Strain-level genetic diversity of *Methylophaga nitratireducenticrescens* confers plasticity to denitrification capacity in a methylotrophic marine denitrifying biofilm

Valérie Geoffroy ¹, ², Geneviève Payette ¹, Florian Mauffrey ³, Livie Lestin ¹, Philippe Constant ¹, Richard Villemur

¹ Institut Armand-Frappier, Institut National de la Recherche Scientifique, Laval, Québec, Canada
² Lallemand, Montreal, Québec, Canada
³ Laboratoire de santé publique du Québec, Ste-Anne-de-Bellevue, Québec, Canada

Corresponding Author: Richard Villemur
Email address: richard.villemur@iaf.inrs.ca

**Background.** The biofilm of a methanol-fed, fluidized denitrification system treating a marine effluent is composed of multi-species microorganisms, among which *Hyphomicrobium nitrativorans* NL23 and *Methylophaga nitratireducenticrescens* JAM1 are the principal bacteria involved in the denitrifying activities. Strain NL23 can carry complete nitrate (NO$_3^-$) reduction to N$_2$, whereas strain JAM1 can perform 3 out of the 4 reduction steps. A small proportion of other denitrifiers exists in the biofilm, suggesting the potential plasticity of the biofilm in adapting to environmental changes. Here, we report the acclimation of the denitrifying biofilm from continuous operating mode to batch operating mode, and the isolation and characterization from the acclimated biofilm of a new denitrifying bacterial strain, named GP59.

**Methods.** The denitrifying biofilm was batch-cultured under anoxic conditions. The acclimated biofilm was plated on *Methylophaga* specific medium to isolate denitrifying *Methylophaga* isolates. Planktonic cultures of strains GP59 and JAM1 were performed, and the growth and the dynamics of NO$_3^-$, nitrite (NO$_2^-$) and N$_2$O were determined. The genomes of strains GP59 and JAM1 were sequenced and compared. The transcriptomes of strains GP59 and JAM1 were derived from anoxic cultures.

**Results.** During batch cultures of the biofilm, we observed the disappearance of *H. nitrativorans* NL23 without affecting the denitrification performance. From the acclimated biofilm, we isolated strain GP59 that can perform, like *H. nitrativorans* NL23, the complete denitrification pathway. The GP59 cell concentration in the acclimated biofilm was 2-3 orders of magnitude higher than *M. nitratireducenticrescens* JAM1 and *H. nitrativorans* NL23. Genome analyses revealed that strain GP59 belongs to the species *M. nitratireducenticrescens*. The GP59 genome shares more than 85% of its coding sequences with those of strain JAM1. Based on transcriptomic analyses of anoxic cultures, most of these common genes in strain GP59 were expressed at similar level than their counterparts in strain JAM1. In contrast to strain JAM1, strain GP59 cannot reduce NO$_3^-$ under oxic culture conditions, and has a 24-h lag time before growth and NO$_3^-$ reduction start to occur in anoxic cultures, suggesting that both strains regulate differently the expression of their denitrification genes. Strain GP59 has the ability to reduce NO$_2^-$ as it carries a gene encoding a NirK-type NO$_2^-$ reductase. Based on the CRISPR sequences, strain GP59 did not emerge from strain JAM1 during the biofilm batch cultures but rather was present in the original biofilm and was enriched during this process.

**Discussion.** These results reinforce the unique trait of the species *M. nitratireducenticrescens* among the *Methylophaga* genus as facultative anaerobic bacterium. These findings also showed the plasticity of denitrifying population of the biofilm in adapting to anoxic marine environments of the bioreactor.
Title

Strain-level genetic diversity of *Methylophaga nitratireducenticrescens* confers plasticity to
denitrification capacity in a methylotrophic marine denitrifying biofilm

Authors

Valérie Geoffroy$^{1,2}$, Geneviève Payette$^1$, Florian Mauffrey$^{1,3}$, Livie Lestin$^1$, Philippe Constant$^1$,
Richard Villemur$^1$

Addresses

$^1$INRS-Institut Armand-Frappier, Laval, Québec, Canada

$^2$Present address: Lallemand, Montreal, Québec, Canada

$^3$Present address: Laboratoire de santé publique du Québec, Ste-Anne-de-Bellevue, Québec, Canada

Corresponding author.

Richard Villemur, richard.villemur@iaf.inrs.ca
Abstract

Background. The biofilm of a methanol-fed, fluidized denitrification system treating a marine effluent is composed of multi-species microorganisms, among which *Hyphomicrobium nitrativorans* NL23 and *Methylophaga nitratireducenticrescens* JAM1 are the principal bacteria involved in the denitrifying activities. Strain NL23 can carry complete nitrate (NO$_3^-$) reduction to N$_2$, whereas strain JAM1 can perform 3 out of the 4 reduction steps. A small proportion of other denitrifiers exists in the biofilm, suggesting the potential plasticity of the biofilm in adapting to environmental changes. Here, we report the acclimation of the denitrifying biofilm from continuous operating mode to batch operating mode, and the isolation and characterization from the acclimated biofilm of a new denitrifying bacterial strain, named GP59.

Methods. The denitrifying biofilm was batch-cultured under anoxic conditions. The acclimated biofilm was plated on *Methylophaga* specific medium to isolate denitrifying *Methylophaga* isolates. Planktonic cultures of strains GP59 and JAM1 were performed, and the growth and the dynamics of NO$_3^-$, nitrite (NO$_2^-$) and N$_2$O were determined. The genomes of strains GP59 and JAM1 were sequenced and compared. The transcriptomes of strains GP59 and JAM1 were derived from anoxic cultures.

Results. During batch cultures of the biofilm, we observed the disappearance of *H. nitrativorans* NL23 without affecting the denitrification performance. From the acclimated biofilm, we isolated strain GP59 that can perform, like *H. nitrativorans* NL23, the complete denitrification pathway. The GP59 cell concentration in the acclimated biofilm was 2-3 orders of magnitude
higher than *M. nitratireducenticrescens* JAM1 and *H. nitrativorans* NL23. Genome analyses revealed that strain GP59 belongs to the species *M. nitratireducenticrescens*. The GP59 genome shares more than 85% of its coding sequences with those of strain JAM1. Based on transcriptomic analyses of anoxic cultures, most of these common genes in strain GP59 were expressed at similar level than their counterparts in strain JAM1. In contrast to strain JAM1, strain GP59 cannot reduce NO$_3^-$ under oxic culture conditions, and has a 24-h lag time before growth and NO$_3^-$ reduction start to occur in anoxic cultures, suggesting that both strains regulate differently the expression of their denitrification genes. Strain GP59 has the ability to reduce NO$_2^-$ as it carries a gene encoding a NirK-type NO$_2^-$ reductase. Based on the CRISPR sequences, strain GP59 did not emerge from strain JAM1 during the biofilm batch cultures but rather was present in the original biofilm and was enriched during this process.

**Discussion.** These results reinforce the unique trait of the species *M. nitratireducenticrescens* among the *Methylophaga* genus as facultative anaerobic bacterium. These findings also showed the plasticity of denitrifying population of the biofilm in adapting to anoxic marine environments of the bioreactor.
Introduction

Denitrification describes the successive reduction of nitrate ($\text{NO}_3^-$) to nitrite ($\text{NO}_2^-$), nitric oxide (NO), nitrous oxide ($\text{N}_2\text{O}$), and nitrogen ($\text{N}_2$) (1). This process is used by bacteria for respiration in environments with low oxygen concentrations and with $\text{NO}_3^-$, $\text{NO}_2^-$, NO, and $\text{N}_2\text{O}$ as electron acceptors. The process is driven by metalloenzymes $\text{NO}_3^-$ reductase, $\text{NO}_2^-$ reductase, NO reductase, and $\text{N}_2\text{O}$ reductase (2). As a facultative trait, denitrification occurs frequently across environments, and is performed by bacteria of diverse origins (3). Furthermore, numerous microorganisms carry incomplete denitrification pathways for either their growth or detoxification, as $\text{NO}_2^-$ and NO are deleterious molecules (4-7).

Several studies were carried out in our laboratory on a naturally-occurring multispecies denitrifying biofilm that has developed in a methanol-fed, fluidized denitrification continuous system that treated recirculating water of a marine aquarium at the Montreal Biodome. The biofilm consists of at least 15 bacterial species and numerous protozoans (8, 9), among which *Methylophaga* spp. and *Hyphomicrobium* spp. compose more than 50% of the biofilm (10). Rissanen et al. (11) also observed the combination of *Methylophaga* spp. and *Hyphomicrobium* spp. in the fluidized-bed type denitrification reactors treating the recirculating seawater of the public fish aquarium SEA LIFE at Helsinki, Finland, suggesting the importance of these two genera in marine denitrification processes.

We isolated *Hyphomicrobium nitrativorans* NL23 and *Methylophaga nitratireducenticrescens* JAM1 from the biofilm of the Biodome denitrification system and showed they are the main actors of the denitrifying activities (8, 12-15). *H. nitrativorans* NL23 can carry out complete $\text{NO}_3^-$ reduction to $\text{N}_2$, whereas *M. nitratireducenticrescens* JAM1 can perform 3 out of the 4 reduction steps; it misses the reduction of $\text{NO}_2^-$ to NO (16-20). In addition
to *H. nitrativorans* NL23 and *M. nitratireducenticrescens* JAM1, we have demonstrated that a
small proportion of other denitrifiers exists in the biofilm of the Biodome denitrification system
(13), suggesting the potential plasticity of the biofilm in adapting to environmental changes.

*Methylophaga* species are halophilic marine methylotrophic gammaproteobacteria that do
not grow in the absence of NaCl and use one-carbon compounds (*e.g.* methanol) as sole carbon
and energy sources with carbon assimilation proceeding via the 2-keto-3-deoxy-6-
phosphogluconate (KDPG)-variant ribulose monophosphate (RuMP) pathway (15, 21). The
operating conditions of the Biodome denitrification process led to the enrichment of *M.
nitratireducenticrescens* JAM1 in the biofilm, which is the only reported isolated *Methylophaga*
species that can grow under anoxic conditions with NO$_3^-$ or N$_2$O as sole electron acceptors (12,
18, 19). This trait is correlated with the presence in the genome of two gene clusters encoding
dissimilatory NO$_3^-$ reductases (*narGHJI*, referred as *nar1* and *nar2*), two gene clusters encoding
cytochrome *bc*-type complex NO reductases (referred as *cnor1* and *cnor2*) and one gene cluster
encoding a dissimilatory N$_2$O reductase (*nos*) (15, 19). *M. nitratireducenticrescens* JAM1 lacks a
gene encoding a dissimilatory copper- (NirK) or cytochrome cd1-type (NirS) NO$_2^-$ reductase.

Our group has also investigated the potential of using the batch operating mode for the
denitrification system instead of the continuous one as it prevailed at the Montreal Biodome.
With this mode, a better real-time control was obtained to avoid sulfate-reduction that can be
easily occurring in seawater biotreatments (22). In this report, we aim to assess the impact of the
batch operating mode on the bacterial population of the denitrifying biofilm. We observed the
disappearance of *H. nitrativorans* NL23 without the loss of the denitrifying activities during the
biofilm acclimation to this mode. We hypothesized that new denitrifiers were enriched during the
biofilm acclimation, which displaced *H. nitrativorans* NL23. From this biofilm, a new
denitrifying bacterial strain, named GP59, was isolated that is closely related to *M. nitratireducenticrescens* JAM1. We compared strain GP59 with *M. nitratireducenticrescens* JAM1 at the physiological, genomic and transcriptomic levels. Genomic data were also compared with available genomes of *Methylophaga* species. Our results present new insights of *M. nitratireducenticrescens* under anoxic environments. They also provide indications that strain-level genomic heterogeneities can influence the dynamism of the biofilm microbial population in adapting to new operating conditions.

**Material and Methods**

**NO\textsubscript{3}^−, NO\textsubscript{2}^− and N\textsubscript{2}O measurements**

Measurements of NO\textsubscript{3}^−, NO\textsubscript{2}^− and N\textsubscript{2}O concentrations in all biofilm and planktonic cultures were performed using ion chromatography as described in Mauffrey et al. (18).

**Biofilm acclimation**

Artificial seawater (ASW) medium was composed of (for 1 liter solution): 27.5 g NaCl, 10.68 g MgCl\textsubscript{2}·6H\textsubscript{2}O, 2 g MgSO\textsubscript{4}·7H\textsubscript{2}O, 1 g KCl, 0.5 g CaCl\textsubscript{2}, 456 µL of FeSO\textsubscript{4}·7H\textsubscript{2}O 4 g/L, 5 mL of KH\textsubscript{2}PO\textsubscript{4} 51.2 g/L, 5 mL of Na\textsubscript{2}HPO\textsubscript{4} 34 g/L. It was supplemented with 1 mL of trace elements (master solution: FeSO\textsubscript{4}·7H\textsubscript{2}O 0.9 g/L, CuSO\textsubscript{4}·5H\textsubscript{2}O 0.03 g/L and MnSO\textsubscript{4}·H\textsubscript{2}O 0.2 g/L) (23) and NaNO\textsubscript{3} (21.4 mM or 300 mg-N/L). The pH was adjusted (NaOH) at 8.0. Three 120-mL vials containing each 20 unused « Bioflow 9 mm » carriers (Rauschert, Steinwiessen, Germany) and 60 mL of ASW were purged of oxygen for 10 min with pure nitrogen gas (N\textsubscript{2}, purity >99.9%; Praxair, Mississauga, ON, Canada), sealed with sterile septum caps and
autoclaved. Prior use, these carriers were washed with HCl 10% (v/v) for 3 h, rinsed with water and autoclaved. A 90 µL volume (72 mg) of filtered-sterilized methanol was then added to the vials (0.15% v/v final concentration).

Carriers (Bioflow 9 mm) with the denitrifying biofilm were taken from the denitrification system at the Montreal Biodome and frozen at -20°C in seawater with 20% glycerol (24) until use. The biomass from several frozen carriers was thawed, scrapped, weighted and dispersed in the ASW medium at 0.08 g (wet weight)/mL. The biomass (5 mL/vial, 0.4 g of biofilm) was then distributed with a syringe and an 18G1½ needle to 3 vials. The vials were incubated at room temperature and 100 rpm with an orbital shaker in the dark. In average once a week, the vial was opened in ambient air, the carriers were taken, gently washed with ASW medium to remove the excess medium and the free bacteria, then transferred into vials with fresh anoxic medium, and incubated in the same conditions (Fig. S1). Methanol and NO₃⁻ were added when needed if NO₃⁻ was completely depleted during the week. After the 5th transfer, the biofilm was collected and referred to in the text as the acclimated biofilm. One part was preserved in 15% glycerol at -70°C for further used such as DNA extraction (see below). The other part was used to isolate bacterial isolates (next subsection).

Isolation of Methylophaga sp. strain GP59

The acclimated biofilm was dispersed in saline solution (3% NaCl, 34.2 mM phosphate buffer pH 7.4), and serial dilutions were made and inoculated onto the Methylophaga medium 1403 (American Type Culture Collection [ATCC], Manassas, VA, USA) supplemented with 1.5% agar and 0.3% v/v methanol. The plates were incubated at 30°C for no more than 7 days, under oxic conditions (no NO₃⁻) or under anoxic conditions in anoxic jars with a gaspak (BBL™)
GasPak™ Plus Anaerobic System). For the latter conditions, the medium was supplemented with 21.4 mM NaNO₃. Several colonies were picked and cultured in the Methylophaga 1403 liquid medium under oxic or anoxic conditions for 2-7 days. After restreaking three times onto agar medium, isolates were tested for growth under anoxic conditions. NO₃⁻ and NO₂⁻ consumption was measured after 4 days of incubation. Identification of the isolates was determined by extracted total DNA from the isolates, followed by PCR amplification of the 16S ribosomal RNA (rRNA) genes (forward primer 5'-AGAGTTTGATCCTGGCTCAG-3' and reverse primer 5'-AAGGAGGTGATCCAGCCGCA-3', which correspond to positions 8 to 27 and 1521 to 1540 in the Escherichia coli 16S rRNA gene) and the sequencing of the resulting PCR products (McGill University and Genome Quebec Innovation Center, Montréal QC, Canada).

Planktonic pure cultures

*M. nitratireducenticrescens* GP59 and JAM1 were cultured in the Methylophaga medium 1403 as described by Villeneuve et al. (15) with 0.3% methanol. When required, NO₃⁻ (NaNO₃) was added to the medium. For the anoxic cultures, 70-ml vials with 30 ml medium were flushed with pure nitrogen gas for 20 min, sealed with sterile septum caps and autoclaved. Culture bottles were incubated at 30°C (unless stated otherwise) in the dark. In the determination of optimal growth conditions, the pH, temperature and NaCl concentrations were adjusted in the Methylophaga medium 1403. For N₂O measurements, 720-mL bottles with 60-ml medium were used. For oxic cultures, cultures were performed in Erlenmeyer flasks and shaken at 150 rpm and 30°C (18). Bacterial growth was monitored by spectrophotometry (OD₆₀₀nm). Bacterial flocs were dispersed with a Potter-Elvehjem homogenizer prior to measurements when needed.
DNA extraction of the biofilm and quantitative PCR

The frozen samples of the biofilm taken from the denitrification system and of the acclimated biofilm taken after the 5th transfer were thawed, and washed 3 times with 0.5 mL the TEN buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA.2Na, NaCl 150 mM) and 5000 g centrifugation for 2 min. The biomass was dispersed in 0.5 mL TEN with lysozyme (2.5 mg/mL) and incubated at 37°C, 30 min. After adding sodium dodecyl sulfate and β-mercaptoethanol (final concentrations of 2% and 1%, respectively), the biomass was subjected to three cycles of freeze-thaw (10 min dry ice/ethanol bath, followed by 5 min at 65°C), then treated with proteinase K (50 µg/mL final concentration) at 45°C for 2 h, and finally with RNAse A (20 µg/mL final concentration) at 37°C for 15 min. The DNA was purified by extractions with phenol/chloroform/isoamyl alcohol (25:24:1 v/v) and chloroform/isoamyl alcohol (49:1), and by precipitation with 0.25 volume of 10 M ammonium acetate and 2 volumes of ethanol 95% (25). DNA concentration was determined by the quantiFluor dsDNA system (Promega, Madison, USA). Quantitative PCR (qPCR) was performed with SYBR green in 20-µL volume with the Fast Start essential DNA Green Master (Roche Diagnostics, Laval, QC, Canada) containing 200 nM of primers (Table 1) and 50 ng biofilm DNA. The amplifications were performed at 95°C for 10 min, followed by 40 cycles at 95°C for 20 s, at the annealing temperature (Table 1) for 20 s, and at 72°C for 15 s. After amplifications, specificity of the PCR products and the presence of primer dimers were verified by performing a melt curve by increasing the temperature from 65°C to 95°C by increments of 1°C per step with a pause of 5 s for each step. Reactions were performed in a Rotor-Gene 6000 real-time rotary analyzer (Qiagen Inc. Toronto, ON, Canada). The amplification efficiencies for all primer pairs varied between 0.9 and 1.1. The copy number of each gene per ng of biofilm was calculated according to standard curves using dilutions of a
gel-purified, PCR-amplified fragment of the corresponding genomic region of strain JAM1
(tagH) or of strain GP59 (nirK) (Table 1), or using dilution of the NL23 genomic DNA (napA) or
the JAM1 genomic DNA (narG1) as the template.

Genome sequencing and annotations

High molecular weight total DNA was extracted from GP59 and JAM1 pure cultures with
the DNeasy Blood & Tissue Extraction Kit according to the manufacturer (Qiagen, Hilden,
Allemagne). DNA samples were sent to the Genome Quebec Innovation Center for genome
sequencing and assembly by using the PacBio technology. Genome assembly used was the
HGAP workflow version April 2016 (26). Coverage was over 260 times with both genomes.
Genome assembly was checked manually with the DNASTAR software (version 14.1)
(Madison, WI, USA) for confirming or correcting pseudogenes. Gene annotations for the JAM1
genome were updated accordingly (Genbank accession number CP003390.3, National Center for
Biotechnology Information [NCBI] https://www.ncbi.nlm.nih.gov/). For the GP59 genome, gene
annotations were performed at NCBI (GenBank accession number CP021973). Annotations were
also performed for both genomes at the Rapid Annotation using Subsystem Technology (RAST)
web site (http://rast.nmpdr.org/seedviewer.cgi) and used for further analyses (27, 28).
Alignment between GP59 and JAM1 genomes was performed at RAST and with the
Mauve Multiple Genome Alignment version 20150226 (progressive alignment) (29). The
resulting alignments were corrected manually to precisely determine the common and unique
genes, but also the intergenic sequences. Riboswitches and non coding RNA (ncRNA) sequences
were determined by the NCBI annotation system. The average nucleotide identity (ANI) analysis
was performed at http://enve-omics.ce.gatech.edu/ani/ with window size 1000 bp, step size 200 bp, minimum identity 70%, minimum length 700 bp, minimum alignment 50 (30, 31). Genomes were analysed for tandem repeats at https://tandem.bu.edu/trf/trf.html (32).

*NirK Phylogenetic analysis*

The deduced amino acid sequence of the GP59 nirK was compared with public protein databases by BLASTP at the NCBI web site to retrieve the closest NirK sequences. These sequences were then aligned with Cobalt (33). Evolutionary analysis was conducted using MEGA6.06 (34) with the Maximum Likelihood method based on the Le_Gascuel_2008 model (35). Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories [+G, parameter = 0.8954]). The analysis involved 21 amino acid sequences. There were a total of 378 positions in the final dataset.

*The transcriptomes of M. nitratireducencrescens GP59 and JAM1*

Anoxic cultures of strains GP59 and JAM1 were performed in triplicate in the *Methylophaga* medium 1403 supplemented with 21.4 mM NaNO₃. Cells were harvested when NO₃⁻ was almost reduced in all cultures. Total RNA extraction, RNA processing, cDNA sequencing and transcriptome analyses were described by Mauffrey et al. (19), with the exception that we used Transcripts Per Million (TPM) for normalization instead of Reads Per Kilobase Million. One replicate culture of strain GP59 generated inconsistent results and was not
considered in the transcriptome analyses. Intergenic regions were also considered in the
transcriptome analyses to detect possible missed annotated genes or non-coding expressed
sequences. An intergenic region includes the sequence from the stop codon of a gene to the start
codon of the adjacent gene. This sequence was then trimmed of 50 nucleotides at both ends to
exclude as much as possible the transcript reads that would belong to the adjacent genes.

Significance for difference in gene expression in the corresponding genes/sequences (defined as
TPM) between strain JAM1 (triplicate cultures) and strain GP59 (duplicate cultures) was
performed with the R Bioconductor NOIseq package v2.14.0 (NOIseqBio with replicates and
q=0.9) (36) and run with the R software v3.2.3 (37). Genes/sequences of one strain that had >2-
fold higher level of TPM than the other strain showed significant differences. RNA-seq reads
were deposited in the Sequence Read Archive (SRA) under the numbers SRP066381 (strain
JAM1) and SRP132510 (strain GP59) at NCBI.

Results and Discussion

Biofilm acclimation to batch cultures

The biofilm of the denitrification system (operated under continuous mode) was dispersed
in vials containing 20 Bioflow carriers, the same used in the denitrification system, and cultured
under anoxic, batch-mode conditions with NO₃⁻. Only the carriers with the growing biofilm were
then transferred each week in fresh medium and cultured under the same conditions (see Material
& Methods section, Fig. S1). At the end of the 5th transfer culture when the NO₃⁻ and NO₂⁻
consumption rates stabilized, the buildup of biofilm was enough (Fig. S2) to extract total DNA
and to perform culture assays on agar plates.

NO₃⁻ and NO₂⁻ concentrations were measured at regular interval times during the 5th
transfer culture. Complete reduction of NO$_3^-$ and NO$_2^-$ (NO$_x$) was achieved within 15 h, at 1.8 mM NO$_x$ h$^{-1}$ (Fig. 1). This is 5-times faster than what was measured by Laurin et al. (24) with 60 Bioflow carriers taken directly from the denitrification reactor that operated under continuous-mode conditions (0.37 mM NO$_x$ h$^{-1}$ or 5.2 NO$_x$-N L$^{-1}$ h$^{-1}$). This result suggests that changes occurred in the microbial population of the batch biofilm cultures. qPCR assays were performed to determine the concentrations of *M. nitratireducenticrescens* JAM1 (*narG1*-targeted primers) and *H. nitrativorans* NL23 (*napA*-targeted primers) in the original and acclimated biofilm. In the original biofilm, strain JAM1 was at 2.5 ± 0.6 x 10$^4$ *narG1*-copies/ng biofilm DNA and *H. nitrativorans* NL23 at 6.2 ± 0.4 x 10$^5$ *napA*-copies/ng biofilm DNA. In the acclimated biofilm, the concentration of strain JAM1 increased by one order of magnitude (3.8 ± 0.9 x 10$^5$ *narG1*-copies/ng biofilm DNA), but decreased substantially for strain NL23 (2.0 ± 0.5 x 10$^2$ *napA*-copies/ng biofilm DNA). With the decrease of strain NL23 and the increase of strain JAM1 in the acclimated biofilm with a gain of denitrification performance, we hypothesized that a subpopulation of *M. nitratireducenticrescens* that can perform the complete denitrification pathway was enriched in the acclimated biofilm.

*Strain GP59 displaced H. nitrativorans NL23 and M. nitratireducenticrescens JAM1*

The acclimated biofilm from the 5$^{th}$ transfer cultures was dispersed and cultured onto a *Methylophaga* specific medium, and several isolates were screened for NO$_3^-$ and NO$_2^-$ consumption. One isolate, which we named strain GP59, was capable to consume NO$_3^-$ and NO$_2^-$, and cultures showed gas production, suggesting full denitrifying activities by this strain. The 16S rRNA gene sequence of strain GP59 has 100% identity to that of *M. nitratireducenticrescens* JAM1.
The concentration of strain GP59 was measured in the acclimated biofilm by qPCR to determine whether this strain is a major denitrifier. To distinguish strain GP59 from strain JAM1, primers targeted strain-specific sequences (tagH for strain JAM1 and nirK for strain GP59) were developed based on their respective genome (Table 1; see below). The GP59 cell concentrations went from undetected (below limit of PCR detection) in the original biofilm to high level in the acclimated biofilm at $2.2 \pm 0.35 \times 10^5$ nirK-copies/ng biofilm DNA. This is 2-3 orders of magnitude higher than that of strain JAM1 ($4.0 \pm 1.7 \times 10^2$ tagH-copies/ng biofilm DNA). These results confirm our hypothesis that a new *M. nitratireducenticrescens* strain with full denitrification capacity occurred in the acclimated biofilm, and displaced *H. nitrativorans* NL23, but also *M. nitratireducenticrescens* JAM1. Comparative analyses were then performed between strain GP59 and *M. nitratireducenticrescens* JAM1 at the physiological, genomic and transcriptomic levels.

**Physiological characterization of strain GP59**

Strain GP59 cultured under anoxic conditions (planktonic pure cultures) showed a 24-h lag before growth occurred (Fig. 2a). The growth yields reached about 1.2 OD$_{600\text{nm}}$ with 42.8 mM NO$_3^-$ exposure and did not increase significantly in cultures exposed with higher NO$_3^-$ concentrations. Strain JAM1 cultured under the same conditions showed no lag phase, and growth yields were 3- to 6-fold lower than those of the GP59 cultures (Fig. 2a). The maximum specific growth rate ($\mu_{\text{max}}$) and the half-saturation constants of NO$_3^-$ for growth (Ks) were calculated (Table 2). The $\mu_{\text{max}}$ for strain GP59 is higher (3.3 times) than that of strain JAM1. To assess the affinity of strain GP59 toward NO$_3^-$ for growth, the $\mu_{\text{max}}$/Ks ratio was calculated (38) (Table 2). This ratio (1.2 $\mu$M$^{-1}$ NO$_3^-$ h$^{-1}$) is not different than the one calculated for strain JAM1.
at 1.3 µM⁻¹ NO₃⁻ h⁻¹ (19). These results concur with the genome sequences (see below) with near 100% identity between the two strains in gene clusters encoding the two Nar systems and the NO₃⁻ transporters (NarK).

As observed with growth, a 24h-lag period was observed in the GP59 cultures before NO₃⁻ started to be consumed, whereas NO₃⁻ consumption appeared in the JAM1 cultures within 24 h (Fig. 2b). The NO₃⁻ reduction rates increased linearly with the increase of NO₃⁻ concentrations in the GP59 cultures (Fig. 3a). The specific NO₃⁻ reduction rates (rates normalized by the biomass) averaged around 1.5 to 2 mM NO₃⁻ h⁻¹ OD⁻¹ and showed no significant changes at any NO₃⁻ concentrations tested (Fig. 3b). In the JAM1 cultures, the NO₃⁻ reduction rates reached a plateau at 24 mM NO₃⁻ (Fig. 3a), and, as observed with the GP59 cultures, showed no significant changes in the specific NO₃⁻ reduction rates (averaged around 4 to 6 mM NO₃⁻ h⁻¹ OD⁻¹) (Fig. 3b). Interestingly, these specific rates are 2.4 to 5.4-fold higher than those of strain GP59, suggesting that the JAM1 cells have a higher dynamism of NO₃⁻ processing (e.g. NO₃⁻ intake and reduction) than the GP59 cells. Close examination of the respective genomes and transcriptomes (see below) did not reveal specific gene(s) that would explain these differences.

N₂O was detected in the GP59 anoxic cultures and reached maximum accumulation (ca. 0.5 µmol N₂O vial⁻¹ or 0.04% N-input) when NO₂⁻ peaked in the medium, and then decreased in concentration afterwards (Fig. 2b). This result concurs with the presence of a complete denitrification pathway in strain GP59 as revealed by its genome sequence (see below).

Strain GP59 was cultured under oxic conditions with different concentrations of NO₃⁻ (9.6 to 142.8 mM). Contrary to the JAM1 cultures that can reduce NO₃⁻ to NO₂⁻ under oxic conditions (19), no NO₃⁻ consumption was observed in the GP59 cultures at any tested concentrations. These latter results, coupled with the 24-h lag period before nitrate reduction
occurs in the GP59 anoxic cultures, suggest that strains GP59 and JAM1 regulate differently the expression of their denitrification genes.

Both strains cannot grow on methylamine and fructose. They showed similar growth profile related to NaCl concentrations (optimal growth at 1 to 5% NaCl; weaker growth at 8%; no growth at 0 and 10%), pH (optimal growth at pH 8; weaker growth at pH 7, 9 and 10; no growth at pH 6 and 11) and temperature (optimal growth at 30°C; variability in culture replicates at 22 and 34 °C; weak growth at 37°C). Sensibility to antibiotics by strain GP59 was also similar to those reported by Villeneuve et al. (15) for strain JAM1 (chloramphenicol, trimethoprim, ampicillin, tetracycline, streptomycin, gentamycin and kanamycin).

The genome of strain GP59 contains nirK

The genome of strain GP59 was sequenced by the PacBio technology. Although strain JAM1 was already sequenced by the pyrosequencing technology, it was resequenced by the PacBio technology to correct some discrepancies in the genome assembly (Supplemental Data1). Analysis of the GP59 genome revealed the presence of two prophages (named GPMu1 and GPMu2) integrated side by side in the same orientation and separated by 11102 nt (Supplemental Data2). The intercalated sequence has a GC content of 40.1%, which is lower than in the overall genome (44.8%). Gene arrangement of the prophages resembles to that of the *E. coli* phage Mu and of the *Haemophilus influenzae* FluMu prophage (39). Gene clusters encoding Mu-type prophage proteins are also present in the genome of two *Methylophaga* species such as *M. frappieri* JAM7 that was also isolated from the original biofilm of the denitrification system (12) (Supplemental Data2). No prophage was found in the JAM1 genome.
The 11102 nt intercalated sequence contains a gene encoding a NirK-type NO$_2^-$ reductase of 363 amino acid residues (Supplemental Data2). Upstream of this nirK, is a gene encoding the NO reductase activation protein NorD. The precise function of NorD is yet to be known, but the norD mutant in Paracoccus denitrificans showed no activity of the cd$_1$-type NO$_2^-$ reductase (NirS) and of the NO reductase (40). Available genomic sequences from other Methylophaga species and from metagenomic studies were screened for the presence of genes encoding NirK highly similar to the GP59 NirK. The closest affiliations of the GP59 NirK are with NirK encoded in three reconstituted Methylophaga sp. genomes from a metagenomic study of marine subsurface aquifer samples (NCBI Bioproject PRJNA391950 by Tully, Wheat, Glazer and Huber, University of Southern California) (Fig. 4). These three NirK have 84-85% similarity with the GP59 NirK sequence. In addition, norD is also present adjacent to nirK in these three Methylophaga sp. genomes.

The previous results suggest that there are Methylophaga strains in natural environments that carry denitrification modules. We further performed data mining to retrieve denitrification genes in available genomes of Methylophaga species or in reconstituted Methylophaga genomes from metagenomics studies. Data provided by the Bioproject PRJNA391950 revealed that the metagenome of the Methylophaga sp. isolate NORP53 (accession number NVVW00000000) contains two gene clusters with one (contig NVVW01000003) having the same gene arrangement of the Nar1 gene cluster in the JAM1 and GP59 genomes (narXL, NarK1K2GHJI), and another one (NVVW01000010) encoding a N$_2$O reductase (nosRZDFYL). At the amino acid sequence level, these gene clusters of isolate NORP53 are 67 to 90% similar with the corresponding gene products in the JAM1 and GP59 genomes. Finally, nirK sequences were found in the genome of M. frappieri JAM7 and of 34 Methylophaga sp. retrieved from
metagenomics studies (Supplemental Data3). Although *M. frappieri* JAM7 does not grow under anoxic conditions with NO$_3^-$, which concurs with the absence gene cluster encoding a dissimilatory NO$_3^-$ reductase, the occurrence of *nirK* may be associated with the detoxification mechanism. All these results suggest that the different denitrification modules are carried by some *Methylophaga* species present in natural environments.

Strain GP59 belongs to the species *Methylophaga nitratireducenti*crescens

Table 3 summaries the overall features of the GP59 and JAM1 genomes. Both genomes share 2790 coding sequences (CDS) and 11 riboswitches (Supplemental Data4). The 67 kbp chromosomic region containing the denitrification gene clusters *nar1, nar2, nor1, nor2* and *nos* (19) is 99.94% identical (43 substitutions, no gap) between the two strains. Based on the genome annotations, functions associated with CDS that are unique to strain JAM1 have in most cases an equivalent in strain GP59. In addition to CDS associated with the two prophages and the intercalated sequence, the GP59 genome has CDS associated with an integrated plasmid (position 724937 to 765476; Supplemental Data4). The GC skew between the two genomes is relatively similar (Fig. S3). Two-way ANI analysis showed 99.30% identity, and >80% conserved sequences between the two genomes, indicating that both strains belong to the same species (30, 31).

Both genomes have small tandem repeats with >10 repeats positioned at equivalent locations. These tandem repeats vary in number of repeats between the two strains (Table 3), and are located in CDS with no putative known function. A long tandem repeat (four 3095-nt repeats) were found in the GP59 genome (Table 3). Genes encoding an aspartate aminotransferase and a TonB-dependent siderophore receptor are among the three genes
repeated 4 times. The equivalent region in the JAM1 genome shows no repetition, with 93% identity with the 3095-nt repeat sequence with most of the divergence in the TonB-dependent siderophore receptor gene.

Strain GP59 contains two plasmids here named pGP32 and pGP34 (Table 3). These plasmids encode type IV secretion system as observed in many plasmids hosted by Gram negative bacteria (Supplemental Data5) (41). They also carry genes encoding for recombinases, nucleases, DNA methyltransferases, toxin/antitoxin, and carbon storage regulator. These two plasmids show no homology with the integrated plasmid in the GP59 genome. *M. frappieri* JAM7 also carries a 48.5 kb plasmid, which resembles in its structure to the IncPalpha plasmids (Supplemental Data5). It shares no homology with the pGP32 and pGP34 plasmids.

*M. nitratireducicrescens GP59 did not evolve from M. nitratireducicrescens JAM1*

Did strain GP59 evolve during the acclimation process of the biofilm from strain JAM1 that underwent series of chromosomic rearrangements and phage infections, or was strain GP59 present in the original biofilm and enriched during the acclimation? To try to answer this question, we analyzed the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) region that has been identified in the GP59 and JAM1 genomes. The CRISPR system is an adaptive immunity system that is present in most archaea and many bacteria and that acts against invading genetic elements, such as bacteriophages and plasmids. The chromosomic arrangement of CRISPR system includes CRISPR associated genes (*cas*), a leader sequence upstream of an array of short repeats interspersed with unique spacers that are almost identical to fragments of bacteriophage and plasmid genes. The transcription of the spacer array provides complementary DNA targeting, for instance, phage genome resulting in target DNA degradation (42-44).
Addition of new spacers in response, for instance, of bacteriophage infection occurs proximal to the leader sequence. Accordingly, CRISPR spacers provide a historical perspective of phage exposure, with spacers at the vicinity of the leader were relatively recently added, and those distal spacers likely originated from previous events (45).

The CRISPR region in GP59 and JAM1 genomes comprises 5 associated CRISPR genes (Fig. 5a). Except for the spacer sequences, the nucleotide sequence of the CRISPR genes and the repeat sequences are identical in both genomes (Fig. 5a, c). There are 105 and 115 spacers in the JAM1 and GP59 CRISPR regions, respectively (Fig. 5b). In the JAM1 CRISPR region, 14 spacer sequences were found twice. The GP59 and JAM1 CRISPR regions share 29 spacer sequences (Fig. 5b); all of them are distal to the leader sequence (the most ancient acquired sequences (45)). No spacer sequence was found twice in the GP59 CRISPR region. As both CRISPR regions do not share sequence between their proximal spacers, strains GP59 and JAM1 underwent a different history of phage infections from a common ancestry. This strongly suggests that strain GP59 did not originate from strain JAM1 that underwent series of chromosomic rearrangements and phage infections in a short period of time and in a closed environment, but rather was present in the original biofilm and was enriched during the acclimation phase of the biofilm.

The GP59 and JAM1 CRISPR regions were also compared to other available *Methylophaga* genomes. *M. frappieri* JAM7 contains 68 repeat units. The repeat sequence has one-nucleotide substitution with the GP59/JAM1 one (Fig 5c). None of its spacer sequences were found in the GP59 and JAM1 CRISPR regions. In the *M. aminisulfidivorans* MP, *M. sulfidovorans* DSM 11578, *M. lonarensis* MPL genomes, 59, 6 and 63 spacer sequences were found, respectively; none were repeated twice or were found in other *Methylophaga* species.
The CRISPR genes and gene arrangement in *M. frappieri* JAM7 are typical of the subtype I-F CRISPR-Cas system, whereas those of *M. aminisulfidivorans* MP, *M. sulfidovorans* DSM 11578, *M. lonarensis* MPL are related to the subtype I-E one (Fig 5a) (44). In strains GP59 and JAM1, the CRISPR genes and gene arrangement cannot be associated to a defined subtype. However, the gene arrangement and the deduced amino acid sequences of *cas3* and of the two downstream CDS are highly similar to CRISPR genes found in several bacterial genomes such as *Legionella brunensis* (GenBank accession number LNXV01000004) and *Moraxella atlantae* (GenBank accession number LZMZ01000013).

Transcriptome analyses

The GP59 and JAM1 genomes share more than 85% of their genes. However, gene variations, the presence of strain-specific genes, and the occurrence of plasmids in strain GP59 could affect the expression pattern of these common genes. To assess whether these genes were expressed differently between strains GP59 and JAM1, the transcriptome of both strains was derived from anoxic cultures at the end of the exponential growth when NO$_3^-$ was nearly completely reduced in both cultures. Among the 2857 common genes (CDS and ncRNA), more than 85% of those in strain GP59 showed no significant difference in the transcription levels (<2-fold difference in TPM) (Supplemental Data4) with their counterparts in strain JAM1. The vast majority of genes involved in the nitrogen pathway were expressed in equivalent levels in both strains (Table 4). The fact that the cultures were sampled for RNA extraction when both strains were at an equivalent physiological phase (anoxic conditions with nitrate almost consumed) could explain this high proportion of similar gene expression between the two strains. Among genes with significant differences are those encoding the synthesis of the osmoprotectant ectoine,
which were 3 to 6-fold higher in the JAM1 cultures than in the GP59 cultures (Fig. S4). This suggests that strain JAM1 could be more resilient to changes in salinity.

The most expressed genes were sorted out from both transcriptomes to assess which specific metabolic pathways are the most solicited under denitrifying conditions. High levels of gene expression were found in both strains for genes involved in growth, energy and respiration (e.g. ribosomal proteins, ATP synthase, pseudoazurin), and the C-1 metabolic carbon such as methanol dehydrogenase, ribulose monophosphate enzymes (D-arabino-3-hexulose 6-phosphate formaldehyde-lyase, 6-phospho-3-hexuloisomerase), formaldehyde activating enzymes. (Supplemental Data6, Fig. S4). One of the highest expressed genes in both genomes is located in the chromosomal region comprising three adjacent genes encoding similar proteins (>88% similarity). Each of these genes has a cyclic di-GMP riboswitch in their respective upstream sequence. The first of the three genes showed high level of TPM in the riboswitch and in the coding sequence (Table 5). The other two genes and their respective riboswitch showed much lower expression levels. The deduced amino acid sequence of these three similar proteins contains a carbohydrate-binding module 6 that is found in carbohydrate-active enzymes such as glycoside hydrolases (46), and three transmembrane domains. Upstream and downstream of these three genes are genes encoding proteins involved in polysaccharide export and in the capsular polysaccharide biosynthesis. These proteins may participate in this synthesis. Finally, the pGP32 plasmid accounted for 23.1% of total TPM in the GP59 transcriptomes, whereas it was 0.7% for the pGP34 plasmid (Supplemental Data6). In the pGP32 plasmid, nine genes accounted for 20% of total TPM, among which are genes involved in restriction-modification systems (Supplemental Data6).
Conclusions

Culturing the denitrifying biofilm under batch conditions has favored the enrichment of a new denitrifying subpopulation, representing by *M. nitratireducenticrescens* GP59 that displaced *H. nitrativorans* NL23 and *M. nitratireducenticrescens* JAM1. In addition of complete denitrification by strain GP59, two important differences were observed between the two strains. Strain GP59 cannot reduce NO$_3^-$ under oxic conditions as does strain JAM1, and a 24h lag period was observed in GP59 anoxic cultures before growth and nitrate reduction occurs. These results suggest differences in regulation of the denitrification genes between the two strains. The genome of strain GP59 showed insertions of two Mu-type prophages that probably brought a *nirK* gene, the missing denitrification gene in strain JAM1. Both genomes share more than 85% of their genes, and these genes were expressed at similar level under anoxic conditions at the end of the exponential growth. Finally, analysis of the CRISPR region suggests that strain GP59 did not originate from strain JAM1, but was present in the original biofilm and was enriched during the acclimation process.

These comparative analyses between strains GP59 and JAM1 unveiled that *M. nitratireducenticrescens* species enriched in the biofilm of the bioreactor encompasses a mosaic population structure. Although plasticity of the genomic landscape of bacterial species resulting to gene rearrangements, point mutations and lateral gene acquisition was observed in nature (47, 48), the observation of *M. nitratireducenticrescens* GP59 reported here shows functional consequences of population variation on process rate and system performances. The reason why strain GP59 did occur in the acclimated biofilm but not in the Biodome denitrification system remains obscure and is under investigation. This could be related to the batch-mode conditions where the transient accumulation of NO$_2^-$ (up to 10 mM; Fig. 1) could have adverse the growth
of *H. nitrativorans* NL23, favoring the enrichment of *M. nitratireducenticrescens* GP59. In the denitrification reactor at the Biodome, such accumulation of NO$_2^-$ was not encountered probably due the continuous-operating mode or the lower denitrification rates that prevailed.

Our study broadens the ecology of *M. nitratireducenticrescens* with the occurrence of a microbial seedbank suited for anoxic conditions. So far among the genus *Methylophaga*, *M. nitratireducenticrescens* is the only reported species that can grow under anoxic conditions. Based on metagenomics studies, however, where denitrification genes were found in *Methylophaga*-reconstituted genomes, this anoxic metabolism within this genus is more widespread than previously thought. Methanol is generated by anthropogenic activities or by biogenic metabolisms such as plant growth and decomposition, and ocean ecosystems are thought to participate actively in its production. Furthermore, oceans are known to be a sink of atmospheric methanol (49, 50). As they are ubiquitous in marine environments, *Methylophaga* spp. are therefore important actors in the methanol turnover on Earth, either under oxic environments (surface ocean) or, based on our studies with *M. nitratireducenticrescens*, under anoxic marine environments (deep ocean).

**Acknowledgments**

We thank Karla Vasquez for her technical assistance.

**References**

1. Van Spanning RJM, Delgado MJ, Richardson DJ. 2005. The nitrogen cycle: denitrification and its relationship to N$_2$ fixation, p. 277-342. *In* Werner DNewton WE
2. **Einsle O, Kroneck PM.** 2004. Structural basis of denitrification. Biol Chem 385:875-883.

3. **Zumft WG.** 1997. Cell biology and molecular basis of denitrification. Microbiol Mol Biol Rev 61:533-616.

4. **Kaspar HF.** 1982. Nitrite reduction to nitrous-oxide by *Propionibacteria* - Detoxication Mechanism. Arch Microbiol 133:126-130.

5. **Poole R.** 2005. Nitric oxide and nitrosative stress tolerance in bacteria. Biochemical Soc T 33:176-180.

6. **Schreiber F, Wunderlin P, Udert KM, Wells GF.** 2012. Nitric oxide and nitrous oxide turnover in natural and engineered microbial communities: biological pathways, chemical reactions, and novel technologies. Front Microbiol 3: 372.

7. **Simon J, Klotz MG.** 2013. Diversity and evolution of bioenergetic systems involved in microbial nitrogen compound transformations. Bba-Bioenergetics 1827:114-135.

8. **Labbé N, Juteau P, Parent S, Villemur R.** 2003. Bacterial diversity in a marine methanol-fed denitrification reactor at the Montreal Biodome, Canada. Microbial Ecol 46:12-21.

9. **Laurin V, Labbé N, Parent S, Juteau P, Villemur R.** 2008. Microeukaryote diversity in a marine methanol-fed fluidized denitrification system. Microbial Ecol 56:637-648.

10. **Labbé N, Laurin V, Juteau P, Parent S, Villemur R.** 2007. Microbiological community structure of the biofilm of a methanol-fed, marine denitrification system, and identification of the methanol-utilizing microorganisms. Microbial Ecol 53:621-630.
11. Rissanen AJ, Ojala A, Dernjatin M, Jaakkola J, Tiirola M. 2016. Methylophaga and Hyphomicrobium can be used as target genera in monitoring saline water methanol-utilizing denitrification. J Ind Microbiol Biotechnol 43:1647-1657.

12. Auclair J, Lepine F, Parent S, Villemur R. 2010. Dissimilatory reduction of nitrate in seawater by a Methylophaga strain containing two highly divergent narG sequences. Isme J 4:1302-1313.

13. Auclair J, Parent S, Villemur R. 2012. Functional diversity in the denitrifying biofilm of the methanol-fed marine denitrification system at the Montreal Biodome. Microbial Ecol 63:726–735.

14. Martineau C, Villeneuve C, Mauffrey F, Villemur R. 2013. Hyphomicrobium nitrativorans sp. nov., isolated from the biofilm of a methanol-fed denitrification system treating seawater at the Montreal Biodome. Int J Syst Evol Microbiol 63:3777-3781.

15. Villeneuve C, Martineau C, Mauffrey F, Villemur R. 2013. Methylophaga nitratireducenticrescens sp. nov. and Methylophaga frappieri sp. nov., isolated from the biofilm of the methanol-fed denitrification system treating the seawater at the Montreal Biodome. Int J Syst Evol Microbiol 63:2216-2222.

16. Martineau C, Mauffrey F, Villemur R. 2015. Comparative analysis of denitrifying activities of Hyphomicrobium nitrativorans, Hyphomicrobium denitrificans, and Hyphomicrobium zavarzinii. Appl Environ Microbiol 81:5003-5014.

17. Martineau C, Villeneuve C, Mauffrey F, Villemur R. 2013. Complete genome sequence of Hyphomicrobium nitrativorans strain NL23, a denitrifying bacterium isolated from biofilm of a methanol-fed denitrification system treating seawater at the Montreal Biodome. Genome Announc 2: e01165-13.
Mauffrey F, Cucaita A, Constant P, Villemur R. 2017. Denitrifying metabolism of the methylotrophic marine bacterium Methylophaga nitratireducenticrescens strain JAM1. PeerJ 5: 4098.

Mauffrey F, Martineau C, Villemur R. 2015. Importance of the two dissimilatory (Nar) nitrate reductases in the growth and nitrate reduction of the methylotrophic marine bacterium Methylophaga nitratireducenticrescens JAM1. Front Microbiol 6:1475.

Villeneuve C, Martineau C, Mauffrey F, Villemur R. 2012. Complete genome sequences of Methylophaga sp. strain JAM1 and Methylophaga sp. strain JAM7. J Bacteriol 194:4126-4127.

Boden R. 2012. Emended description of the genus Methylophaga Janvier et al. 1985. Int J Syst Evol Microbiol 62:1644-1646.

Labelle M-A. 2004. Optimisation d’un bioprocédé de dénitrification d’un système aquacole marin en circuit fermé au Biodôme de Montréal. Master thesis. École Polytechnique de Montréal.

Labbé N, Parent S, Villemur R. 2003. Addition of trace metals increases denitrification rate in closed marine systems. Water Res 37:914-920.

Laurin V, Labbe V, Juteau P, Parent S, Villemur R. 2006. Long-term storage conditions for carriers with denitrifying biomass of the fluidized, methanol-fed denitrification reactor of the Montreal Biodome, and the impact on denitrifying activity and bacterial population. Water Res 40:1836-1840.

Sambrook JF, Russell DW (ed.). 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press.
26. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563-569.

27. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid annotations using subsystems technology. Bmc Genomics 9: 75.

28. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia FF, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206-D214.

29. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394-1403.

30. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81-91.

31. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints 4:e1900v1.

32. Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573-580.
33. Papadopoulos JS, Agarwala R. 2007. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics 23:1073-1079.

34. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 30:2725-2729.

35. Le SQ, Gascuel O. 2008. An improved general amino acid replacement matrix. Mol Biol Evol 25:1307-1320.

36. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. 2011. Differential expression in RNA-seq: A matter of depth. Genome Res 21:2213-2223.

37. Team RC. 2015. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

38. Healey FP. 1980. Slope of the monod equation as an indicator of advantage in nutrient competition. Microbial Ecol 5:281-286.

39. Morgan GJ, Hatfull GF, Casjens S, Hendrix RW. 2002. Bacteriophage Mu genome sequence: Analysis and comparison with Mu-like prophages in Haemophilus, Neisseria and Deinococcus. J Mol Biol 317:337-359.

40. DeBoer APN, VanderOost J, Reijnders WNM, Westerhoff HV, Stouthamer AH, VanSpanning RJM. 1996. Mutational analysis of the nor gene cluster which encodes nitric-oxide reductase from Paracoccus denitrificans. Eur J Biochem 242:592-600.

41. Christie PJ. 2001. Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. Mol Microbiol 40:294-305.

42. Barrangou R, Marraffini LA. 2014. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. Mol Cell 54:234-244.
Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadan AH, Moineau S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468:67-+. Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, Moineau S, Mojica FJM, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. 2011. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9:467-477. Horvath P, Romero DA, Coute-Monvoisin AC, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R. 2008. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. J Bacteriol 190:1401-1412. Henshaw JL, Bolam DN, Pires VMR, Czjzek M, Henrissat B, Ferreira LMA, Fontes CMGA, Gilbert HJ. 2004. The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. J Biol Chem 279:21552-21559. Allen EE, Tyson GW, Whitaker RJ, Detter JC, Richardson PM, Banfield JF. 2007. Genome dynamics in a natural archaeal population. Proc Natl Acad Sci USA 104:1883-1888. Bendall ML, Stevens SLR, Chan LK, Malfatti S, Schwientek P, Tremblay J, Schackwitz W, Martin J, Pati A, Bushnell B, Froula J, Kang DW, Tringe SG, Bertilsson S, Moran MA, Shade A, Newton RJ, McMahon KD, Malmstrom RR. 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. Isme J 10:1589-1601. Galbally IE, Kirstine W. 2002. The production of methanol by flowering plants and the global cycle of methanol. J Atmos Chem 43:195-229.
Millet DB, Jacob DJ, Custer TG, de Gouw JA, Goldstein AH, Karl T, Singh HB, Sive BC, Talbot RW, Warneke C, Williams J. 2008. New constraints on terrestrial and oceanic sources of atmospheric methanol. Atmos Chem Phys 8:6887-6905.
Figure 1 (on next page)

NO$_3^-$ and NO$_2^-$ reduction by the acclimated biofilm

The biofilm was batch-cultured under anoxic conditions. After the 5$^{th}$ transfer, the NO$_3^-$ and NO$_2^-$ concentrations were measured over different time intervals. NO$_x$: combined measurements of NO$_3^-$ and NO$_2^-$. Each point is the average with standard deviation of triplicate cultures.
Figure 2 (on next page)

Growth, and NO$_3^-$, NO$_2^-$ and N$_2$O dynamics by *Methylophaga nitratireducenticrescens* GP59 and JAM1

Panel a. Growth. Strain GP59 and strain JAM1 were cultured under anoxic conditions with 9.6 or 42.8 mM NO$_3^-$. Panel b: NO$_3^-$, NO$_2^-$ and N$_2$O dynamics. Strain GP59 and strain JAM1 were cultured under anoxic conditions with 21.4 mM NO$_3^-$, and NO$_3^-$, NO$_2^-$ and N$_2$O were measured over different time intervals. Each point is the average with standard deviation of triplicate cultures for Panel a, and of duplicate cultures for Panel b.
**Graph a**

- **OD** at 600 nm over time (h)
- Curves for different concentrations and strains:
  - 9.6 mM (GP59)
  - 9.6 mM (JAM1)
  - 42.8 mM (GP59)
  - 42.8 mM (JAM1)

**Graph b**

- **N\textsubscript{2}O** (µmol vial\textsuperscript{-1}) over time (h)
- Concentrations for different species and strains:
  - NO\textsubscript{3}\textsuperscript{-} (GP59)
  - NO\textsubscript{3}\textsuperscript{-} (JAM1)
  - NO\textsubscript{2}\textsuperscript{-} (GP59)
  - NO\textsubscript{2}\textsuperscript{-} (JAM1)
  - N\textsubscript{2}O (GP59)
  - N\textsubscript{2}O (JAM1)
Specific NO$_3^-$ reduction rates

Panel a: NO$_3^-$ reduction rates of strains GP59 and JAM1. Each point is the average with standard deviation of triplicate anoxic cultures. Panel b: Specific NO$_3^-$ reduction rates. These rates were calculated with the NO$_3^-$ reduction rates by the generated culture biomass (OD$_{600nm}$) at the end of the exponential phase.
Figure 1: Nitrate uptake by GP59 and JAM1.

(a) NO$_3^-$ reduction rate (mM h$^{-1}$) as a function of nitrate concentration (mM).

(b) Specific NO$_3^-$ reduction rate (mM h$^{-1}$ OD$^{-1}$) as a function of nitrate concentration (mM).
An unrooted phylogenetic tree demonstrating the evolutionary relationship of NirK deduced amino acid sequences is illustrated. The evolutionary history was inferred by using the Maximum Likelihood and the tree with the highest log likelihood (-7598.4857) is shown. The percentage (values above 70%) of trees in which the associated taxa clustered together is shown next to the branches (bootstrap analysis with 1000 replicates). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The GenBank accession number is provided after each bacterial name. *: The *Cycloclasticus* sp. NORP92, Gammaproteobacterium NORP60, and the *Methylophaga* sp. NORP93, NORP33 and NORP65 are from the NCBI bioproject PRJNA391950 by Tully, Wheat, Glazer and Huber, University of Southern California.
**Figure 5** (on next page)

CRISPR chromosomic arrangement in *Methylophaga* genomes

Panel a. Gene arrangement. Sequences were retrieved from GenBank. Accession numbers: *M. nitratireducenticrescens* JAM1: CP003390 (Q7A_2613 to Q7a_2609) and GP59: CP021973 (CDW43_11545 to CDW43_11525); *M. frappieri* JAM7 CP003380 (Q7C_1045 to Q7C_1040); *M. aminisulfidivorans* MP: NZ_AF1G01000001 (MAMP_RS08465 to MAMP_RS08425), *M. sulfidovorans* DSM 11578: FOSH01000003 (SAMN04488079_1032 to SAMN04488079_10310); *M. lonarensis* MPL, APHR01000054 (MPL1_10022 to MPL1_10062). Nomenclature is based on Makarova et al. (42). HP: Hypothetical protein. Numbers at the right are the number of repeat units with the repeat/spacer nucleotide length in parentheses. Panel b. Spacer/repeat unit arrangement. Blue: The 29 common units. Green: The 14 units repeated twice in strain JAM1. Panel c. CRISPR repeat sequences in different *Methylophaga* species.
M. aminisulfidivorans
M. sulfidovorans (Subtype I-E)
M. lonarensis

csy3
cas3

cse1

cse2

cas7

cas5

cas6

cas1

cas2

HP

54 (29/32)

6 (29/32)

63 (29/32)

M. frappieri JAM7 (Subtype I-F)

JAM1: 105 (28/32)

GP59: 115 (28/32)

68 (28/32)

M. nitratireducenticrescens

M. nitratireducenticrescens JAM1/GP59

M. frappieri JAM7

M. aminisulfidivorans

M. sulfidovorans

M. lonarensis

M. nitratireducenticrescens

M. frappieri

GTTCACTGCCACAGGCAGCTTAGAAA

GTTCACTGCCACAGGCAGCTTAGAAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA

GTGCCGCGACAGCCAGCTTTAGAA
Table 1 (on next page)

PCR primers used for qPCR assays
Table 1: PCR primers used for qPCR assays

| Gene target | Sequence (5'-3') | Annealing temperature (°C) | Fragment length (nt) | Comment |
|-------------|------------------|-----------------------------|----------------------|---------|
| narG1 (strain JAM1/GP59; locus Q7A_446). Nitrate reductase subunit alpha (Nar1 system) | narG1: AGCCCACATCGTATCAAGCA, narG1: CCACGCACCAGCAGTATATTG | 61 | 149 | qPCR |
| TagH (strain JAM1, locus tag: Q7A_1110). Teichoic acid export ATP-binding protein* | TagH Forward: CCGTCATTTCGCTTCAAGAT, TagH Reverse: TCATGGCTTTTCAGCCTTT | 55 | 711 | Standard |
| qTagH Forward: GTTGCAAGGCTATAGTCCGAGT, qTagH Reverse: TGGTACGCAATCCCAGATGAATA | 55 | 119 | qPCR |
| nirK (strain GP59, locus tag: CDW43_15165). Copper containing nitrite reductase | NirK Forward: CGTTCAATACATGGGGTAAAGG, NirK Reverse: TGGGGCACAGTGATAACAA | 55 | 1134 | Standard |
| qNirK Forward: AAGTCGGTAAAGTAGCCGTTGA, qNirK Reverse: TCTCCATTCATCTCAGTGAATA | 55 | 138 | qPCR |
| napA (strain NL23, locus tag W911_13875) periplasmic nitrate reductase | qnapA Forward: AGGACGGGCGGATCAATT, qnapA Reverse: CGGATATGCATCGGACACGA | 61 | 131 | qPCR |

*The function was deduced by RAST annotations.
GenBank accession number of the genome of strain JAM1: CP003390.3, strain GP59: CP021973.1, and strain NL23: CP006912.1.
**Table 2** (on next page)

Kinetics of growth under anoxic conditions
Table 2. Kinetics of growth under anoxic conditions

|                | Strain GP59         | Strain JAM1*       |
|----------------|--------------------|--------------------|
| µmax (h⁻¹)     | 0.0380 (0.0027)    | 0.0116 (0.0008)    |
| Ks (mM)        | 30.7 (5.8)         | 9.2 (1.9)          |
| µmax/Ks (µM⁻¹ h⁻¹) | 1.2 (0.2)       | 1.3 (0.3)          |

µmax, maximum growth rates. Ks, half-saturation constants of NO₃⁻ for growth. Values between parentheses are standard deviation of triplicates. *from Mauffrey et al. (19).
Table 3 (on next page)

Genomic features of *M. nitratireducenticrescens* strains JAM1 and GP59
Table 3. Genomic features of *M. nitratireducenticrescens* strains JAM1 and GP59

|                              | JAM1       | GP59       |
|------------------------------|------------|------------|
| **Genome**                   |            |            |
| Length (nt)                  | 3137100    | 3238484    |
| GC content (%)               | 44.75      | 44.86      |
| CDS                         | 3027       | 3187$^s$   |
| Common CDS                   | 2790       | 2790       |
| tRNA                         | 44         | 44         |
| rRNA (5S-16S-23S)            | 9          | 9          |
| ncRNA                        | 4          | 4          |
| (TmRNA, SRP-RNA, RNase P, 6S RNA) |          |            |
| Riboswitch                   | 11         | 11         |
| **Small Tandem Repeats**     |            |            |
| TCAGYCA                      | 16 (785393)| 12 (710616)|
| CTTCGG                       | 55 (2294123)| 19 (2287110)|
| GGYTCT                       | 39 (2660735)| 37 (2669276)|
| **Large tandem repeat**      |            |            |
| 3095-nt repeats              | 1 (1940231; 93%) | 4 (1904086; 100%) |
| **Plasmids**                 | None       | pGP32      | pGP34      |
| Length (nt)                  | 32421      | 33560      |
| GC content (%)               | 44.29      | 44.13      |
| CDS                          | 51         | 52         |

$^a$Based on RAST annotations

$^s$The possible CDS in the ambiguous region of GPMu1 were not taking into account.

Common CDS: see supplemental Data4.

*Number of copies. Genome location under parenthesis. Only repeats with more than 10 copies are presented.

$^f$Number of copies. Genome location under parenthesis. % identity between the 3095-nt long repeats (see text).
**Table 4 (on next page)**

Relative expression of selected genes under anoxic culture conditions
Table 4. Transcript levels of genes involved in the nitrogen pathway in *Methylophaga nitratireducens* JAM1 and GP59

| Locus Tag       | Gene/function                        | Expression ratio JAM1/GP59 | **mxal=100** GP59 | **mxal=100** JAM1 |
|-----------------|--------------------------------------|----------------------------|-------------------|-------------------|
| Nitric oxide reductase 1 |                                      |                            |                   |                   |
| CDW43_RS01925-40 | norCBDQ                             | 0.93                       | 11.7              | 12.4              |
| CDW43_RS01945   | norR                                | 1.42                       | 4.9               | 6.8               |
| CDW43_RS01950   | norE                                | 0.84                       | 7.8               | 6.5               |
| Nitrate reductase 1 (with transporters and regulators) |                                      |                            |                   |                   |
| CDW43_RS01975-80 | narXL                               | 2.07                       | 3.8               | 7.9               |
| CDW43_RS01985-2010 | narK1K2narGHJI                | 0.66                       | 17.5              | 9.8               |
| Nitrous oxide reductase |                                      |                            |                   |                   |
| CDW43_RS02060-85 | nosRZDFYL                           | 1.51                       | 5.5               | 9.0               |
| Nitrate reductase 2 (with transporter) |                                      |                            |                   |                   |
| CDW43_RS02155   | narK                                | 0.75                       | 1.9               | 1.4               |
| CDW43_RS02165-80 | narGHJI                             | 0.83                       | 1.9               | 1.5               |
| Nitric oxide reductase 2 |                                      |                            |                   |                   |
| CDW43_RS02185-2200 | norCBQD                           | 0.83                       | 2.1               | 1.7               |
| Copper containing nitrite reductase |                                      |                            |                   |                   |
| CDW43_RS15160   | nirK                                | na                         | 1.0               | na                |
| Nitrate assimilatory pathway |                                      |                            |                   |                   |
| CDW43_RS11575   | Nitrate/nitrite transporter         | 1.17                       | 0.2               | 0.2               |
| CDW43_RS11585   | Assimilatory nitrate reductase large sub. | 1.17                       | 0.3               | 0.4               |
| CDW43_RS11590   | Nitrite reductase [NAD(P)H] small sub. | 0.81                       | 0.2               | 0.2               |
| CDW43_RS11595   | Nitrite reductase [NAD(P)H] large sub. | 0.84                       | 0.4               | 0.3               |
| Other genes involved in nitrogen pathway |                                      |                            |                   |                   |
| CDW43_RS01710   | fnr transcriptional regulator       | 1.41                       | 3.9               | 5.5               |
| CDW43_RS00310   | nmrS protein involved in response to NO | 2.71                       | 4.4               | 12.2              |
| CDW43_RS01960   | nmrS                                | 0.99                       | 7.4               | 7.3               |
| CDW43_RS08605   | nmrS                                | 1.61                       | 3.8               | 6.4               |
| CDW43_RS00315   | nsrcR transcriptional regulator     | 1.90                       | 17.4              | 33.8              |
| CDW43_RS01825   | nsrcR                              | 0.86                       | 2.3               | 2.0               |
| CDW43_RS00320   | dur/NorA/yftE NO-dependent regulator | 1.98                       | 25.0              | 48.7              |
| CDW43_RS01015   | Ammonium transporter                | 1.01                       | 1.0               | 1.0               |
| CDW43_RS05605   | Ammonium transporter                | 0.48                       | 0.9               | 0.5               |
| CDW43_RS01835   | Nitric oxide dioxygenase            | 0.79                       | 2.7               | 2.2               |
| CDW43_RS09635   | Nitric oxide dioxygenase            | 0.65                       | 8.0               | 5.0               |
| CDW43_RS01010   | Nitrogen regulatory protein P-II    | 0.85                       | 1.2               | 1.0               |
| CDW43_RS06615   | Nitrogen regulatory protein P-II    | 0.80                       | 7.0               | 5.5               |
| CDW43_RS13530   | Nitrogen regulatory protein P-II    | 0.74                       | 1.2               | 0.8               |
| CDW43_RS01095   | Nitrogen regulation protein NtrY    | 1.30                       | 1.4               | 1.9               |
| CDW43_RS01100   | Nitrogen regulation protein NtrX    | 1.35                       | 3.6               | 4.9               |
| CDW43_RS06315   | Nitrogen regulation protein NtrB    | 0.60                       | 1.1               | 0.6               |
| CDW43_RS06320   | Nitrogen regulation protein NR(I)   | 1.12                       | 0.7               | 0.8               |
| CDW43_RS15625   | PTS IIA-like nitrogen-regulator PtsN | 0.95                       | 12.1              | 11.1              |

*Ratio of JAM1 transcripts per million (TPM) by strain GP59 TPM.
****: Transcript levels relative to the expression of the methanol dehydrogenase small unit (*mxaI*) set to 100 (see supplemental doc6). For polycistronique operons, the values are the average values of individual genes.
na: not applicable
Locus tag refers to the genome sequence of strain GP59 (GenBank accession number CP021973). Gene/function is based on RAST annotations, see supplemental doc4.
Table 5 (on next page)

Relative expression of selected genes under anoxic culture conditions
Table 5: Relative expression of selected genes under anoxic culture conditions

| Locus tag       | Function                                                                 | Strand | GP59 | JAM1 |
|-----------------|--------------------------------------------------------------------------|--------|------|------|
| CDW43_06820     | Mannose-1-phosphate guanylyltransferase                                   | +      | 2.3  | 3.5  |
| CDW43_06825     | Capsular polysaccharide synthesis enzyme CpsA, sugar transferase,         | +      | 3.3  | 5.2  |
|                 | undecaprenyl-phosphate glucose phosphotransferase                        |        |      |      |
| CDW43_06835     | Cyclic di-GMP riboswitch class I                                         |        | 2011 | 1561 |
| CDW43_06835     | Hypothetical protein, carbohydrate-binding module 6                      | +      | 936  | 777  |
| CDW43_06840     | Cyclic di-GMP riboswitch class I                                         |        | 9.1  | 13   |
| CDW43_06840     | Hypothetical protein, carbohydrate-binding module 6                      | +      | 7.3  | 7.6  |
| CDW43_06845     | Cyclic di-GMP riboswitch class I                                         |        | 6.5  | 2.9  |
| CDW43_06845     | Hypothetical protein, carbohydrate-binding module 6                      | +      | 30   | 7.1  |
| CDW43_06850     | Periplasmic protein involved in polysaccharide export                    | +      | 6.8  | 4.6  |
| CDW43_06855     | Hypothetical protein                                                      | +      | 4.7  | 3.6  |
| CDW43_06860     | Exopolysaccharide transport protein, putative                            | +      | 2.6  | 2.0  |

Gene expression is relative to *mxaI* set to 100 (2602 TPM for strain GP59 and 4060 TPM for strain JAM1). Locus tag: from GenBank accession number CP021973. Functions are based on RAST annotations.