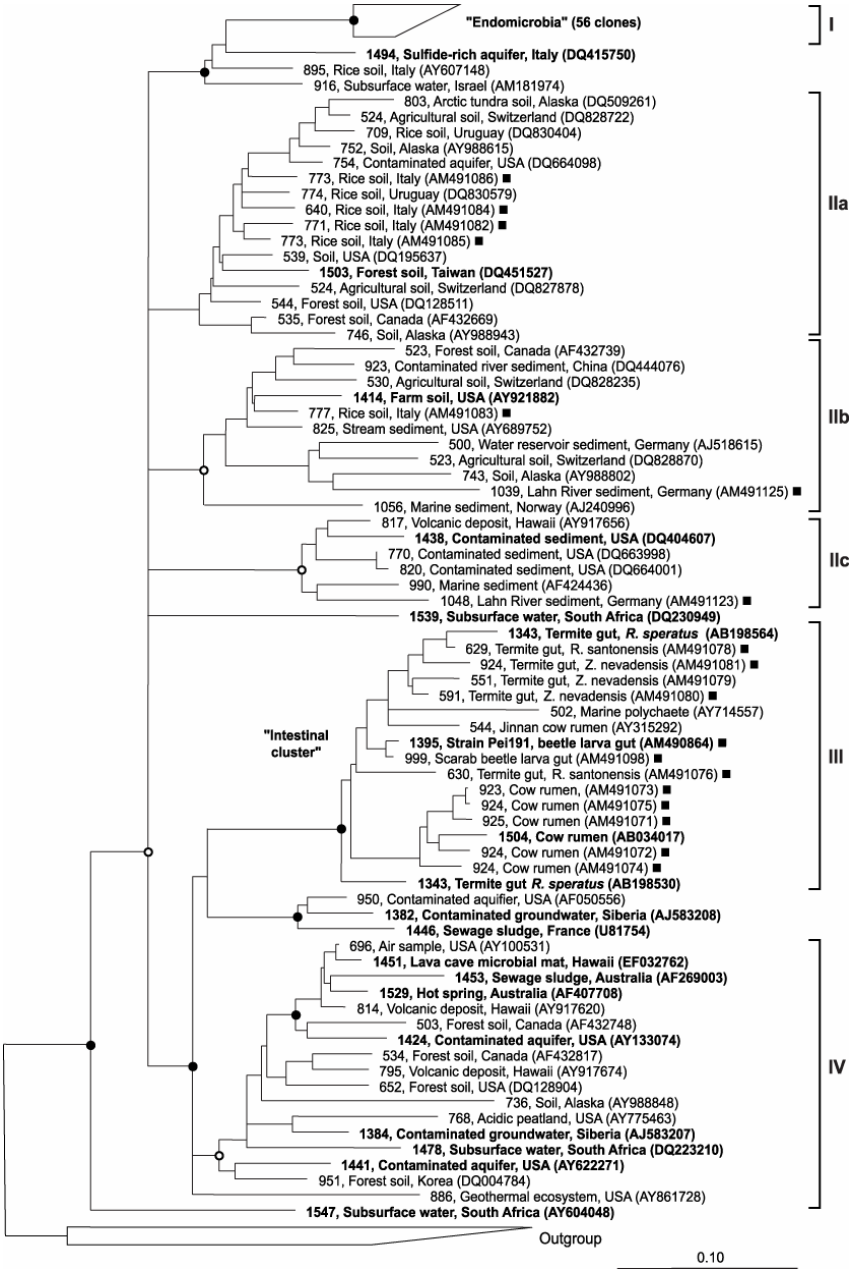


Figure 1. Phylogenetic tree of all sequences in the TG1 phylum obtained in this and previous studies. A core tree was constructed with 1262 unambiguously aligned sequence positions of all near-full-length sequences retrieved from Genbank (in bold), using maximum-likelihood analysis (fastDNAml). Short sequences (> 500 bp), positionally filtered by base frequency (50%), were added without changing the global tree topology using the ARB parsimony tool. The scale bar is only approximate because the procedure distorts branch length. Representatives of 10 other phyla were used as the outgroup. Clones obtained in this study are marked (■). Topology of core tree and individual clusters was tested separately by neighbor-joining and parsimony analysis (DNAPARS) with bootstrapping (seqboot; 1000 bootstraps). Only nodes supported with high bootstraps are marked (●, > 95%; **O**, > 75%); nodes not supported by all analyses are shown as multifurcations. Original sequence definitions in Genbank were replaced with a consistent nomenclature including sequence length, habitat, geographic origin, and accession number. Roman numerals mark the lineages referred to in the text and in (Table 1).

Primer set 2 yielded PCR products also with the termite gut samples, but cloning analysis revealed that they consisted exclusively of endomircrobia sequences, which indicated an insufficient discrimination against this group. Also the termite clone libraries obtained with primer set 3 contained sequences belonging to the endomircrobia (Table 2). Although both primer sets had two or more mismatches to all non-target sequences within the TG1 phylum, they apparently lack differential power for the respective subgroups if too many endomircrobia are present in the sample — a problem encountered only in lower termites.

Table 2. Detection of TG1 bacteria in various habitats using the new designed primer sets for the specific amplification of major lineages in the phylum (see (Table 1). If clone libraries were analyzed, the number of TG1/Endomircrobia phylotypes among the total clones tested (in parentheses) is given.[a]

Habitat	Primer sets[b]			
	1	2	3	4
Lahn River[c] sediment	–	–	–	–
Marburg forest soil[c]	–	–	–	–
Italian rice soil[d]	–	6/0 (14)	–	–
Cow rumen[e]	–	–	5/0 (16)	–
Reticulitermes santonensis gut[f]	+	3/3 (16)	5/3 (17)	–
Zootermopsis nevadensis gut[f]	+	3/3 (19)	5/2 (12)	–
Pachnoda ephippiata gut[f]	–	–	±	–
Strain Pei191[g]	–	–	+	–

[a] The reaction mixture (25 µl) contained reaction buffer (Invitrogen), $MgCl_2$ (Table 1), dNTPs (200 µM each), primers (0.3 µM each), dimethyl sulfoxide (1 µl), DNA extract (200–300 ng), and *Taq* DNA polymerase (1.25 U; Invitrogen). Thermal cycling consisted of an initial denaturation step of 5 min at 96°C, followed by 30 cycles consisting of 30 s at 94°C, 30 s at annealing temperature (Table 1), and 45 s at 72°C.
[b] +, PCR product of the expected size; –, no PCR product; ±, results with different preparations varied.
[c] Topsoil under *Fagus sylvatica* and anoxic sediment of the Lahn River were freshly collected in Marburg, Germany).
[d] Dried rice soil (*Oryza sativa*) from wetland rice fields of the Italian Rice Research Institute in Vercelli (Italy) was regenerated for three days as described by Frenzel et al. (Frenzel et al., 1992).
[e] Rumen content of a freshly slaughtered Holstein cow.
[f] Prepared as previously described (Lemke et al., 2003, Stingl and Brune, 2003).
[g] Isolated from the hindgut of *Pachnoda ephippiata*.

None of the DNA samples used in this study gave a PCR product with primer set 4, designed to detect sequences from lineage IV, which comprises clones from many different habitats (Figure 1). Lahn River sediment yielded PCR products only with primer set T (Table 1), which was designed to detect most TG1 sequences excluding lineage I. It yielded PCR products with all habitats tested, but turned out to be non-specific. Sequence analysis revealed that only about 10% of the clones in each clone library fell into the TG1 phylum (lineages III in *P. ephippiata* gut; lineages IIb and IIc in Lahn River sediment). Other clones were mostly representatives of the Bacteroidetes and the Acidobacteria, which was in agreement with a lack of discrimination of this primer set against a few representatives of these phyla (details not shown).

Abundance in the environment

TG1 bacteria are highly abundant only in bacterial clone libraries from the hindgut of lower termites (6, 7, 8, 18, 25). Fluorescence *in situ* hybridization corroborated that endomircrobia constitute a significant portion of the gut microbiota (11, 22). In contrast, bacterial clone libraries from other habitats generally contain — if any — only small numbers of sequences from the TG1 phylum. In clone libraries of rumen fluid (150 clones) and farm soil (1700 clones), only a single clone each fell into the TG1 phylum (23, 24). Clone libraries of *Pachnoda ephippiata* larva gut homogenates (113 clones; ref. (4) did not contain any clones affiliated with the TG1 phylum, although a representative of the "intestinal cluster" was obtained from this species with primer set T (this study). Possible explanations for these phenomena may lie in either a low relative abundance of such bacteria in the respective communities or a mismatch in the "universal" bacteria primers used in these studies.

Conclusion

The results of this study document that bacteria affiliated with the TG1 phylum form a separate line of descent, as proposed earlier on the basis of a much smaller dataset (10, 18). The phylum consists of numerous diverse and deep-branching lineages comprising bacteria from a wide range of chemically and geographically distinct habitats, including soils, sediment, and intestinal tracts. Although TG1 bacteria seem to be numerically abundant only in the hindgut of lower termites (lineage I; endomircrobia), the large diversity and wide environmental distribution of other lineages suggest a hitherto unrecognized role in the environment.

References

1. **Brune, A., and U. Stingl.** 2005. Prokaryotic symbionts of termite gut flagellates: phylogenetic and metabolic implications of a tripartite symbiosis. *In* J. Overmann (ed.), Molecular Basis of Symbiosis. Springer, Berlin, p. 39–60.

2. **DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen.** 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72:**5069−5072.

3. **Dojka, M. A., P. Hugenholtz, S. K. Haack, and N. R. Pace.** 1998. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl. Environ. Microbiol.* **64:**3869−3877.

4. **Egert, M., B. Wagner, T. Lemke, A. Brune, and M. W. Friedrich.** 2003. Microbial community structure in midgut and hindgut of the humus-feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Appl. Environ. Microbiol.* **69:**6659−6668.

5. **Frenzel, P., F. Rothfuss, and R. Conrad.** 1992. Oxygen profiles and methane turnover in a flooded rice microcosm. *Biol. Fertil. Soils* **14:**84−89.

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

7. **Hongoh, Y., H. Yuzawa, M. Ohkuma, and T. Kudo.** 2003b. Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment. *FEMS Microbiol. Lett.* **221:**299−304.

8. **Hongoh, Y., P. Deevong, T. Inoue, S. Moriya, S. Trakulnaleamsai , M. Ohkuma, C. Vongkaluang, N. Noparatnaraporn, and T. Kudo.** 2005. Intra- and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host. *Appl. Environ. Microbiol.* **71:**6590−6599.

9. **Huber, T., G. Faulkner, and P. Hugenholtz.** 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20:**2317−2319.

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

11. **Ikeda-Ohtsubo, W., M. Desai, U. Stingl, and A. Brune.** 2007. Phylogenetic diversity of Endomircrobia and their specific affiliation with termite gut flagellates. *Microbiology* **150**: 3458-3465.

12. **Lane, D. J.** 1991. 16S/23S rRNA sequencing. *In* E. Stackebrandt and M. Goodfellow (ed.), Nucleic acids techniques in bacterial systematics. John Wiley & Sons, New York, NY, pp. 115−175.

13. **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.

14. **Ludwig, W., S. H. Bauer, M. Bauer, I. Held, G. Kirchhof, R. Schulze, I. Huber, S. Spring, A. Hartmann, and K. H. Schleifer.** 1997. Detection and *in situ* identification of representatives of a widely distributed new bacterial phylum. *FEMS Microbiol. Lett.* **153:**181−190.

15. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, and 27 other authors.** 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32:**1363−1371.

16. **Lueders, T., and M. W. Friedrich.** 2003. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and *mcrA* genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. *Appl. Environ. Microbiol.* **69:**320−326.

17. **Nakajima, H., Y. Hongoh, R. Usamib, T. Kudo, and M. Ohkuma.** 2005. Spatial distribution of bacterial phylotypes in the gut of the termite *Reticulitermes speratus* and the bacterial community colonizing the gut ephitelium. *FEMS Microbiol. Ecol.* **54:**247−255.

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

19. **Reardon, C. L., D. E. Cummings, L. M. Petzke, B. L. Kinsall, D. B. Watson, B. M. Peyton, and G. G. Geesey.** 2004. Composition and diversity of microbial communities recovered from surrogate minerals incubated in an acidic uranium-contaminated aquifer. *Appl. Environ. Microbiol.* **70:**6037−6046.

20. **Spear, J. R., J. J. Walker , T. M. McCollom, and N. R. Pace.** 2005. Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *Proc. Natl. Acad. Sci. U. S. A.* **102:**2555−2560.

21. **Stingl, U., and A. Brune.** 2003. Phylogenetic diversity and whole-cell hybridization of oxymonad flagellates from the hindgut of the wood-feeding lower termite *Reticulitermes flavipes*. *Protist* **154:**147−155.

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

23. **Tajima, K., S. Arai, K. Ogata, T. Nagamine, H. Matsui, M. Nakamura, R. I. Aminov, and Y. Benno.** (2000) Rumen bacterial community transition during adaptation to high-grain diet. *Anaerobe* **6:**273−284.

24. **Tringe, S. G, C., von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E. M. Rubin.** 2004. Comparative metagenomics of microbial communities. *Science* **308:**554−557.

25. **Yang, H., D. Schmitt-Wagner, U. Stingl, and A. Brune.** 2005. Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ. Microbiol.* **7:**916−932.

3. The ultramicrobacterium "*Elusimicrobium minutum*" gen. nov., sp. nov., the first cultivated representative of the Termite Group 1 Phylum

Oliver Geissinger, Daniel P. R. Herlemann, Erhard Mörschel, Uwe G. Maier, and Andreas Brune

Published in Applied and Environmental Microbiology, May 2009, p. 2831-2840, Vol. 75, No. 9

Summary

Insect intestinal tracts harbor several novel, deep-rooting clades of so far uncultivated bacteria, whose biology is typically completely unknown. Here, we report the isolation of the first representative of the Termite Group 1 (TG1) phylum from sterile-filtered gut homogenates of a humivorous scarab beetle larva. Strain Pei191^T is a mesophilic, obligately anaerobic ultramicrobacterium with a Gram-negative cell envelope. Cells are typically rod-shaped but cultures are pleomorphic in all growth phases (0.3–2.5 μm long and 0.17–0.3 μm wide). The isolate grows heterotrophically on sugars and ferments D-galactose, D-glucose, D-fructose, D-glucosamine and *N*-acetyl-D-glucosamine to acetate, ethanol, hydrogen, and alanine as major products, but only if amino acids are present in the medium. PCR-based screening and comparative 16S rRNA gene sequence analysis revealed that strain Pei191^T belongs to the "intestinal cluster", a lineage of hitherto uncultivated bacteria present in arthropod and mammalian gut systems. It is only distantly related to the previously described Endomircrobia lineage, which comprises mainly uncultivated endosymbionts of termite gut flagellates. We propose the name *Elusimicrobium minutum* gen. nov., sp. nov. (type strain: Pei191^T = ATCC BAA-1559^T = JCM 14958^T) for the first isolate of this deep-branching lineage and the name *Elusimicrobia* phyl. nov. for the former TG1 phylum.

Authors' contribution: The physiological analysis was performed and planed by O. G who also prepared the draft manuscript. D. H. prepared the 16S rRNA sequences, performed DAPI staining and was involved in the revision and discussion of the experiments. E. M conducted negative staining electromicrographs and U. M. was involved in the TEM electromicrographs. A.B. and O. G. wrote the final manuscript.

Introduction

Insect intestinal tracts harbor an enormous diversity of so far uncultivated bacteria, which are characterized only by their 16S rRNA gene sequences, and whose biology is typically completely obscure (9, 17, 49). As in other environments (45), many of these sequences form deep-branching phylogenetic lineages that do not contain a single isolate (18, 28). One of these lineages is the Termite Group 1 (TG1), which was originally discovered by Ohkuma and Kudo (37) and recognized as a phylum-level group (candidate division) by Hugenholtz et al. (20). TG1 bacteria form a major proportion of the microbial community in the hindgut of lower termites (16, 69), where they inhabit the cytoplasm of the intestinal flagellates (53, 38). These so-called Endomircrobia are specific for the respective flagellate species (21) and, at least in the case of *Candidatus* "Endomicrobium trichonymphae", are cospeciating with their flagellate host (22).

However, the TG1 phylum comprises also several other deep-rooting lineages (>15% 16S rRNA gene sequence divergence). They are present in a variety of environments, including soils, sediments, and intestinal tracts (14). One of these lineages, the "Intestinal cluster", comprises sequences originating exclusively from intestinal habitats, including the termite gut, but is only distantly related to the lineage comprising the Endomircrobia (14). Here we report the isolation a member of the "Intestinal cluster" from the hindgut of a humivorous scarab beetle larva and its physiological and ultrastructural characterization. We propose a new species, *Elusimicrobium minutum* gen. nov. sp. nov, and define the phylogenetic framework for the first cultivated representative of the TG1 phylum.

Material and methods

Cultivation and inoculum. *Pachnoda ephippiata* (Coleoptera: Scarabaeidae) larvae were from a laboratory population raised on organic soil (30). Hindgut homogenates of four third-instar larvae were prepared under anoxic conditions (30). The homogenates (2 ml) were diluted in 10 ml phosphate buffer solution (58) and centrifuged at 5,000 × *g*. The supernatants — originally intended as a medium

supplement to promote growth of fastidious gut bacteria — were passed twice through 0.2-µm cellulose acetate membrane filters (FP 30, Schleicher & Schüll), and the filtrate was added to a series of rubber-stoppered cultivation tubes containing basal medium amended with D-glucose, D-fructose, and *N*-acetyl-D-glucosamine (each 2 mM). From culture tubes that showed visible growth, pure cultures were isolated by two consecutive agar dilution series in basal medium amended only with glucose (5 mM), following the protocols of Pfennig and Trüper (43).

AM5 basal medium was (3) reduced with 2 mM dithiothreitol (DTT) and kept under a headspace of N_2–CO_2 (80:20, vol/vol). For enrichments, the medium was supplemented with Casamino acids and yeast extract (Becton, Dickinson and Company, Le Pont de Claix, France, 0.4% each). For isolation and growth experiments, each supplement was at 0.02%. Glucose (5 mM) was used as substrate unless indicated otherwise. Cultures were routinely grown in 16-ml rubber-stoppered culture tubes with 5 ml medium at 30 °C and pH 7. To assay pH dependence of growth, the pH of the medium was adjusted by adding sterile solutions of NaOH or HCl (each 1 M). To assay the salt tolerance, the medium was mixed at different ratios with medium amended with 3.5% NaCl + $MgCl_2$ (10:1, by wt.). Growth was determined photometrically by following the increase in optical density at 578 nm (OD_{578}). Growth yields were estimated by using an OD-to-cell-mass conversion factor that was determined with cultures (200 ml). Cell densities were estimated by differential interference contrast microscopy (DIC; see below) using a Thoma counting chamber.

Metabolic profile. Cultures tubes with basal medium AM5 were amended with the respective substrates, inoculated with 1% preculture, and incubated as described above. Supernatants of fully grown cultures were analyzed for substrate utilization and product formation by high-performance liquid chromatography (HPLC) using an ion-exclusion column (Resin-GPZH10812S2508, Grom, Herrenberg, Germany), a UV detector, and a refractive index detector (57). Metabolites were identified by comparing the retention times and the signal ratios of the UV and refractive index

detectors to those of authentic standards. Pyruvate was also verified via enzymatic assay using L-lactate dehydrogenase (EC 1.1.1.27) and standard procedures (6). Amino acids were quantified by precolumn derivatization with *o*-phthaldialdehyde and 3-mercaptopropionic acid followed by reversed-phase HPLC and fluorimetric analysis (330/450 nm) (12) on a SIL OPA-3 column (Grom) using gradient elution according to the manufacturer's instructions. Ammonia was measured colorimetrically (2). Hydrogen was analyzed by gas chromatography using a thermal conductivity detector (57). For computation of electron balances, all metabolites were formally oxidized to CO_2, and the number of electrons theoretically released from the respective amounts of products was compared with that of the amount of substrate consumed. Expressed on a percent basis, this calculation yielded the electron recovery as previously described (58).

Substrate spectra. Utilization of a wide range of substrates was tested with API-50CH test kits (Biomerieux, Marcy-l'Etoile, France) according to the manufacturer's instructions, with the following exceptions: The test strips were incubated in a 4-l anaerobic jar under a N_2/CO_2 atmosphere; the recommended medium was replaced with basal medium containing only 10 mM bicarbonate buffer, 2 mM glucose, and 18 mg l^{-1} phenol red; and the results were read after 4 weeks of incubation.

MALDI-TOF analysis of amino acids. Strain Pei191^T was incubated with 13C-labeled glucose (99 at%; Cambridge Isotope Laboratories, Andover, Mass., USA). The supernatant of a fully grown culture was mixed (1:2:1, by vol) with internal standard solution (1 mM aminoadipic acid) and a matrix solution containing 0.4 % (w/v) α-cyano-4-hydroxycinnamic acid, 0.1% tetrafluoroacetic acid, and 70% acetonitrile. A small drop of the mixture (1 µl) was dried on a gold target and examined in an ABI 4800 Plus MALDI TOF/TOF analyzer using a molecular mass window from 55 to 160 (25). A culture with unlabeled glucose, uninoculated medium, and uninoculated medium containing the 20 proteinogenic amino acids (1 mM each) were treated the same way and served as controls or standards on the same target.

Light microscopy. Light microscopy was done with a Zeiss Axiophot epifluorescence microscope equipped with a cooled CCD camera; filter sets for fluorescence microscopy were described by Schmitt-Wagner et al. (50). Non-stained cultures were routinely examined at 1250× magnification using DIC microscopy.

Gram staining was done according to Süßmuth et al. (54); *Acetobacterium woodii* (DSM1030^T) and *Desulfovibrio vulgaris* (DSM2119) were used as controls. For 4',6-diamidino-2-phenylindol (DAPI) staining, 3-ml samples were acidified with HCl (40 mM), fixed with formaldehyde (3%), mixed with 1 ml DAPI solution (100 μg/ml), and incubated for 20 min at room temperature. The samples were filtered onto a black 0.2-μm polycarbonate membrane (Cyclopore Track-Etched Membrane, Whatman). Color images were recorded using a Chroma 31000 filter set (Chroma, Fürstenfeldbruck, Germany), inverted, and enhanced in contrast with image-editing software, using only the green and blue channels of the image.

Fluorescence *in situ* hybridization (FISH) was performed according to Pernthaler et al. (42) with a specific CY3-labelled probe (Elm1034: 5'- GCA GCA CCT CGG CTG GCT TT -3') and the fluorescein-labeled general eubacterial probe EUB338 (1). *Desulfurella acetivorans* (DSM5264) served as negative control. Optimal hybridization conditions were determined by increasing the formamide concentration from 0 to 50% in steps of 10%.

Electron microscopy. For transmission electron microscopy (TEM), cells were fixed directly in the growth medium with 3% (w/v) glutaraldehyde for 1 h, gently centrifuged, and washed twice in phosphate buffer, followed by a 2% (w/v) osmium tetroxide fixation. Cells were dehydrated and embedded in Epon 812 resin using standard procedures (66).

Negative contrast electron microscopy was done according to Valentine et al. (64), except that the cells were fixed with 2% (w/v) glutaraldehyde directly in the growth medium. After gentle centrifugation, cells were resuspended and placed on hydrophilized or non-hydrophilized carbon-coated 400-mesh grids. After

preparation with 2% (w/v) uranyl acetate, specimens were immediately examined by electron microscopy.

Photomicrographs were taken in a Philips 301G electron microscope using negative film, then digitally scanned and enhanced in contrast with image-editing software.

PCR-based screening. 16S rRNA gene sequences of the "Intestinal cluster" were amplified with primer set 3 (14), using DNA extracted from the following samples: horse feces; rabbit feces; cow rumen; and gut homogenates of lower termites (*Hodotermopsis sjoestedti*, *Kalotermes flavicollis*, *Zootermopsis nevadensis*; pseudergates), higher termites (*Nasutitermes corniger*, *Cubitermes ugandensis*; worker caste), cockroaches (*Blaberus giganteus*, *Nauphoeta cinnerea*; adults), scarab beetle larva (*P. ephippiata*), lepidopteran larva (*Manduca sexta*), crickets (*Achaeta domestica*), and locusts (*Schistocerca gregaria*). PCR products were cloned and sequenced as previously described (14).

16S rRNA genes of the isolates were amplified by PCR using 27f (8) and 1492r (29) bacterial primers, and sequenced on both strands.

Phylogenetic analysis. The 16S rRNA gene sequence of strain Pei191^T was fitted into an alignment of about 270,000 bacterial sequences in the ARB-Silva database (44; version 96, http://www.arb-silva.de), using the automated tools of the ARB software package (33; version 08.03.14org, http://www.arb-home.de). Putative chimerae in the Silva database were removed. Highly variable base positions were removed from the alignment using a frequency-based filter (50% criterion) calculated for all high-quality near-full-length sequences in the TG1 phylum (190); only 20 selected sequences from the Endomircrobia (Cluster I) were included to avoid bias towards this highly overrepresented group. A maximum-likelihood tree was constructed using fastDNAml (39); tree topology was tested using RaxML (52) and maximum-parsimony analysis (1000 bootstraps each); also these tools are implemented in the ARB software package. Tree topologies were challenged by using different data sets varying with respect to positional masking and outgroup. Insufficiently resolved nodes are reported as a multifurcation. The sequences in the

"Intestinal cluster" were analyzed in the same way, except that the filter accommodated only the base positions covered by primer set 3 (14).

Other analyses. Whole-cell fatty acids, respiratory lipoquinones, and polar lipids were analyzed by the Identification Service and B. J. Tindall, Deutsche Sammlung für Mikroorganismen und Zellkulturen (DSMZ), Braunschweig, Germany, using 0.5 g freeze-dried cell material washed with distilled water. Fatty acids were analyzed according to Kämpfer and Kroppenstedt (24). Lipoquinones and polar lipids were extracted and analyzed by two-dimensional thin-layer chromatography (TLC) (60, 61, 62).

Isolation and morphological characterization

During the course of a cultivation-based characterization of the larval gut microbiota of *Pachnoda ephippiata*, we observed that uninoculated culture tubes amended with clarified gut homogenate showed microbial growth after 3–6 months of storage. Since non-supplemented culture tubes from the same batch had remained sterile, we suspected that the contamination was introduced with the sterile-filtered gut homogenate. 16S rRNA gene sequencing of the DNA extracted from the tubes cultures indicated that the contaminating cells were members of the Termite Group 1 (TG1), a bacterial candidate phylum without any cultured representatives.

We isolated the contaminating bacterium using a deep-agar dilution series. Small, white, lentil-shaped colonies appeared after 4 weeks. A pure culture, strain Pei191^T, was obtained from the highest dilution. The 16S rRNA gene sequence of strain Pei191^T was identical to the sequence obtained from the original culture. Repetition of the procedure with sterile-filtered gut homogenates obtained from a different batch of *P. ephippiata* larvae led to the isolation of strain Pei192, which had a 16S rRNA gene sequence identical to that of Pei191^T. A third strain, Pei193, again with an identical sequence, was obtained from a liquid dilution series out of a tube that was inoculated with an equivalent of 10^{-10} ml unfiltered gut homogenate of *P. ephippiata*.

Owing to their extremely small size, the isolates were very difficult to observe by light microscopy. Differential interference contrast (DIC) light microscopy and transmission electron microscopy revealed that growing cultures consisted of rod-shaped cells of variable length (0.3–2.5 μm), but images were poorly resolved. This was confirmed by transmission electron microscopy (Figures. 1A–E), which additionally revealed that the cells had an extremely small diameter (0.17–0.30 μm). However, also larger coccoid cells and cells with undefined shape were observed in all growth stages (see below).

To exclude that this apparent pleomorphism is caused by the presence of a second bacterium, we confirmed the purity of the cultures by simultaneous fluorescence *in situ* hybridization (FISH) with the oligonucleotide probe EUB338, targeting almost all Bacteria, and the probe Elm1034, specifically designed to target the 16S rRNA of strain Pei191^T. A formamide series showed that probe ELM1034 discriminated *Desulfurella acetivorans*, the isolate with the least number of mismatches, in the range of 10–40% formamide (Table S1); the optimal condition for the simultaneous hybridization of strain Pei191^T with both probes was at 30% formamide concentration. Subsequent examination of various cultures of strain Pei191^T in different growth stages revealed that all morphotypes hybridized with both probes, corroborating that — irrespective of their shape — all cells belong to the same phylotype.

Strain Pei191^T was Gram-negative in the classical Gram stain. TEM of ultra-thin sections of Pei191 cultures revealed that the cell envelope consists of two membranes with a thickness of 6 nm and a periplasmic space with a thickness of 6–8 nm (Figure 1B); a thin structure possibly representing a peptidoglycan layer was only poorly resolved in some negative stains (Figure 1D). Sometimes, the outer membrane showed protuberances, which were scarce in exponentially growing cultures but numerous in the late stationary or death phase (details not shown). Structures resembling flagella or spores were never observed. Both TEM and negative-contrast EM showed that the poles of the rod-shaped cells were densely filled with ribosomes, whereas the center of the cells, probably representing the nucleoid, was invariably free of ribosomes (Figures. 1A, D, E). The central location

of the DNA is confirmed by the bright fluorescence of this region in DAPI-stained wet mounts (Figure 1F), where DNA gives a much stronger signal than RNA (56). Based on the negative-contrast images obtained with non-hydrophilized grids (Figure 1D), the average volume of the nucleoid, obtained by geometric approximation, was only 0.010 μm³ (ranging from 0.006 to 0.020 μm³).

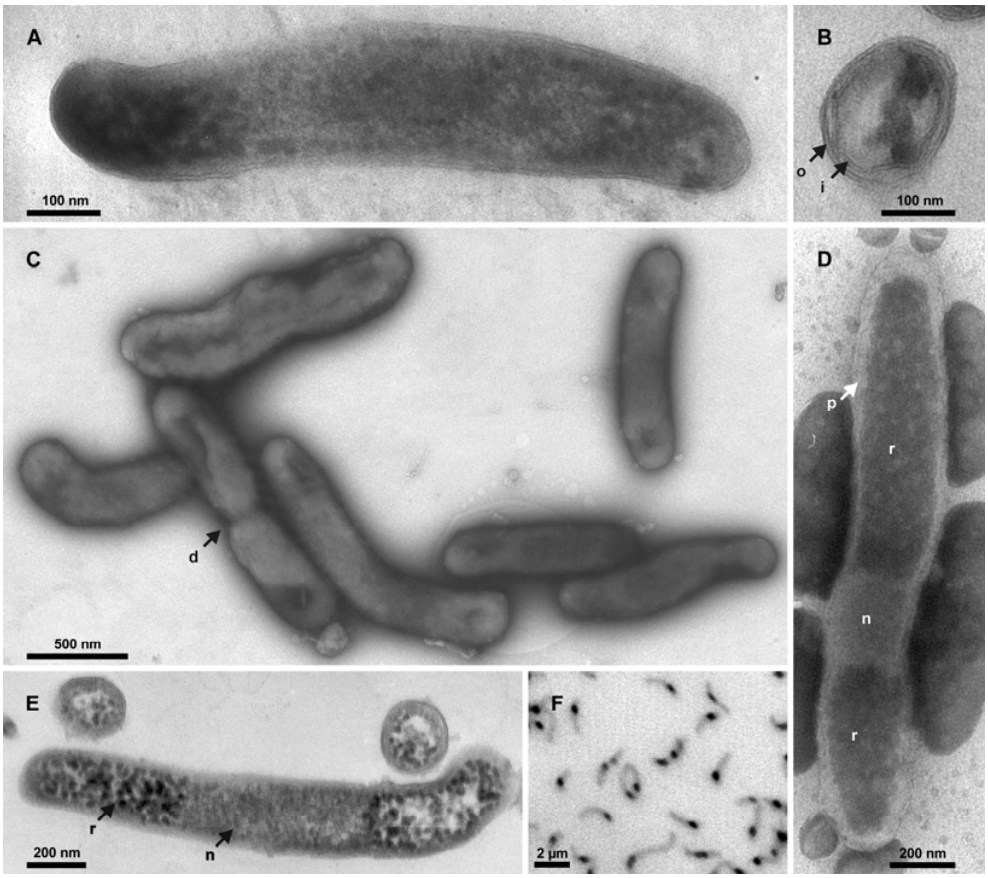


Figure 1. Images of strain Pei191^T. (A) TEM image of a longitudinal section of a smaller cell, showing both cell poles delimited by the outer membrane. (B) TEM image of a radial section resolving both inner and outer membranes and several ribosomes. (C) Negative-contrast TEM image of cells in the mid-exponential growth phase prepared on a hydrophilized grid. (D) Negative-contrast TEM image of a cell prepared on a non-hydrophilized grid. (E) TEM image of longitudinal and radial sections of larger cells. (F) Fluorescence photomicrograph (negative image) of DAPI-stained cells in the mid-exponential growth phase. Arrows denote cell division (d), inner membrane (i), nucleoid (n), outer membrane (o), periplasmic space (p), riboplasm (r).

Growth and nutrition

In the API test, strain Pei191^T showed moderate acid production from D-glucose, D-fructose, D-galactose, and *N*-acetyl-D-glucosamine, and only weak acid production

from D-ribose, methyl-α-D-mannopyranoside, D-turanose, potassium 2-ketogluconate, and potassium 5-ketogluconate. No acid was produced from D-adonitol, amygdalin, D-arabinose, L-arabinose, D-arabitol, L-arabitol, arbutin, D-cellobiose, dulcitol, erythritol, D-fucose, L-fucose, gentiobiose, potassium gluconate, methyl-β-D-glucopyranoside, glycerol, glycogen, inositol, inulin, D-lactose, D-lyxose, D-maltose, D-mannitol, D-mannose, D-melicitose, D-melibiose, D-raffinose, L-rhamnose, salicin, D-sorbitol, L-sorbose, starch, D-sucrose, D-tagatose, D-threhalose, xylitol, methyl-β-D-xylopyranoside, D-xylose, L-xylose. In liquid culture, strain Pei191^T grew on D-glucose, D-fructose, D-galactose, D-glucosamine, and *N*-acetyl-D-glucosamine, which were fermented to acetate, ethanol, lactate, hydrogen, and alanine (Table 1). No growth was observed on lactate, pyruvate, formate, or amino acids. Casein, xylan, arabinogalactan, microcrystalline cellulose (Avicel), and filter paper were not used. Fumarate was not reduced. Oxygen, nitrate, and sulfate were not used as electron acceptors. Thiosulfate (5 mM) inhibited growth.

Table 1. Growth yields and fermentation products of strain Pei191^T on basal medium (0.02% yeast extract) with different substrates. Values are normalized to 1 l medium.

Substrate	Casamino acids	Substrate consumed	Cell biomass [a]	Products formed (mmol) [b]					Electron recovery [c]	Growth yield
	(0.02%)	(mmol)	(mg)	Acetate	Ethanol	Lactate	H_2	Amino acids [e]	(%)	(g mol^{-1})
Glucose	+	5.0	149	3.0	3.1	0.3	7.8	0.8	101	29.8
Glucose	–	4.1 [f]	99	2.2	1.9	0.3	5.7	1.2	97	24.4
Fructose	+	5.0	90	3.3	5.0	0.4	6.5	n.d.[g]	100	18.0
N-acetyl-D-glucosamine	+	5.0	82	8.6	4.1	0.3	8.2	n.d.[g]	96	16.5

[a] Calculated from the increase in optical density using the OD-to-dry-mass ratio of 276 mg l^{-1} at OD_{578} = 1, experimentally determined for glucose-grown cells.
[b,c] Formate, propionate, and succinate were below the detection limit (0.01 mM). Butyrate, 2-methylprionate, and 2- or 3-methylbutyrate, phenylacetate, and 3-hydroyphenylacetate were detected in minor amounts.0.1 mmol of pyruvate were also detected in all samples and pyruvate was confirmed by enzymatic assay.
[d] Electron recovery was calculated assuming the formula <$C_4H_8O_2N$> for cell biomass. Electron recovery in amino acids (produced – consumed) was calculated separately for each amino acid, and the result was added to that obtained for the other metabolites. The contribution of yeast extract to electron recovery was estimated with < 3% (details not shown).
[e] Net increase of total amino acids.
[f] Slow growth and incomplete use of glucose in the absence of Casamino acids.
[g] Not determined.

Growth on glucose (5 mM) required both yeast extract and Casamino acids at a minimum concentration of 0.02% each. No growth occurred in the absence of yeast extract, and at lower concentrations of yeast extract, glucose was not completely

consumed. In the absence of Casamino acids, glucose consumption was incomplete. Casamino acids could be replaced by the addition of individual amino acids, which were consumed during growth, albeit to different extents (Table 2), but not by an equivalent amount of casein. Additives that stimulated growth but were apparently not consumed were L-arginine, L-β-alanine, L-2-aminobutyrate, DL-3-aminobutyrate, L-5-aminovalerate, L-6-aminocaproate, 11-aminoundecanoic acid, DL-threo-β-methylaspartate, L-citrulline, L-ornithine, putrescine, cadaverine, 2-aminopropanol, 3-aminopropanol, and 4-aminobutanol. Growth was also stimulated by the addition of L-cysteine, L-cystine, L-proline, 3-aminobenzoic acid, 2-aminophenol, 3-aminophenol, urea, and uric acid (all 2 mM).

For several amino acids, it was possible to identify the corresponding oxidative decarboxylation products among the fermentation products formed during growth (Table 2). A nitrogen balance of glucose-grown cultures indicated a net formation of amino acids (0.5–1.2 mM), mostly alanine, and a net consumption of ammonia (0.5–1 mM) (Figure 2). Alanine was also the only amino acid that did not stimulate growth on glucose.

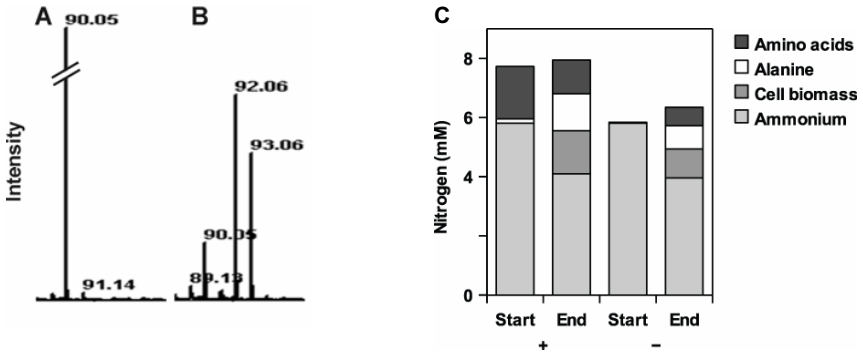


Figure 3. MALDI-TOF spectra of the supernatants of cultures of strain Pei191^T grown with ^{12}C-glucose (A) or ^{13}C-glucose (B). Only the mass range relevant for ^{12}C-alanine (90) and ^{13}C-alanine (92, 93) is shown. (C) Nitrogen content in different metabolites of strain Pei191^T cultures grown with (+) or without (–) Casamino acids (0.02%), immediately after inoculation and at the end of the growth phase.

MALDI-TOF analysis of culture supernatants revealed that cultures grown on uniformly ^{13}C-labeled glucose formed alanine labeled in two and three C-atoms

(Figure 2). Less than 15% of the total alanine was unlabeled, indicating that most of the alanine was formed from the carbon skeleton of glucose.

On basal medium with glucose, the cells grew within a pH range from 6.2 to 8.2. Growth was possible in temperature range from 20 to 32 °C, but not at 15 or 37 °C. Highest growth rates were obtained at 30 °C and pH 7.5. Growth was best in freshwater medium, but strain Pei191^T grew also at 3.5% salt concentration; the lag phase increased greatly (20–40 days) at concentrations above 1.5%.

The growth yield on glucose was 29.8 g (dry wt.) mol^{-1}; substrate-free controls showed that growth on basal medium was negligible. Cultures grown under this condition showed cell densities (normalized to OD_{578} = 1) of 10^{10} cells ml^{-1} or 75.1 µg (dry wt.) ml^{-1}. The doubling time during growth on glucose was 19 h (Figure 4A); higher growth rates (t_d = 11 h) were observed when Casamino acids and yeast extract were increased to 0.1% (w/v) each. When cultures of strain Pei191^T were filtered through a 0.2-µm membrane filter, subcultures on fresh medium grew very slowly. Even after the second transfer, doubling times were still >80 h, but decreased again with subsequent transfers.

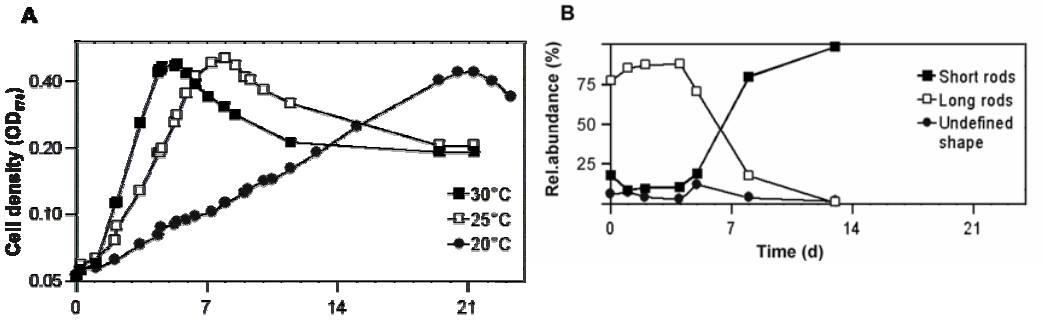


Figure 4. (A) Growth curves of strain Pei191^T at different temperatures (semi-logarithmic plot). (B) Relative abundance of the major cell shapes during the different growth phases, illustrating the increase of short forms in the stationary phase (growth temperature was 30 °C).

Strain Pei191^T required a reduced medium. Although the bacterium did not grow in the presence of oxygen (0.5%) even in static cultures, the redox zonation (visualized by resorufin) showed that cells pre-grown in deep-agar tubes retarded the diffusive influx of oxygen into the tubes considerably longer than uninoculated controls,

indicating the principal capacity to reduce molecular oxygen if it does not accumulate to high concentrations.

Chemotaxonomic analysis

Analysis of the whole-cell fatty acid composition of strain Pei191^T revealed an unusual pattern dominated by *iso*- and *anteiso*-fatty acids (Table S2). There were no similarities to the pattern of any species included in the MIS database (MIDI, Newark, Del., USA). Polar-lipid analysis showed phosphatidylethanolamine and phosphatidylglycerol as major components; numerous unidentified phospholipids and amino-phospholipids were present (Figure S1). Quinones were not detected.

Table 2. Consumption of amino acids added individually to cultures of strain Pei191^T growing on glucose (5 mM) and yeast extract (0.02%). The latter contained no free amino acids. For amino acids that were not consumed, see text.

Amino acid (2 mM) [a]	Consumed (mM)	Products observed [b]
L-Asparagine	2.0	n.d.
L-Serine	2.0	Lactate, pyruvate [c]
L-Glutamine	1.8	Unidentified
L-Threonine	1.4	Glycine
L-Methionine	1.4	Unidentified
L-Leucine	1.4	Butyrate, 3-methylbutyrate
L-Glycine	1.3	Pyruvate
L-Lysine	1.3	Unidentified
L-Isoleucine	1.2	2-Methylbutyrate
L-Tryptophan	1.1	n.d.
L-Phenylalanine	1.0	Phenylacetic acid, unidentified
L-Valine	1.0	2-Methylpropionate
Diaminopimelate (1 mM)	1.0	Lysine
L-Ornithine	0.9	n.d.
L-Glutamic acid	0.7	Unidentified
L-Histidine	0.5	n.d.
L-4-Aminobutyrate	0.4	Lactate, pyruvate
L-Aspartic acid	0.4	n.d.
L-Tyrosine	0.3	4-Hydroxyphenylacetate

[a] Initial concentration
[b] In addition to the products formed from glucose; n.d., not detected
[c] The identity of pyruvate was confirmed by enzymatic assay.

Phylogenetic analysis

Already a preliminary analysis of the 16S rRNA gene sequence of strain Pei191^T had indicated that it belongs to a phylogenetic lineage that so far contains no cultivated representatives, the TG1 phylum (14). Phylogenetic analysis of all near-full-length sequences in the TG1 phylum confirmed that strain Pei191^T falls into the so-called "Intestinal cluster" (cluster III in Figure 5), a stable monophyletic group that comprises clones obtained exclusively from gut-derived samples of termites (14, 35), cows (14, 55), and chimpanzees (31). However, there were also several novel deep-rooting lineages in addition to those already described two years ago (14).

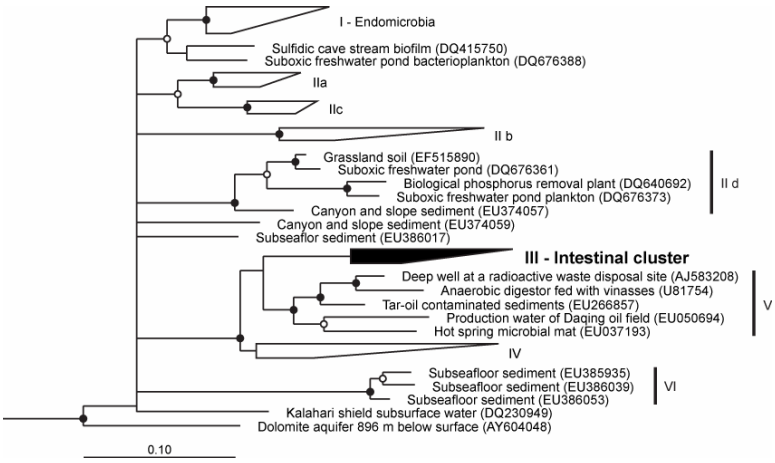


Figure 5. Maximum-likelihood tree of the TG1 phylum, based on an alignment (1260 base positions) of all available near-full-length 16S rRNA gene sequences. Representative members of other phyla were used as outgroup. Roman numerals indicate the major phylogenetic lineages in the TG1 phylum. The tree includes 21 novel sequences that were submitted to GenBank since the first phylogenetic study of the TG1 phylum (14) was published, but details are shown only for those lineages (IId, V, VI) that were not yet defined in that study.

Using a primer set specifically designed to detect members of this cluster (primer set 3; 14), we obtained additional sequences from the guts of numerous wood- and soil-feeding termites and cockroaches, and from horse feces. No TG1 sequences were obtained from the guts of several other insects (*Manduca sexta*, *Acheta domestica*, and *Schistocerca gregaria*) or from rabbit feces. Phylogenetic analysis combining the sequences in cluster III with all partial sequences obtained with

primer set 3 showed that the "Intestinal cluster" consists of two major lineages (Figure 6). One lineage comprises the majority of the clones originating from termites and cockroaches (IIIa). The other lineage comprises several subclusters, containing clones from termites (IIIb), the isolates, the sequence-identical clones from P. ephippiata gut homogenate (AM491098), and a clone from a cockroach (IIIc), and — well separated from the other subclusters — clones from mammalian samples (IIId).

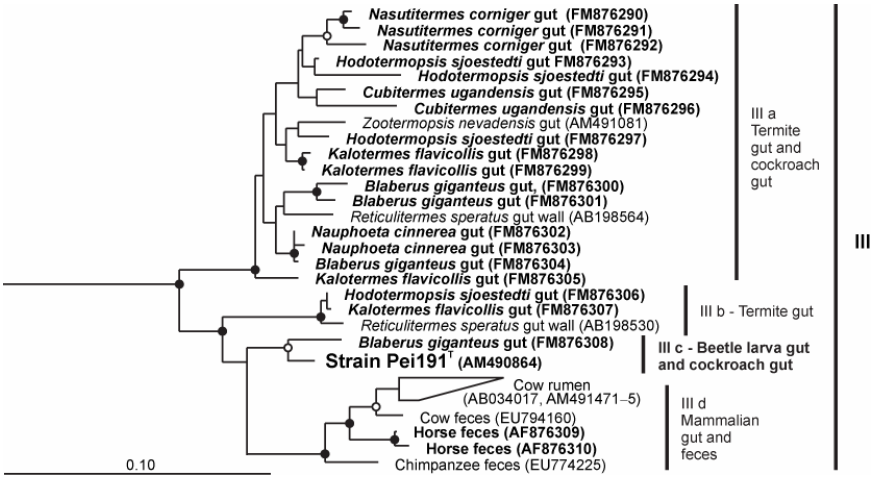


Figure 6. Maximum-likelihood tree of all 16S rRNA gene sequences in the "Intestinal cluster" (III), based on an alignment of 910 base positions. Sequences obtained with primer set 3 (this study) are marked in bold. Sequences in clusters IV and V were included in the analysis and used as outgroup (not shown). In both trees, node support is indicated only if bootstrap values were > 70% (open circles) and > 95% (filled circles) in both ML and MP analyses.

Physiology

We isolated the first representative of the TG1 phylum, so far predicted only by cultivation-independent methods. Strain Pei191^T is a member of a phylogenetic lineage of hitherto uncultivated bacteria occurring in the intestinal tract of insects. It is a strictly anaerobic ultramicrobacterium with a Gram-negative cell wall structure and a purely fermentative energy metabolism.

The scope of substrates used by strain Pei191^T is quite limited. Only a few sugars and amino sugars are fermented to acetate, ethanol, CO_2, and H_2 — products that are typical of many strict anaerobes (16). Unusual is the production of alanine from glucose. The results of the tracer study indicate that alanine is not formed from other amino acids present in the medium but is rather derived from the carbon skeleton of glucose. Alanine is most likely formed by transamination of pyruvate, since it carried the ^{13}C label in two or three C-atoms, a phenomenon caused by the exchange of the carboxyl group of pyruvate with the unlabeled bicarbonate pool in the medium (67).

Formation of alanine as the end product of anaerobic glucose metabolism has been observed in several eukaryotes, e.g., *Sacharomyces cerevisiae* (5), the isopod *Saduria* (*Mesidotea*) *entomon* (13), and *Giardia lamblia*, where alanine production was proposed to be a redox-balancing reaction (41). In prokaryotes, this metabolism has been reported only for a few deep-branching hyperthermophilic archaea (*Thermococcales*; 26, 27) and bacteria (*Thermotogales*; 47). It has even been discussed whether this is a remnant of an ancestral metabolism (47). Alanine is the major end product of glucose fermentation in a thermophilic clostridial strain when incubated at high ammonium concentrations (40). Remarkably high alanine concentrations were observed already in early studies of the cow rumen (68), an environment characterized by an ample supply of amino acids.

Strain Pei191^T did not grow on amino acids, neither by fermentation of single amino acids nor by Stickland reactions, i.e., the fermentation of amino acid pairs (36). Instead, it fermented amino acids only during growth on glucose, presumably by transamination with pyruvate and subsequent oxidative decarboxylation of the resulting 2-oxoacid, as described for *Pyrococcus furiosus* (26). This is indicated by the formation of the corresponding oxidative decarboxylation product from most proteinogenic amino acids and the net formation of alanine concomitant to the disappearance of the respective amino acid from the medium. However, transamination of pyruvate may not be the only mechanism of alanine formation. The consumption of ammonium in the growth medium, which significantly exceeds the nitrogen requirements for biomass formation, indicates that part of the alanine is

formed by a net amination, as reported for *Thermococcus profundus* (27). The strong stimulation of growth by the addition of Casamino acids or individual amino acids and the inhibitory effect of alanine addition underscore that alanine formation is an integral component of the energy metabolism of strain Pei191^T. However, growth on glucose was stimulated also by other metabolites containing amino groups (e.g., urea or amino alcohols), which indicates that the situation may be even more complex.

Ecology

Strain Pei191^T was isolated from the hindgut of *Pachnoda ephippiata*, a humus-feeding beetle larva. Other members of the "Intestinal cluster" have been retrieved from the intestinal tracts of other insects, including the termite gut, but were also detected in the cow rumen and in the feces of horses and chimpanzees. Common features of these habitats are high concentrations of carbohydrates and amino acids. In the case of *P. ephippiata* larvae, the hydrolysis products of peptides and other nitrogenous components of soil organic matter (chitin, peptidoglycan) accumulate to millimolar concentrations (2, 32). With its combined metabolism of carbohydrates and amino acids, strain Pei191^T seems to be well adapted to the supply of nutrients in its habitat.

Another adaptation to the intestinal habitat may be in the apparent ability of strain Pei191^T to remove traces of oxygen. Although it did not grow in liquid medium if the medium was not completely reduced, it was capable of retarding the influx of oxygen into agar tubes containing cells pregrown under anoxic conditions. This feature resembles the situation in *Methanobrevibacter cuticularis*, which colonizes the gut epithelium of termites. *M. cuticularis* is extremely sensitive to the presence of oxygen but can reduce it at astonishingly high rates — a trait that may explain why it remains metabolically active despite the continuous influx of oxygen into its natural habitat (59). The exact location of strain Pei191^T in the gut of *P. ephippiata* remains to be established, but other, hitherto uncultivated members of the "Intestinal cluster" have been obtained specifically from the gut wall of the termite gut (35).

In contrast to the distantly related Endomircrobia, which densely colonize the cytoplasm of termite gut flagellates (21, 38, 53), members of the other clades in the TG1 phylum are rare in environmental clone libraries (14). This may reflect their low abundance in most other environments, but may be related in part also to a low *rrn* gene copy number among the members of this phylum (15, 19). Also, the isolation of a strain with an identical 16S rRNA gene sequence from a high dilution of *P. ephippiata* gut homogenate suggests that strain Pei191^T is more abundant than indicated by molecular approaches. In view of the small size of strain Pei191^T, our failure to detect members of the "Intestinal cluster" by dot blot hybridization (48) of RNA extracted from termite guts (data not shown) may be explained by the small number of ribosomes present in slow-growing ultramicrobacteria (11).

Size

Strain Pei191^T is so small that it passes through a 0.2-µm filter membrane typically used for sterilization, a feature that originally led to the concept of ultramicrobacteria (63). Later definitions restricted the term ultramicrobacteria to cells with a volume < 0.1 µm^3 (10). Most isolates are slow-growing marine oligotrophs (51, 46), but a few non-oligotrophic ultramicrobacteria have been retrieved from terrestrial habitats (65, 23). Strain Pei191^T is the first ultramicrobacterium that has been obtained from a gut environment. The reason for its accidental isolation lies in its extraordinarily small size. Its relatively slow growth on non-selective media — probably combined with a small population size — are likely explanations why strain Pei191^T and other members of the "Intestinal cluster" have so far eluded isolation (14).

The genome of strain Pei191^T, which was recently sequenced, has a size of 1.64 Mbp (15). Since the volume of the nucleoid is extremely small (0.006−0.020 µm³, the range probably reflecting the inevitable increase in DNA content during replication), the DNA of strain Pei191^T must be tightly packed. Using the volume taken up by a DNA molecule in aqueous solution (34) and assuming the average volume of a single copy of the chromosome to be 0.008 µm³, we estimated that the

DNA of strain Pei191^T takes up only fourfold the volume of the hydrated DNA molecule even during the exponential growth phase.

Since a small cell can harbor only a limited number of ribosomes, the increase in cell length during the exponential growth phase may reflect the need to accommodate enough ribosomes to achieve a high growth rate. A tenfold increase in ribosome numbers during the transition from stationary phase to exponential growth phase has been reported for the ultramicrobacterium *Sphingomonas* sp. strain RB2256 (10), and growth rate and ribosome number are linked also in *Escherichia coli* (4).

The minimal cell size is not only limited by the cytoplasmic volume required by the ribosomes, but also by the DNA. Using the model for the estimation of a theoretical lower size limit of living cells summarized by De Duve and Osborn (7) in a workshop on "Size limits of very small microorganisms", corrected for the minimum volume taken up by the hydrated DNA of strain Pei191^T (single copy), we estimated the lower size limit for the cytoplasm of Pei191^T to be 0.008 μm^3. Interestingly, this is in good agreement with the size of the cytoplasm in the smallest cells of Pei191^T determined by geometric approximation (Figure 1A).

Consequently, the reduction of both genome size and ribosome numbers allows bacteria to achieve an extremely small cell size. A small ribosome number results in a low growth rate (4) and would also explain why we found a fourfold longer doubling time in cultures sterile-filtered before inoculation: this mechanism would select for small cells, which have fewer ribosomes and hence grow even more slowly than the average culture.

Taxonomy

Since strain Pei191^T is only remotely related to any other bacterial taxon, we propose a new genus and species, *Elusimicrobium minutum*, with strain Pei191^T as the type strain. *E. minutum* is the first cultivated representative of the candidate phylum TG1. To date, only a putative taxon, "*Candidatus* Endomicrobium", has

been described; it consists of uncultured endosymbionts in flagellate protozoa in the gut of termites (53).

E. minutum is a member of the "Intestinal cluster" (cluster III), which forms a distinct clade (>15% sequence divergence to other clusters) in the TG1 phylum. Members of this cluster are exclusively found in intestinal habitats, ranging from the guts of termites, cockroaches, and beetle larvae to the mammalian intestinal tract. We propose the family *Elusimicrobiaceae* for all members of cluster III, based on the clear monophyly of this cluster and a within-cluster sequence divergence (0−10%) that is typical of the family level in other phyla. In addition, we propose the order *Elusimicrobiales* for the well-supported supercluster formed by clusters III, IV, and V, and the phylum *Elusimicrobia* for all members of the TG1 phylum, with *E. minutum* as the first cultivated representative.

Description of the phylum Elusimicrobia phyl. nov. *Elusimicrobia* (E.lu'si.mi.cro'bi.a. N.L. neut. n. *Elusimicrobium*, a genus of Bacteria; N.L. neut. pl. n. *Elusimicrobia* phylum of the genus *Elusimicrobium*) A lineage of bacteria encompassing the candidate genus "*Endomicrobium*", the genus *Elusimicrobium* and numerous hitherto uncultivated phylotypes that form a stable group according to 16S rRNA gene sequence analysis.
Type order: *Elusimicrobiales* ord. nov.

Description of *Elusimicrobia* classis nov. The description is the same as for the order *Elusimicrobiales*.
Type order: *Elusimicrobiales* ord. nov.
Description of *Elusimicrobiales* ord. nov. *Elusimicrobiales* (E.lu'si.mi.cro.bi.al'es. N.L. neut. pl. n. *Elusimicrobium* type genus of the order; *-ales* ending to denote an order; N.L. neut. pl. n. *Elusimicrobiales* the order of the genus *Elusimicrobium*). It is proposed that the order accommodates bacterial isolates that according to 16S rRNA gene sequence analysis form a stable group within the supercluster formed by clusters III, IV, and V.
Type genus: *Elusimicrobium* gen. nov.

Description of *Elusimicrobiaceae* fam. nov. *Elusimicrobiaceae* (E.lu'si.mi.cro.bi.a.ce'ae. N.L. neut. n. *Elusimicrobium* type genus of the family; -*aceae* ending to denote a family; N.L. neut. n. *Elusimicrobiaceae* family of the genus *Elusimicrobium*). The description is the same as for the genus *Elusimicrobium*. It is proposed that the order accommodates bacterial isolates that according to 16S rRNA gene sequence analysis form a stable group with cluster III. Type genus: *Elusimicrobium* gen. nov.

Description of *Elusimicrobium* gen. nov. *Elusimicrobium* (E.lu'si.mi.cro'bi.um. L. part. adj. *elusus* escaped from capture; N.L. neut. n. *microbium*, microbe; N.L. neut. n. *elusimicrobium* an elusive microbe, hard to find, capture, or isolate). Gram-negative staining, non-motile, rod-shaped cells. Obligately anaerobic and catalase negative. Heterotrophic, purely fermentative metabolism; nitrate, sulfate, and thiosulfate are not reduced.
Type species: *Elusimicrobium minutum* gen. nov., sp. nov.

Description of *Elusimicrobium minutum* sp. nov. *Elusimicrobium minutum* (mi.nu'tum L. neut. part. adj. *minutum*, very small, minute). Cells are typically rod-shaped, but cultures are pleomorphic in all growth phases (0.3–2.5 µm long and 0.17–0.3 µm wide); the smallest cells pass through 0.2-µm membrane filters. Acid production from D-galactose, D-glucose, D-fructose, D-glucosamine, and *N*-acetyl-D-glucosamine. Weak acid production from D-ribose, methyl-α-D-mannopyranoside, D-turanose, potassium 2-ketogluconate, and potassium 5-ketogluconate. Growth on glucose requires yeast extract and Casamino acids (0.02% each); the latter can be replaced by individual amino acids (except alanine). Major end products are acetate, ethanol, lactate, hydrogen, and alanine. Temperature range of growth 20–32 °C, optimum at 30 °C; no growth at 15 and 37 °C. pH range of growth 6.5–8.5; optimum at pH 7.5. The major fatty acids are C15:0 *iso*, C15:0 *anteiso*, and C16:0 *iso*.
Habitat: the intestinal tract of the humus-feeding larva of *Pachnoda ephippiata.*
Type strain: Pei191^T (= ATCC BAA-1559^T = JCM 14958^T). Genbank accession number: AM490846.

Acknowledgments. We thank Jörg Kahnt for MALDI-TOF analysis, Shuning Wang for help with pyruvate measurements, Marianne Johannsen for technical assistance with electron microscopy, and Hans G. Trüper for etymological advice. We are also grateful to Uwe Deggelmann, keeper of the zoological teaching collection of the Biology Department at Konstanz University, for providing *Pachnoda ephippiata* larvae.

References

1. **Amann, R. I., L. Krumholz, and D. A. Stahl.** 1990. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J. Bacteriol.* **172:**762–770.

2. **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. Insect Physiol.* **54:**105–113.

3. **Boga, H., and A. Brune.** 2003. Hydrogen-dependant oxygen reduction by homoacetogenic bacteria isolated from termite gut. *Appl. Environ. Microbiol.* **69:**779–783.

4. **Bremer, H., and P. P. Dennis.** 1996. Modulation of chemical composition and other parameters of the cell by growth rate, *In* Neidhardt, F. C., R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, H. E. Umbarger, eds., *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. ASM Press, Washington, DC, pp. 1553–1569.

5. **Chico, E., J. S. Olavarría, and I. Núnez de Castro.** 1978. L-Alanine as an end product of glycolysis in *Saccharomyces cerevisiae* growing under different hypoxic conditions. *Antonie van Leeuwenhoek* **44:**193–201.

6. **Czok, R. and W. Lamprecht.** 1970. Pyruvat, Phosphoenolpyruvat und d-Glycerat-2-phosphat, *In* H. U. Bergmeyer, ed., Methoden der Enzymatischen Analyse. Verlag Chemie, Weinheim/Bergstr, pp. 1407–1411.

7. **De Duve, C., and M. J. Osborn.** 1999. Discussion summary of panel 1. *In* Knoll, A., M. J. Osborn, J. Baross, H. C. Berg, N. R. Pace, and M. Sogin, Size limits of very small microorganisms, Proceedings of a workshop. National Academy Press, Washington DC, pp. 5–9.

8. **Edwards, U., T. Rogall, H. Blöcker, M. Emde, and E. C. Böttger.** 1989. Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res.* **17:**7843–7853.

9. **Egert, M., B. Wagner, T. Lemke, A. Brune, and M. W. Friedrich.** 2003. Microbial community structure in midgut and hindgut of the humus-feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). Appl. Environ. Microbiol. **69:**6659–6668.

10. **Eguchi, M., M. Ostrowski, F. Fegatella, J. Bowman, D. Nichols, T. Nishino, and R. Cavicchioli.** 2001. *Sphingomonas alaskensis* strain AFO1, an abundant oligotrophic ultramicrobacterium from the North Pacific. Appl. Environ. Microbiol. **67:**4945–4954.

11. **Fegatella, F., J. Lim, S. Kjelleberg, and R. Cavicchioli.** 1998. Implications of rRNA operon copy number and ribosome content in the marine oligotrophic ultramicrobacterium *Sphingomonas* sp. strain RB2256. *Appl. Environ. Microbiol.* **64:**4433–4438.

12. **Godel, H., T. Graser, P. Földi, P. Pfaender, and P. Fürst.** 1984. Measurement of free amino acids in human biological fluids by high-performance liquid chromatography. *J. Chromatogr.* **297:**49–61.

13. **Hagerman, L., and A. Szaniawska.** 1990. Anaerobic metabolic strategy of the glacial relict isopod *Saduria* (*Mesidotea*) *entomon*. *Mar. Ecol. Prog. Ser.* **59:**91–96.

14. **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.

15. **Herlemann, D. P. R., O. Geissinger, W. Ikeda-Ohtsubo, V. Kunin, H. Sun, A. Lapidus, P. Hugenholtz, and A. Brune.** Genome analysis of *Elusimicrobium minutum*, the first cultivated representative of the Elusimicrobia phylum (formerly Termite Group 1). *Appl. Environ. Microbiol.* **75**:2841–2849.

16. **Holdeman, L. V., E. P. Cato,** and **W. E. C. Moore**. 1977. Anaerobe Laboratory Manual. Virginia Polytechnic Institute and State University, Blacksburg, Virginia.

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

18. **Hongoh, Y., P. Deevong, S. Hattori, T. Inoue, S. Noda, N. Noparatnaraporn, T. Kudo, and M. Ohkuma.** 2007. Phylogenetic diversity, localization, and cell morphologies of

members of the candidate phylum TG3 and a subphylum in the phylum *Fibrobacteres*, recently discovered bacterial groups dominant in termite guts. *Appl. Environ. Microbiol.* **72:**6780–6788.

19. **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.

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

21. **Ikeda-Ohtsubo, W., M. Desai, U. Stingl, and A. Brune.** 2007. Phylogenetic diversity of 'Endomircrobia' and their specific affiliation with termite gut flagellates. *Microbiology* **153:**3458–3465.

22. **Ikeda-Ohtsubo, W., and A. Brune.** 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and '*Candidatus* Endomicrobium trichonymphae'. *Mol. Ecol.* **18:**332-342.

23. **Janssen, P. H., A. Schuhmann, E. Mörschel, and F. A. Rainey.** 1997. Novel anaerobic ultramicrobacteria belonging to the *Verrucomicrobiales* lineage of bacterial descent isolated by dilution culture from anoxic rice paddy soil. *Appl. Environ. Microbiol.* **63:**1382–1388.

24. **Kämpfer, P., and R. M. Kroppenstedt.** 1996. Numerical analysis of fatty acid patterns of coryneform bacteria and related taxa. *Can. J. Microbiol.* **42:**989–1005.

25. **Karas, M., D. Bachmann, and F. Hillenkamp.** 1985. Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules. *Anal. Chem.* **57:**2935–2939.

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

27. **Kobayashi, T., S. Higuchi, K. Kimura, T. Kudo, and K. Horikoshi.** 1995. Properties of glutamate dehydrogenase and its involvement in alanine production in a hyperthermophilic archaeon, *Thermococcus profundus*. *J. Biochem.* **118:**587–592.

28. **Köhler, T., U. Stingl, K. Meuser, and A. Brune.** 2008. Novel lineages of *Planctomycetes* densely colonize the alkaline gut of soil-feeding termites (*Cubitermes* spp.). *Environ. Microbiol.* **10:**1260–1270.

29. **Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace.** 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci., U S A.* **82:**6955–6959.

30. **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.

31. **Ley, R. E., M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel, R. Knight, and J. I. Gordon.** 2008. Evolution of mammals and their gut microbes. *Science.* **320:**1647–1651.

32. **Li, X., and A. Brune.** 2007. Transformation and mineralization of soil organic nitrogen by the humivorous larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Plant Soil* **301:**233–244.

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. **Mandelkern, M., J. G. Elias, D. Eden, and D. M. Crothers.** 1981. The dimensions of DNA in solution. *J. Mol. Biol.* **152:**153–161.

35. **Nakajima, H., Y. Hongoh, R. Usami, T. Kudo, and M. Ohkuma.** 2005. Spatial distribution of bacterial phylotypes in the gut of the termite *Reticulitermes speratus* and the bacterial community colonizing the gut epithelium. *FEMS Microbiol. Ecol.* **54:**247–255.

36. **Nisman, B.** 1954. The Stickland reaction. *Bacteriol. Rev.* **18:**16–42.

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

38. **Ohkuma, M., Sato T, S. Noda, S. Ui, T. Kudo, and Y. Hongoh.** 2007. The candidate phylum 'Termite Group 1' of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol. Ecol.* **60:**467–476.

39. **Olsen, G. J., H. Matsuda, R. Hagstrom, and R. Overbeek.** 1994. fastDNAml: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Comput. Appl. Biosci.* **10:**41–48.

40. **Örlygsson, J., R. Anderson, and B. H. Svensson.** 1995. Alanine as an end product during fermentation of monosaccharides by *Clostridium* strain P2. *Antonie van Leeuwenhoek*. **273:**273–280.

41. **Paget, T. A., M. H. Raynor, D. W. Shipp, and D. Lloyd.** 1990. *Giardia lamblia* produces alanine anaerobically but not in the presence of oxygen. *Mol. Biochem. Parasitol.* **42:**63–67.

42. **Pernthaler, J., F. O. Glöckner, W. Schönhuber, and R. Amann.** 2001. Fluorescence *in situ* hybridization with rRNA-targeted oligonucleotide probes. *In* Pernthaler, J., F. O. Glöckner, W. Schönhuber, and R. Amann, Methods in Microbiology: Marine Microbiology. Academic Press, London, pp. 207-226.

43. **Pfennig, N., and Trüper H.G.** 1981. Isolation of members of the families *Chromatiaceae* and *Chlorobiaceae*. *In* Starr, M.P., H. Stolp, H. G. Trüper, A. Balows, H. G. Schlegel (eds.), The Prokaryotes, Springer-Verlag, Berlin, pp. 279–289.

44. **Pruesse, E., C. Quast, K. Knittel, B. Fuchs, W. Ludwig, J. Peplies, and F. O. Glöckner.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* **35:**7188–7196.

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

46. **Rappé, M.S., Connon, S.A., K. L. Vergin, and S. J. Giovannoni.** 2002. Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418:**630–633.

47. **Ravot, G., B. Ollivier, M. L. Fardeau, B. K. Patel, K. T. Andrews, M. Magot and J. L. Garcia** 1996. L-Alanine production from glucose fermentation by hyperthermophilic members of the domains Bacteria and Archaea: a remnant of an ancestral metabolism? *Appl. Environ. Microbiol.* **62:**2657–2659.

48. **Sahm, K., B. J. MacGregor, B. B. Jørgensen, and D. A. Stahl.** 1999. Sulphate reduction and vertical distribution of sulphate-reducing bacteria quantified by rRNA slot-blot hybridization in a coastal marine sediment. *Environ. Microbiol.* **1:**65–74.

49. **Schmitt-Wagner, D., M.W. Friedrich, B. Wagner, and A. Brune.** 2003. Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* **69:**6007–6017.

50. **Schmitt-Wagner, D., M. W. Friedrich, B. Wagner, and A. Brune.** 2003. Axial dynamics, stability, and interspecies similarity of bacterial community structure in the highly compartmentalized gut of soil-feeding termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* **67:**6018–6124.

51. **Schut, F., E. J. de Vries, J. C. Gottschal, B. R. Robertson, W. Harder, R. A. Prins, and D. K. Button.** 1993. Isolation of typical marine bacteria by dilution culture: Growth, maintenance, and characteristics of isolates under laboratory conditions. *Appl. Environ. Microbiol.* **59:**2150–2160.

52. **Stamatakis, A., T. Ludwig, and H. Maier.** 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* **21:**456–463.

53. **Stingl, U., R. Radek, H. Yang., and A. Brune.** 2005. endomircrobia: cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* **71:**1473–1479.

54. **Süßmuth, R., J. Eberspächer, R. Haag, and W. Springer.** 1987. Biochemisch-mikrobiologisches Praktikum, Thieme Verlag, Stuttgart, pp. 57–58.

55. **Tajima, K., S. Arai, K. Ogata, T. Nagamine, H. Matsui, M. Nakamura, R. I. Aminov, and Y. Benno.** 2000. Rumen bacterial community transition during adaptation to high-grain diet. *Anaerobe* **6:**273–284.

56. **Tanious, F. A., J. M. Veal, H. Buczak, L. S. Ratmeyer, and W. D. Wilson.** 1992. DAPI (4',6-diamidino-2-phenylindole) binds differently to DNA and RNA: minor-groove binding at AT sites and intercalation at AU sites. *Biochemistry* **31:**3103–3112.

57. **Tholen, A., and A. Brune.** 1999. Localization and *in situ* activities of homoacetogenic bacteria in the highly compartmentalized hindgut of soil-feeding higher termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* **65:**4497–4505.

58. **Tholen, A., B. Schink, and A. Brune.** 1997. The gut microflora of *Reticulitermes flavipes*, its relation to oxygen, and evidence for oxygen-dependent acetogenesis by the most abundant *Enterococcus* sp. *FEMS Microbiol. Ecol.* **24:**137–149.

59. **Tholen, A., M. Pester, and A. Brune.** 2007. Simultaneous methanogenesis and oxygen reduction by *Methanobrevibacter cuticularis* at low oxygen fluxes. *FEMS Microbiol. Ecol.* **62:**303–312.

60. **Tindall, B. J.** 1990. A comparative study of the lipid composition of *Halobacterium saccharovorum* from various sources. *Syst. Appl. Microbiol.* **13:**128–130.

61. **Tindall, B. J.** 1990. Lipid composition of *Halobacterium lacusprofundi*. *FEMS Microbiol. Lett.* **66:**199–202.

62. **Tindall, B. J., J. Sikorski, R. M. Smibert, and N. R. Krieg.** 2007. Phenotypic characterization and the principles of comparative systematics. *In* C. A. Reddy, T. J. Beveridge, J. A. Breznak, G. Marzluf, T. M. Schmidt and L. R. Snyder, Methods for General and Molecular Microbiology. ASM Press, Washington DC, USA, pp. 330–393.

63. **Torrella, F., and R. Y. Morita.** 1981. Microcultural study of bacterial size changes and microcolony and ultramicrocolony formation by heterotrophic bacteria in seawater. *Appl. Environ. Microbiol.* **41:**518–527.

64. **Valentine, R. C., B. M. Shapiro, and E. R. Stadtman.** 1968. Regulation of glutamine synthetase. XII. Electron microscopy of the enzyme from *Escherichia coli*. Biochemistry **7:**2143–2152.

65. **Verbarg, S., H. Rheims, S. Emus, A. Frühling, R. M. Kroppenstedt, E. Stackebrandt, and P. Schumann.** 2004. *Erysipelothrix inopinata* sp. nov., isolated in the course of sterile filtration of vegetable peptone broth, and description of *Erysipelotrichaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **54:**221–225.

66. **Wehrmeyer, W., and H. Schneider.** 1975. Elektronenmikroskopische Untersuchungen zur reversiblen Veränderung der Chloroplastenfeinstruktur von *Rhodella violacea* bei Stickstoffmangel. *Biochem. Physiol. Pfl.* **7:**519–532.

67. **Wilson, J., L. O. Krampitz, and C. H. Werkman.** 1948. Reversibility of a phosphoroclastic reaction. *Biochem. J.* **42:**598–600.

68. **Wright, D. E., and R. E. Hungate.** 1967. Amino acid concentrations in rumen fluid. *Appl. Microbiol.* **15:**148–151.

69. Yang, H., D. Schmitt-Wagner, U. Stingl, and A. Brune. 2005. Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ. Microbiol.* **7:**916–932.

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

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

Published in Applied and Environmental Microbiology, May 2009, p. 2841-2849, Vol. 75, No. 9

Summary

Organisms of the candidate phylum "Termite Group 1" (TG1), are regularly encountered in termite hindguts but are 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.

Authors' contribution: Genome analysis was performed by D. H. W I.-O. Prepared the DNA for sequencing. V. K., H. S., A. L. and P. H. were responsible for the generation of the genome sequence and the bioinformatic pipeline in the Joint Genome Insititute, including generation of the concatenated tree. O. G. and A. B: were involved in the discussion of the genomic interpretation. The results of the genom were evaluated as draft manuscript. A. B. and D. H. prepared the final manuscript with additions from P. H. and V. K.

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 so-called endomircrobia, 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 *Candidatus* "Endomicrobium trichonymphae" strain Rs-D17, an endosymbiont of

termite gut flagellates (22), we also investigated the phylogenetic position of Elusimicrobia relative to other bacterial phyla.

Materials 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 trichonymphae" 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 trichonymphae" 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).

Genome structure

E. minutum has a relatively small circular chromosome of 1,643,562 bp (Figure 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).

Table 1. Summary of the functional assignment, according to COG domain, of the 1529 protein-coding genes in the *Elusimicrobium minutum* genome. Details are shown in the 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)

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

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 (Figure 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).

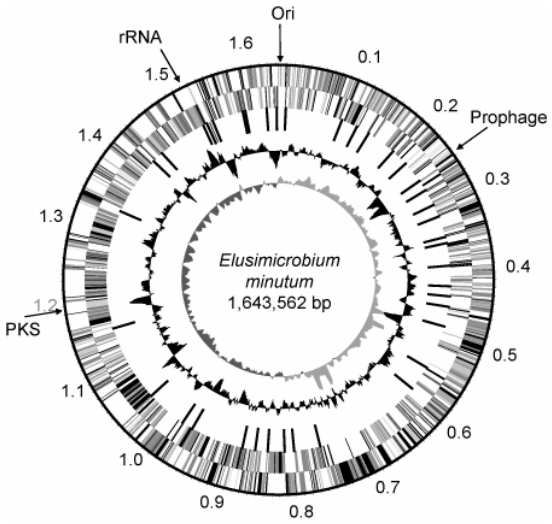


Figure 1. Genomic organization of the *Elusimicrobium minutum* chromosome. The two outermost rings show the genes encoded on the forward and reverse strand (scale in mega base pairs). The third ring depicts the location of tRNA genes. The fourth ring shows the G+C content and the innermost ring the GC 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).

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 trichonymphae" 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 trichonymphae" strain Rs-D17 (Figure 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.

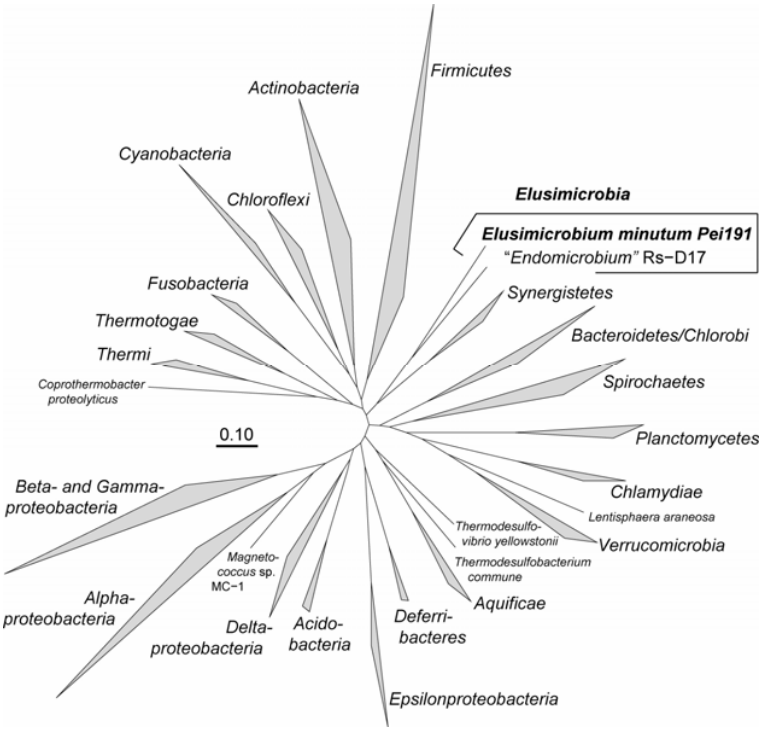


Figure 2. An unrooted maximum-likelihood tree of 280 bacterial genomes, including the two sequenced representatives of the phylum Elusimicrobia, representing the regions of the bacterial domain currently mapped by genome sequences. The tree is based on a concatenated alignment of 22 single-copy genes. 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.

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 (Figure 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.

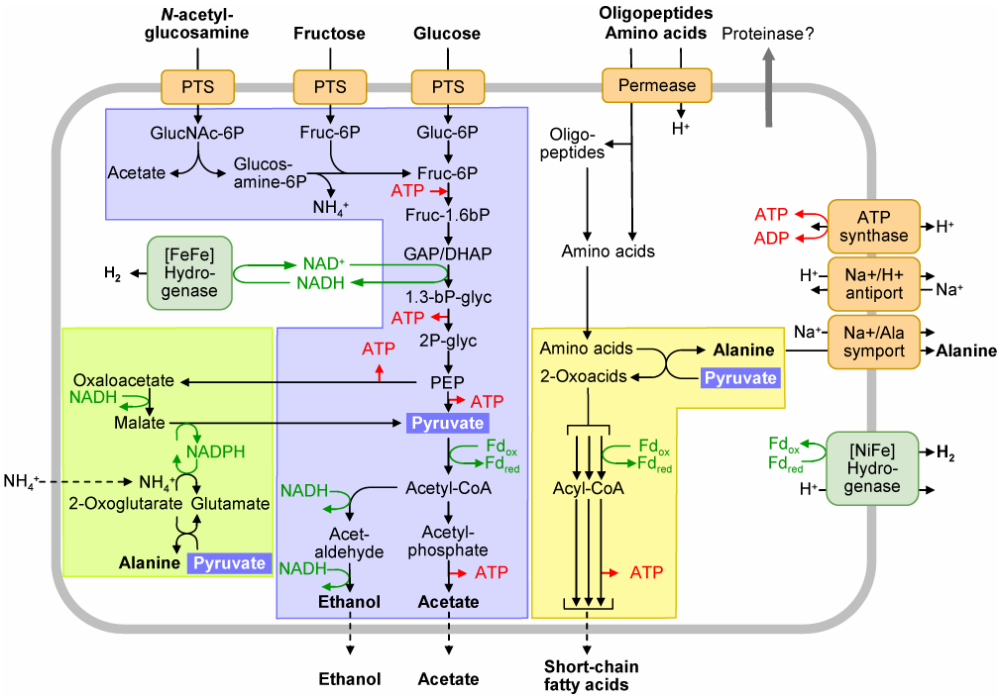


Figure 3. Schematic overview of the energy metabolism in *Elusimicrobium minutum*. Sugars are degraded via the Embden-Meyerhof pathway and pyruvate-ferredoxin oxidoreductase (PFOR) (blue box). NADH is recycled by reduction of acetyl-CoA to ethanol or, at low hydrogen partial pressure, by the cytoplasmic [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).

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

CxxCxnGxCxxxGxmGCPP (*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 (Figure 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* (Figure 3).

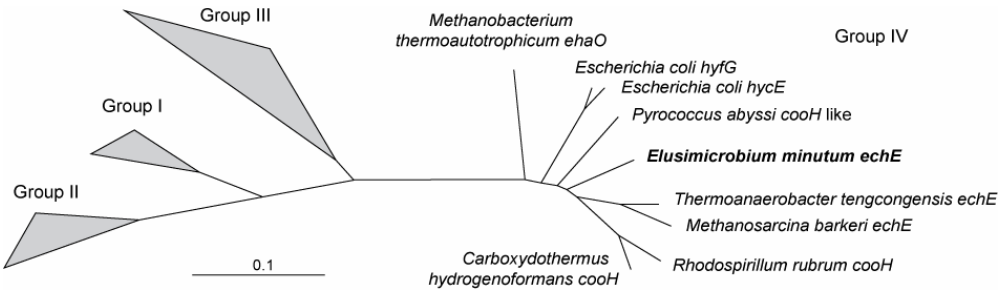


Figure 4. Maximum-likelihood tree of [NiFe] hydrogenases, based deduced amino acid sequences of the large subunit. The sequences of *Elusimicrobium minutum* and *Thermoanaerobacter tengcongensis* fall within the radiation of the sequences assigned to group IV [NiFe] hydrogenases by (54). The topology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

The second hydrogenase shows the typical structure and sequence motifs of a cytosolic NADH-dependent [FeFe] hydrogenase (Figure 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 (Figure 3).

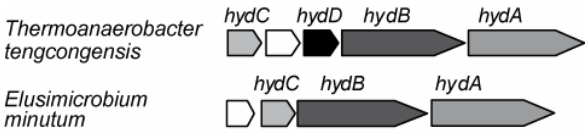


Figure 5. Organization of the genes encoding the subunits of the [FeFe] hydrogenase of *T. tengcongensis* (48) and their predicted homologs in *Elusimicrobium minutum*. The displayed length is proportional to the size of the corresponding ORF. *hydA*, *hydB*, and *hydC* have deduced amino acid sequence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

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

Anabolism

Although the presence of fructose 1,6-bisphosphatase indicates the possibility for gluconeogenesis via the EMP, *E. minutum* requires a hexose for growth (14). The absence of genes coding for 2-oxoglutarate dehydrogenase, succinate dehydrogenase, and succinyl-CoA synthetase is typical for strict anaerobes and documents that *E. minutum* does not possess a complete tricarboxylic acid (TCA) cycle. The reductive branch of the incomplete TCA cycle is initiated by phosphoenol pyruvate (PEP) carboxykinase and allows the interconversion of oxaloacetate, malate, and fumarate. The oxidative branch of the pathway starts with citrate synthase and allows the formation of 2-oxoglutarate. Typical for anaerobic microorganisms, the citrate synthase of *E. minutum* belongs to the *Re*-type (32). The products of the incomplete TCA cycle are precursors of several amino acids. The biosynthetic pathways for the formation of glutamate, glutamine, proline, aspartate, lysine, threonine, and cystathione are present. Also the pathways for the formation of alanine, cysteine, glycine, histidine, and serine, starting with intermediates of the EMP, are almost fully represented by the corresponding genes (Table S1, Figure S1). However, the genes for the synthesis of other proteinogenic amino acids

(arginine, asparagine, isoleucine, leucine, methionine, phenylalanine, tyrosine, tryptophan, and valine) are lacking, which would explain why *E. minutum* requires small amounts of yeast extract in the medium (14).

The genome of *E. minutum* does not possess an oxidative pentose phosphate pathway, which is typically involved in the regeneration of NADPH. This important coenzyme is probably regenerated by the alternative route of pyruvate formation from PEP (formation of oxaloacetate by PEP carboxykinase, NADH-dependent reduction of oxaloacetate by malate dehydrogenase, and $NADP^+$-dependent oxidative decarboxylation of malate by malic enzyme; Figure 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%; Figure 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 (Figure 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-oxoacid:ferredoxin oxidoreductases. Substrate-level phosphorylation is accomplished via an acyl-CoA synthetase (ADP-forming), resulting in the formation of ATP and the corresponding fatty acid. The genome also encodes proton-dependent oligopeptide transporters, ABC-type transport systems for peptides, and numerous proteolytic and peptolytic enzymes, some of which have typical signal peptides, indicating extracellular proteinase activity (Table S1).

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

The operation of this peptide utilization pathway in *E. minutum* is supported by the observation that most proteinogenic (and even some non-proteinogenic) amino acids are converted to their corresponding oxidative decarboxylation products during growth on glucose. Further evidence was provided by ^{13}C-labeling, which demonstrated that the carbon skeleton of the putative transamination product,

alanine, is derived from glucose (14). In principle, *E. minutum* also possesses the capacity for the net amination of pyruvate to alanine (Figure 3, green box), which has been proposed to function as an additional electron sink in *P. 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 (ΔG°' values calculated according to 56; data for isobutyrate from 60).

Glucose + H_2O → Ethanol + $Acetate^-$ + H^+ + 2 H_2 + 2 CO_2

ΔG°' = –225 kJ mol^{-1}

Glucose + Valine + H_2O → Ethanol + Alanine + Isobutyrate– + H^+ + 2 H_2 + 2 CO_2

ΔG°' = –245 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 (Figure 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.

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

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

system with antigen-specific antibodies, it has been reported that a response to an immune challenge can be enhanced by previous exposure (29).

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

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

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).

Oxygen stress

In agreement with the obligately anaerobic nature of *E. minutum*, the genome contains no cytochrome genes and no pathways for the biosynthesis of quinones, corroborating the absence of any respiratory electron transport chains. However, *E. minutum* has a six-gene "oxygen stress protection" cluster consisting of ruberythrin (*rbr*), superoxide reductase (*sor*), rubredoxin:oxygen oxidoreductase (*roo*), and rubredoxin (*rub*) (Figure 6). The *roo* gene of *E. minutum* has similarity to the

corresponding genes of *Desulfovibrio gigas* and *Moorella thermoacetica*, which have been shown to reduce molecular oxygen by reduced rubredoxin (15, 49). The presence of an oxygen-reducing system may explain the ability of *E. minutum* to retard the diffusive influx of oxygen into deep-agar tubes (14) and may play an important role in survival in the intestinal tract of insects, a habitat constantly exposed to the influx of oxygen (4, 31).

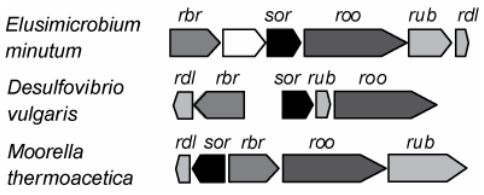


Figure 6. Organization of the genes encoding the "oxidative stress protection" cluster in *Moorella thermoacetica, Desulfovibrio gigas*, and their predicted homologs in *Elusimicrobium minutum.* The displayed length is proportional to the size of the corresponding ORF. The genes for ruberythrin (*rbr*), 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.

Ecological considerations

The genome of *E. minutum* revealed several adaptations of the bacterium to its environment. As a member of the "Intestinal Cluster", *E. minutum* is probably a resident inhabitant of the gut of *P. ephippiata*, which is thought to assist in digestion (31). *P. ephippiata* feeds on a humus-rich diet, and its gut contains high concentrations of glucose, peptides, and amino acids (3). With its putative capacity for proteinase secretion, the potential to maximize ATP yield in a coupled fermentation of sugars and amino acids, and the ability to cope with the exposure to molecular oxygen and reactive oxygen species, *E. minutum* appears to be well adapted to this habitat. As with other intestinal bacteria, it requires complex nutritive supplements and lacks pathways for the synthesis of most vitamins and certain amino acids. Although the genome of *E. minutum* is relatively small, there are no indications for an obligate association with its host. Genes encoding glycosyl hydrolases involved in the degradation of polysaccharides (other than glycogen)

were not identified, indicating that *E. 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.** 2009. *Elusimicrobium minutum* gen. nov., sp. nov., the first isolate of the Termite Group 1 phylum *Appl. Environ. Microbiol.* **75:**2831-2840.

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. *In* Setlow J. K. (ed), *Genetic Engineering*, Kluwer Academic, New York, NY, pp. 225–238.

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.

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

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

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

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

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

28. **Koonin, E. V., and Y. I. Wolf.** 2008. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36:**6688–6719.

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

30. **Lemke, T., T. van Alen, J. H. P. Hackstein, and A. Brune.** 2001. Cross-epithelial hydrogen transfer from the midgut compartment drives methanogenesis in the hindgut 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.

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

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

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

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

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

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

42. **Ohkuma, M., T. Sato, S. Noda, S. Ui, T. Kudo, and Y. Hongoh.** 2007. The candidate 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:**7545–7550.

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

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

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

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

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

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

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

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

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

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

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

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

5. Parallel genomic evolution of *Candidatus* "Endomicrobium trichonymphae" genomes from *Trichonympha* protists in termites

D. P. R. Herlemann, W. Ikeda-Ohtsubo, Alice McHardy, S. Lowry, E. Goltsman, I. Rigoutsos, P. Hugenholtz and A. Brune

Summary

***Candidatus* "Endomicrobium trichonymphae" (CET) are vertically transmitted endosymbionts of cluster I *Trichonympha* protists. The genome sequence of CET strain Rs-D17, the endosymbiont from *Trichonympha agilis* in the termite *Reticulitermes speratus*, has revealed the presence of many pseudogenes and gene duplications, indicating genome reduction in endomicrobia. In this study, we compare the genome of strain Rs-D17 with the genome fragments of other CET lineages from a large metagenome library obtained from the assemblage of *Trichonympha* protists in *Zootermopsis nevadensis*. The genome fragments differed from those of strain Rs-D17 in their levels of pseudogenization and in the positions of individual mutations in homologous genes, but showed the same patterns of functional preservation and gene depletion. This indicates that similar selective pressure led to parallel genomic evolution of CET in different *Trichonympha* hosts. Furthermore, the large genome fragments of the metagenome indicate genome rearrangement when compared to the genome of strain Rs-D17. The presence of an intact *recA* gene in endomicrobia suggests that homologous recombination contributes to their diversification. The large amount of genome rearrangement is unusual for obligate endosymbionts, but the opportunity for recombination may be increased by the horizontal transfer of endosymbionts between *Trichonympha* flagellates during their occasional sexual reproduction.**

Authors' contributions: Genome comparison was performed by D. H., who also prepared the manuscript draft together with A. B. W. I.-O. prepared CET for the metagenome library and contributed the SSU sequences from *Trichonympha* flagellates. S. L., E. G., P. H. were responsible for the generation of the metagenome sequence and the bioinformatic pipeline in the Joint Genome Institute. A. M. and I. G. did the binning of the metagenome fragments.

Introduction

In the evolutionary lower termites the lignocellulose degradation depends on the presence of anaerobic flagellate protozoa, which belong to the orders *Trichomonadida*, *Hypermastigida*, and *Oxymonadida* (12) and occupy the bulk of the hindgut volume (16). Most gut flagellates are regularly colonized by prokaryotic epibionts on the surface and by endosymbionts in the cytoplasm or the nucleus (2, 5, 3, 17). However, little is known about these symbioses (28).

Flagellates of the genus *Trichonympha* are the predominant protozoa in a wide range of lower termites. Many species are abundantly colonized by an endosymbiont of a deep-rooting bacterial lineage, the so-called endomicrobia (14, 28, 24). The cospeciation between *Trichonympha* cluster I flagellates and *Candidatus* "Endomicrobium trichonymphae" (CET) indicates functional significance of such symbioses in the gut ecosystem (15).

Recently, the genomes of CET strain Rs-D17 derived from *Trichonympha agilis* (in the termite *Reticulitermes speratus*; 11) and of the distantly related *Elusimicrobium minutum* (10) were sequenced. The small genome size of strain Rs-D17, its increased A+T content, and the presence of pseudogenes and gene duplications suggests a reductive evolution of endomicrobia in their obligate symbiosis with *Trichonympha* hosts (11). Interestingly, strain Rs-D17 had retained genes for biosynthesis of 15 amino acids, suggesting participation in mutualistic nitrogen upgrading (11).

Since all CET endosymbionts of cluster I *Trichonympha* have evolved from a common ancestor (15), we investigated functional overlaps between strain Rs-D17 and the related CET endosymbionts of the *Trichonympha* spp. from *Zootermopsis nevadensis*. In contrast to *Reticulitermes speratus, Zootermopsis nevadensis* comprises three closely related *Trichonympha* spp. (18). The CET endosymbionts were enriched from a suspension of all *Trichonympha* flagellates from *Zootermopsis nevadensis*, and a metagenome library was constructed. The comparison of the gene content and gene order of the different CET symbiont lineages can give some clues

about specific adaptations to the particular environment and reveal the evolutionary forces acting on the genome. In addition the results provide insights into the lifestyle of *Trichonympha* in *Zootermopsis nevadensis.*

Enrichment of *Candidatus* "Endomicrobium trichonymphae"

Phylogenetic analysis of the flagellate SSU rRNA genes in the DNA extracted from hindgut homogenates revealed that *Zootermopsis nevadensis* harbors four phylotypes of *Trichonympha* flagellates. Two of them had been previously assigned to *Trichonympha collaris* and *Trichonympha sphaerica* (15), the two others both had the morphotype of *Trichonympha campanula* (data not shown). A highly enriched preparation of CET from the *Trichonympha* spp. in the hindgut of *Zootermopsis nevadensis* was obtained by physical separation (Figure1).

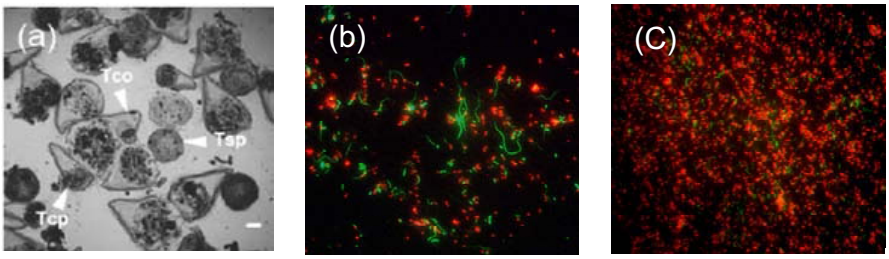


Figure 1. Physical enrichment of *Candidatus* "Endomicrobium trichonymphae". Guts from *Zootermopsis nevadensis* were suspended in solution U (29) and the gut wall fraction was removed. *Trichonympha* cells were sedimented on ice and resuspended in isolation buffer (25). The washed cells (a) were ultrasonicated 10 times for 0.5 s, and *Trichonympha* cell debris was removed by centrifugation at 500 × g. The cytoplasmic fraction was passed through 18-gauge 40-mm syringe needles and then filtered through 80-µm nylon mesh and treated with DNase I for 15 min at 4 °C to remove free DNA of the flagellates. (b). Large bacteria were removed by centrifugation in 5-µm pore size filter tubes at 3000 × g for 20 min (c). Cells hybridizing with an oligonucleotide probe specific for CET (TG1End1023T1) appear in red, cells hybridizing only with a general bacterial probe (EUB338; 1) appear in green. Fluorescence *in situ* hybridization was conducted as described previously (14).

A 16S rRNA clone library of the bacteria associated with this preparation (353 clones) revealed that about half of the library (166 clones) consisted of a variety of highly similar clones falling into the radiation of CET. The other sequences were assigned to Desulfovibrio (59 clones), Mycoplasma (26 clones), and a novel lineage of uncultured Deltaproteobacteria (67 clones). About half of the CET clones represented one dominant phylotype previously described as endosymbiont of

T. collaris, the other clones are subdivided into at least four other phylotypes (Figure 2.).

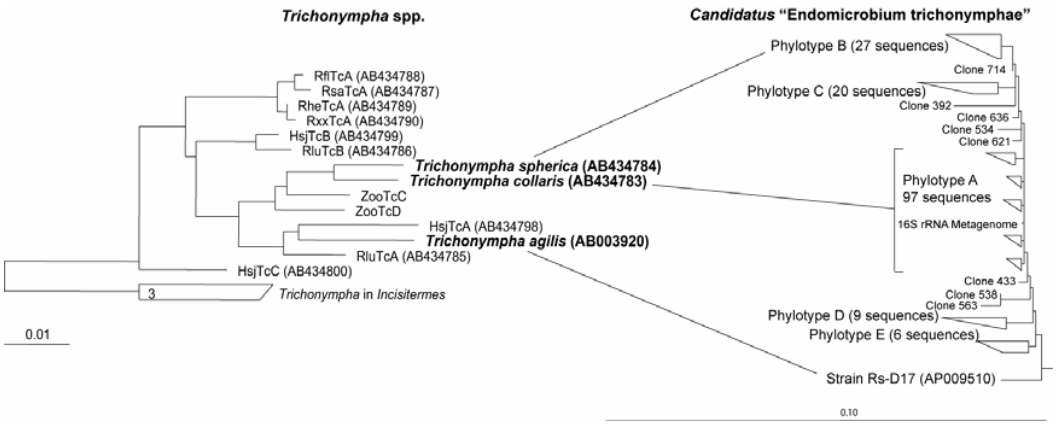


Figure 2. Phylogenetic tanglegram of *Trichonympha* spp. (left) and *Candidatus* "Endomicrobium trichonymphae" (right) The sequences *Trichonympha spherica* and *Trichonympha collaris* as well as ZooTcC and ZooTcD are derived from *Zootermopsis nevadensis* and reflect the *Trichonympha* spp. community enriched by the physical enrichment that were subject for the metagenome (see Figure 1). *Trichonympha agilis* is the *Trichonympha* host in *Reticulitermes speratus*, where the genome sequence of *Candidatus* "Endomicrobium trichonymphae" strain Rs-D17 was constructed from. The *Candidatus* "Endomicrobium trichonymphae" (CET) tree is based on 166 CET 16S rRNA sequences from the physically enriched metagenome library including the 16S rRNA sequence from the genome of strain Rs-D17 and a 16S rRNA sequence from the large metagenome fragment contig1478 (Phylotype A). The assignments of the CET phylotypes A and B to their respective flagellate host is based on Ikeda-Ohtsubo and Brune (15).

Metagenome

A metagenome library was constructed from the enriched CET preparation (9, 30) and subjected to phylogenetic binning using PhyloPythia (19). From a total of 16 Mb of sequence information, 3.3 Mbp were assigned to CET. The longest of the 866 genome fragments in the CET bin (330,800 bp) spanned 420 genes (contig2234), followed by two other fragments with more than 100 genes each (contig1496, 126 genes; contig1525, 131 genes).

Based on the genome size and the amount of tRNA genes in the genome of strain Rs-D17 (Table 1), the total amount of sequence information in the CET metagenome should be equivalent to approximately three complete genomes. However, a large number of relatively small size fragments and a considerable degree of genome fragments encoding for orthologous genes indicated a large

heterogeneity in the genomes represented in the dataset. Even at the level of the assembled genome fragments, which were supported by a high coverage of paired-end reads, various single nucleotide polymorphisms in the individual reads indicated the presence of subpopulations.

Table 1. Comparison of the available Genomes from the phylum Elusimicrobia. To construct the metagenome library, high molecular weight DNA was extracted from the enriched CET cells by lysing the cells in TE buffer with 10% SDS and proteinase K. DNA from the lysate was extracted with an equal volume of phenol-chloroform-isoamyl alcohol (49:49:1) and the phases were separated by three times 10 min centrifugation at 10000 g. The supernatant was washed with 0.6x of isopropanol and ethanol and resolved in TE buffer. Sequencing was performed in the Joint genome Institute (Walnut Creek, Calif.) according to the sequencing pipeline described by (9) Sequences were assigned to a phylogenetic group by PhyloPythia (19).

	Elusimicrobium minutum	Strain Rs-D17	CET metagenome
Size (Mbp)	1.64	1.13	3.37
G+C content	39.0	34.2	35.8
16S rRNA	1	1	5
tRNA	45	45	141
Genes in COG (%)	69	80	56

Parallel evolution

A comparison of the CET metagenome fragments with the genomes of strain Rs-D17 and *Elusimicrobium minutum* showed a general concordance in the distribution of functional groups in the endosymbionts versus their free-living relative. A more detailed analysis indicated that this can be extended to the preservation or loss of particular pathways and the pseudogenization of orthologous genes. Nevertheless, many of these features are apparently not due to the common ancestry of CET, but seem to have originated independently in each strain.

An example for gene loss are the genes coding for proteins involved in outer membrane biosynthesis that are present on the largest genome fragment of the CET metagenome library (contig2234). In contrast to the free-living *E. minutum*, which possesses an intact outer membrane (8), a considerable number of lipid A biogenesis genes are pseudogenized in both strain Rs-D17 (11) and in the large fragments of the metagenome library. In both cases, the genes persist as partially degraded sequences that range from pseudogenes with clear homology to their orthologs in

E. minutum to rudimentary sequences with no recognizable homology to their hypothetical orthologs (Fig. 3A). Although similar functions are disabled in both strain Rs-D17 and the larger CET metagenome fragments, the level of pseudogenization differs and the positions of the respective mutations are generally unique (Fig. 3B). This indicates that the pseudogenes are not ancient but rather reflect a divergent evolution of endomicrobia in the different *Trichonympha* hosts.

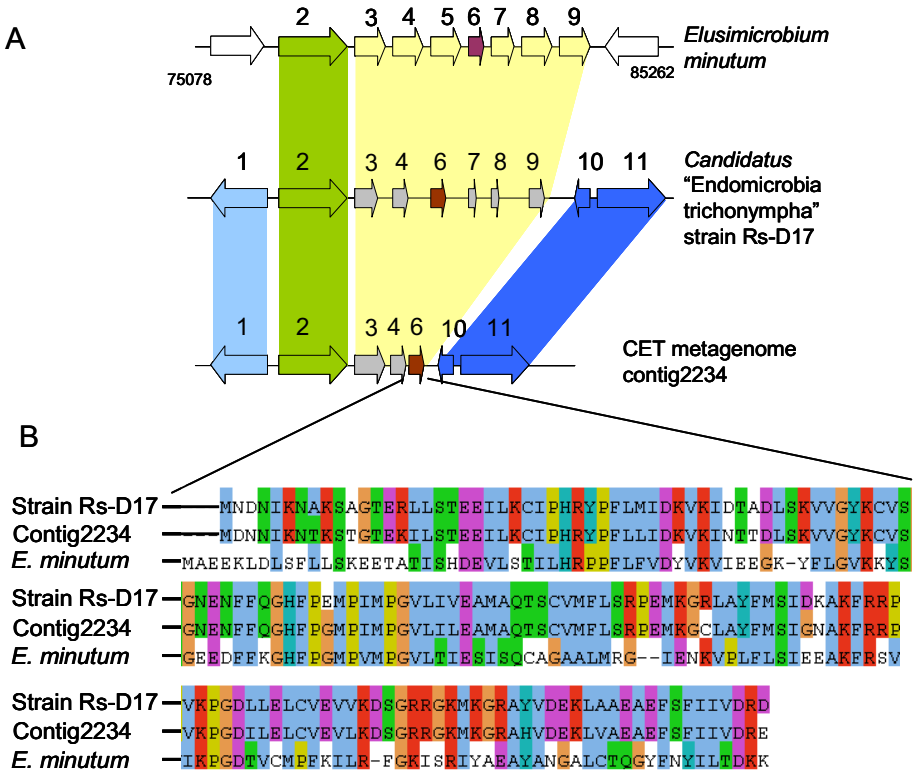


Figure 3. (A) Synthetic regions in the genomes of *Candidatus* "Endomicrobium trichonymphae" strain Rs-D17, *Elusimicrobium minutum* and the metagenome fragment contig2234 showing the pseudogenization of the outer membrane biosyntheses operon (green and yellow). The colors indicate orthologous genes. 1. NAD/FAD-utilizing enzyme 2. outer membrane protein assembly complex, *YaeT* protein (lipid A precursor biosynthesis) 3. Outer membrane protein (*OmpH*-like) 4. UDP-3-O-(3-hydroxymyristoyl) glucosamine *N*-acyltransferase 5. UDP-3-0-acyl N-acetylglucosamine deacetylase 6. *β*-hydroxyacyl-(acp) dehydratase *FabA/FabZ* 7. acyl-(acp)–UDP-*N*- acetylglucosamine *O*-acyltransferase 8. Protein of unknown function DUF1009 9. hypothetical protein 10. LSU ribosomal protein L28P 11. ATP-dependent DNA helicase *RecG*. (B) Alignment of the *β*-hydroxyacyl-(acp) dehydratase *FabA/FabZ* in the respective genome section, amino acids are color coded according to the clustalW scheme, uncolored letters indicate amino acids with different characteristics.

Gene preservation is exemplified by the pathways for amino acid biosynthesis, most of which are present in both strain Rs-D17 (11) and the metagenome but were lost in *Elusimicrobium minutum* (Tab. 2; 10), which thrives in an environment rich in

amino acids (8). This add support to the hypothesis that a primary function of CET in the symbiosis with gut flagellates is the provision of complex amino acids (nitrogen upgrading) and cofactors to the host (11).

The reductive evolution in *Candidatus* "Endomicrobium trichonymphae" is typical for endosymbionts. Endosymbionts live in a protected host environment, where relaxed selection increases the probability for fixation of deleterious mutations. This causes gene loss, which is usually driven by random genetic drift (21, 23) and leads to unique mutations in the genomes. However, host selection in CET causes to the independent development of similar pseudogenes, which is reflected in the preservation of similar amino acid biosynthesis pathways in the different strains. This correlation in relative divergence of proteins indicates identical functional constrains on gene functions in the different *Trichonympha* hosts, leading to parallel evolution.

Table 2. Comparison of the pathways for biosynthesis of amino acids between *Elusimicrobium minutum*, *Candidatus* “Endomicrobium trichomymphae” strain Rs-D17 and the *Candidatus* "Endomicrobium trichomymphae" metagenome.

	Elusimicrobium minutum	Strain Rs-D17	CET metagenome
Alanine	+	+	+
Arginine	−	+	+
Asparagine	−	+	+
Aspartic acid	+	+	+
Cysteine	+	−	−
Glutamic acid	+	+	+
Glutamine	+	−	−
Glycine	(+)	+	+
Histidine	+	+	+
Isoleucine	−	+	+
Leucine	−	+	+
Lysine	−	+	+
Methionine	−	(+)	(+)
Phenylalanine	−	+	+
Proline	+	−	−
Serine	+	+[1]	+[1]
Threonine	+	+	(+)
Tryptophan	−	+	+
Tyrosine	−	+	+
Valine	−	+	+

[1] synthesis via serine hydroxymethyltransferase

Genome rearrangement

A comparative analysis of genome organization in different endomicrobia was conducted using the major fragments (contig2234, contig1496, contig1525) of the CET metagenome and the genome of strain Rs-D17. Contig2234 delineated five major sequence blocks that are syntenic but differ in their relative gene order and length (Figure 4). The breakpoints between the major blocks are supported by a high coverage of mate-pair reads, two of the five major syntenic blocks are flanked by tRNA genes, whereas the other breakpoints are not coupled to gene boundaries (details not shown). These mosaic genome structures indicate that genome rearrangements have occurred during evolution of the endosymbionts.

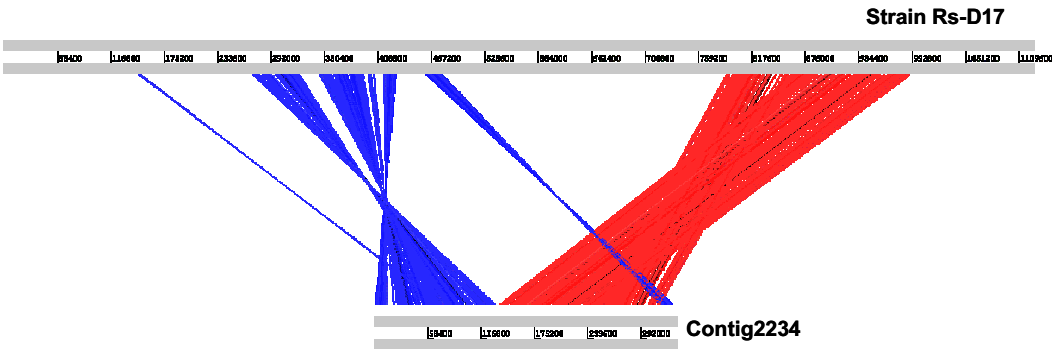


Figure 4. Syntenic overlaps between the genome of *Candidatus* "Endomicrobium trichonymphae" strain Rs-D17 and the genome fragment contig2234. Syntenic regions that are in same direction are connected with red lines; syntenic regions that are inverted are displayed in blue (6).

Genome rearrangement typically caused by homologous recombination (13); the presence of *RecARG* and *RuvABC*, and also single-strand binding proteins, in both the strain Rs-D17 genome (11) and in the CET metagenome (this study) indicates that endomicrobia, in contrast to many other obligate endosymbionts are still capable of homologous recombination. Vertical transmission in a host-enclosed environment normally prevents intracellular bacteria from the acquisition of new genetic material, therefore recombination between different strains cannot occur (22). However, *Trichonympha* flagellates in *Zootermopsis* spp. occasionally undergo sexual reproduction by isogametic fusion of cells (20). This leads to the exchange of cytoplasmic material between nonclonal populations of *Trichonympha*, resulting in horizontal transmission of endosymbionts and a subsequent recombination of genetic material between the respective endomicrobia strains.

With respect to genome rearrangement, the situation in *Candidatus* "Endomicrobium trichonymphae" resembles that of *Wolbachia* symbionts, which show high genome plasticity with ongoing recombination only if they are occasionally transmitted horizontally among the infected arthropods (4). This led to the "intracellular arena" hypothesis, which suggests that genetic material can move in and out of communities of obligate intracellular bacteria that coinfect the same intracellular host environment (4). Homologous recombination among vertically transferred symbionts has also been reported for the endosymbionts of vescomyid clams, in the (rare) event that a clam acquired symbionts by horizontal transfer from a co-occurring host (27). The horizontal exchange of genetic material would influence the evolution of CET considerably since it accelerates genome evolution and adaptation by orthologous replacement of recombinant alleles (7).

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.

References

1. **Amann, R. I., L. Krumholz, and D. A. Stahl.** 1990. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J. Bacteriol.* **172:**762–770.

2. **Ball, G. H.** 1969. Organisms living on and in protozoa. *In* Chen, T. T. (ed.), Research in Protozoology. Pergamon Press, New York, pp. 565–718.

3. **Bloodgood, R. A. and T. P. Fitzharris.** 1976. Specific association of prokaryotes with symbiotic flagellate protozoa from the hindgut of the termite *Reticulitermes* and the wood-eating roach. *Cryptocercus. Cytobios* **17:**103–122.

4. **Bordenstein S. R. and J. J. Wernegreen.** 2004. Bacteriophage flux in endosymbionts (Wolbachia): infection frequency, lateral transfer, and recombination rates. *Mol. Biol. Evol.* **21:**1981–91.

5. **Brune, A.** 2003. Symbionts aiding digestion. *In* R. T. Cardé, and V. H. Resh (ed.), Encyclopedia of Insects. Academic Press, New York, NY, pp. 1102–1107.

6. **Carver T. J., K. M. Rutherford, M. Berriman, M. A. Rajandream, B. G. Barrell and J. Parkhill.** 2005. ACT: the Artemis Comparison Tool. *Bioinformatics* **21:**3422–3.

7. **Cooper T. F. (2007)**. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biol.* **5,**e225.

8. **Geissinger, O., D. P. R. Herlemann, E. Mörschel, U. G. Maier and A. Brune. 2009**. The ultramicrobacterium *Elusimicrobium minutum* gen. nov., sp. nov., the first cultivated representative of the Termite Group 1 phylum. *Appl. Environ. Microbiol.* **75:**2831–2840.

9. **Hauser, L., Larimer F., M. Land, M. Shah, and E. Uberbacher.** 2004. Analysis and annotation of microbial genome sequences. *In* Setlow JK (ed.), Genetic engineering, Kluwer Academic, pp. 225–238.

10. **Herlemann D. P. R., O. Geissinger, W. Ikeda-Ohtsubo, V. Kunin, H. Sun, A. Lapidus, P. Hugenholtz and A. Brune.** 2009. Genomic analysis of "*Elusimicrobium minutum*," the first cultivated representative of the phylum "*Elusimicrobia*" (formerly Termite Group 1). *Appl. Environ. Microbiol.* **75:**2841–2849.

11. **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.

12. **Honigberg, B. M.** 1970. Protozoa associated with termites and their role in digestion, . *In* Krishna, K. and Weesner, F.M. (ed.), Biology of Termites. Academic Press, New York, NY, pp. 1–36

13. **Hughes D.** 2000. Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. *Genome Biol.* **1:**1–8.

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

15. **Ikeda-Ohtsubo, W., M. Desai, U. Stingl, and A. Brune.** 2007. Phylogenetic diversity of endomircrobia and their specific affiliation with termite gut flagellates. *Microbiology* **153**: 3458–3465.

16. **Katzin, L. I. and H. Kirby,** 1939. The relative weights of termites and their protozoa. *J. Parasitol.* **25:**444–445.

17. **Kirby, H. Jr.** 1941. Organisms living on and in protozoa. *In* G. N. Calkins and F.M. Summers (ed.), Protozoa in Biological Research. Columbia Univ. Press, New York, NY, pp. 1009–1113.

18. **Kirby, H., Jr.** 1932. Flagellates of the genus *Trichonympha* in termites. Univ. Calif. Publ. Zool. **37:**349–476.

19. **McHardy A. C., H. G. Martin , A. Tsirigos, P. Hugenholtz, I. Rigoutsos.** 2007. Accurate phylogenetic classification of variable-length DNA fragments. *Nat. Meth.* **4:**63–72.

20. **Messer, A. C. and M. J. Lee.** 1989. Effect of chemical treatments on methane emission by the hindgut microbiota in the termite *Zootermopsis angusticollis*. *Microb. Ecol.* **18:**275–284.

21. **Moran N. A.** 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **93:**2873–8.

22. **Moran, N. A.** 2007. Symbiosis as an adaptive process and source of phenotypic complexity. *Proc. Natl. Acad. Sci. U.S.A.* **104:**8627–8633.

23. **Ochman H. and N. A. Moran.** 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **11:**1096–9.

24. **Ohkuma, M., Sato, T., Noda, S., Ui, S., Kudo, T., and Y. Hongoh.** 2007. The candidate phylum 'Termite Group 1' of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol. Ecol.* **60:**467–476.

25. **Prechtl, J., C. Kneip, P. Lockhart, K. Wenderoth, and U. G. Maier.** 2004. Intracellular spheroid bodies of *Rhopalodia gibba* have nitrogen-fixing apparatus of cyanobacterial origin. *Mol. Biol. Evol.* **21:**1477–1481.

26. **Silva F. J, A. Latorre, and A. Moya.** 2003. Why are the genomes of endosymbiotic bacteria so stable? *Trends Genet.* **19:**176–80.

27. **Stewart F. J., C. R. Young, and C. M. Cavanaugh (2009)**. Evidence for homologous recombination in intracellular chemosynthetic clam symbionts. *Mol. Biol. Evol.* **26:**1391–404.

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

29. **Trager, W.** 1934. The cultivation of a cellulose-digesting flagellate, *Trichomonas termopsidis*, and of certain other termite protozoa. *Biol. Bull.* **66:**182–190.

30. **Warnecke F., P. Luginbühl, N. Ivanova, M. Ghassemian, T. H. Richardson, J. T. Stege, M Cayouette, A. C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S.G. Tringe, M. Podar, H. G. Martin, V. Kunin, D. Dalevi, J. Madejska, V. Kirton, D. Platt, V Szeto, A. Salamov, K. Barry, N. Mikhailova , N. C. Kyrpides, E. G. Matson, E. A. Ottesen, X. Zhang, M. Hernández, C. Murillo, L. G. Acosta, I. Rigoutsos, G. Tamayo, B. D. Green, C. Chang, E. M. Rubin, E. J. Mathur, D. E. Robertson, P. Hugenholtz, and J. R. Leadbetter.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450:**560–565.

6. General Discussion

At the outset of these studies, members of the Elusimicrobia phylum (at that time candidate phylum Termite Group 1) were recognized exclusively as endosymbiotic bacteria in gut flagellates of lower termites and wood-feeding cockroaches (33); beyond that, little was known about the biology of this phylum. The results of this work revealed a much wider environmental distribution of the members of this phylum and increased our understanding of its biology and evolution by a functional analysis of the genome sequence of the first cultivated representative, *Elusimicrobium minutum,* and by comparative genome analysis of the endosymbiotic representatives from this phylum.

The implications of the results reported in the individual chapters were discussed already in their respective context. The following discussion addresses general aspects about the phylum Elusimicrobia and broader implications about its diversity; functions derived from the genome sequence of *E. minutum* and *Candidatus* "Endomicrobium trichonymphae" strain Rs-D17, especially their metabolism and evolutionary aspects of the endomicrobia, are discussed.

The undiscovered diversity of Elusimicrobia

The present studies showed that the phylum Elusimicrobia consists not only of the previously described endomicrobia lineage (33), but also of clades from various habitats including soil, sediment, and intestinal tracts, and is not only present in lower termites (Chapter 2). However, it is still unclear why representatives of the Elusimicrobia were not previously recognized. Since hybridization experiments conducted with "intestinal cluster"-specific probes could not detect 16S rRNA genes of this group in *Zootermopsis nevadensis* (Chapter 3), a possible reason could be a low abundance of members of this phylum. Another reason is the lack of extensive phylogenetic analysis of many 16S rRNA genes in various large-scale diversity studies. The database screening based on Elusimicrobia-specific signatures revealed many 16S rRNA genes, which were assigned wrongly or completely ignored in the

original study (e.g., 1, 36). Since those studies were performed with universal bacteria 16S rRNA primers, which also exactly match with the 16 rRNA genes of genomes in the Elusimicrobia phylum, a primer bias for this bacterial phylum is unlikely. It rather demonstrates that less abundant microorganisms require a more intense investigation. Next generation sequencing technologies are currently able to sequence, for relatively moderate prices, thousands of partial 16S rRNA sequences, allowing the identification of the rare biosphere in the future (32). This huge amount of data will require intensive manual phylogenetic analysis, but will finally lead to an extensive description of diversity of all phyla.

The results reported in Chapter 2 indicate that the Elusimicrobia phylum is much more divergent than previously expected. But describing the diversity of a phylum is only the first step toward a better understanding. Today's microbial diversity studies are no longer only exercises in collecting microbiological taxa; they rather have the goal of gaining a comprehensive overview of microbial diversity and inferring functions of each species in the ecosystem (38). The detailed study of a single cultured representative gives the possibility to obtain insights into its metabolism and putative functions. In combination with a genome sequence, this information can be compared with other bacterial genomes and extrapolated to orthologous genes, i.e., genes that evolved from the same ancestral gene in the common ancestor of the compared genome (18).

Alanine — an unusual fermentation end product for glycolytic organisms?

The physiological investigation of *E. minutum* revealed classical fermentation products (ethanol, hydrogen, acetate, and CO_2) from glucose, but also alanine, which is a rather unusual end product (Chapter 3). The genome sequence analysis indicated that a transfer of the amino group from an amino acid to pyruvate is responsible for alanine production (Chapter 4), which is also in accordance with batch culture experiments (Chapter 3). A similar aminotransfer reaction has been identified in *Pyrococcus furiosus* and *Pyrococcus abyssi* (7, 12, 20). Moreover, other organisms have been shown to excrete alanine when grown on glucose in the presence of amino acids. e.g., *Thermococcus profundus* (17), *Thermotogales* spp.

(30), *Clostridium* strain P2 (27), and also some eukaryotes (6; 28). Therefore, the impact of alanine as an end product of glycolytic pathways in the presence of amino acids may be underestimated in the literature. The identification of a marker gene for this pathway could give an opportunity for a comprehensive analysis of all sequenced genomes and reveal the distribution of this pathway among these organisms. Since aminotransferases are widely distributed in different metabolic pathways, the gene encoding α-ketoacid oxidoreductase, which functions in the decarboxylation of the amino acid residues, may be a possible marker gene. These α-ketoacid oxidoreductases have significant sequence similarity to pyruvate:ferredoxin oxidoreductases, which are widespread among organisms (12, 16, 29). A detailed investigation of the amino acid patterns of α-ketoacid oxidoreductases in all described alanine-producing organisms will be necessary to determine an amino acid motif characteristic for this pathway.

Hydrogenases — key enzymes in the metabolism of *Elusimicrobium minutum*

Hydrogenases catalyze the reversible interconversion of protons, electrons, and H_2. The genome of *E. minutum* revealed gene clusters of two different types of hydrogenases: a [FeFe] hydrogenase and a [NiFe] hydrogenase (Chapter 4). Although NADH-driven hydrogen production is thermodynamically unfavorable (35), the [FeFe] hydrogenases can oxidize NADH, regenerating NAD^+. Very recently, a trimeric [FeFe] hydrogenase from *Thermotoga maritima* has been described to utilize ferredoxin and NADH synergistically for this reaction (31). The [FeFe] hydrogenase of *E. minutum* has sequence similarity to the [FeFe] hydrogenase in *T. maritima*, and consists also of three subunits organized in an operon structure, suggesting a similar mechanism for hydrogen production as described in *T. maritima*. If organisms contain genes encoding a [FeFe] hydrogenase and a [NiFe] hydrogenase in their genome (exemplified by *Thermoanaerobacter tengcongensis*), Schut and Adams suggest that the [FeFe] hydrogenase oxidizes both ferredoxin and NADH at low hydrogen partial pressure, whereas at high partial pressure, NADH is oxidized to produce ethanol, and H_2 production is catalyzed only by the [NiFe] hydrogenase (31). These two hydrogenases allow the organism to adapt to changing environmental situations and

to recycle NADH by the [FeFe] hydrogenase at low hydrogen partial pressure, which avoids production of ethanol and increases the yield of ATP.

In contrast to NADH-dependent hydrogen production, [NiFe] hydrogenases are able to produce hydrogen almost independently of the hydrogen partial pressure (e.g. 11). The [NiFe] hydrogenase in *E. minutum* has the amino acid motifs typical of energy-converting [NiFe] hydrogenases (*ech*), which are suggested to act also as redox-driven ion pumps. Interestingly, this group of hydrogenases has little sequence similarity to other groups of [NiFe] hydrogenases but shares an evolutionary relationship with energy-conserving NADH:quinone oxidoreductases (4). This type of hydrogenase has been identified in organisms with different forms of energy metabolism and phylogenetic association (e.g., methanogens, anaerobic CO-oxidizing bacteria, fermenting organisms; 11). The majority of *ech* genes are found in archaea, where they catalyze the reduction of a low-potential ferredoxin with hydrogen consumption. It was suggested that *Desulfovibrio gigas* and *T. tengcongensis* acquired *ech* by horizontal gene transfer from archaea (37), but in the case of *E. minutum,* this seems rather unlikely because this *ech* gene has no affiliation to an archaeal hydrogenase (Chapter 4).

Orthologous proteins often but not always have the same function

The example of the *ech* gene and the [FeFe] hydrogenase shows the potential of transferring annotations from biochemically investigated enzymes to genomes containing orthologous genes. Many metabolic pathways are constructed with similar sets of enzymes in different organisms; therefore, comparative genomics usually leads to a good overview of putative metabolic pathways of an organism (8). This indirect evidence of functions can be validated if the genome is derived from a batch culture, as in the case of *Elusimicrobium minutum*. Conversely, experimental results from batch-culture experiments can be better understood in context with their respective genes. However, the major fraction of a natural community is not available in culture. Current methods target only those we know how to cultivate (e.g., 2), and even the application of thousands of media and growth conditions, which have been developed over the years, could probably not reveal all

microorganisms from a complex environment like the termite gut and would be labor intensive. Metagenomics provides a means to overcome aspects of the cultivation bias since it provides genome fragments from bacteria in a sample without previous cultivation (see reviews 10, 19). But also metagenomics depends on the annotation of genes that were previously investigated biochemically. The genome of *E. minutum* therefore improves the annotation of related genomes because it adds information about proteins to databases (i.e., Cluster of Orthologous Groups of Proteins; 34) which can be used for the annotation of genomes from related organisms (9, 39).

Although the majority of genes have conserved homologs in other organisms, the transfer of functional information between orthologs from very distantly related genomes requires extreme caution. Projects involving the annotation of genomes for which closely related species, considered as reference genomes, with a valid annotation are already available, yield predictions that are usually reliable. For previously undescribed representatives of a taxonomic group, such as the Elusimicrobia, numerous exceptions and novel features have to be taken into account, often resulting in predictions of only limited use (e.g., unknown enzyme substrate, very short domain). An example of an exceptional feature in the genome of *Elusimicrobium minutum* is the 60 copies of *pilE* genes (Chapter 4). Since the bacterium lacks observable pili or other cell appendages (Chapter 3), some of the *pilE* genes probably belong to a type II secretion system, but it is likely that the majority of these genes are involved in another function, perhaps not related to pili or secretion at all. Other methodical approaches will be necessary to elucidate the functions of unresolved genes in *E. minutum*.

Implications for endomicrobia deduced from the *E. minutum* genome

During my doctoral studies, the complete genome sequence of *Candidatus* "Endomicrobium trichonymphae" strain Rs-D17 was published (13). The genome of this endosymbiont was determined after whole genome amplification of manually collected CET cells from a single cell of *Trichonympha agilis*. Although most genes of *E. minutum* and strain Rs-D17 share quite high sequence similarities, a detailed

comparison of the complete genomes of *E. minutum* and strain Rs-D17 was not possible. This is due to the large evolutionary distance between the organisms, which causes a lack of syntenic regions. Despite this, the genome of strain Rs-D17 encodes genes for similar pathways found to be encoded in the genome of *E. minutum*, for instance for the production of hydrogen, acetate, ethanol, and CO_2. The carbon source for the endosymbiont is probably glucose, which is provided by the host. Only a trimeric [FeFe] hydrogenase and no [NiFe] hydrogenase were identified in the genome of strain Rs-D17. Since *Trichonympha* flagellates, like all parabasalids, have hydrogenosomes, which emit hydrogen into the cytoplasm (5), the concentration of hydrogen in the cytoplasm is above the threshold for hydrogen production by NADH only. Therefore, the trimeric [FeFe] hydrogenase in strain Rs-D17 probably synergistically uses ferredoxin and NADH for hydrogen production.

In contrast to *E. minutum*, the endosymbiont retained genes for synthesis of various cofactors and 15 amino acids, suggesting participation in nitrogen upgrading (13). The presence of amino acid and cofactor pathways also answered the long-standing question about whether CET symbionts are parasites (14). Since strain Rs-D17 obviously benefits from glucose supplied by the host and on the other hand supplies *Trichonympha agilis* with complex amino acids and cofactors, the relationship between protist and endosymbiont is mutualistic.

Does homologous recombination help to escape Muller's ratchet?

Ikeda-Ohtsubo and Brune (15) have shown that CET and cluster I *Trichonympha* protist cospeciate, suggesting vertical transmission of the symbiont among generations. Endosymbionts show typical patterns of genome evolution owing to their protected environment (3). This is reflected in small chromosomes and a biased nucleotide base composition (23). Whereas ancient obligate symbionts have a static genome, strain Rs-D17 displays numbers of pseudogenes, which suggests a continuing process of genome reduction (13). Genome degradation is usually caused by random genetic drift since the selective pressure decreases in the protected host environment and vertical transmission avoids the exchange of genetic

material with other organisms (24, 26). A comparison between the genome of strain Rs-D17 with genome fragments from the metagenome library revealed a parallel evolution of endomicrobia in the different *Trichonympha* hosts and also suggested that the host exerts a strong selective pressure on the endosymbionts (Chapter 5). The comparison between the genome of strain Rs-D17 and the large fragments from the metagenome also indicate major genome rearrangements presumably resulting from homologous recombination. Homologous recombination is facilitated by RecA proteins, which are present in CET. It has been shown that *Trichonympha* in *Z. angusticolis* are able to reproduce sexually (21), which also leads to the exchange of cytoplasmic endosymbionts. This lateral symbiont transfer between hosts permits homologous recombination between closely related organisms by bringing divergent symbiont lineages into contact. This gene exchange has strong influence on the evolution of the endosymbionts, because it accelerates genome evolution and adaptation by orthologous replacement of recombinant alleles. This would also repeal Muller's ratchet, which proposes an evolutionary advantage of recombination, since beneficial mutations that occur in the same population, but in different lineages, must not compete with one another for fixation (25).

Recently, W. Ikeda-Ohtsubo has suggested the presence of free-living representatives of endomicrobia in lower termites (personal communication).These endomicrobia are more closely related to CET than *E. minutum* and would be an interesting subject for the study of evolutionary aspects of the endomicrobia-*Trichonympha* symbiosis. A genome from this organism would allow the reconstruction of the genome of the last common ancestor of the free-living endomicrobium and CET, similar to the reconstruction with genomes from *Buchnera aphidicola* and *Escherichia coli* (22). This will yield further insights into the mechanisms of reductive evolution and the physiological role in the symbiosis with *Trichonympha* flagellates and could also support the hypothesis of homologous recombination in the endosymbionts.

References

[1] **Abulencia, C. B., D. L. Wyborski, J. A. Garcia, M. Podar, W. Chen, S. H. Chang, H. W. Chang, D. Watson, E. L. Brodie, T. C. Hazen, and M. Keller.** 2006. Environmental whole-genome amplification to access microbial populations in contaminated sediments. *Appl. Environ. Microbiol.* **72:**3291–3301.

[2] **Amann, R. I., W. Ludwig, and K. H. Schleifer.** 1995. Phylogenetic identification, and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59:**143–169.

[3] **Andersson S. G., and C. G. Kurland.** 1998. Reductive evolution of resident genomes. *Trends Microbiol.* **6:**236–8.

[4] **Brandt, U., S. Kerscher, S. Dröse, K. Zwicker, and V. Zickermann.** 2003. Proton pumping by NADH:ubiquinone oxidoreductase. A redox driven conformational change mechanism? *FEBS Lett.* **545:**9–17.

[5] **Carpenter, K. J., and P. J. Keeling.** 2007. Morphology and Phylogenetic Position of *Eucomonympha imla* (Parabasalia: Hypermastigida). *J. Eukaryot. Microbiol.* **54:**325–332.

[6] **Chico, E., J. S. Olavarría, and I. Núnez de Castro.** 1978. *L*-Alanine as an end product of glycolysis in *Saccharomyces cerevisiae* growing under different hypoxic conditions. *Antonie v. Leeuwenhoek*. **44:**193–201.

[7] **Cohen, G. N., V. Barbe, D. Flament, M. Galperin, R. Heilig, O. Lecompte, O. Poch, D. Prieur, J. Quérellou, R. Ripp, J. Weissenbach, Y. Zivanovic, and P. Forterre.** 2003. An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi*. *Mol. Microbiol.* **47:**1495–1512.

[8] **Durot, M., P. Y. Bourguignon, and V. Schachter.** 2009. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol. Reviews* **33:**164–90.

[9] **Eisen, J. A.** 1998. Phylogenomics: Improving Functional Predictions for Uncharacterized Genes by Evolutionary Analysis. *Genome Res.* **8:**163–7.

[10] **Handelsman, J.** 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68:**669–685.

[11] **Hedderich, R., and L. Forzi.** 2005. Energy-converting [NiFe] hydrogenases: more than just H_2 activation. *J. Mol. Microbiol. Biotechnol.* **10:**92–104.

[12] **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.

[13] **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.

[14] **Ikeda-Ohtsubo, W.** 2007. Endomicrobia in termite guts: symbionts within symbiont. Doctoral thesis, Philipps-Universität Marburg.

[15] **Ikeda-Ohtsubo, W., and A. Brune.** 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and '*Candidatus* Endomicrobium trichonymphae'. *Mol. Ecol.* **18:**332–342.

[16] **Kletzin, A., and M. W. Adams.** 1996. Molecular and phylogenetic characterization of pyruvate and 2-ketoisovalerate ferredoxin oxidoreductases from *Pyrococcus furiosus* and pyruvate ferredoxin oxidoreductase from *Thermotoga maritima*. J. Bacteriol. **178:**248–57.

[17] **Kobayashi, T., S. Higuchi, K. Kimura, T. Kudo, and K. Horikoshi.** 1995. Properties of glutamate dehydrogenase and its involvement in alanine production in a hyperthermophilic archaeon, *Thermococcus profundus*. *J. Biochem.* **118:**592.

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

[19] **Kunin, V., A. Copeland, A. Lapidus, K. Mavromatis, and P. Hugenholtz.** 2008. A bioinformatician's guide to metagenomics. *Microbiol. Mol. Biol. Rev.* **72:**557–78.

[20] **Mai, X., and M. 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.

[21] **Messer, A. C., and M. J. Lee.** 1989. Effect of chemical treatments on methane emission by the hindgut microbiota in the termite *Zootermopsis angusticollis*. *Microb. Ecol.* **18:**275–284.

[22] **Moran, N. A., and A. Mira.** 2001. The process of genome shrinkage in the obligate symbiont *Buchnera aphidicola*. *Genome Biol.* 2:0054.

[23] **Moran, N. A., J. P. McCutcheon, and A. Nakabachi.** 2008. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42:**165–90.

[24] **Moran, N. A.** 1996. Accelerated evolution and Muller's rachet in endosymbiotic bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **93:**2873–8.

[25] **Muller, H. J.** 1963. The relation of recombination to mutual advance. *Mutat. Res.* **106:**2–9.

[26] **Ochman, H., and N. A. Moran.** 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **11:**1096–9.

[27] **Örlygsson, J., R. Anderson, and B. H. Svensson.** 1995. Alanine as an end product during fermentation of monosaccharides by *Clostridium* strain P2. *Antonie v. Leeuwenhoek*. **273:**273–280.

[28] **Paget, T. A., M. H. Raynor, D. W. Shipp, and D. Lloyd.** 1990. *Giardia lamblia* produces alanine anaerobically but not in the presence of oxygen. *Mol. Biochem. Parasitol.* **42:**63–67.

[29] **Ragsdale, S. W.** 2003. Pyruvate:ferredoxin oxidoreductase and its radical intermediate. *Chem. Rev.* **103:**2333–46.

[30] **Ravot, G., B. Ollivier, M. L. Fardeau, B. K. Patel, K. T. Andrews, M. Magot, and J. L. Garcia.** 1996. L-Alanine production from glucose fermentation by hyperthermophilic members of the domains Bacteria and Archaea: a remnant of an ancestral metabolism? *Appl. Environ. Microbiol.* **62:**2657–2659.

[31] **Schut, G. J., and M. W. Adams.** 2009. The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *Appl. Environ. Microbiol.* **191:**4451–7.

[32] **Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl.** 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. U. S. A.* **103:**12115–20.

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

[34] **Tatusov, R. L., M. Y. Galperin, D. A. Natale, and E. V. Koonin.** 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28:**33–36.

[35] **Thauer, R. K., K. Jungermann, and K. Decker.** 1977. Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol. Rev.* **41:**100–180.

[36] **Tsai, S. H., A. Selvam, and S. Yang.** 2007. Microbial diversity of topographical gradient profiles in Fushan forest soils of Taiwan . *Ecol. Res.* **22:**814–24.+

[37] **Vignais, P. M., and B. Billoud.** 2007. Occurrence, classification, and biological function of hydrogenases: an overview. *Chem. Rev.* **107:**4206–72.

[38] **Woese, C. R.** 2002. Microbiology in transition, *In* Staley J. T. and A.-L. Reysenbach (ed.), Biodiversity of microbial life, Wiley-Liss, New York, NY pp. 17–31.

[39] **Xu, J.** 2006. Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol. Ecol.* **15:**1713–31.

7. Summary

This thesis summarizes a series of studies of the phylum Elusimicrobia (formerly candidate phylum "Termite Group 1"). The environmental distribution of members of this phylum, and the genome sequence of the first and only cultivated representative, *Elusimicrobium minutum,* were the focus of these studies. In addition, a comparative genome analysis of the endosymbiotic representatives from this phylum was conducted.

Intestinal tracts of insects harbor several novel deep-rooting lineages of hitherto uncultivated bacteria, whose physiology is obscure. One of these groups is the phylum Elusimicrobia. Originally, the Elusimicrobia were represented only by the so-called endomicrobia, which had been identified as endosymbionts of flagellate protists of lower termites and wood-feeding cockroaches, but public databases contain a growing number of distantly related 16S rRNA gene sequences that fall into the radiation of the Elusimicrobia phylum. In the first study, group-specific primers were used to demonstrate that members of the Elusimicrobia are widespread in the environment and, in addition to the endomicrobia, comprise numerous other monophyletic lineages occurring in various habitats.

One of these lineages consisted of sequences obtained from various intestinal habitats. It comprised also the 16S rRNA gene of *Elusimicrobium minutum*, the first cultivated representative of the Elusimicrobia phylum, which was isolated from the gut of the scarab beetle larva *Pachnoda ephippiata*. This pure culture was the basis for the second study, which included the physiological and morphological characterization of *E. minutum* and the analysis of its genome. The genome allowed reconstruction of the metabolic pathways for the most important growth substrates (glucose, fructose, *N*-acetyl-glucosamine), revealing all genes required for uptake and fermentation of sugars via the Embden-Meyerhoff pathway and production of ethanol, acetate, H_2, and CO_2. Based on the genome, it could be predicted that a [NiFe] hydrogenase is responsible for the production of hydrogen, depending on the hydrogen partial pressure. This enzyme probably couples the production of

hydrogen with the generation of a proton-motive force (energy-converting hydrogenase). At low hydrogen partial pressure, hydrogen production may shift to a [FeFe] hydrogenase that synergistically uses NADH and reduced ferredoxin to produce hydrogen. In addition the genome also encodes for an unusual peptide degradation pathway that comprises transamination reactions and leads to the formation of alanine from pyruvate, an explanation for why *E. minutum* excretes alanine in substantial amounts. The function of the exceptionally high number of *pilE* genes in the genome is still elusive, and since electron micrographs showed no cell appendages, their participation in pilus assembly is uncertain. The presence of a rubredoxin:oxygen oxidoreductase operon in the genome of *E. minutum* indicated that the strictly anaerobic bacterium may also be able to reduce small amounts of molecular oxygen, which is also in accordance with physiological observations.

The phylogenetic analysis of 22 concatenated single-copy marker genes from the genome reinforced the phylum-level status of Elusimicrobia and confirmed the reproducible relationship between *E. minutum* and endomicrobia (represented by *Candidatus* "Endomicrobium trichonymphae" strain Rs-D17) as already predicted at the 16S rRNA level. This analysis was possible because of the recent publication of the genome of strain Rs-D17 from *Trichonympha agilis*. In parallel, a metagenome library from an enriched CET population derived from diverse flagellates of *Zootermopsis nevadensis* was prepared. The last part of this thesis consists of a comparative analysis of the large genome fragments from the metagenome library with the genome of strain Rs-D17. It revealed parallel evolution of endomicrobia strains in different termites and indicated the presence of genome rearrangements. The genome rearrangements suggest the ability for homologous recombination, which could be facilitated by the horizontal transfer of endosymbionts between *Trichonympha* spp. during occasional sexual reproduction.

Together, these studies contribute fundamental insights into the diversity, distribution, metabolism, and evolution of representatives from the phylum Elusimicrobia. The genome revealed along with the metabolic capacities of *E. minutum* numerous genes with only hypothetical functions, e.g., polyketide synthesis, indicating the presence of hitherto undiscovered physiological traits

which could be the subject for further investigations. The detailed annotation of *E. minutum* will also serve as reference for future annotations of other closely related members of this phylum.

Together, these studies contribute fundamental insights into the diversity, distribution, metabolism and evolution of representatives from the phylum Elusimicrobia. The genome revealed along with the metabolic capacities of *E. minutum* numerous genes with only hypothetical functions, e.g., polyketide synthesis indicating the presence of hitherto undiscovered physiological traits which could be the subject for further investigation. The detailed annotation of *E. minutum* will also serve as reference for future annotations of other closely related members of this phylum.

8. Zusammenfassung

Die vorliegende Arbeit umfasst eine Reihe unterschiedlicher Studien, die sich mit dem Phylum Elusimicrobia (ehemaliges Candidatus Phylum "Termite Group 1") befassen. Neben der Verbreitung des Phylums in der Umwelt stand dabei die Genomsequenz des ersten und bisher einzig kultivierten Vertreters, *Elusimicrobium minutum*, im Mittelpunkt dieser Arbeit. Weiterhin wurde eine vergleichende Genomanalyse von endosymbiontischen Vertretern dieses Phylums durchgeführt.

Im Verdauungstrakt von Insekten befinden sich viele phylogenetisch tief-wurzelnde Gruppen von Bakterien, über deren Physiologie und Funktionen bislang nur wenig bekannt war. Eine dieser phylogenetischen Gruppen ist das Phylum Elusimicrobia. Obwohl Vertreter der Elusimicrobia, die so genannten Endomicrobia, bisher nur als Flagellat-Endosymbionten in Därmen von niederen Termiten und holzfressenden-Schaben bekannt waren, enthielten öffentliche Datenbanken eine Anzahl von nahe verwandten 16S rRNA Sequenzen aus unterschiedlichen Lebensräumen. Im ersten Teil dieser Arbeit wurde mittels gruppen-spezifischer Primer gezeigt, dass Elusimicrobia in der Umwelt weit verbreitet sind und dass neben den bereits bekannten Endomicrobia viele monophyletische Abstammungslinien aus unterschiedlichsten Habitaten existieren.

Eine dieser Abstammungslinien besteht aus Sequenzen, die in unterschiedlichen interstinalen Habitaten nachgewiesen wurden. Darunter befand sich auch ein 16S rRNA Gen von *E. minutum*, dem ersten kultivierten Vertreter des Elusimicrobia phylums, der aus dem Darm der Rosenkäferlarve *Pachnoda ephippiata* isoliert wurde. Diese Reinkultur war Grundlage für den zweiten Teil der Studie, welcher die physiologische und morphologische Charakterisierung sowie eine Genomanalyse vorsah. Mittels der Genomsequenz konnten die Stoffwechselwege für die wichtigsten Wachstumsubstrate (Glukose, Fruktose, *N*-acteyl-Glucosamine) nachvollzogen werden. *E. minutum* besitzt alle Gene für die Aufnahme und Vergärung dieser Zucker mittels des Embden-Meyerhoff-Weges und die anschließende Produktion von Acetat, Ethanol, H_2 und CO_2. Basierend auf dem

Genom wird bei hohem Wasserstoffpartialdruck H_2 durch eine [NiFe]-Hydrogenase gebildet. Dieses Enzym kann diese Reaktion wahrscheinlich mit dem Aufbau eines Protonen-Gradienten verbinden (Energie-konvertierende Hydrogenase). Bei niedrigem Wasserstoffpartialdruck nutzt eine [FeFe]-Hydrogenase synergistisch NADH und reduziertes Ferredoxin, um Wasserstoff zu bilden.

Das Genom kodiert außerdem einen ungewöhnlichen Peptidabbauweg, bei dem mittels Transamminierung von Aminosäuren aus Pyruvat Alanin entsteht, was auch erklärt, warum Alanine von *E. minutum* in großen Mengen ausgeschieden wird. Die Funktion der ungewöhnlich hohen Anzahl von *pilE* Genen (60), die sich im Genom befinden, ist noch unklar, da auch elektronenmikroskopische Bilder keine Art von Zellanhängen aufzeigen. Das Genom des strikt anaeroben Bakteriums besitzt außerdem ein Rubredoxin: Oxygen Oxidoreduktase Operon, welches auf die Möglichkeit zur Reduktion von geringen Mengen Sauerstoff hinweist, was auch in physiologischen Experimenten gezeigt werden konnte.

Eine phylogenetische Analyse von 22 verketteten Marker-Genen unterstützte, dass Elusimicrobia ein eigenes Phylum darstellt, und bestätigte die Verwandtschaft zwischen *E. minutum* und Endomicrobia (repräsentiert durch *Candidatus* "Endomicrobium trichonymphae" Stamm Rs-D17), die bereits schon anhand der 16S rRNA Analyse beschrieben wurde. Diese Analyse war möglich, da vor kurzer Zeit das Genom von *Candidatus* "Endomicrobium trichonymphae" Stamm Rs-D17 aus *Trichonympha agilis* veröffentlicht wurde. Parallel dazu wurde eine Metagenom-Bibliothek von angereicherten CET aus einer *Zootermopsis nevadensis*-Flagellatensuspension erstellt.

Der letzte Teil dieser Dissertation besteht aus dem Vergleich von großen Genomfragmenten der Metagenome-Bibliothek mit dem Genom vom Stamm Rs-D17. Dieser hat gezeigt, dass in unterschiedlichen CET Stämmen parallele Evolution stattgefunden hat, und weist auf Genomneuordnungen ("genome rearrangements") hin. Diese Genomneuordnungen sind wahrscheinlich auf homologe Rekombination zurückzuführen, welche durch horizontalen Transfer von Endomicrobia zwischen Protozoen erleichtert wird.

Die Ergebnisse diese Arbeit haben Erkenntnisse über Vielfalt, Verteilung, Metabolismus und Evolution einiger Vertreter des Elusimicrobia Phylums erbracht. Das Genom weist neben den metabolischen Fähigkeiten von *E. minutum* viele Gene auf, denen nur hypothetische Funktionen, wie eine Polyketidsynthase, zugeordnet werden konnten. Daher gibt es viele bislang noch unentdeckten physiologischen Eigenschaften, deren Untersuchung noch aussteht. Die ausführliche Annotation von *E. minutum* dient außerdem als Referenz für künftige Annotationen von nahe verwandten Vertretern dieses Phylums.

9. Curiculum Vitae

Name	Daniel HERLEMANN
Address	Kirchweg 2
	D-35043 Marburg
Born:	22.02.1978 Offenburg, Germany

Academic studies

2006 –2009	Max Planck Institute for Terrestrial Microbiology Marburg German Doctor of natural science
2005	External diploma thesis at the Helmholtz Centre for Environmental research (UFZ), Leipzig. "Isolation and characterization of nonylphenol-degrading bacteria and fungi out of sediments from the Saale river near Halle"
2003	Semester abroad (ERASMUS) at the University of Klaipeda, Lithuania
2000 - 2005	Ernst-Moritz-Arndt University Greifswald Diploma studies in Biology Main subjects: microbiology, biochemistry, ecology

Alternative military service

1999- 2000	National park Wadden Sea at the "Schutzstation Wattenmeer"

School education

1996 –1999	Abitur (GCE) at the Ernährungswissenschaftliches Gymnasium (Grammar School of nutrition science) Main subjects: chemistry of nutrition, biology
1995 - 1996	Foreign education year at Britton High School, South Dakota/USA

Supplementary material

TABLE S1. Fluorescence signal strength at different formamide concentrations obtained in double hybridizations with oligonucleotide probes ELM1034 (specifically designed for strain Pei191^T; Cy3-labeled) and EUB338 (detecting most bacteria; fluorescein-labeled).[a]

Formamide	Strain Pei191^T		*Desulfurella acetivorans*	
(%)	ELM1034	EUB338	ELM1034	EUB338
0	+++	+	–	++
10	+++	+	–	++
20	+++	+	–	++
30 [b]	+++	+	–	++
40	++	W	–	++
50	–	–	–	++

[a] Symbols indicate fluorescence signal strength in decreasing order: +++, ++, +, w (weak), – (absent).

[b] Optimal formamide concentration used for purity control (see text).

TABLE S2. Whole-cell fatty acid composition of strain Pei191^T.

Fatty acid	Abundance (%)
C15:0 *iso*	25.3
C15:0 *anteiso*	16.3
C16:0 *iso*	12.1
C17:0 *iso*	8.0
C17:0 *anteiso*	7.3
C14:0 *iso*	6.1
C17:0 *iso* 3-OH	4.9
C16:0 *iso* 3-OH	3.3
C13:0 *iso*	2.9
C14:0 *iso* 3-OH	2.3
C16:0	2.1
C13:0 *anteiso*	1.7
C15:0 *iso* 3-OH	1.6
C17:0 2-OH	1.3

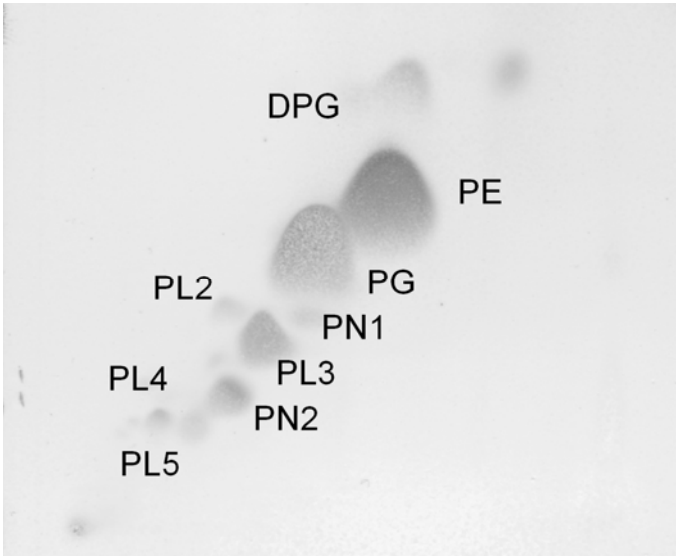
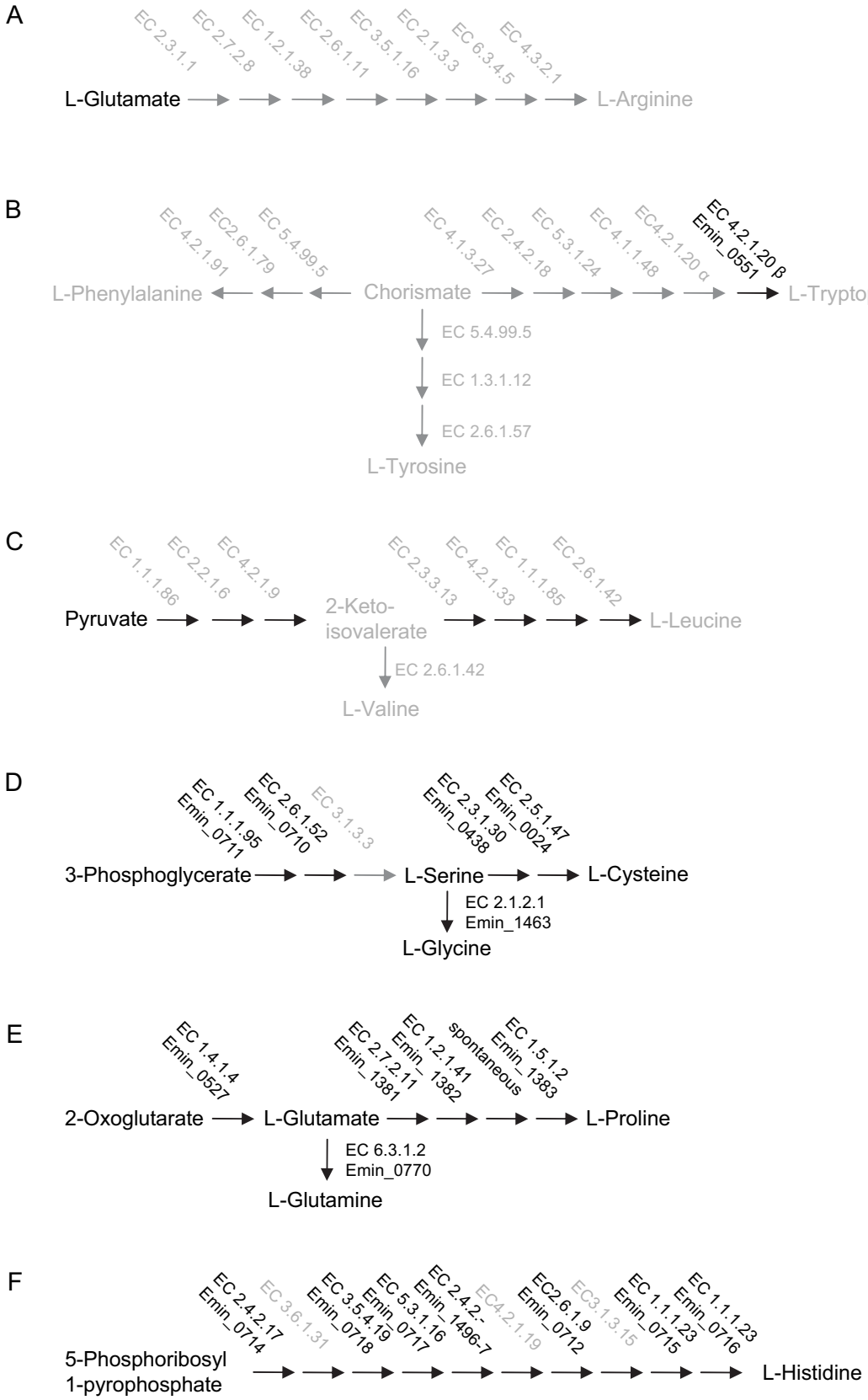


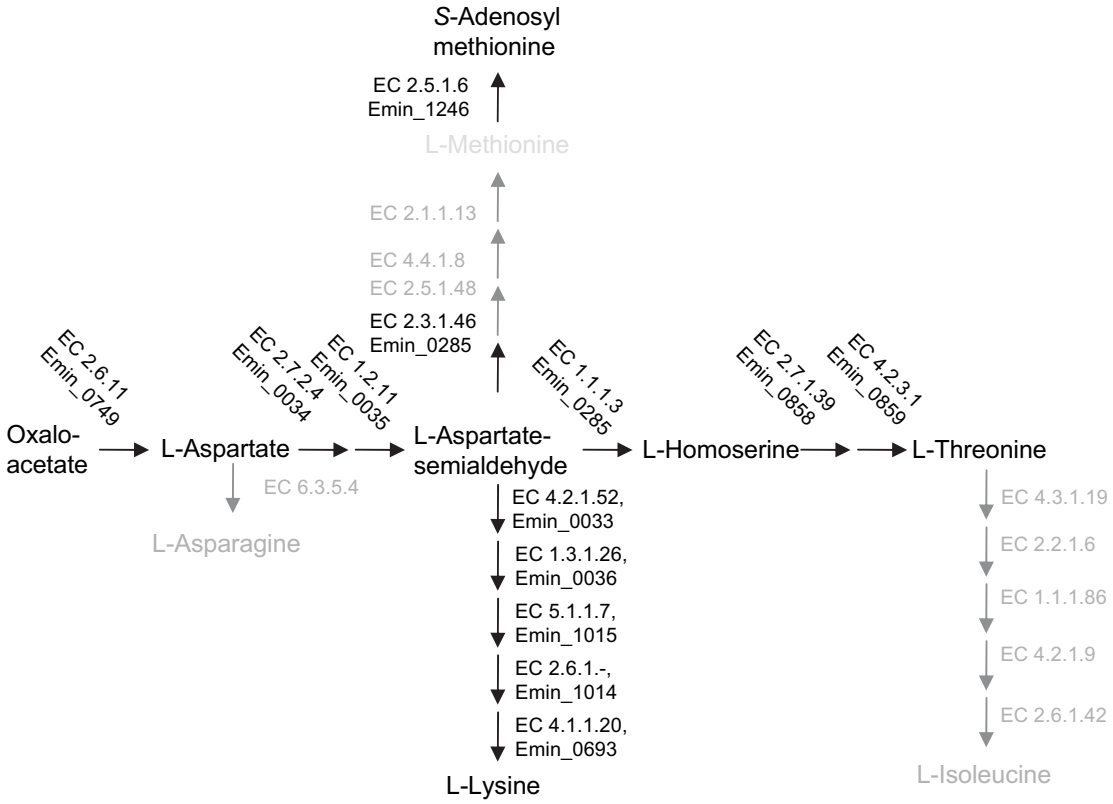
FIG. S1. Polar lipid analysis of strain Pei191^T by two-dimensional thin-layer chromatography. DPG, diphosphatidylglycerol; PG, phosphatidylglycerol; PE, phosphatidylethanolamine; PN1, PN2, unidentified amino-phospholipids; PL1, PL2, PL3, PL4, PL5, unidentified phospholipids. The first dimension was developed in chloroform/methanol/water (65:25:4, by vol.) and the second dimension in chloroform/methanol/acetic acid/water (80:12:15:4, by vol.).

Herlemann, D. P. R., O. Geissinger, W. Ikeda-Ohtsubo, V. Kunin, H. Sun, A. Lapidus, P. Hugenholtz and A. Brune. 2009. Genome analysis of *Elusimicrobium minutum*, the first cultivated representative of the *Elusimicrobia* phylum (formerly Termite Group 1).

Supplementary material



G



Supplementary Figure 1 A −G. Schematic overview of the amino acid anabolism of *Elusimicrobium minutum* based on the presence (black) and absence (grey) of genes. Amino acids that most likely can not be synthesized are grey.

Herlemann, D. P. R., O. Geissinger, W. Ikeda-Ohtsubo, V. Kunin, H. Sun, A. Lapidus, P. Hugenholtz and A. Brune. 2009. Genome analysis of *Elusimicrobium minutum*, the first cultivated representative of the *Elusimicrobia* phylum (formerly Termite Group 1).

Supplementary material

Table S1. List of all genes in the genome of *Elusimicrobium minutum* and the predicted function of their products according to a manually curated annotation (see Materials and Methods). The genes are sorted by COG groups. The length of the deduced amino acid sequence, the enzyme, the COG assignment, and the presence of signal peptides (S) and transmembrane helices (T) are indicated.

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
D	**Cell cycle control, cell division, chromosome partitioning**					
	Emin_0380	Sporulation protein and related proteins	686		COG2385	S
	Emin_0412	Chromosome segregation ATPases	1148		COG1196	
	Emin_0439	DNA segregation ATPase FtsK/SpoIIIE and related proteins	763		COG1674	S,T
	Emin_0486	Sporulation protein and related proteins	714		COG2385	S,T
	Emin_0506	Sporulation protein and related proteins	366		COG2385	S
	Emin_0541	Cell division protein FtsA	416		COG0849	
	Emin_0542	Cell division protein FtsZ	381		COG0206	
	Emin_0602	Nucleotide-binding protein implicated in inhibition of septum formation	184		COG0424	
	Emin_0846	Integral membrane protein possibly involved in chromosome condensation	125		COG0239	S,T
	Emin_0943	Putative chromosome segregation ATPases	776		COG1196	
	Emin_1091	Chromosome segregation ATPases	625		COG1196	
	Emin_1218	Membrane-bound metallopeptidase	387		COG4942	S
	Emin_1251	Cell cycle protein	383		COG0772	S,T
	Emin_1259	Glucose-inhibited division protein A	587		COG0445	T
	Emin_1273	ATPases involved in chromosome partitioning	245		COG0489	
	Emin_1377	Cell cycle protein	451		COG0772	S,T
	Emin_1388	Cell-shape-determining protein, MreB/Mrl family	343		COG1077	
	Emin_1543	ATPases involved in chromosome partitioning	274		COG1192	
	Emin_1551	tRNA(Ile)-lysidine synthetase	241		COG0037	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
E	**Amino acid transport and metabolism**					
	Emin_0033	Dihydrodipicolinate synthase	289		COG0329	
	Emin_0034	Aspartate kinase	243	EC 2.7.2.4	COG0527	
	Emin_0035	Aspartate-semialdehyde dehydrogenase	292	EC 1.2.1.11	COG0136	
	Emin_0036	Dihydrodipicolinate reductase	218	EC 1.3.1.26	COG0289	
	Emin_0133	Bacterial extracellular solute-binding protein family	576		COG0747	S
	Emin_0146	Putative threonine dehydrogenase	345		COG1063	
	Emin_0167	NADPH-dependent glutamate synthase	465		COG0493	
	Emin_0234	Succinyl-diaminopimelate desuccinylase	411		COG0624	
	Emin_0251	Permeases of the drug/metabolite transporter (DMT) superfamily protein	305		COG0697	S,T
	Emin_0284	Homoserine dehydrogenase	441	EC 1.1.1.3	COG0460	
	Emin_0285	Homoserine *O*-succinyltransferase	268	EC 2.3.1.46	COG1897	
	Emin_0286	*O*-acetylhomoserine/*O*-acetylserine sulfhydrylase	421	EC 2.5.1.47	COG2873	
	Emin_0394	Lipolytic protein family	204		COG2755	S,T
	Emin_0401	Pyruvate carboxyltransferase family protein	416		COG0119	
	Emin_0404	Isocitrate dehydrogenase (NAD^+)	334	EC 1.1.1.41	COG0473	
	Emin_0421	Aminotransferase class I and II	403		COG1167	
	Emin_0433	Permeases of the drug/metabolite transporter (DMT) superfamily	303		COG0697	S,T
	Emin_0437	Cysteine desulfurase NifS	386	EC 2.8.1.7	COG1104	
	Emin_0438	Serine *O*-acetyltransferase	284	EC 2.3.1.30	COG1045	
	Emin_0527	Glutamate dehydrogenase ($NADP^+$)	450	EC 1.4.1.4	COG0334	
	Emin_0528	Glutamate formiminotransferase	319	EC 4.3.1.4	COG3643	
	Emin_0529	Peptidase T	408	EC 3.4.11.4	COG2195	
	Emin_0551	Tryptophan synthase, beta subunit	412	EC 4.2.1.20	COG0133	
	Emin_0673	Na^+/alanine symporter	470		COG1115	S,T
	Emin_0708	Amino acid permease	480		COG0531	S,T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0710	Putative phosphoglycerate dehydrogenase	315		COG0111	
	Emin_0711	Phosphoserine transaminase	383	EC 2.6.1.52	COG1932	
	Emin_0715	Histidinol dehydrogenase	429	EC 1.1.1.23	COG0141	
	Emin_0716	Histidinol phosphatase	357	EC 3.1.3.15	COG0131	
	Emin_0718	Phosphoribosyl-ATP diphosphatase	202		COG0139	
	Emin_0749	Aspartate transaminase	401		COG0436	S
	Emin_0770	Glutamine synthetase, type I	450		COG0174	
	Emin_0834	Dipeptide/tripeptide permease	432	EC 3.1.3.1	COG3104	T
	Emin_0859	Threonine synthase	460	EC 4.2.3.1	COG0498	
	Emin_0860	Homoserine dehydrogenase	438	EC 1.1.1.3	COG0460	S
	Emin_0900	Histidinol-phosphate phosphatase family protein	184		COG0241	
	Emin_0917	Putative alanine transaminase	390		COG1168	
	Emin_0918	Peptidase M20	383		COG2195	
	Emin_0920	Dipeptide/tripeptide permease	436	EC 3.1.3.1	COG3104	T
	Emin_0922	Dipeptide/tripeptide permease	477	EC 3.1.3.1	COG3104	S,T
	Emin_0932	Tryptophanase	479	EC 4.1.99.1	COG3033	
	Emin_1009	Lysine decarboxylase	626	EC 4.1.1.18	COG1982	
	Emin_1014	LL-diaminopimelate aminotransferase EC 2.6.1.83	409		COG0436	
	Emin_1015	Diaminopimelate epimerase	274	EC 5.1.1.7	COG0253	
	Emin_1032	Carbamoyl-phosphate synthase, small subunit	390		COG0505	
	Emin_1033	Carbamoyl-phosphate synthase, large subunit	1102	EC 6.3.4.16	COG0458	
	Emin_1048	Semialdehyde dehydrogenase	291		COG1748	
	Emin_1057	High-affinity basic amino acid transporter	490		COG0531	T
	Emin_1104	Putative amino acid transporter	676		COG0531	S,T
	Emin_1192	Peptidase M29 aminopeptidase II	399		COG2309	
	Emin_1229	Plp-dependent enzyme	416		COG0498	
	Emin_1231	Putative peptidase	485		COG0624	S
	Emin_1233	ABC-type dipeptide/oligopeptide/nickel transport system	324		COG0444	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1235	ABC-type dipeptide/oligopeptide/nickel transport system	341		COG0601	S,T
	Emin_1236	ABC-type dipeptide/oligopeptide/nickel transport system	288		COG1173	S,T
	Emin_1240	Methenyl tetrahydrofolate cyclohydrolase-like protein	201		COG3404	
	Emin_1268	Zn-dependent oligopeptidases	681	EC 3.4.24.15	COG0339	S,T
	Emin_1292	Ribose-phosphate pyrophosphokinase	316	EC 2.7.6.1	COG0462	
	Emin_1313	Dipeptidyl aminopeptidases/acylaminoacyl-peptidases	294		COG1506	S,T
	Emin_1327	Serine dehydratase alpha chain	410		COG1760	S
	Emin_1351	Glycine dehydrogenase (decarboxylating)	482	EC 1.4.4.2	COG1003	
	Emin_1352	Glycine dehydrogenase (decarboxylating)	440	EC 1.4.4.2	COG0403	
	Emin_1353	Glycine cleavage system H protein	125		COG0509	
	Emin_1354	Glycine cleavage system T protein	371		COG0404	
	Emin_1381	Glutamate 5-kinase	270	EC 2.7.2.11	COG0263	
	Emin_1383	Pyrroline-5-carboxylate reductase	263	EC 1.5.1.2	COG0345	S
	Emin_1463	Glycine hydroxymethyltransferase	415	EC 2.1.2.1	COG0112	
	Emin_1496	Imidazoleglycerol phosphate synthase, cyclase subunit	250		COG0107	
	Emin_1497	Imidazole glycerol phosphate synthase, glutamine amidotransferase subunit	194		COG0118	
	Emin_1505	Dipeptide/tripeptide permease	444	EC 3.1.3.1	COG3104	T
	Emin_1536	Zn-dependent oligopeptidases	676	EC 3.4.24.15	COG0339	S
	Emin_0024	Cysteine synthase A	307	EC 2.5.1.47	COG0031	
	Emin_0402	Putative aconitase	418	EC 4.2.1.33	COG0065	
	Emin_0403	(re) Citrate synthase	164		COG0066	
	Emin_0693	Diaminopimelate decarboxylase	417		COG0019	
	Emin_0712	Histidinol-phosphate aminotransferase	344		COG0079	S
	Emin_0714	ATP phosphoribosyltransferase	295	EC 2.4.2.17	COG0040	
	Emin_0717	Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase	240	EC 5.3.1.16	COG0106	
	Emin_0731	Xaa-Pro aminopeptidase	352		COG0006	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0732	Xaa-Pro aminopeptidase	351		COG0006	
	Emin_1382	Gamma-glutamyl phosphate reductase	415	EC 1.2.1.41	COG0014	
	Emin_1538	Agmatinase	289		COG0010	
G	**Carbohydrate transport and metabolism**					
	Emin_0018	Alpha amylase	485		COG0366	
	Emin_0021	Glycoside hydrolase	786		COG1449	
	Emin_0063	Glucose-1-phosphate adenylyltransferase	403		COG0448	
	Emin_0085	Putative Beta-glucosidase	541		COG1472	S
	Emin_0114	Phosphomannomutase	467	EC 5.4.2.8	COG1109	
	Emin_0126	PTS system fructose subfamily IIA component	132		COG2893	
	Emin_0127	Phosphotransferase system, mannose/fructose/*N*-acetylgalactosamine	154		COG3444	
	Emin_0129	Phosphotransferase system, mannose/fructose/*N*-acetylgalactosamine	265		COG3716	T
	Emin_0130	Phosphotransferase system, mannose/fructose/*N*-acetylgalactosamine	89		COG1925	
	Emin_0131	PTS system EI component	580		COG1080	
	Emin_0135	Nucleoside H^+ symporter	414		COG2211	S,T
	Emin_0151	Glycolysis	252	EC 5.3.1.1	COG0149	
	Emin_0160	Repressor, ORF, kinase protein family	322		COG1940	
	Emin_0180	Ribokinase family	302		COG0524	
	Emin_0251	Permeases of the drug/metabolite transporter (DMT) family protein	305		COG0697	S,T
	Emin_0273	Alpha-glucan phosphorylase	850	EC 2.4.1.1	COG0058	
	Emin_0280	Phosphopentomutase	397	EC 5.4.2.7	COG1015	S
	Emin_0287	Phosphoglycerate mutase 1 family	248	EC 5.4.2.1	COG0588	
	Emin_0315	Histidine triad (HIT) protein	112		COG0537	
	Emin_0344	Pyruvate, phosphate dikinase	929	EC 2.7.9.1	COG0574	
	Emin_0353	Fructose-1,6-bisphosphate aldolase, class II	325	EC 4.1.2.13	COG0191	
	Emin_0354	Peptidase M42 family protein	346		COG1363	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0355	Transketolase central region	776		COG0021	
	Emin_0367	Extracellular sugar kinase family	296		COG0524	S
	Emin_0368	Sugar transport family	448	EC 1.3.1.74	COG2814	S,T
	Emin_0369	Phosphomannomutase	501	EC 5.4.2.8	COG1109	
	Emin_0370	Phosphopyruvate hydratase	423	EC 4.2.1.11	COG0148	
	Emin_0386	4-Alpha-glucanotransferase	518	EC 2.4.1.25	COG1640	S
	Emin_0410	6-Phosphofructokinase	368	EC 2.7.1.11	COG0205	
	Emin_0414	Major facilitator superfamily MFS_1	473		COG2814	S,T
	Emin_0423	Glycerate kinase	369	EC 2.7.1.31	COG1929	S
	Emin_0427	ATP-NAD kinase	279		COG0061	
	Emin_0433	Permeases of the drug/metabolite transporter (DMT) superfamily	303		COG0697	S,T
	Emin_0447	Nucleoside-diphosphate-sugar epimerases	316		COG0451	T
	Emin_0488	Putative fucose permease	2741		COG0738	
	Emin_0558	Pyruvate kinase	473	EC 2.7.1.40	COG0469	
	Emin_0607	UDP-glucose/hexose-1-phosphate uridylyltransferase	507	EC 2.7.7.12	COG4468	
	Emin_0609	Galactokinase	387	EC 2.7.1.6	COG0153	
	Emin_0611	Putative galactose mutarotase	312		COG2017	
	Emin_0616	Phosphoglycerate kinase	399	EC 2.7.2.3	COG0126	
	Emin_0618	Glyceraldehyde-3-phosphate dehydrogenase, type I	338	EC 1.2.1.12	COG0057	
	Emin_0625	Putative chitin deacetylase	266		COG0726	S
	Emin_0640	Sugar transporter	480		COG2271	S,T
	Emin_0703	Methylglyoxal synthase	151	EC 4.2.3.3	COG1803	
	Emin_0743	Ribulose-phosphate 3-epimerase	226	EC 5.1.3.1	COG0036	
	Emin_0776	Fucose permease	326		COG0738	S,T
	Emin_0809	Channel protein family	234		COG0580	S,T
	Emin_0810	Arabinose efflux permease	416		COG2814	S,T
	Emin_0882	Major facilitator superfamily	462		COG2814	S,T
	Emin_0921	PTS system, glucose-like IIB subunint	551		COG1263	T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0997	Putative mannose-6-phosphate isomerase	137		COG0662	
	Emin_1007	6-Phosphofructokinase	562		COG0205	
	Emin_1040	Dehydrogenase family protein	308		COG0451	
	Emin_1067	Putative phosphotransferase system EIIC	446		COG1263	T
	Emin_1103	PTS IIA-family protein	155		COG1762	
	Emin_1116	Phosphoglucose isomerase (PGI)	552		COG0166	
	Emin_1117	Ribose-5-phosphate isomerase	150		COG0698	
	Emin_1118	Phosphoheptose isomerase	194		COG0279	
	Emin_1215	Extracellular solute-binding protein family	402		COG1653	
	Emin_1221	Glycogen/starch synthase, ADP-glucose type	482	EC 2.4.1.21	COG0297	
	Emin_1249	Polysaccharide deacetylase	254		COG0726	S,T
	Emin_1269	Phosphoglucomutase/phosphomannomutase alpha/beta/alpha domain I	518		COG0033	
	Emin_1317	Carbohydrate kinase family protein	287		COG0063	
	Emin_1321	Phosphoglucosamine mutase	444		COG1109	
	Emin_1340	Phosphoglycerate mutase, 2,3-bisphosphoglycerate-independent	518	EC 5.4.2.1	COG0696	
	Emin_1347	*N*-Acetylglucosamine-6-phosphate deacetylase	387	EC 3.5.1.25	COG1820	
	Emin_1348	Glucosamine-6-phosphate deaminase	261	EC 3.5.99.6	COG0363	
	Emin_1349	PTS system, *N*-acetylglucosamine-specific IIBC subunit	481		COG1263	T
	Emin_1350	PTS system, glucose subfamily, IIA subunit; phosphocarrier	835		COG1080	
	Emin_1479	Sugar phosphate permease	405		COG2271	T
	Emin_1520	ADP-L-glycero-D-manno-heptose-6-epimerase	312		COG0451	S
	Emin_1531	Glycoside hydrolase family	478		COG1472	
N	**Cell motility**					
	Emin_0015	PilE-like protein	193		COG4968	S,T
	Emin_0057	PilE-like protein	169		COG4968	S,T
	Emin_0059	PilE-like protein	151		COG4968	S,T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0093	PilE-like protein	169		COG4968	S,T
	Emin_0096	PilE-like protein	163		COG4968	S,T
	Emin_0148	PilE-like protein	174		COG4968	S,T
	Emin_0204	PilE-like protein	155		COG4968	S,T
	Emin_0210	PilE-like protein	154		COG4968	S,T
	Emin_0216	PilE-like protein	163		COG4968	S,T
	Emin_0236	Type II secretion system subunit O	260		COG1989	S,T
	Emin_0265	PilE-like protein	166		COG4968	S,T
	Emin_0274	Type II secretion system subunit G	174		COG2165	S,T
	Emin_0301	Putative methyl-accepting chemotaxis sensory transducer	161		COG0840	S,T
	Emin_0331	PilE-like protein	172		COG4968	S,T
	Emin_0340	PilE-like protein	165		COG4968	S,T
	Emin_0345	PilE-like protein	156		COG4968	S,T
	Emin_0349	PilE-like protein	179		COG4968	S,T
	Emin_0376	PilE-like protein	174		COG4968	S,T
	Emin_0383	PilE-like protein	168		COG4968	S,T
	Emin_0392	PilE-like protein	175		COG4968	S,T
	Emin_0407	Putativeley involved in type II secretion system	423		COG3063	S,T
	Emin_0422	PilE-like protein	163		COG4968	S,T
	Emin_0449	PilE-like protein	167		COG4968	S,T
	Emin_0475	PilE-like protein	171		COG4968	S,T
	Emin_0482	PilE-like protein	171		COG4968	S,T
	Emin_0483	Type II secretion system subunit G	589		COG2165	S,T
	Emin_0495	PilE-like protein	172		COG4968	S,T
	Emin_0526	PilE-like protein	158		COG4968	S
	Emin_0535	PilE-like protein	634		COG4968	S,T
	Emin_0536	Type II secretion system subunit G	533		COG2165	S,T
	Emin_0543	Putative type II secretion system protein	362		COG2805	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0548	Type II secretion system subunit H, I, J	169		COG4969	S,T
	Emin_0567	PilE-like protein	147		COG4968	S,T
	Emin_0571	PilE-like protein	163		COG4968	S,T
	Emin_0575	Putatively involved in type II secretion system	536		COG4972	T
	Emin_0578	Putatively involved in type II secretion system	527		COG1450	S,T
	Emin_0579	Type II secretion system subunit E	576		COG2804	
	Emin_0580	Type II secretion system subunit F	412		COG1459	T
	Emin_0581	PilE-like protein	165		COG4968	S,T
	Emin_0582	PilE-like protein	173		COG4968	S,T
	Emin_0583	PilE-like protein	166		COG4968	S,T
	Emin_0606	PilE-like protein	156		COG4968	S,T
	Emin_0612	PilE-like protein	161		COG4968	S,T
	Emin_0613	Type II secretion system subunit G	165		COG2165	S,T
	Emin_0617	PilE-like protein	158		COG4968	S,T
	Emin_0631	Putatively involved in type II secretion system	200		COG3063	S
	Emin_0651	PilE-like protein	165		COG4968	S,T
	Emin_0668	PilE-like protein	155		COG4968	S,T
	Emin_0669	PilE-like protein	48		COG4968	S,T
	Emin_0681	Type II secretion system subunit G	567		COG2165	S,T
	Emin_0682	PilE-like protein	165		COG4968	S,T
	Emin_0692	Type II secretion system subunit G	548		COG2165	S,T
	Emin_0699	Putatively involved in type II secretion system	563		COG3063	S
	Emin_0765	PilE-like protein	170		COG4968	S,T
	Emin_0779	Type II secretion system subunit G	136		COG2165	S,T
	Emin_0780	PilE-like protein	157		COG4968	S,T
	Emin_0788	PilE-like protein	79		COG4968	S,T
	Emin_0800	PilE-like protein	167		COG4968	S,T
	Emin_0802	PilE-like protein	152		COG4968	S,T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0807	PilE-like protein	182		COG4968	S,T
	Emin_0926	Type II secretion system subunit H, I, J	171		COG4969	S,T
	Emin_0933	PilE-like protein	527		COG4968	S,T
	Emin_1018	PilE-like protein	161		COG4968	S,T
	Emin_1058	Type II secretion system subunit H, I, J	176		COG4969	S,T
	Emin_1096	PilE-like protein	157		COG4968	S,T
	Emin_1097	PilE-like protein	151		COG4968	S,T
	Emin_1142	Putatively involved in type II secretion system	331		COG3063	S,T
	Emin_1161	PilE-like protein	157		COG4968	S,T
	Emin_1204	TadC subunit	296		COG2064	S,T
	Emin_1230	PilE-like protein	178		COG4968	S,T
	Emin_1261	PilE-like protein	168		COG4968	S,T
	Emin_1280	PilE-like protein	152		COG4968	S,T
	Emin_1300	PilE-like protein	158		COG4968	S,T
	Emin_1371	PilE-like protein	174		COG4968	S,T
	Emin_1387	PilE-like protein	175		COG4968	S,T
	Emin_1421	Type II secretion system subunit G	172		COG2165	S,T
	Emin_1422	PilE-like protein	169		COG4968	S,T
	Emin_1436	PilE-like protein	171		COG4968	S,T
	Emin_1442	Type II secretion system subunit H, I, J	155		COG4969	S,T
	Emin_1487	PilE-like protein	164		COG4968	S,T
	Emin_1488	PilE-like protein	158		COG4968	S,T
	Emin_1495	PilE-like protein	167		COG4968	S,T
	Emin_1524	PilE-like protein	159		COG4968	S,T
	Emin_1548	PilE-like protein	167		COG4968	S,T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
M	**Cell wall/membrane/envelope biosynthesis**					
	Emin_0017	Glycosyl transferase group 1	356		COG0438	
	Emin_0033	Lysine biosynthesis	289		COG0329	
	Emin_0037	Putative UDP-*N*-acetylmuramate-alanine ligase	475		COG0773	S
	Emin_0053	Apolipoprotein *N*-acyltransferase	553		COG0815	T
	Emin_0073	Outer membrane protein assembly complex, YaeT protein	772		COG4775	S
	Emin_0075	UDP-3-*O*-(3-hydroxymyristoyl) glucosamine *N*-acyltransferase	341		COG1044	
	Emin_0076	UDP-3-*O*-acyl *N*-acetylglucosamine deacetylase	277		COG0774	
	Emin_0078	Acyl-(acyl-carrier-protein)/UDP-*N*- acetylglucosamine *O*-acyltransferase	267	EC 2.3.1.129	COG1043	
	Emin_0108	Large conductance mechano-sensitive channel protein	156		COG1970	T
	Emin_0120	Outer membrane efflux protein	448		COG1538	S
	Emin_0121	Efflux transporter, RND family, MFP subunit	295		COG0845	S,T
	Emin_0132	GTP-binding protein LepA	605		COG0481	
	Emin_0163	Putative peptidoglycan glycosyltransferase	713	EC 2.4.1.129	COG5009	S,T
	Emin_0177	Membrane proteins related to metalloendopeptidases	298		COG0739	T
	Emin_0178	Integral membrane protein CcmA involved in cell shape determination	122		COG1664	
	Emin_0232	*N*-acetylmuramoyl-L-alanine amidase	641		COG0860	S
	Emin_0276	Methyltransferase family protein	223		COG2230	
	Emin_0277	Glycosyl transferase family 8	320		COG1442	S
	Emin_0278	Glutamate racemase	264	EC 5.1.1.3	COG0796	
	Emin_0302	Outer membrane efflux protein	431		COG1538	S
	Emin_0318	Outer membrane protein	174		COG2885	S
	Emin_0321	Putative mechano-sensitive ion channel	424		COG0668	T
	Emin_0324	Putative protein of poly-gamma-glutamate biosynthesis	346		COG2843	S
	Emin_0326	Alanine racemase	369	EC 5.1.1.1	COG0787	
	Emin_0338	ADP-heptose:LPS heptosyltransferase	320		COG0859	
	Emin_0341	Glycosyl transferase family	307		COG0463	T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0351	Putative outer membrane protein	452		COG4775	S
	Emin_0374	Prolipoprotein diacylglyceryl transferase	263		COG0682	T
	Emin_0377	Alkaline phosphatase superfamily	671		COG1368	S,T
	Emin_0409	Hydroxymethylbutenyl pyrophosphate reductase	564	EC 1.17.1.2	COG0761	T
	Emin_0440	Outer membrane lipoprotein carrier	326		COG2834	S
	Emin_0445	Putative periplasmic protein TonB	244		COG0810	S,T
	Emin_0447	Nucleoside-diphosphate-sugar epimerases	316		COG0451	T
	Emin_0453	Outer membrane protein	993		COG2885	
	Emin_0487	Glycosyltransferase	394		COG0438	T
	Emin_0502	Uncharacterized protein, involved in the regulation of septum location	98		COG2088	
	Emin_0552	Lipopolysaccharide heptosyltransferase II	516		COG0859	S
	Emin_0553	3-Deoxy-D-manno-octulosonic-acid transferase domain protein	423		COG1519	S,T
	Emin_0565	Sulfatase family protein	581		COG1368	S,T
	Emin_0591	Arabinose-5-phosphate isomerase	329	EC 5.3.1.13	COG0794	
	Emin_0593	2-Dehydro-3-deoxyphosphooctonate aldolase	275	EC 2.5.1.55	COG2877	S
	Emin_0608	UDP-glucose 4-epimerase	324	EC 5.1.3.2	COG1087	
	Emin_0620	Tetraacyldisaccharide 4'-kinase	374	EC 2.7.1.130	COG1663	
	Emin_0626	Glycosyl transferase family 9	340		COG0859	T
	Emin_0689	Putative membrane-associated zinc metalloprotease	376		COG0750	S,T
	Emin_0702	D-ala-D-ala carboxypeptidase	496	EC 3.4.16.4	COG2027	S
	Emin_0754	Aminoglycoside phosphotransferase	249		COG0510	
	Emin_0760	Putative enzyme of poly-gamma-glutamate biosynthesis	363		COG2843	
	Emin_0763	Outer membrane protein and related peptidoglycan-associated (lipo)proteins	189		COG2885	S
	Emin_0822	Outer membrane protein	214		COG3047	S,T
	Emin_0835	D-Alanyl-D-alanine dipeptidase	263		COG2173	S,T
	Emin_0836	Membrane carboxypeptidase	719		COG4953	S,T
	Emin_0854	Uncharacterized protein involved in outer membrane biogenesis	715		COG2982	S,T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0861	DTDP-glucose 4,6-dehydratase	337	EC 4.2.1.46	COG1088	S,T
	Emin_0863	DegT/DnrJ/EryC1/StrS aminotransferase	377		COG0399	
	Emin_0864	DTDP-4-dehydrorhamnose 3,5-epimerase	185	EC 5.1.3.13	COG1898	
	Emin_0865	Glucose-1-phosphate thymidylyltransferase	294	EC 2.7.7.24	COG1209	
	Emin_0866	Lipopolysaccharide biosynthesis protein-like protein	317		COG3754	
	Emin_0867	ADP-heptose:LPS heptosyltransferase-like protein	805		COG0859	
	Emin_0892	UDP-*N*-acetylmuramyl pentapeptide phosphotransferase/UDP-*N*-acetylglucosamine-1-phosphate transferas	342		COG0472	S,T
	Emin_0899	Lipid A biosynthesis acyltransferase	304		COG1560	S,T
	Emin_0902	ADP-heptose synthase	332		COG2870	
	Emin_0903	3-Deoxy-D-manno-octulosonate cytidylyltransferase	245		COG1212	
	Emin_0910	Outer membrane protein	487		COG1538	S
	Emin_0916	Uncharacterized protein involved in outer membrane biogenesis	767		COG2982	T
	Emin_0934	Outer membrane protein	217		COG2885	S,T
	Emin_1027	Outer membrane efflux protein	452		COG1538	S
	Emin_1040	Dehydrogenase family protein	308		COG0451	
	Emin_1042	ABC transporter related	447		COG1538	S
	Emin_1043	ABC transporter related	239		COG0845	S,T
	Emin_1050	Mechano-sensitive ion channel	481		COG3264	S,T
	Emin_1069	Putative phosphoglycerol transferase	641		COG1368	S,T
	Emin_1087	O-antigen polymerase	612		COG3307	S,T
	Emin_1088	O-antigen polymerase	737		COG3307	S,T
	Emin_1090	D-Alanine/D-alanine ligase	353	EC 6.3.2.4	COG1181	
	Emin_1093	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	422	EC 2.5.1.7	COG0766	
	Emin_1143	Lipoprotein signal peptidase	165		COG0597	T
	Emin_1177	Outer membrane efflux protein	427		COG1538	S,T
	Emin_1178	Efflux transporter, RND family, MFP subunit	390		COG0845	T
	Emin_1184	RfaE bifunctional protein	154		COG2870	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1217	Periplasmic protease	444	EC 3.4.21.102	COG0793	S,T
	Emin_1248	Glycosyl transferase family 2	249		COG0463	T
	Emin_1250	Undecaprenyldiphospho-muramoylpentapeptide beta-*N*-acetylglucosaminyltransferase	361	EC 2.4.1.227	COG0707	
	Emin_1252	Phospho-*N*-acetylmuramoyl-pentapeptide- transferase	361	EC 2.7.8.13	COG0472	S,T
	Emin_1253	UDP-*N*-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate/D-alanyl-D-alanyl ligase	450		COG0770	
	Emin_1254	Peptidoglycan glycosyltransferase	562	EC 2.4.1.129	COG0768	S,T
	Emin_1256	Predicted *S*-adenosylmethionine-dependent methyltransferase involved in cell envelope biogenesis	293		COG0275	
	Emin_1258	Methyltransferase GidB	202		COG0357	
	Emin_1271	UDP-*N*-acetylenolpyruvoylglucosamine reductase	298	EC 1.1.1.158	COG0812	
	Emin_1277	Outer membrane protein	172		COG2885	S
	Emin_1293	UDP-*N*-acetylglucosamine pyrophosphorylase	484	EC 2.3.1.157	COG1207	
	Emin_1301	Predicted pyridoxal phosphate-dependent enzyme	373		COG0399	
	Emin_1306	Glycosyl transferase	355		COG0438	S
	Emin_1307	Glycosyl transferase	335		COG0463	
	Emin_1310	Putative lipoprotein releasing system	409		COG4591	S,T
	Emin_1319	Glucosamine/fructose-6-phosphate aminotransferase, isomerizing	614	EC 2.6.1.16	COG0449	
	Emin_1331	Putative mechano-sensitive ion channel	541		COG0668	S,T
	Emin_1344	Uncharacterized protein involved in outer membrane biogenesis	776		COG2982	S,T
	Emin_1368	Membrane proteins related to metalloendopeptidases	434		COG0739	S,T
	Emin_1374	*N*-Acetylmuramoyl-L-alanine amidase	529	EC 3.5.1.28	COG0860	S
	Emin_1378	Penicillin-binding protein 2	598	EC 2.4.1.129	COG0768	S,T
	Emin_1380	Rod-shape-determining protein MreC	276		COG1792	S,T
	Emin_1508	Lipid-A-disaccharide synthase	382	EC 2.4.1.182	COG0763	
	Emin_1511	UDP-*N*-acetylmuramyl tripeptide synthase	479		COG0769	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1512	UDP-*N*-acetylmuramoylalanine/D-glutamate ligase	452		COG0771	S
	Emin_1520	ADP-L-glycero-D-manno-heptose-6-epimerase	312		COG0451	S
	Emin_1550	Glycosyltransferase	398		COG0438	
H	**Coenzyme transport and metabolism**					
	Emin_0055	Putative oxygen-independent coproporphyrinogen III oxidase	365	EC 1.3.99.22	COG0635	
	Emin_0106	Pantetheine-phosphate adenylyltransferase	161	EC 2.7.7.3	COG0669	
	Emin_0147	Putative 2-amino-3-ketobutyrate CoA ligase	395	EC 2.3.1.29	COG0156	
	Emin_0168	Oxidoreductase FAD/NAD(P)-binding domain protein	286		COG0543	
	Emin_0248	Methyltransferase family protein	195		COG2226	
	Emin_0264	Methenyltetrahydrofolate cyclohydrolase	288	EC 3.5.4.9	COG0190	
	Emin_0267	Geranylgeranyl pyrophosphate synthase	334		COG0142	
	Emin_0268	Deoxyxylulose-5-phosphate synthase	621	EC 2.2.1.7	COG1154	
	Emin_0330	FAD dependent oxidoreductase	511	EC 1.4.3.16	COG0029	
	Emin_0381	Lipoic acid synthetase	297		COG0320	
	Emin_0419	Quinolinate synthetase complex, A subunit	344	EC 2.1.1.63	COG0379	
	Emin_0420	Nicotinate-nucleotide pyrophosphorylase	277	EC 2.4.2.19	COG0157	
	Emin_0498	Dephospho-CoA kinase	187	EC 2.7.1.24	COG0237	
	Emin_0549	Methyltransferase domain	255		COG2227	
	Emin_0633	Dihydropteroate synthase	258	EC 2.5.1.15	COG0294	
	Emin_0634	GTP cyclohydrolase I	172	EC 3.5.4.16	COG0302	
	Emin_0636	Putative 6-pyruvoyl tetrahydropterin synthase	115		COG0720	
	Emin_0652	NAD^+ synthetase	271	EC 6.3.1.5	COG0171	
	Emin_0677	Riboflavin biosynthesis protein RibF	302		COG0196	
	Emin_0710	Putative phosphoglycerate dehydrogenase	315		COG0111	
	Emin_0711	Phosphoserine transaminase	383	EC 2.6.1.52	COG1932	
	Emin_0728	Oxidoreductase	257		COG0543	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0758	Methyltransferase family protein	261		COG2226	S,T
	Emin_0761	Putative 6-pyruvoyl-tetrahydropterin synthase	120		COG0720	
	Emin_0762	2-Amino-4-hydroxy-6- hydroxymethyldihydropteridine pyrophosphokinase	163	EC 2.7.6.3	COG0801	
	Emin_0893	Amine oxidase family protein	431		COG1232	S,T
	Emin_0894	Nicotinate nucleotide adenylyltransferase family protein	375		COG1713	
	Emin_1013	Putative lactate dehydrogenase	338		COG1052	
	Emin_1092	Putative folylpoly-gamma-glutamate synthetase	410		COG0285	
	Emin_1100	4'-Phosphopantetheinyl transferase	206		COG2091	
	Emin_1122	Von Willebrand factor type A	335		COG1240	S,T
	Emin_1123	Von Willebrand factor type A	373		COG1240	S,T
	Emin_1146	Na^+/pantothenate symporter	469		COG4145	S,T
	Emin_1147	Thiamine pyrophosphokinase	205		COG1564	
	Emin_1185	Putative 8-amino-7-oxononanoate synthase	396	EC 2.3.1.29	COG0156	
	Emin_1246	Methionine adenosyltransferase	386	EC 2.5.1.6	COG0192	
	Emin_1320	Putative pyridoxine 5 phosphate synthase	242		COG0854	S
	Emin_1329	Radical SAM domain protein	346		COG0502	
	Emin_1333	Biotin and thiamin synthesis associated	458		COG1060	
	Emin_1346	Biotin/lipoate A/B protein ligase	238		COG0095	
	Emin_1482	4-Hydroxythreonine-4-phosphate dehydrogenase	309	EC 1.1.1.262	COG1995	
	Emin_1535	Biotin/acetyl-CoA-carboxylase ligase	186		COG0340	
V	**Defense mechanism**					
	Emin_0122	The hydrophobe/amphiphile Efflux-1 (HAE1) Family	1073		COG0841	S,T
	Emin_0226	Restriction endonuclease	887		COG3587	
	Emin_0400	MATE efflux family protein	456		COG0534	T
	Emin_0494	Na^+-driven multidrug efflux pump	455		COG0534	S,T
	Emin_0855	Restriction endonuclease	230		COG1403	
	Emin_0879	ABC-type multidrug transport system	592		COG1132	T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0880	ABC-type multidrug transport system	579		COG1132	T
	Emin_0888	ABC transporter	593		COG1132	T
	Emin_1025	ABC-2 type transporter	391		COG0842	T
	Emin_1026	Secretion protein HlyD family protein	326		COG1566	T
	Emin_1044	ABC transporter related	239		COG1131	
	Emin_1045	ABC transporter related	311		COG1131	
	Emin_1046	ABC transporter related	383		COG0842	S,T
	Emin_1047	ABC transporter related	371		COG0842	T
	Emin_1175	Inner membrane protein involved in colicin E2 resistance	469		COG4452	S,T
	Emin_1179	ABC transporter related	235		COG1136	
	Emin_1180	ABC-type efflux carrier	412		COG0577	S,T
	Emin_1289	Undecaprenyl-diphosphatase	258	EC 3.6.1.27	COG1968	S,T
	Emin_1311	Putative lipoprotein-releasing system	221		COG1136	
C	**Energy production and conversion**					
	Emin_0009	Ruberythrin rbr	191		COG1592	
	Emin_0010	redoxin	168		COG2077	
	Emin_0011	Superoxid reductase sor	131	EC 1.15.1.1	COG2033	
	Emin_0012	Rubredoxine:oxygen oxidoreductase roo	400		COG0426	
	Emin_0014	Rubredoxine rub	52		COG1773	
	Emin_0049	FAD/FMN-containing dehydrogenase	656	EC 1.1.2.4	COG0277	
	Emin_0050	Fe-S oxidoreductase-like protein	406		COG0247	
	Emin_0168	Oxidoreductase FAD/NAD(P)-binding domain protein	286		COG0543	
	Emin_0217	Nitroreductase	202		COG0778	S
	Emin_0262	Glycerophosphodiester phosphodiesterase	224	EC 3.1.4.46	COG0584	
	Emin_0282	Dehydrogenase family protein	353		COG0644	S
	Emin_0296	Malic enzyme, $NADP^+$-dependent	489	EC 1.1.1.40	COG0281	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0297	Putative fumarate hydratase	279	EC 4.2.1.2	COG1951	S
	Emin_0298	Putative fumarate hydratase	171		COG1838	
	Emin_0310	4Fe-4S ferredoxin iron-sulfur binding domain protein	255		COG2221	S
	Emin_0404	Isocitrate dehydrogenase (NAD^+)	334	EC 1.1.1.41	COG0473	
	Emin_0436	Fe-S cluster assembly protein NifU	284		COG0822	
	Emin_0469	Iron-containing alcohol dehydrogenase	387		COG1979	
	Emin_0531	Putative FeFe hydrogenase subunit HydC (NuoE)	163		COG1905	
	Emin_0532	FeFe Hydrogenase HydB (NuoF)	620	EC 1.6.99.5	COG1894	
	Emin_0694	putative Nitroreductase	183		COG0778	S
	Emin_0720	Acetate kinase	452	EC 2.7.2.1	COG0282	
	Emin_0728	Oxidoreductase	257		COG0543	
	Emin_0748	Phosphoenolpyruvate carboxykinase (ATP)	534	EC 4.1.1.49	COG1866	
	Emin_0753	Predicted novel	79		COG1905	
	Emin_0755	Iron-containing alcohol dehydrogenase	872	EC 1.2.1.10	COG1454	
	Emin_0772	Iron-containing alcohol dehydrogenase	887	EC 1.2.1.10	COG1454	
	Emin_0797	NifU-related domain-containing protein	238		COG0822	
	Emin_0806	Pyruvate:ferredoxin oxidoreductase	1185		COG0674	
	Emin_0829	Inorganic pyrophosphatase	782	EC 3.6.1.1	COG3808	S,T
	Emin_0852	Putative ADP acetyl-CoA synthetase	732		COG1042	
	Emin_0878	Flavodoxin	141		COG0716	
	Emin_0912	4Fe-4S ferredoxin iron-sulfur binding domain protein	236		COG1149	
	Emin_1013	Putative lactate dehydrogenase	338		COG1052	
	Emin_1036	Rubrerythrin	179		COG1592	
	Emin_1070	Glycerol kinase	481		COG0554	
	Emin_1072	Putative indolepyruvate ferredoxin oxidoreductase	56		COG4231	
	Emin_1127	Hydrogenase maturation protease	158		COG0680	
	Emin_1128	Ech hydrogenase subunit F	140		COG1143	
	Emin_1129	Ech hydrogenase putative subunit B	322	EC 1.6.99.5	COG1005	T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1130	NiFe-hydrogenase III large subunit	402		COG3261	
	Emin_1131	Putative Ech hydrogenase component	175		COG0852	
	Emin_1132	NiFe-hydrogenase III small subunit	149		COG3260	T
	Emin_1134	Putative Ech hydrogenase component	114		COG0713	S,T
	Emin_1136	Putative Ech hydrogenase component	476	EC 1.6.99.5	COG1008	S,T
	Emin_1137	Putative Ech hydrogenase component	595		COG1009	S,T
	Emin_1138	Putative Ech hydrogenase component	470	EC 1.6.99.5	COG1009	T
	Emin_1283	Predicted oxidoreductases	306		COG0667	
	Emin_1328	Putative sepiapterin reductase	170		COG0778	
	Emin_1363	Dihydrolipoamide dehydrogenase	448		COG1249	
	Emin_1366	Acetate kinase	370	EC 2.7.2.7	COG3426	
	Emin_1367	Putative phosphotransacetylase	305	EC 2.3.1.19	COG0280	
	Emin_1431	4Fe-4S ferredoxin iron-sulfur binding domain protein	264		COG0716	
	Emin_1475	Putative 2-ketovalerate ferredoxin oxidoreductase subunit	67		COG1149	
	Emin_1476	2-Ketovalerate ferredoxin oxidoreductase	375		COG0674	
	Emin_1477	2-Ketovalerate ferredoxin oxidoreductase subunit	247		COG1013	
	Emin_1478	2-Ketovalerate ferredoxin oxidoreductase subunit	178		COG1014	S
	Emin_1484	Electron transfer flavoprotein alpha subunit	397		COG2025	
	Emin_1485	Electron transfer flavoprotein alpha/beta subunit	262		COG2086	
	Emin_1513	Nitroreductase	180		COG0778	
	Emin_1515	ATP synthase F0, A subunit	248		COG0356	T
	Emin_1517	ATP synthase F0, B subunit	165		COG0711	S,T
	Emin_1518	ATP synthase F1, delta subunit	175		COG0712	
	Emin_1523	ATP synthase F1, epsilon subunit	135		COG0355	
	Emin_1110	Glycerol-3-phosphate dehydrogenase	327	EC 1.1.1.94	COG0240	S
	Emin_1519	ATP synthase F1, alpha subunit	512	EC 3.6.3.14	COG0056	
	Emin_1521	ATP synthase F1, gamma subunit	303	EC 3.6.3.14	COG0224	S
	Emin_1522	ATP synthase F1, beta subunit	484		COG0055	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
P	**Inorganic transport and metabolism**					
	Emin_0019	Putative divalent metal ion transporter	303		COG0598	T
	Emin_0043	Phosphate transporter	331		COG0306	S,T
	Emin_0044	Phosphate transport regulator	209		COG1392	
	Emin_0405	Sulfate permease	560		COG0659	T
	Emin_0448	Zinc transporter	265		COG0428	S,T
	Emin_0476	Na^+/phosphate symporter	586		COG1283	S,T
	Emin_0489	Cation transport ATPase	908		COG0474	T
	Emin_0555	Putative Mg/Co/Ni transporter	252		COG2239	
	Emin_0909	Cation diffusion facilitator family transporter	327		COG1230	T
	Emin_1011	Carbonic anhydrase	215		COG0288	
	Emin_1037	Ferric uptake regulator	133		COG0735	
	Emin_1051	Hemerythrin-like metal-binding protein	137		COG2703	
	Emin_1059	Magnesium-translocating ATPase	896	EC 3.6.3.2	COG0474	S,T
	Emin_1073	Putative indolepyruvate ferredoxin oxidoreductase	650		COG0370	T
	Emin_1137	Putative Ech hydrogenase component	595		COG1009	S,T
	Emin_1138	Putative Ech hydrogenase component	470	EC 1.6.99.5	COG1009	T
	Emin_1141	Cation transport ATPase	635		COG2217	S,T
	Emin_1160	Chloride channel	494		COG0038	S,T
	Emin_1220	Putative dihydroneopterin monophosphate dephosphorylase	454		COG1785	S
	Emin_1233	ABC-type dipeptide/oligopeptide/nickel transport system	324		COG0444	
	Emin_1235	ABC-type dipeptide/oligopeptide/nickel transport system	341		COG0601	S,T
	Emin_1236	ABC-type dipeptide/oligopeptide/nickel transport system	288		COG1173	S,T
	Emin_1304	Na+/H+ exchanger	435		COG0025	S,T
	Emin_1443	Integral membrane protein TerC	305		COG0861	S,T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
U	**Intracellular trafficking, secretion, and vesicular transport**					
	Emin_0015	PilE-like protein	193		COG4968	S,T
	Emin_0016	Signal peptidase I	324	EC 3.4.21.89	COG0681	S,T
	Emin_0025	Flp pilus assembly protein TadD, contains TPR repeats	445		COG5010	S
	Emin_0032	Signal peptide peptidase SppA, 36K type	371		COG0616	S,T
	Emin_0052	Preprotein translocase, SecA subunit	866		COG0653	
	Emin_0057	PilE-like protein	169		COG4968	S,T
	Emin_0059	PilE-like protein	151		COG4968	S,T
	Emin_0093	PilE-like protein	169		COG4968	S,T
	Emin_0096	PilE-like protein	163		COG4968	S,T
	Emin_0120	Outer membrane efflux protein	448		COG1538	S
	Emin_0148	PilE-like protein	174		COG4968	S,T
	Emin_0157	DNA protecting protein DprA (DNA processing chain A)	372		COG0758	
	Emin_0188	Multiple antibiotic resistance (MarC)-related protein	197		COG2095	T
	Emin_0204	PilE-like protein	155		COG4968	S,T
	Emin_0210	PilE-like protein	154		COG4968	S,T
	Emin_0216	PilE-like protein	163		COG4968	S,T
	Emin_0236	Type II secretion system subunit O	260		COG1989	S,T
	Emin_0265	PilE-like protein	166		COG4968	S,T
	Emin_0274	Type II secretion system subunit G	174		COG2165	S,T
	Emin_0302	Outer membrane efflux protein	431		COG1538	S
	Emin_0331	PilE-like protein	172		COG4968	S,T
	Emin_0340	PilE-like protein	165		COG4968	S,T
	Emin_0345	PilE-like protein	156		COG4968	S,T
	Emin_0349	PilE-like protein	179		COG4968	S,T
	Emin_0361	Membrane protein implicated in regulation of membrane protease activity	141		COG1585	T
	Emin_0376	PilE-like protein	174		COG4968	S,T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0383	PilE-like protein	168		COG4968	S,T
	Emin_0392	PilE-like protein	175		COG4968	S,T
	Emin_0407	Putatively involved in type II secretion system	423		COG3063	S,T
	Emin_0422	PilE-like protein	163		COG4968	S,T
	Emin_0443	Biopolymer transport protein	209		COG0811	S,T
	Emin_0444	Biopolymer transport protein	134		COG0848	T
	Emin_0449	PilE-like protein	167		COG4968	S,T
	Emin_0475	PilE-like protein	171		COG4968	S,T
	Emin_0482	PilE-like protein	171		COG4968	S,T
	Emin_0483	Type II secretion system subunit G	589		COG2165	S,T
	Emin_0495	PilE-like protein	172		COG4968	S,T
	Emin_0526	PilE-like protein	158		COG4968	S
	Emin_0535	PilE-like protein	634		COG4968	S,T
	Emin_0536	Type II secretion system subunit G	533		COG2165	S,T
	Emin_0543	Putatively involved in type II secretion system	362		COG2805	
	Emin_0548	Type II secretion system subunit H, I, J	169		COG4969	S,T
	Emin_0562	Protein-export membrane protein SecF	321		COG0341	S,T
	Emin_0563	Protein-export membrane protein SecD	473		COG0342	S,T
	Emin_0564	Pre-protein translocase subunit YajC	93		COG1862	S,T
	Emin_0567	PilE-like protein	147		COG4968	S,T
	Emin_0571	PilE-like protein	163		COG4968	S,T
	Emin_0575	Putatively involved in type II secretion system	536		COG4972	T
	Emin_0578	Putatively involved in type II secretion system	527		COG1450	S,T
	Emin_0579	Type II secretion system subunit E	576		COG2804	
	Emin_0580	Type II secretion system subunit F	412		COG1459	T
	Emin_0581	PilE-like protein	165		COG4968	S,T
	Emin_0582	PilE-like protein	173		COG4968	S,T
	Emin_0583	PilE-like protein	166		COG4968	S,T

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0606	PilE-like protein	156		COG4968	S,T
	Emin_0612	PilE-like protein	161		COG4968	S,T
	Emin_0613	Type II secretion system subunit G	165		COG2165	S,T
	Emin_0617	PilE-like protein	158		COG4968	S,T
	Emin_0631	Putatively involved in type II secretion system	200		COG3063	S
	Emin_0651	PilE-like protein	165		COG4968	S,T
	Emin_0668	PilE-like protein	155		COG4968	S,T
	Emin_0669	PilE-like protein	48		COG4968	S,T
	Emin_0681	Type II secretion system subunit G	567		COG2165	S,T
	Emin_0682	PilE-like protein	165		COG4968	S,T
	Emin_0692	Type II secretion system subunit G	548		COG2165	S,T
	Emin_0699	Putatively involved in type II secretion system	563		COG3063	S
	Emin_0765	PilE-like protein	170		COG4968	S,T
	Emin_0766	Endopeptidase Clp	195	EC 3.4.21.92	COG0740	
	Emin_0779	Type II secretion system subunit G	136		COG2165	S,T
	Emin_0780	PilE-like protein	157		COG4968	S,T
	Emin_0788	PilE-like protein	79		COG4968	S,T
	Emin_0800	PilE-like protein	167		COG4968	S,T
	Emin_0802	PilE-like protein	152		COG4968	S,T
	Emin_0807	PilE-like protein	182		COG4968	S,T
	Emin_0910	Outer membrane protein	487		COG1538	S
	Emin_0926	Type II secretion system subunit H, I, J	171		COG4969	S,T
	Emin_0933	PilE-like protein	527		COG4968	S,T
	Emin_0954	Clp protease	355		COG0740	
	Emin_1018	PilE-like protein	161		COG4968	S,T
	Emin_1027	Outer membrane efflux protein	452		COG1538	S
	Emin_1042	ABC transporter related	447		COG1538	S
	Emin_1058	Type II secretion system subunits H, I, J	176		COG4969	S,T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1096	PilE-like protein	157		COG4968	S,T
	Emin_1097	PilE-like protein	151		COG4968	S,T
	Emin_1142	Putatively involved in type II secretion system	331		COG3063	S,T
	Emin_1143	Lipoprotein signal peptidase	165		COG0597	T
	Emin_1154	Tetratricopeptide repeat protein	511		COG5010	S
	Emin_1161	PilE-like protein	157		COG4968	S,T
	Emin_1177	Outer membrane efflux protein	427		COG1538	S,T
	Emin_1181	Pre-protein translocase, SecG subunit	100		COG1314	S,T
	Emin_1204	TadC subunit	296		COG2064	S,T
	Emin_1205	TadB	535		COG4965	S,T
	Emin_1206	TadA subunit	742		COG4962	
	Emin_1207	RcpA subunit	248		COG4964	S
	Emin_1208	Pilus assembly protein	253		COG3745	S,T
	Emin_1230	PilE-like protein	178		COG4968	S,T
	Emin_1261	PilE-like protein	168		COG4968	S,T
	Emin_1263	Pre-protein translocase subunit	531		COG0706	S,T
	Emin_1279	Putative component of the Tol biopolymer transport system	428		COG0823	S,T
	Emin_1280	PilE-like protein	152		COG4968	S,T
	Emin_1300	PilE-like protein	158		COG4968	S,T
	Emin_1371	PilE-like protein	174		COG4968	S,T
	Emin_1387	PilE-like protein	175		COG4968	S,T
	Emin_1399	Pre-protein translocase, SecY subunit	446		COG0201	S,T
	Emin_1421	Type II secretion system subunit G	172		COG2165	S,T
	Emin_1422	PilE-like protein	169		COG4968	S,T
	Emin_1436	PilE-like protein	171		COG4968	S,T
	Emin_1442	Type II secretion system subunits H, I, J	155		COG4969	S,T
	Emin_1487	PilE-like protein	164		COG4968	S,T
	Emin_1488	PilE-like protein	158		COG4968	S,T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1495	PilE-like protein	167		COG4968	S,T
	Emin_1524	PilE-like protein	159		COG4968	S,T
	Emin_1532	Multiple antibiotic transporter	228		COG2095	T
	Emin_1548	PilE-like protein	167		COG4968	S,T
I	**Lipid transport and metabolism**					
	Emin_0026	Squalene synthase	268		COG1562	
	Emin_0030	Kinase family protein	288		COG1597	
	Emin_0077	Beta-hydroxyacyl-(acyl-carrier protein) dehydratase FabA/FabZ	158		COG0764	
	Emin_0116	Putative phosphatidylserine decarboxylase	222		COG0688	S,T
	Emin_0117	CDP-diacylglycerol/serine *O*-phosphatidyltransferase	259		COG1183	T
	Emin_0169	Squalene synthetase	340		COG1562	T
	Emin_0268	Deoxyxylulose-5-phosphate synthase	621	EC 2.2.1.7	COG1154	
	Emin_0337	Phosphatidylglycerophosphate synthase	485		COG1502	S,T
	Emin_0409	Hydroxymethylbutenyl pyrophosphate reductase	564	EC 1.17.1.2	COG0761	T
	Emin_0415	Acetyl-CoA carboxylase, biotin carboxyl carrier protein	172		COG0511	
	Emin_0441	Phosphatidylglycerophosphate synthase	216		COG0558	S,T
	Emin_0442	Phosphatidylglycerophosphate synthase	162		COG1267	T
	Emin_0454	Lysophospholipase	523		COG2267	S,T
	Emin_0490	Phosphatidylserine/phosphatidylglycerophosphate/cardiolipin synthase-like protein	677		COG1502	S,T
	Emin_0501	4-Diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate synthase	301		COG1947	
	Emin_0545	1-Acyl-sn-glycerol-3-phosphate acyltransferase	195		COG0204	S
	Emin_0658	Fatty acid/phospholipid synthesis protein PlsX	341		COG0416	
	Emin_0659	3-Oxoacyl-(acyl-carrier protein) reductase	245	EC 1.1.1.100	COG1028	S
	Emin_0660	Acyl carrier protein	81		COG0236	
	Emin_0687	1-Hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase	363	EC 1.17.4.3	COG0821	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0690	1-Deoxy-D-xylulose 5-phosphate reductoisomerase	395	EC 1.1.1.267	COG0743	
	Emin_0691	Phosphatidate cytidylyltransferase	271		COG0575	S,T
	Emin_0705	Undecaprenyl diphosphate synthase	228	EC 2.5.1.31	COG0020	
	Emin_0844	Phospholipase D/transphosphatidylase	486		COG1502	S,T
	Emin_0907	Membrane protein putatively involved in long-chain fatty acid transport	430		COG2067	S,T
	Emin_0996	Putative phosphoribosylglycinamide synthetase	541		COG0439	
	Emin_1000	Membrane protein putatively involved in long-chain fatty acid transport	402		COG2067	S,T
	Emin_1012	AMP-dependent synthetase and ligase	1133		COG0318	T
	Emin_1099	Putative biotin/[propionyl-CoA-carboxylase (ATP-hydrolyzing)] ligase	535	EC 6.4.1.3	COG4799	
	Emin_1153	Acyl-ACP thioesterase	243		COG3884	
	Emin_1165	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	382		COG1211	
	Emin_1182	Biotin carboxyl carrier protein	124		COG0511	
	Emin_1183	Acetyl-CoA carboxylase, biotin carboxylase	455	EC 6.3.4.14	COG4770	
	Emin_1237	Putative fatty acid Co-A ligase	533		COG0318	T
	Emin_1318	Holo-acyl-carrier-protein synthase	120		COG0736	
	Emin_1486	Acyl-CoA dehydrogenase domain protein	547		COG1960	
	Emin_1528	Putative lysophospholipase	285		COG2267	S
F	**Nucleotide transport and metabolism**					
	Emin_0020	Purine-nucleoside phosphorylase	248		COG0813	T
	Emin_0062	Thymidylate synthase, flavin-dependent	228	EC 2.1.1.148	COG1351	
	Emin_0069	ADP-ribose pyrophosphatase	147		COG1051	
	Emin_0105	Putative deoxyguanosinetriphosphate triphosphohydrolase	376		COG0232	
	Emin_0107	Cytosine/adenosine deaminases	163	EC 3.5.4.3	COG0590	
	Emin_0152	Deoxyribose-phosphate aldolase	215	EC 4.1.2.4	COG0274	
	Emin_0220	Phosphoribosylaminoimidazolesuccinocarboxamide synthase	226	EC 6.3.2.6	COG0152	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0288	Malate dehydrogenase	486	EC 1.1.1.205 EC 1.1.1.37	COG0516	
	Emin_0315	Histidine triad (HIT) protein	112		COG0537	
	Emin_0417	Orotate phosphoribosyltransferase	196		COG0461	
	Emin_0514	Phosphoribosylaminoimidazole-succinocarboxamide synthase	292	EC 6.3.2.6	COG0152	
	Emin_0517	Phosphoribosylformylglycinamidine synthase I	248	EC 6.3.5.3	COG0047	
	Emin_0518	Amidophosphoribosyltransferase	458		COG0034	
	Emin_0519	Phosphoribosylamine/glycine ligase	415	EC 6.3.4.13	COG0151	S
	Emin_0520	Phosphoribosylaminoimidazole carboxylase, catalytic subunit	158	EC 4.1.1.21	COG0041	
	Emin_0521	Phosphoribosylformylglycinamidine cyclo-ligase	319	EC 6.3.3.1	COG0150	
	Emin_0522	Formyl transferase domain protein	187		COG0299	
	Emin_0523	Phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	517	EC 2.1.2.3 EC 3.5.4.10	COG0138	
	Emin_0544	Cytidylate kinase	220	EC 2.7.4.14	COG0283	
	Emin_0554	Deoxyuridine 5'-triphosphate nucleotidohydrolase Dut	141		COG0756	
	Emin_0630	Inosine guanosine and xanthosine phosphorylase family	273		COG0005	
	Emin_0707	Uridylate kinase	237		COG0528	
	Emin_0724	Uracil phosphoribosyltransferase	179	EC 2.4.2.9	COG2065	
	Emin_0725	Aspartate carbamoyltransferase	312	EC 2.1.3.2	COG0540	
	Emin_0726	Dihydroorotase	432	EC 3.5.2.3	COG0044	
	Emin_0729	Dihydroorotate dehydrogenase	303		COG0167	
	Emin_0730	Orotidine-5'-phosphate decarboxylase	234	EC 4.1.1.23	COG0284	
	Emin_0769	GMP synthase, large subunit	510		COG0519	
	Emin_0781	Adenine phosphoribosyltransferase	172	EC 2.4.2.7	COG0503	
	Emin_0837	Guanylate kinase	201	EC 2.7.4.8	COG0194	
	Emin_0904	CTP synthase	521	EC 6.3.4.2	COG0504	S
	Emin_1019	Thymidine kinase	186	EC 2.7.1.21	COG1435	
	Emin_1029	IMP dehydrogenase/GMP reductase	367		COG0516	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1031	Metallophosphoesterase	492		COG0737	S
	Emin_1032	Carbamoyl-phosphate synthase, small subunit	390		COG0505	
	Emin_1033	Carbamoyl-phosphate synthase, large subunit	1102	EC 6.3.4.16	COG0458	
	Emin_1035	Amidophosphoribosyltransferase	456		COG0034	
	Emin_1052	Putative formyltetrahydrofolate synthetase	556	EC 6.3.4.3	COG2759	
	Emin_1054	Anaerobic ribonucleoside-triphosphate reductase	746	EC 1.17.4.2	COG1328	S
	Emin_1162	Adenosine deaminase	402	EC 3.5.4.4	COG1816	
	Emin_1174	Deoxycytidylate deaminase	168		COG2131	S
	Emin_1190	Anaerobic ribonucleoside-triphosphate reductase	690		COG1328	
	Emin_1292	Ribose-phosphate pyrophosphokinase	316	EC 2.7.6.1	COG0462	
	Emin_1305	Non-canonical purine NTP pyrophosphatase	197		COG0127	
	Emin_1323	Hypoxanthine phosphoribosyltransferase	176	EC 2.4.2.8	COG0634	
	Emin_1343	Orotate phosphoribosyltransferase	192		COG0461	
	Emin_1359	Adenylosuccinate synthase	338	EC 6.3.4.4	COG0104	
	Emin_1360	Adenylosuccinate lyase	463	EC 4.3.2.2	COG0015	
	Emin_1398	Adenylate kinase	214	EC 2.7.4.3	COG0563	
	Emin_1461	Thymidylate kinase	203	EC 2.7.4.9	COG0125	
	Emin_1473	5'-Nucleotidase domain protein	494		COG0737	S
	Emin_1474	5'-Nucleotidase domain protein	491		COG0737	S
	Emin_1527	Nucleoside-diphosphate kinase	148	EC 2.7.4.6	COG0105	
O	**Posttranslational modification, protein turnover, chaperones**					
	Emin_0010	redoxin	168		COG2077	
	Emin_0032	Signal peptide peptidase SppA, 36K type	371		COG0616	S,T
	Emin_0100	Chaperone protein DnaK	619	EC 1.3.1.74	COG0443	
	Emin_0101	Molecular chaperone GrpE (heat shock protein)	186		COG0576	
	Emin_0156	Predicted ATPase with chaperone activity	508		COG0606	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0192	Peptidylprolyl isomerase FKBP-type	240		COG0545	S
	Emin_0236	Type II secretion system subunit O	260		COG1989	S,T
	Emin_0260	Thioredoxin-like proteins and domains	74		COG0694	
	Emin_0304	Tetratricopeptide TPR_2 repeat protein	324		COG4235	S,T
	Emin_0360	Chaperone DnaJ domain protein	327		COG0330	S,T
	Emin_0361	Membrane protein implicated in regulation of membrane protease activity	141		COG1585	T
	Emin_0379	Peptidase family M48	330		COG0501	S,T
	Emin_0388	(NiFe) hydrogenase maturation protein	71		COG0298	
	Emin_0389	(NiFe) hydrogenase maturation protein	346		COG0409	
	Emin_0390	(NiFe) hydrogenase maturation protein	335		COG0309	
	Emin_0391	(NiFe) hydrogenase maturation protein	736		COG0068	
	Emin_0398	DNA repair protein RadA	454	EC 2.1.1.63	COG1066	
	Emin_0485	Endopeptidase La	830	EC 3.4.21.53	COG0466	S
	Emin_0496	Thioredoxin	106		COG3118	
	Emin_0524	SsrA-binding protein	157		COG0691	
	Emin_0603	Thioredoxin reductase	309		COG0492	
	Emin_0635	Putative 6- pyruvoyltetrahydropterin 2-reductase	195		COG0602	
	Emin_0653	Subtilisin-like serine proteases	360		COG1404	S
	Emin_0722	Trigger factor domain	426		COG0544	
	Emin_0766	Endopeptidase Clp	195	EC 3.4.21.92	COG0740	
	Emin_0791	Chaperone DnaJ domain protein	323		COG0484	
	Emin_0817	Peptidylprolyl isomerase	138		COG1047	
	Emin_0845	Putative redox protein	132		COG1765	
	Emin_0924	Chaperonin GroEL	542		COG0459	
	Emin_0925	Chaperonin Cpn10	97		COG0234	
	Emin_0954	Clp protease	355		COG0740	
	Emin_1002	Heat shock protein	166		COG3187	S,T
	Emin_1016	Thioredoxin domain	105		COG3118	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1055	Anaerobic ribonucleoside-triphosphate reductase activating protein	171		COG1180	
	Emin_1156	Glycoprotease family protein	227		COG1214	S
	Emin_1188	Anaerobic ribonucleoside-triphosphate reductase activating protein	205		COG1180	S
	Emin_1216	Metalloendopeptidase, glycoprotease family	342	EC 3.4.24.57	COG0533	
	Emin_1299	Chaperone protein DnaJ	371		COG0484	
	Emin_1322	ATP-dependent metalloprotease FtsH	631		COG0465	S,T
	Emin_1385	Methionine-R-sulfoxide reductase	359		COG0225	S,T
	Emin_1439	Peptidylprolyl isomerase FKBP-type	125	EC 5.2.1.8	COG0545	
	Emin_1464	Deoxyhypusine synthase	375		COG1899	
	Emin_1526	ATPases with chaperone activity, ATP-binding subunit	839		COG0542	
	Emin_1530	Putative chaperone	354		COG2377	
L	**Replication, recombination, and repair**					
	Emin_0002	ATPase involved in DNA replication initiation	452		COG0593	
	Emin_0003	DNA polymerase III, beta subunit	369	EC 2.7.7.7	COG0592	
	Emin_0005	DNA gyrase, B subunit	825	EC 5.99.1.3	COG0187	
	Emin_0006	DNA gyrase, A subunit	840	EC 5.99.1.3	COG0188	
	Emin_0023	Integrase family protein	291		COG4974	
	Emin_0042	Exonuclease VII, large subunit	396	EC 3.1.11.6	COG1570	
	Emin_0103	Mg-dependent DNase	269		COG0084	
	Emin_0141	ATPase family protein	416		COG2256	
	Emin_0150	Endonuclease III	215	EC 4.2.99.18	COG0177	
	Emin_0153	ATPase involved in DNA replication initiation	1097		COG0593	
	Emin_0157	DNA protecting protein DprA (DNA processing chain A,)	372		COG0758	
	Emin_0158	DNA topoisomerase I	765	EC 5.99.1.2	COG0550	
	Emin_0159	Integrase family protein	295		COG4974	
	Emin_0162	Transcription-repair coupling factor	1044		COG1197	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0186	DNA repair protein RecO	232		COG1381	
	Emin_0193	Putative nuclease	182		COG1525	S,T
	Emin_0197	Excinuclease ABC C subunit domain protein	437		COG0322	
	Emin_0198	ATP-dependent DNA helicase RecG	691		COG1200	
	Emin_0203	DNA polymerase III, delta subunit	334		COG1466	
	Emin_0209	Superfamily I DNA and RNA helicases	786		COG0210	
	Emin_0212	DNA polymerase III, alpha subunit	1167	EC 2.7.7.7	COG0587	
	Emin_0213	NUDIX hydrolase	178		COG0494	
	Emin_0214	DNA mismatch repair protein MutS	851		COG0249	
	Emin_0215	DNA mismatch repair protein MutL	608		COG0323	
	Emin_0222	DNA repair photolyase	274		COG1533	
	Emin_0224	DNA methylase	656		COG2189	
	Emin_0225	Predicted ATP-dependent endonuclease	578		COG3593	
	Emin_0231	Resolvase domain	519		COG1961	
	Emin_0244	CRISPR-associated protein Cas1	298		COG1518	
	Emin_0250	DNA alkylation repair enzyme	235		COG4912	
	Emin_0307	Putative methyltransferase	226		COG0742	
	Emin_0308	RecB family exonuclease	244		COG2887	
	Emin_0313	DNA primase	578		COG0358	
	Emin_0325	Replicative DNA helicase	465		COG0305	
	Emin_0336	NUDIX hydrolase	261		COG2816	
	Emin_0426	ATPase involved in DNA repair	553		COG0497	
	Emin_0497	DNA polymerase I	862	EC 2.7.7.7	COG0749	
	Emin_0504	Holliday junction resolvasome, DNA binding subunit	202		COG0632	
	Emin_0505	Holliday junction resolvasome, helicase subunit	340		COG2255	
	Emin_0595	Transposase and inactivated derivatives	258		COG1943	
	Emin_0596	Excinuclease ATPase subunit	943		COG0178	
	Emin_0628	Ribonuclease H	210	EC 3.1.26.4	COG0164	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0629	Uncharacterized protein family UPF0102	122		COG0792	
	Emin_0678	DNA-3-methyladenine glycosylase I	189	EC 3.2.2.20	COG2818	
	Emin_0686	DNA methylase N-4/N-6 domain protein	266		COG0863	
	Emin_0773	Superfamily II DNA and RNA helicases	502		COG0513	S
	Emin_0839	Uracil-DNA glycosylase	297		COG1573	
	Emin_0840	Primosomal protein N	630		COG1198	
	Emin_0905	Methylated DNA-protein cysteine methyltransferase	157	EC 2.1.1.63	COG0350	
	Emin_0928	Serine/threonine protein kinase	761		COG0515	S,T
	Emin_0959	DNA modification methylase	396		COG0863	
	Emin_0962	Radical SAM domain protein	235		COG1533	
	Emin_0969	DNA modification methylase	222		COG0863	
	Emin_0974	ATPase involved in DNA replication initiation	215		COG0593	
	Emin_0989	Integrase family protein	355		COG4974	
	Emin_1030	Putative DNA repair photolyase	295		COG1533	
	Emin_1066	G:T/U mismatch-specific DNA glycosylase	162		COG3663	
	Emin_1068	Putative RecB family exonuclease	411		COG2887	
	Emin_1109	Bacterial DNA-binding protein	86		COG0776	
	Emin_1152	Crossover junction endodeoxyribonuclease RuvC	189	EC 3.1.22.4	COG0817	
	Emin_1173	Ribonuclease HI	194		COG0328	
	Emin_1227	DNA ligase, NAD-dependent	656	EC 6.5.1.2	COG0272	
	Emin_1247	Exodeoxyribonuclease III Xth	255	EC 4.2.99.18	COG0708	
	Emin_1282	DNA repair photolyase-like protein	426		COG1533	
	Emin_1286	UvrD/REP helicase	1074		COG1074	
	Emin_1287	ATP-dependent nuclease subunit B-like protein	994		COG3857	
	Emin_1309	Protein kinase	517		COG0515	S,T
	Emin_1339	DNA polymerase III, epsilon subunit	201		COG2176	
	Emin_1357	Single-strand binding protein	144		COG0629	
	Emin_1362	DNA repair protein RadC	228		COG2003	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1456	Integrase family protein	337		COG4974	
	Emin_1460	DNA polymerase III, delta prime subunit	320		COG2812	S
	Emin_1468	DNA polymerase III, subunits gamma and tau	627	EC 2.7.7.7	COG2812	
	Emin_1469	Recombinational DNA repair protein	197		COG0353	S
	Emin_1525	RecA protein	353	EC 3.6.3.8	COG0468	
	Emin_1534	Excinuclease ABC, B subunit	665		COG0556	
Q	**Secondary metabolites biosynthesis, transport, and catabolism**					
	Emin_0327	ABC transporter related	258		COG0767	T
	Emin_0328	ABC transporter related	249		COG1127	
	Emin_0329	ABC-type transport system involved in resistance to organic solvents, periplasmic component	430		COG1463	T
	Emin_0659	3-Oxoacyl-(acyl-carrier-protein) reductase	245	EC 1.1.1.100	COG1028	S
	Emin_0660	Acyl carrier protein	81		COG0236	
	Emin_0995	Non-ribosomal peptide synthetase	1284		COG1020	
	Emin_1012	AMP-dependent synthetase and ligase	1133		COG0318	T
	Emin_1101	Polyketide synthase	3008	EC 2.3.1.111	COG3321	
	Emin_1237	Putative fatty acid Co-A ligase	533		COG0318	T
	Emin_1334	Thioesterase family protein	157		COG2050	
T	**Signal transduction mechanisms**					
	Emin_0001	Response regulator receiver protein	118		COG2204	
	Emin_0022	Predicted ATPase related to phosphate starvation-inducible protein	442		COG1875	
	Emin_0051	Response regulator containing a CheY-like receiver domain	407		COG3437	
	Emin_0064	Phosphate starvation-inducible protein family	309		COG1702	
	Emin_0087	Response regulator receiver modulated diguanylate cyclase	308		COG3706	
	Emin_0097	Putative GAF sensor protein	147		COG1956	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0113	Response regulator receiver protein	119		COG2197	
	Emin_0124	Serine kinase of the HPr protein, regulates carbohydrate metabolism	330		COG1493	
	Emin_0174	Putative two-component system	248		COG0642	S
	Emin_0301	Putative methyl-accepting chemotaxis sensory transducer	161		COG0840	S,T
	Emin_0305	Histidine kinase	360		COG4585	
	Emin_0306	Two-component transcriptional regulator, LuxR family	215		COG2197	
	Emin_0455	SOS-response transcriptional repressors (RecA-mediated autopeptidases)	214		COG1974	
	Emin_0510	Putative two-component system	320		COG0642	S,T
	Emin_0573	GTP-binding protein TypA	612		COG1217	
	Emin_0599	Response regulator receiver protein	123		COG0745	
	Emin_0812	Putative PAS/PAC sensor protein	1543		COG2202	
	Emin_0911	Cyclic nucleotide-binding protein	226		COG0664	
	Emin_0927	Protein serine/threonine phosphatase	273		COG0631	
	Emin_0928	Serine/threonine protein kinase	761		COG0515	S,T
	Emin_0987	Transcriptional regulator, XRE family	261		COG1974	
	Emin_1020	Transcriptional repressor, LexA family	201		COG1974	
	Emin_1103	PTS IIA- family protein	155		COG1762	
	Emin_1197	Two-component transcriptional regulator, LuxR family	225		COG2197	
	Emin_1198	Response regulator receiver sensor signal transduction histidine kinase	375		COG0642	
	Emin_1199	Integral membrane sensor signal transduction histidine kinase	604		COG4191	S,T
	Emin_1214	Diguanylate cyclase with GAF sensor	496		COG2199	
	Emin_1267	DnaK suppressor protein	258		COG1734	
	Emin_1288	Putative phosphohistidine phosphatase, SixA	165		COG2062	S
	Emin_1309	Protein kinase	517		COG0515	S,T
	Emin_1337	Putative two-component system	562		COG0642	S,T
	Emin_1507	Guanosine polyphosphate pyrophosphohydrolases/synthetases	482	EC 2.7.6.5 EC 3.1.7.2	COG0317	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
K	**Transcription**					
	Emin_0051	Response regulator containing a CheY-like receiver domain	407		COG3437	
	Emin_0102	Heat-inducible transcription repressor HrcA	363		COG1420	
	Emin_0113	Response regulator receiver protein	119		COG2197	
	Emin_0119	Transcriptional regulator	201		COG1309	
	Emin_0160	Repressor, ORF, kinase protein family	322		COG1940	
	Emin_0162	Transcription-repair coupling factor	1044		COG1197	
	Emin_0181	Sigma-70 family	196		COG1595	
	Emin_0198	ATP-dependent DNA helicase RecG	691		COG1200	
	Emin_0306	Two-component transcriptional regulator, LuxR family	215		COG2197	
	Emin_0312	RNA polymerase, sigma 70 subunit, RpoD family	571		COG0568	
	Emin_0395	Putative transcriptional activator	258		COG1521	
	Emin_0397	Transcription elongation factor	156		COG0782	
	Emin_0421	Aminotransferase class I and II	403		COG1167	
	Emin_0455	SOS-response transcriptional repressors (RecA-mediated autopeptidases)	214		COG1974	
	Emin_0493	Cold-shock DNA-binding domain protein	66		COG1278	
	Emin_0499	Transcription termination factor	497		COG1158	
	Emin_0525	Mn-dependent transcriptional regulator	140		COG1321	
	Emin_0557	Putative transcriptional regulator	360		COG2865	
	Emin_0599	Response regulator receiver protein	123		COG0745	
	Emin_0661	Ribonuclease III	221	EC 3.1.26.3	COG0571	
	Emin_0666	Transcription termination factor	463		COG0195	
	Emin_0757	Transcriptional regulator	307		COG0583	
	Emin_0773	Superfamily II DNA and RNA helicases	502		COG0513	S
	Emin_0782	DNA-directed RNA polymerase, sigma subunit RpoH	380		COG0568	
	Emin_0783	Transcriptional regulator	158		COG1846	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0851	Predicted transcriptional regulators	129		COG0640	
	Emin_0877	AraC-type DNA-binding domain-containing protein	298		COG2207	
	Emin_0887	Predicted transcriptional regulators	112		COG0789	
	Emin_0928	Serine/threonine protein kinase	761		COG0515	S,T
	Emin_0987	Transcriptional regulator, XRE family	261		COG1974	
	Emin_0998	Predicted transcriptional regulator	106		COG1733	
	Emin_1020	Transcriptional repressor, LexA family	201		COG1974	
	Emin_1028	Transcriptional regulator	167		COG1846	
	Emin_1041	Transcriptional regulator, TetR family	208		COG1309	
	Emin_1056	Rrf2 family protein (putative transcriptional regulator)	150		COG1959	S
	Emin_1197	Two-component transcriptional regulator, LuxR family	225		COG2197	
	Emin_1222	Segregation and condensation protein B	282		COG1386	
	Emin_1228	NusB anti-termination factor	140		COG0781	
	Emin_1309	Protein kinase	517		COG0515	S,T
	Emin_1390	DNA-directed RNA polymerase, alpha subunit	329		COG0202	
	Emin_1427	DNA-directed RNA polymerase, beta' subunit	1385		COG0086	
	Emin_1428	DNA-directed RNA polymerase, beta subunit	1270		COG0085	
	Emin_1440	Cold-shock DNA-binding domain protein	67		COG1278	
	Emin_1492	NusG anti-termination factor	178		COG0250	
	Emin_1507	Guanosine polyphosphate pyrophosphohydrolases/synthetases	482	EC 2.7.6.5 EC 3.1.7.2	COG0317	
	Emin_1542	Predicted transcriptional regulators	294		COG1475	
J	**Translation, ribosomal structure, and biogenesis**					
	Emin_0028	Alanyl-tRNA synthetase	868	EC 6.1.1.7	COG0013	
	Emin_0038	Predicted rRNA methylase	258		COG1189	S
	Emin_0054	Peptide chain release factor 2	371		COG1186	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0058	Lysyl-tRNA synthetase	498	EC 6.1.1.6	COG1190	
	Emin_0067	Glutamyl-tRNA(Gln) amidotransferase, C subunit	90		COG0721	
	Emin_0068	Glutamyl-tRNA(Gln) amidotransferase, A subunit	474	EC 3.5.1.4	COG0154	
	Emin_0072	Ribosomal protein L28	64		COG0227	
	Emin_0083	Glutamyl-tRNA synthetase	488	EC 6.1.1.17	COG0008	
	Emin_0088	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	354	EC 2.1.1.61	COG0482	
	Emin_0090	Ribosomal protein L21	104		COG0261	
	Emin_0091	Ribosomal protein L27	114		COG0211	
	Emin_0107	Cytosine/adenosine deaminases	163	EC 3.5.4.3	COG0590	
	Emin_0123	Sigma 54 modulation protein	178		COG1544	
	Emin_0137	Phenylalanyl-tRNA synthetase, alpha subunit	349		COG0016	
	Emin_0138	Phenylalanyl-tRNA synthetase, beta subunit	791	EC 6.1.1.20	COG0072	
	Emin_0184	Glycyl-tRNA synthetase, beta subunit	687	EC 6.1.1.14	COG0751	
	Emin_0185	Glycyl-tRNA synthetase, alpha subunit	285	EC 6.1.1.14	COG0752	
	Emin_0201	Leucyl-tRNA synthetase	835		COG0495	
	Emin_0206	GTP-binding conserved hypothetical protein	357		COG0012	S
	Emin_0253	Valyl-tRNA synthetase	904	EC 6.1.1.9	COG0525	
	Emin_0256	RNA modification enzyme, MiaB family	430	EC 1.3.1.74	COG0621	
	Emin_0272	Putative tRNA/rRNA methyltransferase (SpoU)	148		COG0219	
	Emin_0346	Translation initiation factor IF-3	177		COG0290	
	Emin_0347	Ribosomal protein L35	67		COG0291	
	Emin_0348	Ribosomal protein L20	118		COG0292	
	Emin_0375	Pseudouridine synthase	299	EC 3.2.1.17	COG0564	
	Emin_0384	Aspartyl-tRNA synthetase	583		COG0173	
	Emin_0385	Histidyl-tRNA synthetase	410	EC 6.1.1.21	COG0124	
	Emin_0387	Putative endoribonuclease	126		COG0251	
	Emin_0411	rRNA methylases	249		COG0566	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0507	*S*-Adenosylmethionine:tRNA-ribosyltransferase-isomerase (queuine synthetase)	341		COG0809	
	Emin_0508	Queuine tRNA-ribosyltransferase	388	EC 2.4.2.29	COG0343	
	Emin_0511	Putative translation factor	210		COG0009	
	Emin_0546	Ribosomal protein S1	399		COG0539	
	Emin_0566	Threonyl-tRNA synthetase	583		COG0441	
	Emin_0642	tRNA nucleotidyltransferase/poly(A) polymerase	374		COG0617	
	Emin_0648	Ribosomal protein S2	324		COG0052	
	Emin_0649	Translation elongation factor Ts	278		COG0264	
	Emin_0657	Ribosomal protein L32	68		COG0333	
	Emin_0667	Translation initiation factor 2	826		COG0532	
	Emin_0674	Ribosome-binding factor A	119		COG0858	
	Emin_0676	tRNA pseudouridine synthase B	235		COG0130	
	Emin_0684	Prolyl-tRNA synthetase	573	EC 6.1.1.15	COG0442	
	Emin_0700	Dimethyladenosine transferase	261		COG0030	
	Emin_0706	Ribosome recycling factor	187		COG0233	
	Emin_0733	Translation elongation factor P	186		COG0231	
	Emin_0739	Ribosomal protein L19	141		COG0335	
	Emin_0740	tRNA-(guanine-N1)-methyltransferase	237	EC 2.1.1.31	COG0336	
	Emin_0742	Ribosomal protein S16	82		COG0228	
	Emin_0745	Methionyl-tRNA formyltransferase	333		COG0223	
	Emin_0773	Superfamily II DNA and RNA helicases	502		COG0513	S
	Emin_0799	Asparaginyl-tRNA synthetase	431	EC 6.1.1.22	COG0017	
	Emin_0841	Tyrosyl-tRNA synthetase	387	EC 6.1.1.1	COG0162	
	Emin_0862	GCN5-related N-acetyltransferase	170		COG1670	
	Emin_0897	Arginyl-tRNA synthetase	532	EC 6.1.1.19	COG0018	
	Emin_0906	*N*-Formylmethionyl-tRNA deformylase	176	EC 3.5.1.88	COG0242	
	Emin_0929	Ribosomal protein L31	115		COG0254	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1094	Protein-(glutamine-N5) methyltransferase	277		COG2890	
	Emin_1098	Peptide chain release factor 1	358		COG0216	
	Emin_1113	tRNA delta(2)-isopentenylpyrophosphate transferase	307	EC 2.5.1.8	COG0324	
	Emin_1115	RNA modification enzyme, MiaB family	410	EC 1.3.1.74	COG0621	
	Emin_1144	Isoleucyl-tRNA synthetase	942		COG0060	
	Emin_1163	Cysteinyl-tRNA synthetase	460	EC 6.1.1.16	COG0215	
	Emin_1167	Polyribonucleotide nucleotidyltransferase	698	EC 2.7.7.8	COG1185	
	Emin_1168	Ribosomal protein S15	89		COG0184	
	Emin_1224	Tryptophanyl-tRNA synthetase	373	EC 6.1.1.2	COG0180	
	Emin_1238	Metal-dependent phosphohydrolase	500		COG0617	
	Emin_1265	Ribonuclease P protein component	127		COG0594	
	Emin_1276	Glutamyl-tRNA(Gln) amidotransferase, B subunit	472		COG0064	
	Emin_1290	Aminoacyl-tRNA hydrolase	187	EC 3.1.1.29	COG0193	
	Emin_1291	Ribosomal 5S rRNA E-loop binding protein Ctc/L25/TL5	219		COG1825	
	Emin_1294	RNA modification enzyme, MiaB family	409	EC 1.3.1.74	COG0621	
	Emin_1355	Ribosomal protein L9	152		COG0359	
	Emin_1356	Ribosomal protein S18	113		COG0238	
	Emin_1358	Ribosomal protein S6	104		COG0360	
	Emin_1365	tRNA/rRNA methyltransferase	225		COG0565	S
	Emin_1375	Ribonuclease, Rne/Rng family	483		COG1530	
	Emin_1389	Ribosomal protein L17	114		COG0203	
	Emin_1391	RNA-binding S4 domain protein	201		COG0522	
	Emin_1392	Ribosomal protein S11	142		COG0100	
	Emin_1393	Ribosomal protein S13	124		COG0099	
	Emin_1395	Translation initiation factor IF-1	70		COG0361	
	Emin_1396	Methionine aminopeptidase, type I	249	EC 3.4.11.18	COG0024	
	Emin_1400	Ribosomal protein L15	151		COG0200	
	Emin_1401	Ribosomal protein S5	192		COG0098	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1402	Ribosomal protein L18	121		COG0256	
	Emin_1403	Ribosomal protein L6	184		COG0097	
	Emin_1404	Ribosomal protein S8	130		COG0096	
	Emin_1405	Ribosomal protein S14	61		COG0199	
	Emin_1406	Ribosomal protein L5	192		COG0094	
	Emin_1407	Ribosomal protein L24	101		COG0198	
	Emin_1408	Ribosomal protein L14	122		COG0093	
	Emin_1409	Ribosomal protein S17	89		COG0186	
	Emin_1411	Ribosomal protein L16	136		COG0197	S
	Emin_1412	Ribosomal protein S3	239		COG0092	
	Emin_1413	Ribosomal protein L22	118		COG0091	
	Emin_1414	Ribosomal protein S19	94		COG0185	
	Emin_1415	Ribosomal protein L2	279		COG0090	
	Emin_1416	Ribosomal protein L25/L23	98		COG0089	
	Emin_1417	Ribosomal protein L4/L1e	211		COG0088	
	Emin_1418	Ribosomal protein L3	259		COG0087	
	Emin_1419	Ribosomal protein S10	117		COG0051	
	Emin_1423	Translation elongation factor Tu	395		COG0050	
	Emin_1424	Translation elongation factor G	700		COG0480	
	Emin_1425	Ribosomal protein S7	159		COG0049	
	Emin_1426	Ribosomal protein S12	126		COG0048	
	Emin_1429	Ribosomal protein L7/L12	126		COG0222	
	Emin_1430	Ribosomal protein L10	173		COG0244	
	Emin_1434	Threonyl- and alanyl-tRNA synthetase second additional domain	151		COG0013	
	Emin_1459	Methionyl-tRNA synthetase	500		COG0143	
	Emin_1472	Dihydrouridine synthase DuS	318		COG0042	
	Emin_1489	Ribosomal protein L1	226		COG0081	
	Emin_1490	Ribosomal protein L11	144		COG0080	S

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1498	GTPases - translation elongation factors	395		COG0050	
	Emin_1503	Seryl-tRNA synthetase	424	EC 6.1.1.11	COG0172	
	Emin_1546	Ribosomal protein S9	131		COG0103	
	Emin_1547	Ribosomal protein L13	144		COG0102	
S	**Function unknown**					
	Emin_0047	Glycoside hydrolase family protein	537		COG1543	
	Emin_0056	Uncharacterized conserved protein	234		COG1478	
	Emin_0079	Uncharacterized protein conserved in bacteria DUF1009	273		COG3494	
	Emin_0086	Uncharacterized membrane-associated protein	216		COG0586	T
	Emin_0136	Uncharacterized protein conserved in bacteria DUF520	164		COG1666	
	Emin_0154	Uncharacterized protein conserved in bacteria DUF400	146		COG3018	S
	Emin_0165	LemA family protein	182		COG1704	S,T
	Emin_0170	RDD domain containing protein	145		COG1714	S,T
	Emin_0171	RDD domain containing protein	140		COG1714	T
	Emin_0233	Conserved hypothetical protein DUF34	248		COG0327	
	Emin_0243	CRISPR-associated protein, Csn1 family	1195		COG3513	
	Emin_0245	CRISPR-associated protein Cas2	108		COG3512	S
	Emin_0258	Uncharacterized protein conserved in bacteria DUF1058	157		COG3807	S,T
	Emin_0279	Conserved hypothetical protein DUF152	230		COG1496	
	Emin_0332	Uncharacterized conserved protein	177		COG3247	S,T
	Emin_0378	LemA family protein	185		COG1704	S,T
	Emin_0396	Uncharacterized protein conserved in bacteria DUF195	402		COG1322	S,T
	Emin_0399	RDD domain containing protein	278		COG1714	T
	Emin_0435	CoA-substrate-specific enzyme activase	1406		COG3581	
	Emin_0471	Bacteriophage protein gp37	238		COG4422	
	Emin_0479	Uncharacterized conserved protein related to dihydrodipicolinate reductase	358		COG3804	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0481	Uncharacterized protein conserved in bacteria DUF208	180		COG1636	S
	Emin_0492	Uncharacterized protein conserved in bacteria	226		COG2849	
	Emin_0500	Uncharacterized protein conserved in bacteria	354		COG2849	
	Emin_0550	Putative Trp repressor protein	116		COG4496	
	Emin_0560	Uncharacterized protein conserved in bacteria DUF374	225		COG2121	S,T
	Emin_0597	Predicted integral membrane protein	392		COG0392	T
	Emin_0665	Uncharacterized protein conserved in bacteria	152		COG0779	
	Emin_0697	Phage portal protein, HK97 family	399		COG4695	
	Emin_0709	Uncharacterized conserved protein DUF1015	410		COG4198	
	Emin_0713	Trp operon repressor	90		COG4496	
	Emin_0721	Uncharacterized protein conserved in bacteria DUF885	585		COG4805	S,T
	Emin_0735	Uncharacterized conserved protein	112		COG1430	
	Emin_0744	PASTA domain containing protein	334		COG2815	S,T
	Emin_0785	Predicted membrane protein	199		COG3647	S,T
	Emin_0790	Conserved hypothetical protein DUF167	93		COG1872	
	Emin_0811	Uncharacterized conserved protein DUF1063	423		COG3681	S
	Emin_0819	LemA family protein	176		COG1704	S,T
	Emin_0821	Uncharacterized conserved protein	151		COG2606	
	Emin_0853	Uncharacterized conserved protein	108		COG3339	T
	Emin_0875	Decarboxylase family protein	168		COG0599	S,T
	Emin_0883	Cobalamin (vitamin B12) biosynthesis CbiX protein	220		COG2138	S
	Emin_0896	Domain of unknown function DUF143	204		COG0799	
	Emin_0936	Peptidase M15A	151		COG3108	
	Emin_0957	Phage uncharacterized protein (putative large terminase), C-terminal domain	469		COG5410	
	Emin_1004	Uncharacterized conserved protein	108		COG3937	
	Emin_1049	Uncharacterized membrane protein	293		COG1814	T
	Emin_1111	Conserved hypothetical integral membrane protein DUF205	212		COG0344	S,T

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1145	Antiporter family protein	464		COG1288	S,T
	Emin_1151	Conserved hypothetical protein DUF28	250		COG0217	
	Emin_1223	Chromosome segregation and condensation protein ScpA	261		COG1354	
	Emin_1232	Uncharacterized conserved protein DUF192	168		COG1430	S,T
	Emin_1278	Putative Tol-Pal system protein	229		COG1729	S
	Emin_1295	Conserved hypothetical protein	230		COG1385	
	Emin_1296	Uncharacterized membrane-anchored protein	161		COG4929	S,T
	Emin_1298	Membrane protein-like protein	312		COG4984	T
	Emin_1302	Putative Mg transporter	210		COG1285	S,T
	Emin_1325	Predicted membrane protein	191		COG1971	S,T
	Emin_1326	Predicted membrane protein DUF204	185		COG1971	S,T
	Emin_1335	Adenylate cyclase family protein	192		COG3025	
	Emin_1341	Uncharacterized conserved protein DUF167	71		COG1872	
	Emin_1376	Hypothetical protein	224		COG5011	
	Emin_1445	Conserved hypothetical protein	138		COG0432	
	Emin_1471	Carbohydrate kinase, YjeF related protein	221		COG0062	
	Emin_1537	Arginine decarboxylase, pyruvoyl-dependent	189	EC 4.1.1.19	COG1945	T
R	**General function, prediction only**					
	Emin_0013	FprA subunit	166		COG1853	
	Emin_0027	Hydrolase family protein	184		COG2316	
	Emin_0030	Kinase family protein	288		COG1597	
	Emin_0045	Peptidase U62 modulator of DNA gyrase	556		COG0312	S
	Emin_0046	Zn-dependent protease-like protein	534		COG0312	S,T
	Emin_0060	Protein containing tetrapyrrole methyltransferase domain and MazG-like (predicted pyrophosphatase) domain	127		COG3956	
	Emin_0065	Hydrolase family protein	143		COG0319	S

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0084	Phosphohydrolases family gene	361		COG1408	S,T
	Emin_0092	GTP-binding protein Obg/CgtA	458		COG0536	
	Emin_0109	Phosphoribosyltransferase	245		COG1040	
	Emin_0111	Permease YjgP/YjgQ family protein	364		COG0795	S,T
	Emin_0112	Permease YjgP/YjgQ family protein	370		COG0795	S,T
	Emin_0125	Phosphate binding loop	285		COG1660	
	Emin_0146	Putative threonine dehydrogenase	345		COG1063	
	Emin_0167	NADPH-dependent glutamate synthase	465		COG0493	
	Emin_0172	Predicted permease	347		COG0628	S,T
	Emin_0176	Predicted phosphohydrolases	392		COG1409	S
	Emin_0189	Uncharacterized proteobacterial lipoprotein	196		COG3417	S
	Emin_0199	Oxidoreductase family protein	374		COG0673	S
	Emin_0213	NUDIX hydrolase	178		COG0494	
	Emin_0218	Predicted permeases	250		COG0730	S,T
	Emin_0251	Permeases of the metabolite transporter (DMT) family protein	305		COG0697	S,T
	Emin_0261	Predicted metal-dependent phosphoesterases	270		COG0613	
	Emin_0266	Small GTP-binding protein	377		COG1160	S
	Emin_0269	Conserved hypothetical protein, DprA/Smf-related	270		COG1611	
	Emin_0270	Ankyrin repeat	445		COG0666	S,T
	Emin_0281	ATPase components of ABC transporters with duplicated ATPase domains	469		COG0488	
	Emin_0290	Beta-lactamase domain protein	285		COG0491	S,T
	Emin_0303	Patatin	314		COG1752	S
	Emin_0309	Dehydrogenase family protein	249		COG0300	
	Emin_0311	Conserved domain	160		COG1853	T
	Emin_0315	Histidine triad (HIT) protein	112		COG0537	
	Emin_0339	Pirin domain protein	289		COG1741	
	Emin_0342	Channel protein, hemolysin III family	219		COG1272	T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0352	Hydrolase of alkaline phosphatase superfamily	600		COG3083	S,T
	Emin_0356	Predicted DNA-binding protein with PD1-like DNA-binding motif	142		COG1661	
	Emin_0358	GTP-binding conserved hypothetical protein	300		COG1159	S
	Emin_0366	Predicted phosphatases	212		COG0546	
	Emin_0372	Beta-lactamase family protein	203		COG0491	
	Emin_0393	Putative peptidase	248		COG4783	S,T
	Emin_0413	Metallophosphoesterase	246		COG0622	
	Emin_0431	Predicted phosphatases	284		COG0546	
	Emin_0433	Permeases of the drug/metabolite transporter (DMT) superfamily	303		COG0697	S,T
	Emin_0434	Sel1 domain protein repeat-containing protein	464		COG0790	S
	Emin_0468	Predicted hydrolase	202		COG1011	
	Emin_0472	4Fe-4S ferredoxin iron-sulfur binding domain protein	257		COG2768	S
	Emin_0477	Fe-hydrogenase large subunit family protein	478		COG4624	
	Emin_0491	Predicted flavoproteins	411		COG2081	
	Emin_0512	Uroporphyrin-III C/tetrapyrrole (corrin/porphyrin) methyltransferase	231		COG0313	S
	Emin_0533	FeFe hydrogenase HydA (NuoG)	582		COG4624	
	Emin_0539	Putative Zn-dependent protease	389		COG4783	S
	Emin_0572	ABC transporter protein family	628		COG0488	S
	Emin_0588	Leucine/isoleucine/valine porter	239		COG1137	
	Emin_0598	Putative acetyltransferase	171		COG0663	
	Emin_0604	Putative cation transporter	349		COG4756	S,T
	Emin_0605	Polysaccharide biosynthesis protein involved in cell wall biogenesis	474		COG2244	T
	Emin_0610	Putative transporter	551		COG4146	S,T
	Emin_0615	Hypothetical protein	182		COG1853	
	Emin_0627	Hemolysins and related proteins containing CBS domains	440		COG1253	S,T
	Emin_0638	Large extracellular alpha-helical protein	1872		COG2373	
	Emin_0654	Sel1 domain protein repeat-containing protein	276		COG0790	S
	Emin_0659	3-Oxoacyl-(acyl-carrier-protein) reductase	245	EC 1.1.1.100	COG1028	S

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0662	Ankyrin	358		COG0666	S,T
	Emin_0675	Exopolyphosphatase-related proteins	331		COG0618	
	Emin_0719	Predicted nuclease (RecB family)	495		COG2251	
	Emin_0741	Predicted RNA-binding protein	76		COG1837	S
	Emin_0750	Hydrogenase large subunit domain protein	482		COG4624	
	Emin_0752	Putative PAS/PAC sensor protein	559		COG4624	
	Emin_0756	DNA-binding protein	216		COG2344	
	Emin_0764	Predicted flavoproteins	384		COG2081	
	Emin_0775	Predicted hydrolases or acyltransferases	247		COG0596	S
	Emin_0789	Alanine racemase domain protein	229		COG0325	
	Emin_0793	Ferritin family protein	167		COG2406	
	Emin_0795	Hydrolase family protein	229		COG1011	
	Emin_0798	uncharacterized protein containing SIS (Sugar isomerase) phosphosugar binding domain	246		COG4821	
	Emin_0820	Predicted phosphatase/phosphohexomutase	221		COG0637	
	Emin_0832	Putative intracellular protease/amidase	180		COG0693	
	Emin_0842	Putative aminodeoxychorismate lyase	327		COG1559	S,T
	Emin_0848	Phage terminase, large subunit	397		COG1783	
	Emin_0849	Predicted permeases	316		COG0701	T
	Emin_0858	Homoserine kinase	317	EC 2.7.1.39	COG2334	
	Emin_0876	Hydrolases family protein	347		COG1073	S,T
	Emin_0898	Large extracellular alpha-helical protein	2019		COG2373	
	Emin_0928	Serine/threonine protein kinase	761		COG0515	S,T
	Emin_0952	Mu-like protein prophage major head subunit gpT-like protein	307		COG4397	
	Emin_1005	Predicted unusual protein kinase	564		COG0661	T
	Emin_1013	Putative lactate dehydrogenase	338		COG1052	
	Emin_1017	Predicted oxidoreductase	199		COG3560	
	Emin_1062	Glyoxalase	144		COG3607	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1071	Ankyrin	342		COG0666	S
	Emin_1086	Phage major capsid protein	400		COG4653	
	Emin_1107	Acetolactate synthase small subunit	136		COG4747	
	Emin_1112	GTP-binding conserved hypothetical protein	415		COG2262	
	Emin_1125	Uncharacterized conserved protein DUF58	286		COG1721	
	Emin_1126	ATPase associated with various cellular activities	327		COG0714	
	Emin_1140	GTP cyclohydrolase family protein	132		COG0780	
	Emin_1150	Predicted permeases	275		COG0730	S,T
	Emin_1155	Ribosomal-protein-alanine acetyltransferase	145		COG0456	
	Emin_1157	Conserved hypothetical nucleotide-binding protein	152		COG0802	
	Emin_1159	Radical SAM domain protein	215		COG0535	
	Emin_1170	ExsB protein	484		COG0603	
	Emin_1172	Radical SAM enzyme, Cfr family	343		COG0820	
	Emin_1225	Zn-dependent proteases	226		COG1994	S,T
	Emin_1234	ABC-type dipeptide/oligopeptide/nickel transport system	270		COG1123	
	Emin_1260	Predicted GTPase	449		COG0486	S
	Emin_1262	Single-stranded nucleic acid binding R3H domain protein	470		COG1847	
	Emin_1272	Hydrogenase expression/synthesis HypA family	134		COG0375	
	Emin_1275	Uncharacterized membrane protein, putative virulence factor	495		COG0728	S,T
	Emin_1284	CoA-binding domain protein	124		COG1832	
	Emin_1303	Putative Mg transporter	350		COG4756	S,T
	Emin_1309	Protein kinase	517		COG0515	S,T
	Emin_1333	Biotin and thiamine synthesis associated	458		COG1060	
	Emin_1342	Predicted SAM-dependent methyltransferases	390		COG1092	
	Emin_1370	Predicted metal-binding protein related to the C-terminal domain of SecA	135		COG3318	
	Emin_1432	Uncharacterized stress protein	151		COG3871	
	Emin_1444	Thioesterase superfamily protein	129		COG0824	
	Emin_1462	3-Deoxy-D-manno-octulosonate 8-phosphate phosphatase, YrbI family	198		COG1778	

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1470	Predicted membrane metal-binding protein	609		COG0658	S,T
	Emin_1500	3-Deoxy-D-manno-octulosonate 8-phosphate phosphatase, YrbI family	206		COG1778	
	Emin_1509	Oxidoreductase family protein	355		COG0673	
	Emin_1529	Glucokinase regulatory-like protein	295		COG2103	
	Predicted novel					
	Emin_0004	Predicted novel	98			
	Emin_0007	Predicted novel	428			
	Emin_0008	Predicted novel	250			S
	Emin_0029	Predicted novel	108			S,T
	Emin_0031	Predicted novel	190			T
	Emin_0040	Predicted novel	156			S,T
	Emin_0041	Predicted novel	146			S,T
	Emin_0061	Predicted novel	451			S,T
	Emin_0066	Predicted novel	103			S,T
	Emin_0080	Predicted novel	222			S,T
	Emin_0082	Predicted novel	625			T
	Emin_0089	Predicted novel	219			S
	Emin_0094	Predicted novel	193			S,T
	Emin_0095	Predicted novel	408			
	Emin_0098	Predicted novel	324			S
	Emin_0099	Predicted novel	336			S
	Emin_0104	Predicted novel	187			S
	Emin_0110	Predicted novel	1752			
	Emin_0115	Predicted novel	135			
	Emin_0128	Predicted novel	200			S,T
	Emin_0134	Predicted novel	321			

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0139	Predicted novel	77			
	Emin_0140	Predicted novel	100			
	Emin_0142	Predicted novel	97			
	Emin_0143	Predicted novel	46			
	Emin_0145	Predicted novel	633			S,T
	Emin_0149	Predicted novel	352			S,T
	Emin_0155	Predicted novel	369			S
	Emin_0166	Predicted novel	92			S,T
	Emin_0173	Predicted novel	171			
	Emin_0175	Predicted novel	281			S
	Emin_0179	Predicted novel	231			S,T
	Emin_0182	Predicted novel	127			T
	Emin_0183	Predicted novel	191			S
	Emin_0187	Predicted novel	377			S,T
	Emin_0190	Predicted novel	434			S
	Emin_0191	Predicted novel	306			S
	Emin_0194	Predicted novel	189			S,T
	Emin_0195	Predicted novel	213			S,T
	Emin_0202	Predicted novel	190			S
	Emin_0205	Predicted novel	264			S
	Emin_0207	Predicted novel	357			T
	Emin_0211	Predicted novel	74			
	Emin_0219	Predicted novel	140			S,T
	Emin_0221	Predicted novel	30			
	Emin_0223	Predicted novel	349			
	Emin_0227	Predicted novel	394			
	Emin_0228	Predicted novel	356			
	Emin_0229	Predicted novel	54			

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0235	Predicted novel	151			T
	Emin_0237	Predicted novel	277			S,T
	Emin_0238	Predicted novel	291			S
	Emin_0239	Predicted novel	87			T
	Emin_0240	Predicted novel	245			
	Emin_0241	Predicted novel	209			
	Emin_0242	Predicted novel	59			S
	Emin_0246	Predicted novel	57			
	Emin_0247	Predicted novel	41			T
	Emin_0249	Predicted novel	172			S,T
	Emin_0252	Predicted novel	88			T
	Emin_0254	Predicted novel	376			S,T
	Emin_0255	Predicted novel	100			S,T
	Emin_0257	Predicted novel	155			S,T
	Emin_0259	Predicted novel	149			
	Emin_0263	Predicted novel	145			S
	Emin_0275	Predicted novel	67			S
	Emin_0283	Predicted novel	367			
	Emin_0289	Predicted novel	159			S,T
	Emin_0291	Predicted novel	236			S,T
	Emin_0293	Predicted novel	619			S,T
	Emin_0294	Predicted novel	641			S,T
	Emin_0295	Predicted novel	602			S,T
	Emin_0299	Predicted novel	185			S
	Emin_0300	Predicted novel	231			S
	Emin_0317	Predicted novel	316			S
	Emin_0319	Predicted novel	559			S
	Emin_0320	Predicted novel	296			S

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0322	Predicted novel	251			S
	Emin_0323	Predicted novel	44			
	Emin_0334	Predicted novel	97			
	Emin_0343	Predicted novel	470			S,T
	Emin_0350	Predicted novel	82			
	Emin_0359	Predicted novel	448			S,T
	Emin_0362	Predicted novel	278			
	Emin_0364	Predicted novel	195			S
	Emin_0365	Predicted novel	251			S,T
	Emin_0373	Predicted novel	206			
	Emin_0382	Predicted novel	96			S
	Emin_0406	Predicted novel	145			S
	Emin_0416	Predicted novel	65			
	Emin_0418	Predicted novel	86			S
	Emin_0424	Predicted novel	416			S,T
	Emin_0425	Predicted novel	122			T
	Emin_0428	Predicted novel	319			S,T
	Emin_0429	Predicted novel	324			S
	Emin_0430	Predicted novel	324			S
	Emin_0446	Predicted novel	146			T
	Emin_0450	Predicted novel	158			S,T
	Emin_0451	Predicted novel	303			S
	Emin_0452	Predicted novel	268			S,T
	Emin_0456	Predicted novel	78			
	Emin_0457	Predicted novel	64			S,T
	Emin_0458	Predicted novel	116			
	Emin_0459	Predicted novel	135			
	Emin_0460	Predicted novel	198			S

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0461	Predicted novel	192			S
	Emin_0462	Predicted novel	114			
	Emin_0463	Predicted novel	128			
	Emin_0464	Predicted novel	319			T
	Emin_0465	Predicted novel	186			
	Emin_0466	Predicted novel	93			
	Emin_0467	Predicted novel	290			S
	Emin_0470	Predicted novel	1062			S
	Emin_0473	Predicted novel	37			S
	Emin_0474	Predicted novel	180			
	Emin_0478	Predicted novel	262			S,T
	Emin_0484	Predicted novel	552			S,T
	Emin_0509	Predicted novel	215			S,T
	Emin_0516	Predicted novel	201			S
	Emin_0530	Predicted novel	153			S,T
	Emin_0534	Predicted novel	225			S
	Emin_0537	Predicted novel	49			
	Emin_0538	Predicted novel	233			S,T
	Emin_0540	Predicted novel	282			S,T
	Emin_0547	Predicted novel	402			S
	Emin_0556	Predicted novel	279			S
	Emin_0559	Predicted novel	146			
	Emin_0561	Predicted novel	210			S
	Emin_0569	Predicted novel	95			
	Emin_0570	Predicted novel	132			
	Emin_0574	Predicted novel	549			S,T
	Emin_0576	Predicted novel	229			T
	Emin_0577	Predicted novel	275			S,T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0586	Predicted novel	118			S,T
	Emin_0587	Predicted novel	504			S
	Emin_0589	Predicted novel	256			S
	Emin_0592	Predicted novel	57			S,T
	Emin_0594	Predicted novel	1353			T
	Emin_0601	Predicted novel	74			S,T
	Emin_0614	Predicted novel	76			S,T
	Emin_0619	Predicted novel	241			T
	Emin_0621	Predicted novel	438			S,T
	Emin_0622	Predicted novel	347			T
	Emin_0623	Predicted novel	513			S,T
	Emin_0624	Predicted novel	686			S
	Emin_0632	Predicted novel	110			S,T
	Emin_0637	Predicted novel	261			S,T
	Emin_0639	Predicted novel	2530			
	Emin_0641	Predicted novel	211			
	Emin_0643	Predicted novel	194			T
	Emin_0644	Predicted novel	367			S,T
	Emin_0645	Predicted novel	496			T
	Emin_0646	Predicted novel	82			S,T
	Emin_0647	Predicted novel	166			S,T
	Emin_0650	Predicted novel	219			S
	Emin_0655	Predicted novel	265			S
	Emin_0656	Predicted novel	159			
	Emin_0663	Predicted novel	98			S,T
	Emin_0664	Predicted novel	3965			
	Emin_0670	Predicted novel	101			
	Emin_0671	Predicted novel	188			S

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0672	Predicted novel	295			S
	Emin_0679	Predicted novel	90			S,T
	Emin_0680	Predicted novel	371			S
	Emin_0683	Predicted novel	262			T
	Emin_0685	Predicted novel	174			
	Emin_0688	Predicted novel	78			T
	Emin_0695	Predicted novel	156			
	Emin_0698	Predicted novel	449			S,T
	Emin_0701	Predicted novel	167			S,T
	Emin_0704	Predicted novel	167			S,T
	Emin_0723	Predicted novel	228			S
	Emin_0727	Predicted novel	189			S,T
	Emin_0734	Predicted novel	172			S
	Emin_0736	Predicted novel	242			S
	Emin_0737	Predicted novel	243			S,T
	Emin_0738	Predicted novel	246			S,T
	Emin_0746	Predicted novel	1486			S
	Emin_0747	Predicted novel	84			S
	Emin_0753	Predicted novel	79			
	Emin_0759	Predicted novel	491			S,T
	Emin_0767	Predicted novel	597			S,T
	Emin_0768	Predicted novel	629			S,T
	Emin_0771	Predicted novel	468			S
	Emin_0778	Predicted novel	37			
	Emin_0787	Predicted novel	100			
	Emin_0792	Predicted novel	45			
	Emin_0794	Predicted novel	326			S
	Emin_0796	Predicted novel	332			

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0801	Predicted novel	130			S
	Emin_0803	Predicted novel	263			S
	Emin_0804	Predicted novel	222			S
	Emin_0805	Predicted novel	255			T
	Emin_0808	Predicted novel	380			S,T
	Emin_0813	Predicted novel	52			
	Emin_0814	Predicted novel	782			S,T
	Emin_0815	Predicted novel	176			S,T
	Emin_0816	Predicted novel	234			S
	Emin_0818	Predicted novel	543			S,T
	Emin_0823	Predicted novel	225			
	Emin_0824	Predicted novel	43			
	Emin_0825	Predicted novel	430			S,T
	Emin_0826	Predicted novel	319			S,T
	Emin_0827	Predicted novel	458			S
	Emin_0828	Predicted novel	396			S
	Emin_0831	Predicted novel	213			S
	Emin_0833	Predicted novel	72			S,T
	Emin_0843	Predicted novel	1721			
	Emin_0847	Predicted novel	123			
	Emin_0850	Predicted novel	32			
	Emin_0856	Predicted novel	278			S
	Emin_0857	Predicted novel	157			
	Emin_0868	Predicted novel	57			
	Emin_0869	Predicted novel	147			
	Emin_0870	Predicted novel	209			
	Emin_0871	Predicted novel	200			S
	Emin_0873	Predicted novel	326			S

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0874	Predicted novel	124			
	Emin_0884	Predicted novel	77			S
	Emin_0885	Predicted novel	203			S,T
	Emin_0886	Predicted novel	200			S,T
	Emin_0890	Predicted novel	144			S,T
	Emin_0891	Predicted novel	207			S,T
	Emin_0895	Predicted novel	305			S,T
	Emin_0901	Predicted novel	370			S
	Emin_0908	Predicted novel	209			T
	Emin_0915	Predicted novel	227			S,T
	Emin_0919	Predicted novel	189			S
	Emin_0923	Predicted novel	68			S
	Emin_0930	Predicted novel	402			S
	Emin_0931	Predicted novel	402			S
	Emin_0935	Predicted novel	118			S,T
	Emin_0937	Predicted novel	133			S,T
	Emin_0938	Predicted novel	81			S
	Emin_0939	Predicted novel	184			
	Emin_0940	Predicted novel	383			
	Emin_0941	Predicted novel	747			
	Emin_0942	Predicted novel	264			S
	Emin_0944	Predicted novel	35			S
	Emin_0945	Predicted novel	160			
	Emin_0946	Predicted novel	286			
	Emin_0947	Predicted novel	97			S
	Emin_0948	Predicted novel	155			
	Emin_0949	Predicted novel	171			
	Emin_0950	Predicted novel	136			

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0951	Predicted novel	138			
	Emin_0953	Predicted novel	126			
	Emin_0955	Predicted novel	56			
	Emin_0958	Predicted novel	151			
	Emin_0960	Predicted novel	44			
	Emin_0961	Predicted novel	125			
	Emin_0963	Predicted novel	77			S
	Emin_0964	Predicted novel	48			
	Emin_0965	Predicted novel	62			
	Emin_0966	Predicted novel	126			
	Emin_0968	Predicted novel	540			
	Emin_0970	Predicted novel	84			
	Emin_0971	Predicted novel	40			
	Emin_0972	Predicted novel	58			
	Emin_0973	Predicted novel	83			
	Emin_0975	Predicted novel	229			
	Emin_0978	Predicted novel	288			
	Emin_0979	Predicted novel	118			
	Emin_0980	Predicted novel	250			
	Emin_0981	Predicted novel	419			
	Emin_0982	Predicted novel	61			
	Emin_0983	Predicted novel	74			
	Emin_0984	Predicted novel	59			S
	Emin_0985	Predicted novel	83			
	Emin_0986	Predicted novel	79			
	Emin_0988	Predicted novel	288			S
	Emin_0990	Predicted novel	198			S
	Emin_0992	Predicted novel	414			S

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_0993	Predicted novel	298			S,T
	Emin_0999	Predicted novel	72			S,T
	Emin_1001	Predicted novel	130			S
	Emin_1003	Predicted novel	155			S,T
	Emin_1006	Predicted novel	356			S
	Emin_1008	Predicted novel	96			S
	Emin_1021	Predicted novel	420			S,T
	Emin_1022	Predicted novel	120			S
	Emin_1023	Predicted novel	87			S,T
	Emin_1024	Predicted novel	293			T
	Emin_1034	Predicted novel	128			
	Emin_1038	Predicted novel	275			S,T
	Emin_1039	Predicted novel	147			T
	Emin_1053	Predicted novel	49			
	Emin_1060	Predicted novel	236			
	Emin_1061	Predicted novel	55			T
	Emin_1063	Predicted novel	40			S
	Emin_1065	Predicted novel	284			
	Emin_1068	Predicted novel	411			
	Emin_1074	Predicted novel	93			
	Emin_1076	Predicted novel	84			T
	Emin_1077	Predicted novel	790			
	Emin_1078	Predicted novel	217			
	Emin_1079	Predicted novel	232			
	Emin_1080	Predicted novel	120			
	Emin_1081	Predicted novel	168			
	Emin_1082	Predicted novel	131			
	Emin_1083	Predicted novel	50			

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1085	Predicted novel	182			
	Emin_1089	Predicted novel	139			S,T
	Emin_1102	Predicted novel	310			S,T
	Emin_1106	Predicted novel	528			S,T
	Emin_1108	Predicted novel	123			
	Emin_1120	Predicted novel	601			S
	Emin_1124	Predicted novel	306			S,T
	Emin_1133	Predicted novel	112			S,T
	Emin_1139	Predicted novel	262			S,T
	Emin_1148	Predicted novel	172			S,T
	Emin_1149	Predicted novel	85			S,T
	Emin_1164	Predicted novel	169			S
	Emin_1169	Predicted novel	179			S
	Emin_1171	Predicted novel	145			S,T
	Emin_1186	Predicted novel	370			
	Emin_1187	Predicted novel	258			S,T
	Emin_1189	Predicted novel	62			
	Emin_1191	Predicted novel	84			S,T
	Emin_1193	Predicted novel	251			S,T
	Emin_1195	Predicted novel	369			T
	Emin_1196	Predicted novel	59			T
	Emin_1200	Predicted novel	320			S
	Emin_1201	Predicted novel	89			S,T
	Emin_1210	Predicted novel	65			S,T
	Emin_1211	Predicted novel	59			S,T
	Emin_1213	Predicted novel	1397			S,T
	Emin_1219	Predicted novel	267			S,T
	Emin_1226	Predicted novel	62			S

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1239	Predicted novel	207			T
	Emin_1241	Predicted novel	530			
	Emin_1242	Predicted novel	715			T
	Emin_1243	Predicted novel	93			
	Emin_1244	Predicted novel	155			S,T
	Emin_1245	Predicted novel	113			S,T
	Emin_1257	Predicted novel	365			S,T
	Emin_1264	Predicted novel	77			S
	Emin_1270	Predicted novel	240			S,T
	Emin_1274	Predicted novel	192			S
	Emin_1285	Predicted novel	97			
	Emin_1297	Predicted novel	352			T
	Emin_1308	Predicted novel	227			T
	Emin_1312	Predicted novel	251			S
	Emin_1316	Predicted novel	173			S
	Emin_1324	Predicted novel	457			S,T
	Emin_1330	Predicted novel	204			S,T
	Emin_1332	Predicted novel	326			S,T
	Emin_1336	Predicted novel	69			
	Emin_1338	Predicted novel	242			S,T
	Emin_1345	Predicted novel	231			S,T
	Emin_1361	Predicted novel	350			S
	Emin_1364	Predicted novel	99			T
	Emin_1369	Predicted novel	190			
	Emin_1372	Predicted novel	342			S
	Emin_1373	Predicted novel	52			
	Emin_1386	Predicted novel	360			S
	Emin_1397	Predicted novel	135			S,T

Supplementary material

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1420	Predicted novel	870			S
	Emin_1433	Predicted novel	201			S,T
	Emin_1435	Predicted novel	131			T
	Emin_1437	Predicted novel	97			S,T
	Emin_1438	Predicted novel	87			
	Emin_1446	Predicted novel	247			S,T
	Emin_1447	Predicted novel	139			
	Emin_1448	Predicted novel	384			
	Emin_1449	Predicted novel	129			T
	Emin_1450	Predicted novel	302			S
	Emin_1452	Predicted novel	359			
	Emin_1454	Predicted novel	295			
	Emin_1455	Predicted novel	208			S
	Emin_1458	Predicted novel	58			T
	Emin_1465	Predicted novel	618			S
	Emin_1466	Predicted novel	280			S
	Emin_1467	Predicted novel	148			S,T
	Emin_1480	Predicted novel	286			S
	Emin_1483	Predicted novel	151			
	Emin_1491	Predicted novel	203			S
	Emin_1499	Predicted novel	203			S
	Emin_1501	Predicted novel	228			S
	Emin_1502	Predicted novel	77			S,T
	Emin_1504	Predicted novel	393			S
	Emin_1506	Predicted novel	125			T
	Emin_1533	Predicted novel	308			S
	Emin_1539	Predicted novel	211			S
	Emin_1540	Predicted novel	209			S

COG group	Locus tag	Product	Length (aa)	Enzymes	COGs	Signal peptides, transmembrane helices
	Emin_1541	Predicted novel	42			
	Emin_1544	Predicted novel	74			S
	Emin_1549	Predicted novel	222			S,T
	Pseudogenes					
	Emin_0048					
	Emin_0081					
	Emin_0144					
	Emin_0164					
	Emin_0200					
	Emin_0208					
	Emin_0271					
	Emin_0292					
	Emin_0316					
	Emin_0357					
	Emin_0363					
	Emin_0480					
	Emin_0515					
	Emin_0913					
	Emin_1064					
	Emin_1166					
	Emin_1176					
	Emin_1384					
	Emin_1514					
	Emin_1545					

Herlemann, D. P. R., O. Geissinger, W. Ikeda-Ohtsubo, V. Kunin, H. Sun, A. Lapidus, P. Hugenholtz and A. Brune. 2009. Genome analysis of *Elusimicrobium minutum*, the first cultivated representative of the *Elusimicrobia* phylum (formerly Termite Group 1).

Supplementary material

Table S2. List of insertion sequences found in the genome of *Elusimicrobium minutum* by IS finder using BLASTX.

Sequences producing significant alignments	IS Family	IS Group	Origin	Score	e-value
ISAzo26	IS91		*Azoarcus* sp.	107	2×10^{-21}
ISSde12	IS91		*Shewanella denitrificans*	92	6 x 10-^{17}
ISBce8	IS4	IS231	*Bacillus cereus* ATCC 10987	78	1×10^{-12}
ISSod25	IS91		*Shewanella oneidensis*	74	2×10^{-11}
ISPpu12	ISL3		*Pseudomonas putida*	74	2×10^{-11}
ISCARN110	IS91		Metagenomic data from Carnoules	73	4×10^{-11}
ISCgl1	IS481		*Corynebacterium glutamicum* (pXZ10145.1)	55	1×10^{-05}
IS5564	IS481		*Corynebacterium striatum* (pTP10) M82B	55	1×10^{-05}
ISBce16	IS21		*Bacillus cereus*	51	2×10^{-04}
IS231K	IS4	IS231	*Bacillus cereus*	49	6×10^{-04}
IS643	IS21		*Bacillus halodurans* C-125	49	7×10^{-04}
ISCARN89	IS21		Metagenomic data from Carnoules	46	0.005
ISMac3	IS21		Methanosarcina acetivorans	45	0.011
ISAzo17	IS21		*Azoarcus* sp.	44	0.024
IS21	IS21		*Pseudomonas aeruginosa* PAO25 (pR68.45)	44	0.031
IS100kyp	IS21		*Yersinia pseudotuberculosis* (pKYP1)	43	0.053
ISThsp10	IS21		*Thiomonas* sp.	43	0.053
ISThsp5	IS21		*Thiomonas* sp.	42	0.09
IS1612	IS21		*Mycobacterium avium* subsp. *silvaticum*	41	0.15
IS1326	IS21		*Pseudomonas aeruginosa* (In0 from pVS1)	41	0.15
ISChy4	IS21		*Carboxydothermus hydrogenoformans*	41	0.2
ISMac9	IS21		*Methanosarcina acetivorans*	41	0.2
IS5377	IS4	IS4Sa	*Bacillus stearothermophilus* CU21	41	0.2
IS232A	IS21		*Bacillus thuringiensis* subsp. *thuringiensis* berliner 1715 (p65kb)	40	0.26
ISAfe9	IS21		*Acidithiobacillus ferrooxidans*	40	0.34

Supplementary material

Sequences producing significant alignments	IS Family	IS Group	Origin	Score	e-value
ISAli13	IS21		*Azospirillum lipoferum*	40	0.45
ISBlo1	IS21		*Bifidobacterium longum* NCC2705	40	0.45
ISC1926	IS607		*Sulfolobus* sp.	40	0.45
ISGdi17	IS21		*Gluconacetobacter diazotrophicus*	39	0.58
ISGme4	IS21		*Geobacter metallireducens*	39	0.58
ISCARN3	IS21		Metagenomic data from Carnoules	39	1
ISC1913	IS607		*Sulfolobus solfataricus*	39	1
IS1600	IS21		*Ralstonia eutropha* pENH91	39	1
ISCARN49	IS21		Metagenomic data from Carnoules	38	1.3
ISNpu6	IS701		*Nostoc punctiforme*	38	1.3
IS1221A	IS3	IS150	*Mycoplasma hyorhinis* GDL-1	38	1.3
ISSpu5	IS21		*Shewanella putrefaciens*	38	1.7
ISRso19	IS21		*Ralstonia solanacearum*	37	2.2
ISBf1	IS21		*Bacteroides fragilis* RBF103	37	2.2
ISSau9	IS21		*Staphylococcus aureus*	37	2.2
ISGur8	IS21		*Geobacter uraniumreducens*	37	2.9
ISAli20	Tn3		*Azospirillum lipoferum*	37	3.8
ISSba11	IS21		*Shewanella baltica*	37	3.8
ISRme4	IS21		*Cupriavidus metallidurans*	37	3.8
ISLxx1	IS21		*Leifsonia xyli* subsp *xyli*	37	3.8
ISLhe9	IS607		*Lactobacillus helveticus*	36	4.9
IS5376	IS21		*Bacillus stearothermophilus* CU21	36	4.9
ISSpu15	IS21		*Shewanella putrefaciens*	36	6.5
ISPst3	IS21		*Pseudomonas stutzeri*	36	6.5
ISKpn7	IS21		*Klebsiella pneumoniae*	36	6.5
ISFsp3	IS21		*Frankia* sp. Ar15	36	6.5
ISNsp3	IS200/IS605	IS608	*Nostoc* sp. PCC 7120	36	6.5
ISNsp2	IS200/IS605	IS608	*Nostoc* sp. PCC 7120	36	6.5
ISRso6	IS21		*Ralstonia solanacearum*	36	6.5
IS100	IS21		*Yersinia pestis* 106 Otten	36	6.5

Table S2-2

Supplementary material

Sequences producing significant alignments	IS Family	IS Group	Origin	Score	e-value
ISRso19	IS21		*Ralstonia solanacearum*	36	6.5
IS1165	ISL3		*Leuconostoc mesenteroides* subsp. *cremoris* DB1165 (p30kb)	36	6.5
ISAba8	IS21		*Acinetobacter baumannii*	35	8.4
ISBcen13	IS21		*Burkholderia cenocepacia* strain J2315	35	8.4
ISFlsp1	IS21		*Flavobacterium* sp.	35	8.4