"*Candidatus* Cloacamonas Acidaminovorans": Genome Sequence Reconstruction Provides a First Glimpse of a New Bacterial Division[∇][†]

Eric Pelletier,[1,2][¶] Annett Kreimeyer,[1,2][¶] Stéphanie Bocs,[1,2][‡] Zoé Rouy,[1,2] Gábor Gyapay,[1,2]
Rakia Chouari,[1,2,3][§] Delphine Rivière,[1,2,3,4] Akila Ganesan,[1,2] Patrick Daegelen,[1,2]
Abdelghani Sghir,[1,2,3] Georges N. Cohen,[5] Claudine Médigue,[1,2]
Jean Weissenbach,[1,2,3] and Denis Le Paslier[1,2]*

*CEA/Genoscope,[1] CNRS-UMR8030,[2] and Université d'Evry Val d'Essonne,[3] 91057 Evry, France; Suez-Environnement CIRSEE,
78230 Le Pecq, France[4]; and Institut Pasteur, 75015 Paris, France[5]*

Many microorganisms live in anaerobic environments. Most of these microorganisms have not yet been cultivated. Here, we present, from a metagenomic analysis of an anaerobic digester of a municipal wastewater treatment plant, a reconstruction of the complete genome of a bacterium belonging to the WWE1 candidate division. In silico proteome analysis indicated that this bacterium might derive most of its carbon and energy from the fermentation of amino acids, and hence, it was provisionally classified as "*Candidatus* Cloacamonas acidaminovorans." "*Candidatus* Cloacamonas acidaminovorans" is probably a syntrophic bacterium that is present in many anaerobic digesters. This report highlights how environmental sequence data might provide genomic and functional information about a new bacterial clade whose members are involved in anaerobic digestion.

The use of molecular techniques over the past few decades has shown the extent of microbial diversity and that the majority of archaeal and bacterial phyla still lack a cultivable representative (18, 24). The recent advent of whole-community genome sequencing, or metagenomics, is rapidly changing this view and will revolutionize our understanding of the functional diversity of complex environments. For instance, a recent study has surprisingly revealed that ammonia-oxidizing *Crenarchaeota* are very abundant in soils (20). However, apart from the pioneering work in the Sargasso Sea (31) and the recent addition of millions of sequences from the Global Ocean Sampling expedition that revealed the extent of the ocean microbial diversity (25), most of the community sequencing programs have focused on relatively simple ecosystems. However, most microbial ecosystems are complex. Anaerobic digestion is a complex biological process that involves several metabolic pathways for the decomposition of organic matter into methane and carbon dioxide. The overall reactions—depolymerization, primary and secondary fermentation, acidogenesis, acetogenesis, and methanogenesis—are performed by a complex microbial community. Despite the industrial, technological, economic, and ecological importance of this com-

munity, little is known about the roles and activities of the microorganisms that inhabit anaerobic niches. The anaerobic digestion of organic matter involved in wastewater processing represents a good example of a complex and active microflora. During exploration of the bacterial diversity of an anaerobic mesophilic digester, a new bacterial candidate division called WWE1 was discovered (8). It was found that WWE1 bacteria could represent up to 10% of the bacterial microflora and thus could be a subdominant group. Using metagenomic sequence data and a specific genome assembly procedure, we were able to reconstruct the genome of a representative bacterium of the WWE1 division. We have for the first time obtained the complete genome sequence from a complex environment and from a bacterial candidate division with no cultivated representative. Because the metabolic pathways of anaerobic bacteria are less well known than those of aerobes, we focused our analysis on the presence of enzymes typical of an anoxic lifestyle and on the metabolic capabilities of the bacterium. This analysis revealed that the bacterium uses principally fermentation processes as its carbon and energy sources. It was provisionally classified as "*Candidatus* Cloacamonas acidaminovorans." Finally, the gene content suggests that "*Candidatus* Cloacamonas acidaminovorans" could be a hydrogen-producing syntroph. The availability of the genomic sequence, as well as genomic clones, of uncultivated bacteria is of the utmost importance for further functional studies and analysis of their metabolic potentials.

* Corresponding author. Mailing address: CEA/Genoscope, CNRS-UMR8030, 2 rue Gaston Crémieux, 91057 Evry, France. Phone: 33 1 60 87 25 98. Fax: 33 1 60 87 25 14. E-mail: denis@genoscope.cns.fr.
† Supplemental material for this article may be found at http://jb.asm.org/.
‡ Present address: UMR DAP-CIRAD, 34398 Montpellier Cedex 5, France.
§ Present address: Faculté des Sciences de Bizerte, 7021 Zarzouna, Bizerte, Tunisia.
¶ E.P. and A.K. contributed equally to this work.
∇ Published ahead of print on 1 February 2008.

MATERIALS AND METHODS

Metagenomic library. The sludge sample was obtained from an anaerobic mesophilic digester (7,000 m³) at the Evry wastewater treatment plant (250,000 population equivalents), located about 35 km south of Paris, France (48°37′33.65″N, 2°27′53.34″E). The digester temperature was 33°C, the pH was 7.2, and

the retention time was 39 ± 10 days. Sampling was performed on 18 March 2002. Total genomic DNA was extracted as described previously (28) and then sheared by repeated pipetting until the majority of DNA fragments were 50 kb in mean size. These DNA fragments were gel purified using pulsed-field gel electrophoresis and cloned using the EpiFOS and the CopyControl fosmid library production kits (Epicenter, Madison, WI) following the manufacturer's protocols. A total of 45 DNA ligations and 119 packagings were performed (Gigapack; Stratagene Europe, Amsterdam, The Netherlands). The resulting library contained 1 million fosmids. Terminal fosmid end sequences (FES) were systematically determined using standard methods and filtered for low-quality bases. Each FES was trimmed from both ends until the first base had exhibited a Phred quality greater than 15, and only sequences with a remaining length of at least 300 bp were retained for subsequent analysis. A total of 1,741,439 reads (1.12 Gb) were obtained, with a mean size of 633 bp, and 82% of the fosmid clones had both ends sequenced.

Small-subunit rRNA screening and sequencing. Fosmid DNA was extracted from 27,648 clones (384 × 72) and spotted in duplicate onto 20- by 20-cm nylon membranes (Hybond N+; GE Healthcare Europe GmbH, Saclay, France). The membranes were successively hybridized with ^{32}P-labeled complex 16S rRNA gene probe mixtures. The probes were either the Eub338 (I, II, and III) oligonucleotide probes (9) or mixtures of 16S rRNA gene PCR products representative of the different archaeal and bacterial lineages described in the wastewater treatment plant of Evry (6, 7, 8). Positive clones were picked, and their 16S rRNA genes were sequenced using internal primers SSM-8F (5′ AGAGTTTG ATCCTGGCTCAG 3′), TTM-330F (5′ ACTCCTACGGGAGGCAGC 3′), ADM-1110R (5′ GCAACGAGCGCAACCC 3′), and ACM-1517R (5′ GGGC CTTGTACACACCG 3′).

"*Candidatus* Cloacamonas acidaminovorans" genome reconstruction. An iterative assembly process was set up to allow genomic extension. The first iteration was initiated with a given seed sequence as the query and the complete FES resource, while the subsequent iterations used a selection of contigs generated by the previous iteration as query sequences At each iteration, FES were assigned on the query sequences if they shared more than 99% nucleic acid identity over at least 90% of the length of the read. Assigned FES constituted the pool of sequences to be assembled for this iteration. Existing paired FES of the assigned FES were then added to the pool of sequences to be assembled, which was subjected to standard assembly (Phrap 0.990319 [http://www.phrap.org/]). After assembly, contigs larger than 1,500 bp and constituted of at least eight FES were selected as query sequences for the next cycle. Iterations were stopped when no new sequences were added to the pool of FES to be assembled compared with the pool from the previous iteration. An overview of this iterative assembly process is illustrated in Fig. 1.

WWE1 and "*Candidatus* Cloacamonas acidaminovorans" PCR detection and quantification. 16S rRNA sequences for WWE1 were obtained after PCR amplification with Pla46F (5′ GGATTAGGCATGCAAGTC 3′) and 1181R (5′ CTTCCTCTGCGTTGTTAC 3′), as described previously (8). "*Candidatus* Cloacamonas acidaminovorans" 16S rRNA gene was amplified with 18F (5′ AGT GTTAATGGTTGAGAG 3′) or 25F (5′ GCTATTCACGGAAGTGAGTAG TGTT 3′) and 1181R. As these two primer pairs could also amplify "*Candidatus* Cloacamonas acidaminovorans"-like bacteria, a set of "*Candidatus* Cloacamonas acidaminovorans"-specific primers was developed. Oligonucleotides were chosen in intergenic regions that were larger than 800 bp or in intergenic regions larger than 200 bp flanking a gene ranging in size from 300 to 800 bp. A total of 180 "*Candidatus* Cloacamonas acidaminovorans"-specific regions were found. A set of 40 regions was used for further analysis, and finally, 30 specific PCR primer pairs, well spaced on the "*Candidatus* Cloacamonas acidaminovorans" genome, were used to confirm the presence of the bacterium in DNA extracted from digester sludge and enrichment cultures (see Table S1 in the supplemental material for a complete list and sequences of oligonucleotides). After testing several "primer couple-probe" systems, the following were retained for quantitative PCR experiments: Cloam1554-Cloam_23s_1, forward (5′ ATACTCTGCGA AATCTGCGTAAAC 3′) and reverse (5′ GATGGTACTGCACGGGCAAC 3′), and a Taqman probe (5′ AGTAGGTCGCCGCCATCGCTAACG 3′).

Genome annotation. Gene prediction was conducted using AMIGene software (4). A total of 1,820 coding sequences (CDSs) were predicted (and assigned a unique identifier prefixed with "Cloam") and subjected to automatic functional annotation (2, 30). Sequence data for comparative analyses were obtained from the NCBI data bank (RefSeq and WGS [whole-genome shotgun] sections). Putative orthologs and groups of neighbor orthologs were computed between "*Candidatus* Cloacamonas acidaminovorans" and all other complete genomes as previously described (30). All the data were stored in a relational database called CloacamonaScope, which is publicly available (http://www.genoscope.cns.fr/agc/mage).
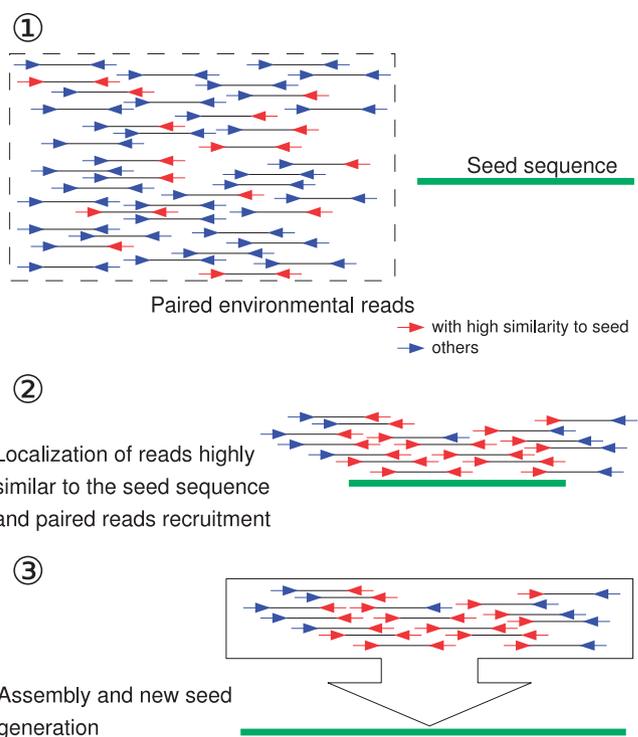

FIG. 1. Iterative assembly process schema. Each iteration of the assembly process consists of the following steps. (Step 1) Nucleotide similarities between environmental reads and seed sequences are computed, and highly similar reads are selected. (Step 2) Paired reads are recruited, along with the highly similar reads. (Step 3) Both similar and paired reads are subjected to the assembly process. (Step 4) The resulting contigs are screened both for length and coverage, and those selected are used as seed sequences for the next iteration.

Enrichment of "*Candidatus* Cloacamonas acidaminovorans." The medium used for the enrichment of "*Candidatus* Cloacamonas acidaminovorans" was a synthetic medium in which lysine (25×10^{-3} M) served as the carbon source. The other medium components were $Na_2HPO_4 \cdot 12H_2O$ (3.1×10^{-2} M), KH_2PO_4 (2.5×10^{-2} M), $MgSO_4 \cdot 7H_2O$ (8×10^{-4} M), $FeSO_4 \cdot 7H_2O$ (1.8×10^{-6} M), $CaCl_2 \cdot 2H_2O$ (4.1×10^{-4} M); the amino acids leucine, isoleucine, valine, threonine, methionine, proline, arginine, histidine, phenylalanine, cysteine (each at a concentration of 10^{-4} M), and tryptophan (2×10^{-5} M); and the cofactors DL-pantothenate (2×10^{-6} M), biotin (10^{-8} M), *p*-aminobenzoate, pyridoxine hydrochloride, thiamine hydrochloride, lipoic acid, hematin, spermidine (the last six at a concentration of 10^{-6} M), coenzyme B12 (10^{-7} M), *S*-adenosylmethionine (10^{-7} M), resazurin (0.001% [vol/vol]), and dithiothionate (10^{-4} M). The amino acids and the cofactors were freshly prepared, filter sterilized, and added to the medium after being autoclaved. The pH was adjusted to 7.0 with KOH (1 M). The headspace was flushed with nitrogen (2 atm). The enrichments were maintained in 100-ml serum vials containing 50 ml of the medium and inoculated with the digester sludge (1% [vol/vol]). The vials were incubated at 33°C in the dark without agitation. Growth of "*Candidatus* Cloacamonas acidaminovorans" was followed by optical density measurements and PCR amplification with 16S rRNA gene-specific primers and the set of "*Candidatus* Cloacamonas acidaminovorans"-specific oligonucleotides (see Table S1 in the supplemental material).

Nucleotide sequence accession number. The "*Candidatus* Cloacamonas acidaminovorans" nucleotide sequence and annotation data have been deposited in the EMBL data bank under accession number CU466930.

RESULTS

Reconstruction of the complete genome of a WWE1 bacterium, "*Candidatus* Cloacamonas acidaminovorans." WWE1 is a recently characterized candidate division deeply branching

from the *Spirochaetes*. Dot blot hybridization and fluorescence in situ hybridization analysis have shown that WWE1 bacteria are one of the subdominant groups in several anaerobic digesters (8). As part of a metagenomic project, we investigated the presence and diversity of the WWE1 bacteria. Sludge samples were obtained from a full-scale anaerobic mesophilic digester located in the municipal wastewater treatment plant in Evry, France. Hybridization screening of a subset of a fosmid library constructed from this sludge revealed that 10% of the 16S rRNA gene-containing fosmids (64/599 clones) were affiliated with the WWE1 division. The 16S rRNA genes of these fosmids were classified into four phylotypes (based on a 97% identity threshold). In order to evaluate their genomic diversity, we fully shotgun sequenced 33 of the fosmids. In addition, we systematically determined the FES from the library and identified overlapping fosmids under the above-mentioned stringent matching criteria. The sequence of a phylotype 1 fully shotgun-sequenced fosmid (clone TCI) was found densely covered with FES (10- to 12-fold in terms of sequence coverage) and prompted us to reconstruct the complete tentative genome sequence of a representative bacterium of division WWE1.

An iterative assembly process was set up (see Materials and Methods) and applied starting from the phylotype 1 TCI fosmid insert sequence. After 31 iterations, the process stopped gathering new FES, and a scaffold was obtained. Completion and circularization of the genome sequence of phylotype 1 ("*Candidatus* Cloacamonas acidaminovorans") was achieved by standard finishing techniques using fosmid clones as a matrix for PCR amplification and/or direct sequencing.

When applied to fosmids not affiliated with phylotype 1, the iterative assembly process failed to significantly extend the seed contig. Since "*Candidatus* Cloacamonas acidaminovorans" was not obtained in pure culture, it is difficult to check whether the reconstructed genome is a composite one. However, its reconstruction was probably made possible because of the extremely low sequence polymorphism (1.7×10^{-4}) of the "*Candidatus* Cloacamonas acidaminovorans" dominant phylotype within the WWE1 population, as computed from the occurrence frequency of high-quality individual-read bases of "*Candidatus* Cloacamonas acidaminovorans" assigned FES differing from the consensus genome sequence.

Genome assembly robustness assessment. Genome assembly robustness was assessed using fosmid coverage coherence (the relative orientation and spacing of the two FES of each clone localized). From the 1,741,439 available FES, 38,709 reads (2.2%, representing a cumulative 24,998,287 bp, that is, a mean genome coverage of 10.9) coming from 21,621 distinct fosmid clones were localized on the "*Candidatus* Cloacamonas acidaminovorans" genome using the above-mentioned criteria. Among those clones, 17,088 had their two FES localized. Only a few fosmids seemed to have incoherent features in regard to the "*Candidatus* Cloacamonas acidaminovorans" genome (172 of 199 rejected clones had one of the extremities localized more than one time [repeated regions], 21 had their end sequences in an incorrect orientation, and 6 were inconsistent in size). The 16,889 correctly localized fosmids allowed us to compute a clone coverage (with a mean value of 302.6 covering clones per base) showing no gap, with a minimum value of 4 over a 507-bp region. The 27 clones displaying incoherent pair end localization on the genome (either incorrect orientation or

an insert size out of the normal fosmid range) were neither colocalized nor overlapping and matched in regions otherwise covered by several coherently positioned inserts. On the other hand, the 172 clones with multiply positioned ends were clustered, with all the ambiguous ends corresponding to a repeated region, and therefore were validated during the finishing of the genome. As, after reconstruction, the genome sequence of "*Candidatus* Cloacamonas acidaminovorans" was subjected to standard finishing, the impact of these reads on the structure of the sequence was carefully examined and ruled out.

Two percent of the (complete) FES library was assigned to the "*Candidatus* Cloacamonas acidaminovorans" genome, suggesting that the prevalence of the genome in this environment is quite low. It should, however, be noted that FES prevalence as defined gives only an approximation, as the number is dependent on (i) the complete genome size and (ii) the possible cloning biases.

"*Candidatus* Cloacamonas acidaminovorans" genome properties. The genome of "*Candidatus* Cloacamonas acidaminovorans" is represented by a single circular chromosome consisting of 2,246,820 nucleotides with a GC content of 37.9% (Fig. 2). We do not know if "*Candidatus* Cloacamonas acidaminovorans" harbors any plasmid. Two rRNA operons (5S, 16S, and 23S) were identified, together with two small noncoding RNA genes (*ssrS*, encoding a 6S RNA, and *rnpB*, encoding the RNA moiety of RNase P). A total of 47 tRNA genes representing all amino acids were annotated. Although only four insertion sequences were found, the fraction of repeated sequences was quite high (8.4%). One of these corresponded to a prophagic region (34.7 kb in length; Cloam0195-0237 and Cloam0474-0509) and displayed a strong GC percent deviation (Fig. 2) and an atypical codon usage compared to the rest of the genome (see Fig. S1 in the supplemental material). Moreover, two large clustered regularly interspaced short palindromic repeats (CRISPR) (12.4 and 6 kb) (17) have been identified. CRISPR, together with associated *cas* genes, have been found in the majority of bacterial genera and are ubiquitous in *Archaea*. They provide resistance against phages (3, 5, 21). These regions are free of coding sequences and could be partly responsible for the atypical "*Candidatus* Cloacamonas acidaminovorans" GC skew (not shown). "*Candidatus* Cloacamonas acidaminovorans" has a rather low density of protein coding sequences (81%). Among the 1,820 predicted genes, 926 (51%) were assigned a probable biological function. Some 163 (9%) were similar to proteins with unknown functions, and 731 (40%) were unique to "*Candidatus* Cloacamonas acidaminovorans" (Table 1). The percentage of genes found in groups of colocalized orthologs shared with other bacteria was generally very small (the maximum value was 15.6%, obtained with *Pelobacter carbinolicus*) (see Table S2 in the supplemental material), showing that the "*Candidatus* Cloacamonas acidaminovorans" genome is remote from known microbial genomes. The highest proportion of orthologs was found with anaerobic bacteria, such as *Kuenenia stuttgartiensis*, *Geobacter metallireducens*, and *Thermoanaerobacter tengcongensis* (see Table S2 in the supplemental material). The genetic information necessary for self-maintenance and reproduction in the presence of a full complement of essential nutrients (13) seemed to be present in "*Candidatus* Cloacamonas acidaminovorans" (see Table S3 in
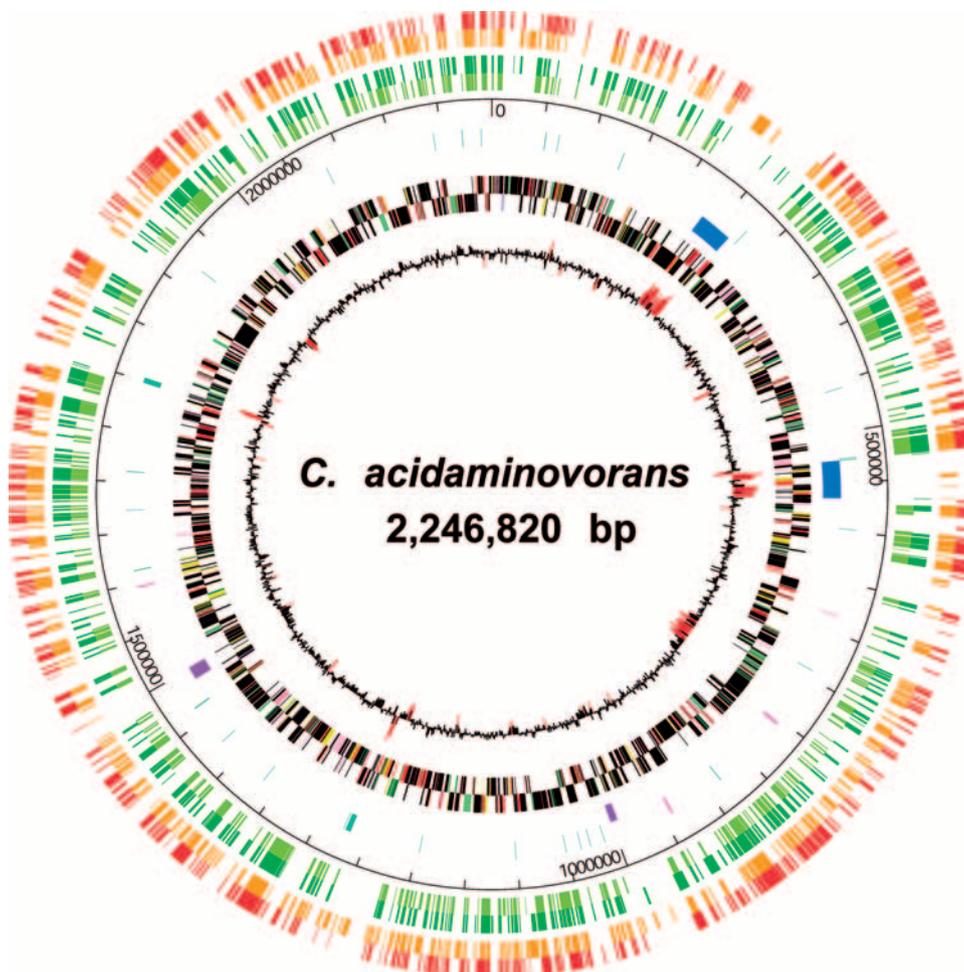
FIG. 2. Circular representation of the "*Candidatus* Cloacamonas acidaminovorans" chromosome. The circles, from the inside out, show the following. (Circle 1) GC deviation (GC$_{window}$ − average GC of the genome, using a 1-kb window); regions with a GC deviation of less than twice the standard deviation are highlighted in red. (Circles 2 and 3) Predicted CDSs transcribed in clockwise/counterclockwise directions. Color coding for genes is as follows: amino acid biosynthesis, salmon; nucleotides/nucleosides, orange; central metabolism, brown; DNA metabolism, yellow; energy, green; fatty acid metabolism, purple; cell envelope, light green; protein synthesis/fate, pink; transcription, gray; transport, peach; regulatory functions, slate blue; extrachromosomal origin, violet-red; hypothetical and conserved hypothetical proteins, black. (Circle 4) CRISPR (violet-brown) and prophagic regions, (dark blue). (Circle 5) tRNAs (green), rRNA (violet-blue), insertion elements and transposases (pink). (Circle 6) Coordinates in Mb beginning at the *dnaA* gene. Circles 6 and 7 show the gene content comparison between the "*Candidatus* Cloacamonas acidaminovorans" and *P. carbinolicus* (light green), *Synthrophus aciditrophicus* (dark green), *K. stuttgartiensis* (orange), and *T. tengcongensis* (red) genomes using a similarity threshold of 30% identity and a ratio of 0.8 of the length of the smallest protein.

TABLE 1. General features of the "*Candidatus* Cloacamonas acidaminovorans" genome

General characteristic	Value
Size (bp)	2,246,820
GC content (%)	37.9
Coding density (%)	81
16S-23S-5S rRNA operon	2
tRNA	47
Predicted CDS	1820
Proteins with predicted function (%)	51
Cell wall/membrane/envelope biogenesis (no. of CDSs)	147
Energy production and conversion (no. of CDSs)	108
Amino acid transport and metabolism (no. of CDSs)	148
Proteins without predicted function (%)	49
Conserved hypothetical proteins (%)	9
Hypothetical proteins (%)	40

the supplemental material for a detailed analysis) and supported the tentative reconstruction of the present genome assembly. Some interesting features of the "*Candidatus* Cloacamonas acidaminovorans" genome were noted. First, it harbors proteins with sequences that are more related to their eukaryotic than their prokaryotic counterparts. For example, Cloam0177 is highly similar to the *ftdD* gene of rat, mouse, and human (identity of 40% on the total length of the protein), a bifunctional enzyme involved in histidine degradation (glutamate formimidoyltransferase fused with formimidoyltetrahydrofolate cyclodeaminase). Secondly, it uses pyrophosphate-dependent enzymes: fructose-6-phosphate 1-phosphotransferase (EC 2.7.1.90) and pyruvate-phosphate dikinase (EC 2.7.9.1). Finally, the presence of the enzymes involved in the biosynthesis of lipopolysaccharides indicates that "*Candidatus* Cloacamonas acidaminovorans" is a gram-negative bacterium. This is further

confirmed by the detection of the typical signature of gram-negative bacteria found in the *dnaK* gene (15).

Metabolism. "*Candidatus* Cloacamonas acidaminovorans" does not possess the enzymes necessary for the synthesis of 12 amino acids: leucine, isoleucine, valine, threonine, methionine, proline, arginine, histidine, phenylalanine, tyrosine, tryptophan, and cysteine. Moreover, it cannot produce polyamines and several cofactors: thiamine, biotin, lipoic acid, pyrroloquinolinequinone, coenzyme B12, folic acid, pyridoxine, and heme. Thus, all these compounds must be obtained from the environment. The in silico proteome analysis suggests that "*Candidatus* Cloacamonas acidaminovorans" could grow on a medium that includes glucose, alanine, asparagine, aspartate, glutamate, histidine, lysine, proline, serine, and certain aliphatic carboxylic acids (succinate, lactate, and acetate).

Since "*Candidatus* Cloacamonas acidaminovorans" possesses several proteins typical of anaerobic bacteria (the oxygen-sensitive class III ribonucleoside triphosphate reductase, several ferredoxin oxidoreductases, and radical *S*-adenosylmethionine-dependent proteins), it seems to be adapted to an anaerobic lifestyle (Table 2; see Table S4 in the supplemental material). However, the presence of the aerotolerant class II ribonucleotide reductase and some enzymes involved in protection against oxidative stress (i.e., superoxide reductase, ruberythrin, and thioredoxin reductase) may enable the bacterium to survive under minimal oxygen concentrations (Table 2).

No electron transfer chain necessary for respiration was found in "*Candidatus* Cloacamonas acidaminovorans," and none of the classical terminal electron acceptors of anaerobic bacteria appeared to be used. Therefore, the bacterium seems to get its energy through the glycolytic Embden-Meyerhof pathway and the fermentation of amino acids, sugars, and carboxylic acids, as well as by the excretion of protons and sodium ions coupled with the reaction of the hydrogenase, the pyrophosphate-energized proton pump, and the methylmalonyl-coenzyme A (CoA) decarboxylase, respectively (Table 2). This protein catalyzes a reaction characteristic of anaerobic prokaryotes. ATP synthesis is performed by a membrane-bound ATP synthase. More than 30 proteases and peptidases were identified. "*Candidatus* Cloacamonas acidaminovorans" derives most of its carbon and nitrogen sources from the degradation of proteins. In addition, the bacterium possesses five different ferredoxin oxidoreductases, which are principally involved in the fermentation process of amino acids (Table 2). Among these processes, the best identified was the lysine fermentation pathway. In fact, all the enzymes of this pathway except one have been detected. In contrast, the exact outcome of the fermentation of the other amino acids is unknown. These catabolic-degradation pathways are generally coupled with substrate level phosphorylation (ATP formation) and synthesis of reductants, mainly ferredoxins. Most of these ferredoxins are shared with anaerobic bacteria, known or not, to develop syntrophic interactions (see Table S4 in the supplemental material).

Besides these reducing agents, "*Candidatus* Cloacamonas acidaminovorans" possesses a remarkable variety of redoxins (thioredoxin, rubredoxin, glutaredoxin, peroxiredoxin, and polyferredoxin). They principally act as hydrogen donors for the reduction of metabolites and of disulfide bonds in proteins and also function as electron carriers for the hydrogenase to

TABLE 2. Genetic determinants of "*Candidatus* Cloacamonas acidaminovorans"

Function	Gene	Gene identifier
Life in the presence of oxygen		
Superoxide reductase	*dfx*	Cloam 1366
Ruberythrin	*rbr*	Cloam 1367
Rubredoxin	*rub*	Cloam 1368
Peroxiredoxin	*ahpC*	Cloam 1564
Thioredoxin	*trx*	Cloam 0816
Thioredoxin reductase	*trxB*	Cloam 1697
Rubredoxin-oxygen oxidoreductase or nitric oxide reductase (putative)	*fprA*	Cloam 1876
Class II ribonucleotide reductase (putative)		Cloam 0913
Anaerobic lifestyle		
Class III ribonucleotide reductase	*nrdD*	Cloam 1253
Class III ribonucleotide reductase-activating enzyme	*nrdG*	Cloam 1254
Pyruvate ferredoxin oxidoreductase	*por*	Cloam 0324
2-Ketoglutarate ferredoxin oxidoreductase	*korC*	Cloam 0295 (γ-subunit)
	korB	Cloam 0296 (β-subunit)
	korA	Cloam 0297 (α-subunit)
	korD	Cloam 0298 (δ-subunit, putative)
Aldehyde ferredoxin oxidoreductase, tungsten containing	*aorB*	Cloam 0708
	aorA	Cloam 0709
Branched-chain ketoacid ferredoxin oxidoreductase	*vorD*	Cloam 1001 (δ-subunit)
	vorC	Cloam 1004 (γ-subunit)
	vorB	Cloam 1003 (β-subunit)
	vorA	Cloam 1002 (α-subunit)
Indolepyruvate ferredoxin oxidoreductase	*iorB*	Cloam 1037 (β-subunit)
	iorA	Cloam 1038 (α-subunit)
Energy		
Fe-only hydrogenase	*hymC*	Cloam 1180 (α-subunit)
	hymB	Cloam 1183 (β-subunit)
	hymA	Cloam 1186 (γ-subunit)
Fe-only hydrogenase, assembly protein	*hydEF*	Cloam 0448
	hydG	Cloam 0921
Methylmalonyl-CoA decarboxylase (putative)		Cloam 1768 (α-subunit)
		Cloam 1770 (γ-subunit)
		Cloam 1771 (β-subunit)
ATP synthase		Cloam 1052 (subunit A1)
		Cloam1053 (subunit B1)
		Cloam 1054 (subunit D)
		Cloam 1055 (subunit I, putative)
		Cloam 1056 (subunit K, putative)
Pyrophosphate-energized proton pump	*hppA*	Cloam 1022

form hydrogen. "*Candidatus* Cloacamonas acidaminovorans" possesses one Fe-only hydrogenase and one putative Fe-only hydrogenase. In addition, an association of another putative hydrogenase with a putative formate dehydrogenase was detected. This complex might function as a simple type of formate hydrogen lyase, generating CO_2 and H_2 from formate (14). Figure 3 highlights the metabolic specificities likely to be used by "*Candidatus* Cloacamonas acidaminovorans."
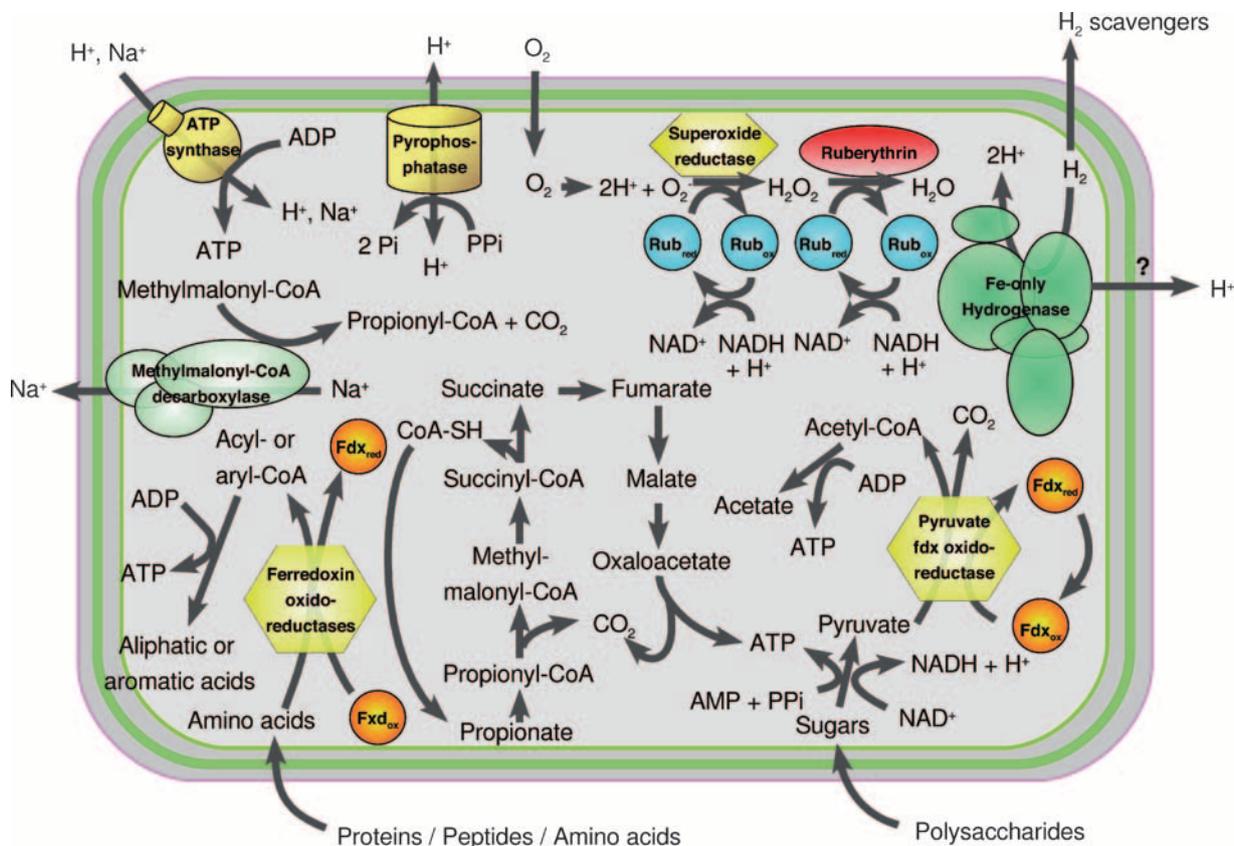
FIG. 3. Overview of the main metabolic pathways in "*Candidatus* Cloacamonas acidaminovorans." Shown are the oxygen defense process, glycolysis and pyruvate degradation, amino acid fermentation, and systems involved in energy production. The question mark means that the presence of the reaction was not clearly identified.

A characteristic metabolic pathway in obligate syntrophic bacteria is the oxidation of propionate into acetate and carbon dioxide via methylmalonyl-CoA, succinate, fumarate, malate, oxaloacetate, pyruvate, and acetyl-CoA as intermediates (26). This pathway is thermodynamically favorable only when hydrogen pressure is low. Thus, propionate-oxidizing bacteria form syntrophic consortia with hydrogen-scavenging bacteria, such as methanogenic, sulfate-reducing, and acetogenic bacteria (1, 22). All the genes involved in this pathway have been identified in "*Candidatus* Cloacamonas acidaminovorans." One step, the carboxylation of propionyl-CoA, is coupled with the decarboxylation of oxaloacetate by means of a transcarboxylase. This protein consists of three subunits: 12S, 5S, and 1.3S. Interestingly, the two genes encoding the 5S and the 1.3S subunit were fused. The presence of the oxidative propionate degradation pathway in "*Candidatus* Cloacamonas acidaminovorans," as well as the ability of the bacterium to produce hydrogen via fermentation and its Fe-only hydrogenases, makes it likely that "*Candidatus* Cloacamonas acidaminovorans" is a syntrophic bacterium. Furthermore, compared to the other sequenced microbial genomes, the largest groups of colocalized orthologous genes with "*Candidatus* Cloacamonas acidaminovorans" were found in *P. carbinolicus*, *Syntrophobacter fumaroxidans*, and *Syntrophus aciditrophicus* (see Table S2 in the supplementary material), which are known to develop syntrophic interactions (10, 23, 27).

Detection and culture enrichment. Trials of enrichment cultures of "*Candidatus* Cloacamonas acidaminovorans" from activated sludge obtained from the anaerobic digester of Evry were carried out using a synthetic growth medium in which lysine served as a carbon and energy source (see Materials and Methods). A number of experiments were set up in order to obtain enrichment cultures of "*Candidatus* Cloacamonas acidaminovorans." Although "*Candidatus* Cloacamonas acidaminovorans" was still detectable in some cases after 1 year of cultivation, quantitative PCR assays showed no detectable increase in the "*Candidatus* Cloacamonas acidaminovorans" cell numbers.

Moreover, "*Candidatus* Cloacamonas acidaminovorans" had disappeared or at least had become undetectable by PCR in the digester of Evry at the time (July 2005) we started enrichment cultures. The presence of these bacteria was then checked for monthly. One year later, "*Candidatus* Cloacamonas acidaminovorans" reappeared. This phenomenon was also observed in another digester (in Creil, France).

The presence of WWE1 and "*Candidatus* Cloacamonas acidaminovorans" bacteria was checked for by PCR on DNAs extracted from 43 anaerobic digesters (19; see Table S5 in the supplemental material). In addition to a few pilot or laboratory scale digesters, most of them were industrial installations processing wastewater with different amounts of industrial wastes. While WWE1 bacteria were detected in 32 anaerobic digest-

ers, "*Candidatus* Cloacamonas acidaminovorans" was found in 13 of them. WWE1 and "*Candidatus* Cloacamonas acidaminovorans" bacteria thus seem to be widespread in anaerobic digesters.

DISCUSSION

Complete reconstruction of the genomes of uncultivated microorganisms remains a challenge, and the majority of the hundred or so bacterial divisions (11) are still known only from their 16S rRNA gene sequences. In a few cases, nearly complete genomes were obtained after shotgun sequencing of enrichment cultures (12, 28), symbionts (16, 32), or environmental samples (29), but in these studies, the predominant microorganisms represented at least 60% of the microbial population. Reconstructing a genome from a large metagenomic library requires several conditions, such as (i) limited intraspecies variation (29), as genome rearrangement can severely interfere with the iterative process; (ii) sufficient genome coverage in the metagenomic library to allow the assembly step; and (iii) availability of a large clone library with long inserts in order to allow efficient genome walking through the iterations and to prevent problems due to the presence of repeats in the genome sequence. Consequently, this iterative approach is limited to only a subset of genomes of a given metagenomic study. In this report, we present for the first time the genome reconstruction and analysis of an uncultivated bacterium from a candidate bacterial division that is not predominant in a complex environment. The "*Candidatus* Cloacamonas acidaminovorans" genome represents 2% of the FES generated from the studied anaerobic digester. The availability of the complete "*Candidatus* Cloacamonas acidaminovorans" genome gives us the opportunity to investigate the metabolic potential of this anaerobic bacterium. However, we have seen that "*Candidatus* Cloacamonas acidaminovorans," but not the WWE1 bacteria, disappeared for long periods from two of the studied digesters, which raises questions about their respective roles in anaerobic digestion.

This achievement suggests that it should be possible to recover partial or entire microbial genomes from diverse complex environments. This could become an interesting way to circumvent cultivation or enrichment difficulties.

Genome analysis has shown that "*Candidatus* Cloacamonas acidaminovorans" has a large proportion of genes of unknown function and a low percentage of genes found in groups of colocalized orthologs shared with other species. "*Candidatus* Cloacamonas acidaminovorans" and WWE1 bacteria are remote from known microbial genomes. In silico proteome analysis suggests that "*Candidatus* Cloacamonas acidaminovorans" is probably a syntrophic bacterium. However, despite the fact that "*Candidatus* Cloacamonas acidaminovorans" survived for a 1-year period in some cultures, we have not yet been able to enrich or isolate the bacterium or any other representative of the WWE1 candidate division. The selective enrichment of "*Candidatus* Cloacamonas acidaminovorans" was also unsuccessfully tried in a stable mixed culture with potential syntrophic partners, such as *Desulfovibrio desulfuricans* and *Methanobacterium formicicum*. Perhaps the syntrophic partners of "*Candidatus* Cloacamonas acidaminovorans" are microorganisms that have not yet been cultivated.

WWE1 bacteria in general and "*Candidatus* Cloacamonas acidaminovorans" in particular were detected in a number of anaerobic digesters. An extension of the investigation to many other digesters and also to diverse anaerobic environments, such as sediments and soils, should be worthwhile, allowing us to gain a better knowledge of WWE1 bacteria and "*Candidatus* Cloacamonas acidaminovorans" obligate companions.

REFERENCES

1. **Ariesyady, H. D., T. Ito, K. Yoshiguchi, and S. Okabe.** 2007. Phylogenetic and functional diversity of propionate-oxidizing bacteria in an anaerobic digester sludge. Appl. Microbiol. Biotechnol. **75:**673–683.
2. **Barbe, V., D. Vallenet, N. Fonknechten, A. Kreimeyer, S. Oztas, L. Labarre, S. Cruveiller, C. Robert, S. Duprat, P. Wincker, L. N. Ornston, J. Weissenbach, P. Marlière, G. N. Cohen, and C. Médigue.** 2004. Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. Nucleic Acids Res. **32:**5766–5779.
3. **Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero, and P. Horvath.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709–1712.
4. **Bocs, S., S. Cruveiller, D. Vallenet, G. Nuel, and C. Médigue.** 2003. AMIGene: Annotation of MIcrobial Genes. Nucleic Acids Res. **31:**3723–3726.
5. **Bolotin, A., B. Quinquis, A. Sorokin, and S. D. Ehrlich.** 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology **151:**2551–2561.
6. **Chouari, R., D. Le Paslier, P. Daegelen, P. Ginestet, J. Weissenbach, and A. Sghir.** 2003. Molecular evidence for novel planctomycete diversity in a municipal wastewater treatment plant. Appl. Environ. Microbiol. **69:**7354–7363.
7. **Chouari, R., D. Le Paslier, P. Daegelen, P. Ginestet, J. Weissenbach, and A. Sghir.** 2005. Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. Environ. Microbiol. **7:**1104–1115.
8. **Chouari, R., D. Le Paslier, C. Dauga, P. Daegelen, J. Weissenbach, and A. Sghir.** 2005. Novel major bacterial candidate division within a municipal anaerobic sludge digester. Appl. Environ. Microbiol. **71:**2145–2153.
9. **Daims, H., A. Bruhl, R. Amann, K. H. Schleifer, and M. Wagner.** 1999. The domain-specific probe EUB338 is insufficient for the detection of all *Bacteria*: development and evaluation of a more comprehensive probe set. Syst. Appl. Microbiol. **22:**434–444.
10. **de Bok, F. A., M. L. Luijten, and A. J. Stams.** 2002. Biochemical evidence for formate transfer in syntrophic propionate-oxidizing cocultures of *Syntrophobacter fumaroxidans* and *Methanospirillum hungatei*. Appl. Environ. Microbiol. **68:**4247–4252.
11. **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.
12. **Garcia Martin, H., N. Ivanova, V. Kunin, F. Warnecke, K. W. Barry, A. C. McHardy, C. Yeates, S. He, A. A. Salamov, E. Szeto, E. Dalin, N. H. Putnam, H. J. Shapiro, J. L. Pangilinan, I. Rigoutsos, N. C. Kyrpides, L. L. Blackall, K. D. McMahon, and P. Hugenholtz.** 2006. Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. Nat. Biotechnol. **24:**1263–1269.
13. **Gil, R., F. J. Silva, J. Pereto, and A. Moya.** 2004. Determination of the core of a minimal bacterial gene set. Microbiol. Mol. Biol. Rev. **68:**518–537.
14. **Graentzdoerffer, A., D. Rauh, A. Pich, and J. R. Andreesen.** 2003. Molecular and biochemical characterization of two tungsten- and selenium-containing formate dehydrogenases from *Eubacterium acidaminophilum* that are associated with components of an iron-only hydrogenase. Arch. Microbiol. **179:**116–130.
15. **Gupta, R. S.** 1998. Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. Microbiol. Mol. Biol. Rev. **62:**1435–1491.
16. **Hallam, S. J., K. T. Konstantinidis, N. Putnam, C. Schleper, Y. Watanabe, J. Sugahara, C. Preston, J. de la Torre, P. M. Richardson, and E. F. DeLong.** 2006. Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. Proc. Natl. Acad. Sci. USA **103:**18296–18301.

17. **Jansen, R., J. D. Embden, W. Gaastra, and L. M. Schouls.** 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. **43:**1565–1575.
18. **Keller, M., and K. Zengler.** 2004. Tapping into microbial diversity. Nat. Rev. Microbiol. **2:**141–150.
19. **Leclerc, M., J. P. Delgenes, and J. J. Godon.** 2004. Diversity of the archaeal community in 44 anaerobic digesters as determined by single strand conformation polymorphism analysis and 16S rDNA sequencing. Environ. Microbiol. **6:**809–819.
20. **Leininger, S., T. Urich, M. Schloter, L. Schwark, J. Qi, G. W. Nicol, J. I. Prosser, S. C. Schuster, and C. Schleper.** 2006. Archaea predominate among ammonia-oxidizing prokaryotes in soils. Nature **442:**806–809.
21. **Makarova, K. S., N. V. Grishin, S. A. Shabalina, Y. I. Wolf, and E. V. Koonin.** 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.
22. **Megonigal, J. P., M. E. Hines, and P. T. Visscher.** 2004. Anaerobic metabolism: linkages to trace gases and aerobic processes, p. 317–424. *In* W. H. Schlesinger (ed.), Biochemistry. Elsevier-Pergamon, Oxford, United Kingdom.
23. **Peters, F., Y. Shinoda, M. J. McInerney, and M. Boll.** 2007. Cyclohexa-1,5-diene-1-carbonyl-coenzyme A (CoA) hydratases of *Geobacter metallireducens* and *Syntrophus aciditrophicus*: evidence for a common benzoyl-CoA degradation pathway in facultative and strict anaerobes. J. Bacteriol. **189:**1055–1060.
24. **Rappé, M. S., and S. J. Giovannoni.** 2003. The uncultured microbial majority. Annu. Rev. Microbiol. **57:**369–394.
25. **Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. E. Venter, K. Li, S. Kravitz, J. F. Heidelberg, T. Utterback, Y. H. Rogers, L. I. Falcon, V. Souza, G. Bonilla-Rosso, L. E. Eguiarte, D. M. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M. R. Ferrari, R. L. Strausberg, K. Nealson, R. Friedman, M. Frazier, and J. C. Venter.** 2007. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol. **5:**e77.
26. **Schink, B.** 1997. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol. Mol. Biol. Rev. **61:**262–280.
27. **Scholten, J. C., D. E. Culley, F. J. Brockman, G. Wu, and W. Zhang.** 2007. Evolution of the syntrophic interaction between *Desulfovibrio vulgaris* and *Methanosarcina barkeri*: involvement of an ancient horizontal gene transfer. Biochem. Biophys. Res. Commun. **352:**48–54.
28. **Strous, M., E. Pelletier, S. Mangenot, T. Rattei, A. Lehner, M. W. Taylor, M. Horn, H. Daims, D. Bartol-Mavel, P. Wincker, V. Barbe, N. Fonknechten, D. Vallenet, B. Ségurens, C. Schenowitz-Truong, C. Médigue, A. Collingro, B. Snel, B. E. Dutilh, H. J. Op den Camp, C. van der Drift, I. Cirpus, K. T. van de Pas-Schoonen, H. R. Harhangi, L. van Niftrik, M. Schmid, J. Keltjens, J. van de Vossenberg, B. Kartal, H. Meier, D. Frishman, M. A. Huynen, H. W. Mewes, J. Weissenbach, M. S. Jetten, M. Wagner, and D. Le Paslier.** 2006. Deciphering the evolution and metabolism of an anammox bacterium from a community genome. Nature **440:**790–794.
29. **Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, and J. F. Banfield.** 2004. Community structure and metabolism through reconstruction of microbial genomes from the environment. Nature **428:**37–43.
30. **Vallenet, D., L. Labarre, Z. Rouy, V. Barbe, S. Bocs, S. Cruveiller, A. Lajus, G. Pascal, C. Scarpelli, and C. Médigue.** 2006. MaGe: a microbial genome annotation system supported by synteny results. Nucleic Acids Res. **34:**53–65.
31. **Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:**66–74.
32. **Woyke, T., H. Teeling, N. N. Ivanova, M. Huntemann, M. Richter, F. O. Gloeckner, D. Boffelli, I. J. Anderson, K. W. Barry, H. J. Shapiro, E. Szeto, N. C. Kyrpides, M. Mussmann, R. Amann, C. Bergin, C. Ruehland, E. M. Rubin, and N. Dubilier.** 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. Nature **443:**950–955.