

**SEQUENCE DIVERSITY AND EXPRESSION OF NOVEL BACTERIAL
NITROUS OXIDE REDUCTASE (*nosZ*) GENES IN TROPICAL
ENVIRONMENTS**

By
Gina Paola Rodríguez-Castaño

A thesis submitted in partial fulfillment of the requirements for the degree of
MASTER OF SCIENCE
in
BIOLOGY
UNIVERSITY OF PUERTO RICO
MAYAGÜEZ CAMPUS

June, 2008

Approved by:

Frank E.Loeffler, Ph.D.
Member, Graduate Committee

Date

Rafael R. Montalvo-Rodríguez, Ph.D.
Member, Graduate Committee

Date

Carlos A. Muñoz, Ph.D.
Member, Graduate Committee

Date

Arturo A. Massol-Deyá, Ph.D.
President, Graduate Committee

Date

Jorge E. Corredor, Ph.D.
Representative of Graduate Studies

Date

Lucy Bunkley-Williams, Ph.D.
Chairperson of the Department

Date

ABSTRACT

Nitrous oxide (N_2O) is produced in different microbial processes including denitrification, nitrification and dissimilatory nitrate reduction to ammonia (DNRA) (Kelso et al. 1999). It has been suggested that nitrification is the main source of N_2O . Understanding the mechanisms that control the flux of N_2O is crucial to predict and manage emissions of this powerful greenhouse gas. Amplification of genes (*nosZ*) coding for nitrous oxide reductases (N_2ORs) from denitrifiers and N_2 - fixers has been obtained by PCR methods. In this thesis, we refer to these sequences as “traditional” *nosZ* sequences; on the other hand, the *nosZ* genes from microaerophilic *Anaeromyxobacter* spp. and *Magnetospirillum* spp., and other obligate anaerobic microorganisms, such as *Wolinella* spp., *Desulfitobacterium* spp and *Dechloromonas* spp. have not been well studied. Their primary N_2OR sequences diverge from the traditional ones and therefore, different primer sets must be developed to better understand their diversity and distribution in nature. This study developed oligonucleotides for amplifying a broader range of *nosZ* genes to assess their diversity in soil and bioreactors by culture-independent techniques. *nosZ* sequences obtained from environmental samples were different from traditional *nosZ* sequences. None of the clone sequences shared more than 62% amino acid similarity with traditional NosZ. Additionally, this analysis revealed the presence of conserved histidine residues essential for function of a mature N_2OR protein. Through a phylogenetic analysis using Neighbor-Joining, Maximum Parsimony, Maximum Likelihood, and Bayesian Inference methods, nine clades of NosZ variants were identified. Neither of the clone sequences fell into the traditional NosZ phylogenetic

clades, but grouped with *Anaeromyxobacter* spp., *Magnetospirillum* spp., *Desulfitobacterium hafniense*, and *Dechloromonas aromatica*. The detection of *nosZ* genes was achieved by ISRT-PCR/FISH using an internal fluorescently-labeled NosZ943 probe. This constitutes the first published report of probing *nosZ* amplicons inside of active microbial cells using an optimized *In Situ* Reverse Transcriptase-PCR/Fluorescent *In Situ* Hybridization (ISRT-PCR/FISH) protocol. Our results show that a clone carrying the partial *Anaeromyxobacter nosZ* gene emits a strong fluorescent signal when detected with NosZ943 probe, but not with Nos1527 probe; while a clone carrying the partial *Pseudomonas* gene did not emit any fluorescent signal with the NosZ943 probe. ISRT-PCR was further applied to natural samples from an anoxic bioreactor. Approximately 4% of total cell counts were expressing the novel *nosZ* genes. We demonstrate the existence of many variants of *nosZ* gene that are not yet represented by cultured organisms. These variants could represent a high functional diversity for reducing N₂O in the environment. Therefore, the diversity of *nosZ* genes in nature and their contribution to the N₂O budget warrants further exploration.

RESUMEN

El óxido nitroso (N_2O) es producido en diferentes procesos microbianos incluyendo desnitrificación, nitrificación y reducción desasimilativa del nitrato a amonio (DNRA) (Kelso et al. 1999). La desnitrificación se ha sugerido es la fuente principal de N_2O (Wolf and Russow, 2000). Entender los mecanismos que controlan el flujo de N_2O es crucial para predecir y tratar las emisiones de este poderoso gas de invernadero. La amplificación de los genes (*nosZ*) que codifican para las óxido nitroso reductasas (N_2ORs) de desnitrificadores y fijadores de N_2 ha sido obtenida por métodos de PCR. En esta tesis se refiere a estas secuencias como las secuencias *nosZ* “tradicionales”; por otro lado, los genes *nosZ* de microerófilos como *Anaeromyxobacter* spp., *Magnetospirillum* spp. y otros microorganismos anaerobios obligados *Wolinella* spp., *Desulfitobacterium* spp. y *Dechloromonas* spp. no han sido bien estudiados. La secuencia primaria de sus N_2ORs diverge de las tradicionales y, por lo tanto, diferentes parejas de cebadores deben ser desarrollados para entender mejor su diversidad y distribución en la naturaleza. Este estudio desarrolló oligonucleótidos para amplificar un rango más amplio de genes *nosZ* para evaluar su diversidad en suelo y bioreactores por técnicas independientes de cultivo. Las secuencias *nosZ* obtenidas de muestras ambientales fueron diferentes de las secuencias *nosZ* tradicionales. Ninguna secuencia de aminoácidos de los clones compartió más del 62% de similaridad con las *NosZ* tradicionales. Adicionalmente, este análisis reveló la presencia de residuos de histidina conservados esenciales para el funcionamiento de la proteína N_2OR madura. A través de un análisis filogenético utilizando los métodos del vecino más cercano, máxima parsimonia, máxima

verosimilitud e inferencia bayesiana, nueve clados de variantes del NosZ fueron identificados. Ninguna de las secuencias de los clones se ubicó dentro de los clados de los NosZ tradicionales, pero se agruparon con *Anaeromyxobacter* spp., *Magnetospirillum* spp., *Desulfitobacterium hafniense*, y *Dechloromonas aromatica*. La detección de los genes *nosZ* fue lograda por ISRT-PCR/FISH usando una sonda interna NosZ943 marcada fluorescentemente. Este representa el primer informe publicado del sondaje de amplificaciones del *nosZ* dentro de células microbianas activas usando un protocolo optimizado de *In Situ* PCR de Transcripción Reversa/Hibridización *In Situ* Fluorescente (ISRT-PCR/FISH). Nuestros resultados muestran que un clon con la secuencia parcial del gen *nosZ* de *Anaeromyxobacter* emitía una fuerte señal fluorescente cuando era detectado con la sonda NosZ943 pero no con la sonda Nos1527; mientras que un clon con la secuencia parcial del gen *nosZ* de *Pseudomonas* no emitió ninguna señal fluorescente con la sonda NosZ943. El ISRT-PCR fue también empleado para muestras naturales de un bioreactor anóxico. Aproximadamente 4% del conteo total de células estaban expresando los genes *nosZ novel*. Nosotros demostramos la existencia de numerosas variantes del gen *nosZ* que aún no están representadas por organismos cultivables. Estas variantes pueden representar una alta diversidad funcional para reducir N₂O en el ambiente. Por lo tanto, la diversidad de los genes *nosZ* en la naturaleza y su contribución al monto de N₂O requiere más exploración.

DEDICATORY

To the heaven in my hands, Eva and Alejandro

ACKNOWLEDGMENTS

Specially thanks to my husband Alejandro Caro for his priceless contribution to this work.

Thanks to my parents that gave me their comprehension, love and a lot of time.

I would like to express gratitude to Dr. Arturo Massol-Deyá for his advice, critically reading of this manuscript, and his support in difficult months.

I would like to express gratitude to Mrs. Gladys Toro for her technical support, research assistance and frienship.

Thanks to Dr. Frank Loeffler for his mentoring during my summer internship in GeorgiaTech, for his guidance and critical comments of the manuscript.

Thanks to Dr. Carlos Muñoz for being part of my graduate committee and all his suggestions to this manuscript.

I would like to show my appreciation to Dr. Rafael R. Montalvo-Rodríguez for all the things I learnt from him.

Specially thanks to Alexis Valentín, Dubiezel Medina, Paola Bracho, Katherine Carrero, and Ivette Ludeña for kindly providing me with DNA extractions and for their friendship.

Thanks to José Almodovar for his assistance with fluorescence microscopy.

Thanks to Idaris from Dr. Nanette Diffoot lab for making available some of the equipment needed for this research.

TABLE OF CONTENTS

Abbreviations.....	1
List of Tables.....	3
List of Figures.....	4
List of Appendixes.....	15
Introduction.....	16
Objectives.....	17
Literature review	
Nitrogen Cycle.....	18
Nitrous oxide (N ₂ O) gas.....	19
Nitrous oxide reductase (N ₂ OR).....	20
Strains harboring a “divergent” N ₂ OR.....	22
Molecular biology techniques.....	24
Phylogenetic analysis.....	28
Materials and methods	
Bacterial cultures and growth conditions.....	32
DNA and RNA extractions.....	33
Development of PCR primers.....	34
Amplification of <i>nosZ</i> genes.....	35
Cloning of <i>nosZ</i> PCR products and sequencing.....	36
Sequence and Phylogenetic analysis.....	38
Terminal Restriction Fragment Length Polymorphism of <i>nosZ</i> amplicons.....	40
<i>In Situ</i> RT-PCR.....	41
Bacterial cell fixation.....	42
Cell permeabilization.....	42
ISRT-PCR protocol.....	44
Detection of amplified gene products by fluorescent <i>in situ</i> hybridization (FISH).....	44
ISRT-PCR protocol-Environmental samples.....	46
Results	
Development of PCR primers.....	47
Primers specific to <i>nosZ</i> from <i>Anaeromyxobacter</i> spp.....	47
Primers specific to novel <i>nosZ</i> genes.....	49
Environmental <i>nosZ</i> -gene libraries.....	52
Phylogenetic analysis of <i>nosZ</i> genes.....	55
Diversity indexes and Richness estimators of <i>nosZ</i> -gene libraries.....	58
Terminal Restriction Fragment Length Polymorphism of <i>nosZ</i> Amplicons.....	59
<i>In Situ</i> -PCR.....	60
Cell Permeabilization.....	61
<i>In Situ</i> RT-PCR.....	62
Discussion	
Primers specific to <i>nosZ</i> from <i>Anaeromyxobacter</i> spp.....	65
Primers Specific to novel <i>nosZ</i> Genes and Environmental Analysis.....	66

Phylogenetic analysis of <i>nosZ</i> genes.....	68
Diversity indexes and Richness estimators of <i>nosZ</i> -gene libraries.....	73
Terminal Restriction Fragment Length Polymorphism of <i>nosZ</i> Amplicons.....	75
<i>In Situ</i> -PCR.....	77
Cell Permeabilization.....	77
<i>In Situ</i> RT-PCR.....	78
Literature cited.....	81
Figures.....	93
Appendix 1.....	148
Appendix 2.....	171

ABBREVIATIONS

4',6'-diamidino-2-phenylindole (DAPI)

Abundance-based coverage estimator (ACE)

Aerobic station of Mayagüez Wastewater Treatment Plant (AES)

Ammonium (NH_4^+)

Anaerobic station of Mayagüez Wastewater Treatment Plant (ANS)

Basepair (bp)

Basic local alignment search tool (BLAST)

Bayesian Inference (BI)

Commonwealth Refining Company (CORCO)

Cyanine dye 3 (Cy3)

Dinitrogen (N_2)

Dinucleotide-triphosphate (dNTP)

Dissimilatory nitrate reduction to ammonia (DNRA)

Distance-based OTU and Richness (DOTUR)

DNA nuclease (DNase)

Double distilled water (ddH_2O)

Fluidized bed reactor (FBR)

Fluorescent *In Situ* Hybridization (FISH)

InfraRedDye (IRDye)

In Situ -PCR (IS-PCR)

In Situ Reverse Transcriptase-PCR/Fluorescent *In Situ* Hybridization (IS RT-PCR/FISH)

Joint Genome Institute (JGI)

Luria Bertani (LB)

Maximum Likelihood (ML)

Maximum Parsimony (MP)

Mayagüez wastewater treatment plant (MWWT)

Melting temperature (T_m)

Magnesium (Mg)

National Center for Biotechnology Information (NCBI)

Neighbor-Joining (NJ)

Nitrate (NO_3^-)

Nitric oxide (NO)

Nitrite (NO_2^-)

Nitrogen (N)

Nitrous oxide (N_2O)

Nitrous oxide reductase (N_2Or or NosZ)

Operational taxonomic unit (OTU)

Optical Density (OD)

Phosphate-buffered saline (PBS)

Pipeline Functional Gene/ Repository (PFGR)

Polymerase Chain Reaction (PCR)

Polyvinylpyrrolidone (PVPP)

Reverse Transcriptase-PCR (RT-PCR)

RNA nuclease (RNase)

Terminal restriction fragment length polymorphism (T-RFLP)

LIST OF TABLES

Table 1. Environmental samples used in this study

Table 2. Primer sequences and designation

Table 3. Characteristics and PCR conditions for primer pairs.

Table 4. Similarity matrix of amino acid sequences from non-denitrifying *nosZ* group and environmental clones.

Table 5. Similarity matrix of amino acid sequences from traditional-*nosZ* group.

Table 6. Diversity indexes and Richness estimators.

Table 7. Expected *nosZ* amplicons based on *HinPI*, *NlaIII*, and *RsaI* T-RF fragment lengths for MWWT clones.

Table 8. Expected *nosZ*-amplicon based *HinPI*, *NlaIII*, and *RsaI* T-RF fragments lengths for reference organisms carrying a non-denitrifying *nosZ* gene (NosZ943F/NosZ1789R primer pair).

LIST OF APPENDIXES

Appendix 1. Partial sequences of the *nosZ*-gene of environmental clones and *A. dehalogenans* strain 2CP-3 and *Anaeromyxobacter* sp. strain R.

Appendix 2. Nucleotide accession numbers for reference organisms used in this study (protein and nucleotide sequences were obtained from the Pipeline Functional Gene/Repository [PFGR] [<http://flyingcloud.cme.msu.edu/fungene/>]).

LIST OF FIGURES

Figure 1. Denitrification, DNRA and nitrification: three known pathways involved in nitrous oxide production (modified from Kelso et al., 1997).

Figure 2. Graphic depiction of primer binding sites along the 1863 bp *nosZ* gene sequence of *A. dehalogenans* strain 2CP-C.

Figure 3. Partial multiple *nosZ* nucleotide sequence alignments showing priming sites for forward (upper) and reverse (lower) primers.

Figure 4. Partial multiple *nosZ* amino acid sequences alignments showing priming sites for forward (upper) and reverse (lower) primers.

Figure 5. Optimization of PCR conditions for amplification with primer pair NosZ943F/1789R. Results from *A. dehalogenans* 2CP-C genomic DNA amplifications using PCR annealing temperatures of 40, 41.2, 43.3, 46.3, 49.8, 52.8, 54.9, and 56°C (Lanes 1-8). Five microlitres of each amplification reaction were loaded onto an agarose gel (1%). MW, 1Kb DNA ladder (Promega). NosZ943F/1789R amplicons with the expected size of 873bp were obtained at 52.8, 54.9, and 56°C (Lanes 6-8). Unspecific amplification occurred at lower temperatures (Lanes 1-5).

Figure 6. Optimization of PCR conditions for primer pair NosZ943F/1789R. Results from *A. dehalogenans* 2CP-C DNA amplifications using PCR annealing temperatures of 56, 57.3, 58.4, 59.7, 60.8, 61.6 and 62 °C (Lanes 1-7). Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW, 1Kb DNA ladder (Promega). Primer pair NosZ943F/1789R generated an amplicon of 873bp.

Figure 7. Analysis of PCR amplifications with primer pair NosZ256F/718R. Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW, 1Kb plus DNA ladder (Invitrogen). Primer pair NosZ256F/718R generated an amplicon of 483bp. A PCR annealing temperature of 68°C and a magnesium concentration of 2mM were used. Lanes 1-4, *Anaeromyxobacter dehalogenans* strains Lane 1, strain 2CP-C DNA amplification. Lane 2, strain 2CP-1 DNA amplification. Lane 3, strain 2CP-3 DNA amplification. Lane 4, strain R DNA amplification. Lane 5, *Desulfitobacterium* sp. PCE-1 DNA amplification. Lane 6, *D. hafniense* DNA amplification. Lane 7, *Desulfitobacterium chlororespirans* Co23 DNA amplification. Lane 8, *W.succinogenes* DNA amplification. Lane 9, *Dehalococcoides ethenogenes* BAV1 DNA amplification. Lane 10, *Myxococcus xanthus* DNA amplification.

Figure 8. PCR amplifications with primer pair NosZ903F/1789R targeting the *Anaeromyxobacter* spp. *nosZ* gene and species of the non-denitrifying group. Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW, 1Kb plus DNA ladder (Invitrogen). Primer pair NosZ903F/1789R generated an amplicon of 913bp. PCR annealing temperatures of 63°C and magnesium concentration of 2.5mM

were used. Lane 1, *A. dehalogenans* 2CP-C DNA amplification. Lane 2, *A. dehalogenans* 2CP-1 DNA amplification. Lane 3, *A. dehalogenans* 2CP-3 DNA amplification. Lane 4, *Anaeromyxobacter* sp. strain R DNA amplification. Lane 5, *D. hafniense* DNA amplification. Lane 6, *Desulfitobacterium chlororespirans* Co23 DNA amplification. Lane 7, *W. succinogenes* DNA amplification. Lane 8, *Dehalococcoides* sp. BAV1 DNA amplification.

Figure 9. Partial multiple non-denitrifying (first 8 sequences) and traditional- *nosZ* (last 10 sequences) nucleotide sequences alignments showing priming sites for degenerate primers.

Figure 10. Analysis of PCR amplification of primer pair NosZ943F/1789R with DNA of species of the non-denitrifying group. The PCR annealing temperature was 59°C and the reactions contained 2.5mM magnesium. Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW, Lambda DNA plus *EcoRI/HindIII* molecular weight marker. Primer pair NosZ943F/1789R generated an amplicon of 873bp. Lane 1, *A. dehalogenans* 2CP-C DNA amplification. Lane 2, *A. dehalogenans* 2CP-3 DNA amplification. Lane 3, *Anaeromyxobacter* sp. strain R DNA amplification. Lane 4, *Pseudomonas stutzeri* OX1. Lane 5, *Desulfitobacterium chlororespirans* Co23 DNA amplification. Lane 6, *D. hafniense* DNA amplification. Lane 7, *Desulfitobacterium* sp. strain PCE-1 DNA amplification.

Figure 11. Optimization of PCR conditions for *nosZ* amplification with primer pair NosZ334F/1789R. Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW1, Lambda DNA plus *EcoRI/HindIII* molecular weight marker. MW2, 1Kb plus DNA ladder (Invitrogen). Primer pair NosZ334F/1789R generated an amplicon of 1482bp. A PCR annealing temperature of 57°C and a magnesium concentration of 2.5mM were used. Lane 1, *A. dehalogenans* 2CP-C DNA amplification. Lane 2, *A. dehalogenans* 2CP-3 DNA amplification. Lane 3, *Anaeromyxobacter* sp. strain R DNA amplification. Lane 4, *D. chlororespirans* Co23 DNA amplification. Lane 5, *D. hafniense* DNA amplification. Lane 6, *Desulfitobacterium* sp. PCE-1 DNA second PCR amplification using 1 µl of product from the first round of PCR amplification. Lane 7, *A. dehalogenans* 2CP-C DNA amplification. Lane 8, *A. dehalogenans* 2CP-3 DNA amplification. Lane 9, *Anaeromyxobacter* sp. strain R DNA amplification. Lane 10, *D. chlororespirans* Co23 DNA amplification. Lane 11, *D. hafniense* DNA amplification. Lane 12, *W. succinogenes* DNA amplification.

Figure 12. Analysis of PCR amplification with total community DNA of environmental samples. Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW1, Lambda DNA plus *HindIII/EcoRI* molecular weight marker. MW2, 1Kb DNA ladder (Promega). Lanes 1-8, PCR amplification with primer pair NosZ334F/1789R Mayagüez Wastewater Treatment Plant (MWWT). Primer pair NosZ334F/1789R generated an amplicon of 1482bp. A PCR with 25 cycles and magnesium concentration of 2.5mM were used. AES, aerobic station; ANS; anaerobic station. Lane 1, AES-January. Lane 2, ANS-January. Lane 3, AES-February. Lane 4,

ANS-February. Lane 5, AES-April. Lane 6, ANS-April. Lane 7, AES-May. Lane 8, ANS-May. Lanes 9-20, PCR amplification with primer pair NosZ943F/1789R using DNA of different tropical environments. Primer pair NosZ943F/1789R generated an amplicon of 873bp. A PCR annealing temperature of 57 °C and 2.5mM magnesium were used. Even lanes are replicates of the previous odd lane but with addition of *Anaeromyxobacter dehalogenans* strain 2CP-C genomic DNA. Lane 9-10, Pueblo Forest (bottom core 5-10 cm). Lane 11-12, Pueblo Forest (top core 0-5 cm). Lane 13-14, CORCO soil. Lane 15-16, CORCO soil with addition of hydrocarbons. Lane 17-18, Mayagüez Bay; Lane 19-20, Fluidized Bed Reactor (May).

Figure 13. Analysis of colony PCR amplification of primer pair NosZ943F/1789R with environmental clones from CORCO DNA sample. PCR was performed with T7 promoter and SP6 promoter primer pair. Five µl of each amplification reaction were loaded onto the agarose gel (1%). MW, Lambda DNA plus EcoRI/HindIII molecular weight marker. Primer pair T7 promoter/SP6 promoter generated an amplicon of approx. 1034bp. Lane 1 to 11 correspond to CORCO18, CORCO22, CORCO24, CORCO25, CORCO26, CORCO28, CORCO29, CORCO30, CORCO31, CORCO32, and CORCO33 clones.

Figure 14. Alignment of NosZ amino acid sequences generated with Clustal W (Thompson et al. 1994) and edited using BioEdit Sequence Alignment Editor (Hall 1999). The shaded graphic view show only similar residues within non-denitrifying *nosZ* sequences and *Pseudomonas denitrificans*. All *nosZ* sequences used in the phylogenetic analysis were included in the analysis but not all are included in the alignment shown.

Pseudomonas, *Ps. denitrificans*; *Anaeromyxobacter* 1, *A. dehalogenans* 2CP-C; *Anaeromyxobacter* 2, *Anaeromyxobacter* sp. Fw109-5; *Desulfitobacterium*, *D. hafniense*; *Geobacillus*, *G. thermodenitrificans*; *Psychroflexus*, *P. torquis*; *Robiginitalea*, *R. biformata*; *Magnetospirillum* 1, *M. gryphiswaldense*; *Magnetospirillum* 2, *M. magneticum*; *Magnetospirillum* 3, *M. magnetotacticum*; *Campylobacter*, *C. fetus*; *Thiomicrospira*, *Tm. denitrificans*; *Wolinella*, *W. succinogenes*. Symbols ‡ and * show the calcium and chloride ion ligands and the conserve histidine residues, respectively.

Figure15. Rooted neighbor-joining (NJ) tree of partial *nosZ* gene sequences (1000 bootstraps). The trees were generated based on the alignment of 180 nucleotides of truncated *nosZ* sequences retrieved from Pipeline Functional Gene and environmental clone sequences from FBR, CORCO, and MWWT libraries. *C. crenilabis* served as an outgroup. Branch lines have distinct colors, in blue are traditional-*nosZ* sequences, in green are non-denitrifying *nosZ* sequences, in red are archaeal *nosZ* sequences, and in yellow are *nosZ* sequences from Flavobacteria. The bar represents 0.05 nucleotide substitutions per nucleotide.

Figure 16. Strict consensus (Lenght = 1898, Consistency Index=0.48, Retention Index=0.75) of 8 most parsimonious trees (Lenght =1895, Consistency Index=0.48, Retention Index=0.75) from *nosZ* gene dataset. The sequence of *C. crenilabis* served as the root. Numbers above the branches represent the percentage of 1,000 bootstrap replications. Branch lengths correspond with numbers of amino acid replacements. The

bar represents 20 replacements. Tree is based on the non-gapped multiple sequence alignment of 90 amino acids.

Figure 17. Rooted maximum likelihood tree from *nosZ* gene dataset. The sequence of *C.crenilabis* served as the root. Numbers above the branches represent the percentage of 1000 bootstrap replications, values less than 50% are not shown, and those nodes are collapsed. Scale bar represents 0.1 substitutions per amino acid position. Tree is based on the non-gapped multiple sequence alignment of 90 amino acids.

Figure 18. Bayesian phylogenetic tree from *nosZ* gene dataset after 500,000 generations of MCMCMC chains. Values to the right of the branches correspond to the posterior probabilities. The horizontal bar at the base represents 0.1 substitutions/site. Trees are based on the non-gapped multiple sequence alignment of 90 amino acids.

Figure 19. Rarefaction curves of *nosZ* clone sequences for the clone libraries FBR, CORCO and MWWT at 27-29% distance level.

Figure 20. *Pseudomonas stutzeri* cells permeabilized by electroporation and stained with DAPI (blue) were observed by confocal microscopy. (Upper panels) Permeabilized cells showing cell integrity. (Lower panels) Permeabilized cells with RNase and DNase treatment. (A and D) Cells electroporated during 50 ms. (B and E) Cells electroporated during 100 ms. (C and F) Cells electroporated during 200 ms.

Figure 21. Confocal images of ISRT-PCR FISH of *nirS* in a co-culture of *P. stutzeri* with a non-denitrifying rod, Cy3 signal (orange) was superposed on the corresponding DAPI signal (blue). (A) Detection of *nirS* gene with a Cy3 labeled probe in a co-culture of *P. stutzeri* with a non-denitrifying rod. B, *P. stutzeri* cells treated with 1mg/ml RNase A for 1h after permeabilization.

Figure 22. Analysis of RT-PCR amplification. PCR annealing temperatures of 56°C. Ten microlitres of each amplification reaction were loaded into the wells of the agarose gel (1%). MW1, 1Kb DNA ladder (Promega); MW2, Lambda DNA plus *Hind*III/*Eco*RI molecular weight marker. Primer pair NosZ1366F/1773R generated an amplicon of 407bp; primer pair Nos661F/1773R generated an amplicon of 1113bp; primer pair 27F/1392R generated an amplicon of 1394bp. Lane 1, 16S rRNA RT-PCR with primers 27F/1392R with RNA from *P. stutzeri* grown Nitrate broth; Lane 2, 16S rRNA RT-PCR with primers 27F/1392R with RNA from *P. stutzeri* grown in BLK medium with histidine; Lane 3, *nosZ* RT-PCR with primers NosZ1366F/1773R from *P. stutzeri* grown in Nitrate broth; Lane 4, *nosZ* RT-PCR with primers NosZ1366F/1773R from *P. stutzeri* grown in BLK medium with histidine; Lane 5, *nosZ* RT-PCR with primers Nos661F/1773R (Scala and Kerkhof, 1998) from *P. stutzeri* grown in Nitrate broth; Lane 6, *nosZ* RT-PCR with primers Nos661F/1773R (Scala and Kerkhof, 1998) from *E. coli* grown in Nitrate broth.

Figure 23. ISRT-PCR/FISH of *nosZ* gene of *P. stutzeri*. A-B; C-D; E-F, Confocal images corresponding to the same microscopic section. A, C, E, DAPI signal emitting-cells

(blue). B, D, F, Cy3 signal emitting-cells (orange). A-B, ISRT-PCR/FISH of *nosZ* gene of *P. stutzeri* grown in BLK medium with histidine as only nitrogen source. C-D, ISRT-PCR/FISH of *nosZ* gene of *P. stutzeri* grown in Nitrate broth. E-F, ISRT-PCR/FISH of *nosZ* gene of *P. stutzeri* grown in Nitrate broth but with a RNase treatment after the enzymatic permeabilization and before PCR reaction.

Figure 24. IS-PCR/FISH of *E.coli* DH5 α clones harboring *nosZ* genes in pGEM. A-B; C-D; E-F; G-H, Confocal images corresponding to the same microscopic section. A, C, E, G, DAPI signal emitting-cells (blue). B, D, F, H, Cy3 signal emitting-cells (orange). A-B, IS-PCR/FISH of *E.coli* DH5 α clone harboring in pGEM the partial *nosZ* gene of *A. dehalogenans* 2CP-C obtained by PCR with primers NosZ256F/1807R, Probe943 was used for FISH and primers NosZ334F/1789R were used for PCR. C-D, IS-PCR/FISH of *E.coli* DH5 α clone harboring in pGEM the partial *nosZ* gene of *A.dehalogenans* 2CP-C obtained by PCR with primers NosZ256F/1807, Probe1527 was used for FISH and primers NosZ334F/1789R were used for PCR. E-F, IS-PCR/FISH of *E.coli* DH5 α , Probe943 was used for FISH and primers NosZ334F/1789R were used for PCR. G-H, IS-PCR/FISH of *E.coli* DH5 α clone harboring in pGEM the partial *nosZ* gene of *P. stutzeri* obtained by PCR with primers Nos661/1773R, Probe943 was used for FISH and primers Nos661F/1773R were used for PCR.

Figure 25. ISRT-PCR/FISH of *nosZ* genes in environmental samples from an anaerobic bioreactor. A-B, 1000X epifluorescence images corresponding to the same microscopic section. A and B, DAPI (blue) and Cy3 (orange) signal emitting-cells,

respectively, of a ISRT-PCR/FISH of *nosZ* gene in a MWTP sample immobilized in a 0.2µm-pore-size filter. C-D, Confocal 1000X images corresponding to the same microscopic section. C section showing only DAPI signal emitting-cells (blue), while D section showing Cy3 signal (orange) superposed on the corresponding DAPI signal (blue). E-F, Confocal 600X images corresponding to the same microscopic section. E section shows DAPI signal emitting-cells (blue), while F section shows Cy3 signal (orange). G-H, 1000X epifluorescence images corresponding to the same microscopic section. G and H, DAPI (blue) and Cy3 (orange) signal, respectively, of a ISRT-PCR/FISH with a RNase treatment after the enzymatic permeabilization and before PCR reaction. C-D, E-F and G-H, ISRT-PCR/FISH of *nosZ* gene in a MWWT sample immobilized in glass slide.

Figure 26. *nosZ*-based *HinPI* (1, 4, 7), *NlaIII* (2, 5, 8), and *RsaI* (3, 6, 9) T-RFLP fingerprint patterns of Mayaguez Wastewater Treatment Plant (MWWT) anaerobic reactor sample (1-6) and MWWT34 clone (7-9). MW, molecular sizing standard. KBPlus-LICOR (50-700bp).

Figure 27. Comparison of electropherograms profiles generated with *HinPI* for Aerobic station (A), Anaerobic station (B) of Mayagüez Wastewater Treatment Plant samples and clone MWWT34 (C).

Figure 28. Comparison of electropherograms profiles generated with *NlaIII* for Aerobic station (A), Anaerobic station (B) of Mayagüez Wastewater Treatment Plant samples and clone MWWT34 (C).

Figure 29. Comparison of electropherograms profiles generated with *RsaI* for Aerobic station (A), Anaerobic station (B) of Mayagüez Wastewater Treatment Plant samples and clone MWWT34 (C).

INTRODUCTION

Denitrification contributes to greenhouse gas production through the release of nitrous oxide (N_2O) and nitric oxide (NO). Key enzymes involved in catalyzing various steps of the denitrification process are nitrate reductase (coded by *nar* and *nap* genes, which catalyze the reduction of $\text{NO}_3^- \rightarrow \text{NO}_2^-$), nitrite reductase (coded by *nirK* and *nirS* genes, which catalyze the reduction of $\text{NO}_2^- \rightarrow \text{NO}$), nitric oxide reductase (coded by *norB* gene, which catalyzes the reduction of $\text{NO} \rightarrow \text{N}_2\text{O}$), and nitrous oxide reductase (N_2OR) (coded by *nosZ* gene, which catalyzes the reduction of $\text{N}_2\text{O} \rightarrow \text{N}_2$). However, denitrification is often incomplete and NO and N_2O are released to the atmosphere (Wolf and Russow, 2000). Figure 1 shows that N_2O is also produced in other pathways, such as nitrification and dissimilatory nitrate reduction to ammonia (DNRA).

N_2O reduction has been evaluated in bacterial cultures and manipulated soil systems (Kelso et al., 1997; Wolf and Russo, 2000; Trimmer et al., 2003; Trimmer et al., 2005). Generally, these experiments use the kinetics of ^{15}N and ^{14}N isotopes to quantify the formation of different N species. The study of N-cycling with molecular biology approaches promises to provide important insights, to characterize the genetic diversity of N-cycling genes, and to elucidate the activity of N_2OR s under certain environmental conditions.

The Polymerase Chain Reaction (PCR) technique has allowed the amplification of *nosZ* genes from denitrifiers, nitrifiers and N_2 -fixers (Scala and Kerkhof, 1998; Rösch et al.,

2002). Nonetheless, biased PCR amplification of proteobacterial *nosZ* genes has limited the discovery of NosZ variants. *nosZ* from some dissimilatory nitrate reduction to ammonia (DNRA) bacteria such as *Anaeromyxobacter* spp., *Wolinella succinogenes*, and *Desulfitobacterium* spp. differ significantly in their sequences from those sequences obtained by PCR technique. Teraguchi and Hollocher (1989) described the enzyme N₂OR of these non-denitrifiers as "divergent" from other known N₂OR. In order to determine the significance of NosZ diversity, new priming sites to amplify preferentially this functional group are necessary as well as new tools for better understanding the distribution of a broad range of *nosZ* genes in nature.

OBJECTIVES

- To identify, design and test PCR primers and linear hybridization probes for detecting of novel *nosZ* genes.
- To asses *nosZ* gene diversity using clone libraries, DNA sequence analysis, and T-RFLP.
- To develop and validate an *In Situ* PCR technique to prove the presence and expression of novel *nosZ* genes inside microbial cells.

LITERATURE REVIEW

Nitrogen Cycle

Nitrogen (N) exists in eight oxidation states (+5, +4, +3, +2, +1, 0, -3, -4) and it can be present as gaseous (NO, N₂O, N₂) or dissolved organic (chemicals in which N is bound to carbon) and inorganic (NO₃⁻, NO₂⁻, NH₃/NH₄⁺) species. Among these species, nitrate (NO₃⁻), nitrite (NO₂⁻), ammonia (NH₃), nitric oxide (NO), and nitrous oxide (N₂O) are forms of N of environmental concern, while dinitrogen (N₂) is an innocuous gas (Zumft, 1997). The abundance of NO₃⁻ in water bodies is challenging because it promotes eutrophication, as does the deposition of atmospheric NH₃. The Clean Water Act regulates nitrite because it is toxic for many organisms. The concentration of NO₂⁻ in freshwaters worldwide was estimated to be near 1µg per litre. However, recent studies in German and Northern Ireland rivers revealed much higher concentrations between 100 and 200µg of N per litre (Kelso et al., 1997). Other forms of nitrogen such as NO and N₂O are gaseous species are greenhouse gases and contribute to global warming.

Microbial denitrification, nitrification, dissimilatory nitrate reduction to ammonia (DNRA), anaerobic ammonium oxidation (ANAMMOX), N-fixation, and ammonification are key processes involved in the N cycle. Several of these processes can be coupled and are occurring simultaneously, adding complexity to the understanding of biogeochemical N cycling (Trimmer et al., 2003).

Nitrous oxide (N₂O) gas

During the last ice age the atmospheric concentrations of N₂O rose to ~275ppb and stayed constant until the 19th century (Trogler, 1999). Currently, the level of N₂O has reached 315 ppb and increases 0.25% per year (Dong et al., 2002). Even a small change in its concentration could cause a big perturbation that will likely last for centuries. The N₂O lifetime is around 150 years with a high warming potential 310 times greater than that of CO₂ (Trogler, 1999). N₂O has many anthropogenic sources including its use as a propellant (e.g., in canned whipping cream), fuel additive in racing cars, anesthetic, and prime material for the production of nylon, fertilizers and explosives (Bouwman et al., 1995; Trogler, 1999). However, rather than direct inputs due to manufacture and use, it has been suggested that the main source of N₂O is from microbial activity. Major changes in land-use and associated changes in microbial activity may be the main reason for increased N₂O concentrations in the atmosphere (Avrahami et al., 2002).

Agricultural production (i.e., feedstock waste and fertilizers) have influenced the atmospheric levels of N₂O (Bouwman et al., 1995). Fertilization increases the availability of nitrogen in soils and waters, thus promoting biological processes demanding NO₃⁻, NO₂⁻, and ammonium (NH₄⁺) as substrates and that produce N₂O (García-Ruiz, Pattinson, and Whitton, 1998). After winter, freezing and thawing periods can account for more than 70% of the total N₂O loss from soil in temperate climates (Müller et al., 2002). Estuarine and coastal sources contribute nearly 60% of the global marine N₂O flux (Dong et al., 2002), which is up to 20% of the global budget (Barnes and Owens, 1998).

Nitrous oxide is emitted from at least two microbial processes: denitrification and nitrification. Denitrification is often incomplete and large amounts of N_2O are released to the atmosphere. In nitrification, N_2O is a by-product. There is general agreement that a third process, DNRA, produces N_2O as a by-product when NO_2^- is reduced to NH_4^+ (Figure 1) (Smith and Zimmerman, 1981; Smith, 1982; Kelso et al., 1997), but Simon (2002) has recently reported that NH_4^+ is the only product of DNRA. Clearly, the intermediates and by-products between NO_2^- and N_2 have not been conclusively established.

Nitrous oxide reductase (N_2OR)

Microbial nitrogen metabolism may also serve as a sink for atmospheric N_2O . Firestone and Tiedje (1979) found that the antibiotic chloramphenicol could affect the rate of N_2O reduction in soils incubated anaerobically. This chemical inhibits bacterial protein biosynthesis suggesting that *de novo* synthesis of enzymes is involved in the N_2O -reducing activity. Nowadays, the enzyme that catalyzes the two-electron reduction of N_2O to N_2 is known as nitrous oxide reductase (N_2OR), and was first isolated in 1982 by Zumft and Matsubara (1982). Microorganisms that harbor this enzymatic capability carry out denitrification, N-fixation or DNRA (Rösch et al., 2002; Simon et al., 2004).

In denitrification, NO_3^- is reduced to N_2 via NO_2^- , NO , and N_2O catalyzed by nitrite reductase (cytochrome *cd*₁- or Cu-containing, depending on the species), NO reductase, and N_2O reductase, respectively. When NO_3^- or NO_2^- are not completely reduced, NO and

N₂O are released to the atmosphere (Rösch et al., 2002). Wolf and Russow (2000) report that N₂ emissions exceed N₂O emissions by a factor of 3. However, there are some denitrifiers that reduce NO₃⁻ to N₂O as an end product. In DNRA, cytochrome *c* nitrite reductase reduces NO₂⁻ to NH₄⁺. Additionally some DNRA bacteria, for example *Wolinella succinogenes*, possess the ability to reduce N₂O to N₂ (Simon et al., 2004). *Bradyrhizobium japonicum*, *Sinorhizobium meliloti*, and *Rhizobium meliloti* are N-fixing bacteria, which also possess a N₂O-reducing activity (Sameshima-Saito, Chiba, and Minamisawa, 2006; Chan and Wheatcroft, 1993; Chan and McCormick, 2004). Although these diazotrophs possess the gene that encodes N₂OR (*nosZ*), it is also possible that nitrogenase, which catalyzes N₂ reduction, has N₂O as an alternative substrate (Rösch et al., 2002).

Nitrous oxide reductase is a periplasmic homodimeric multicopper enzyme with a molecular mass of 65kDa per subunit. Early studies showed that N₂OR has certain unusual properties compared to other copper proteins (Matsubara et al., 1982; Coyle et al., 1985). The crystal structure of N₂OR reveals that each subunit consists of two domains, designated as Cu_A and Cu_Z. The former is a binuclear copper centre and the electron entry site, and it is structurally homologous to the Cu_A centre of cytochrome *c* oxidase. The two enzymes revealed a conserved set of potential copper ligands, but there is no overall sequence homology. Cu_Z is a novel tetranuclear copper site that might be coordinated by multiple histidine residues, and it is the catalytic site (Rasmussen et al., 2002).

The N₂OR of the non-denitrifier *W. succinogenes* diverges from other known N₂ORs

(Teraguchi and Hollocher, 1989). This enzyme possesses only three metal transition atoms per subunit, and has a carboxy terminal extension that carries a heme cytochrome *c* motif (CNGCH) that could be an electron donor. All the important ligands of the Cu_A and Cu_Z of traditional N₂OR are found in the primary sequence of *W. succinogenes* protein, except for a tryptophane residue between two cysteines in Cu_A. Comparative genome analysis revealed sequence similarities to this N₂OR in databases. For example, amino acid sequences from *Dechloromonas aromatica*, *Magnetospirillum magneticum*, *Desulfitobacterium hafniense*, and *Anaeromyxobacter dehalogenans* are 67, 66, 44, 43% similar, respectively. NosZ sequences from this group clearly diverge from the traditional NosZ protein sequences, especially after the third and fourth histidine residues in Cu_Z. The conserved histidine after the two first histidines in Cu_Z is replaced by asparagine and alanine in the amino acid sequences of these organisms (Simon et al., 2004).

Strains harboring a “divergent” N₂OR

W. succinogenes, *D. aromatica*, and *D. hafniense* are obligate anaerobes, while *M. magneticum* and *A. dehalogenans* are microaerophiles. *W. succinogenes* is an ϵ -proteobacterium that lives in rumen of cows and grows by different modes of anaerobic respiration, as well as in the presence of 2% of oxygen (Baar et al., 2003). The study of this DNRA bacterium has been of high relevance because it has the capacity to reduce N₂O, even though it does not have N₂O producing capacity. It is possible that free-living microorganisms with similar physiological capabilities may exist in soils; this hypothesis merits further investigation. Furthermore the sequences that could code for a nitrite

reductase or NO reductase have not been found (Simon et al., 2004). *D. aromatica* is a β -proteobacterium involved in the anaerobic degradation of benzene using NO_3^- as the electron acceptor (Coates et al., 2001). Members of the *Desulfitobacterium* genus are gram positive bacteria involved in respiratory reductive dechlorination of halophenols and chlorinated ethenes, and include *Desulfitobacterium hafniense* strain PCE1, *Desulfitobacterium chlororespirans*, and *Desulfitobacterium dehalogenans* (Sanford et al., 1996; Mackiewicz and Wiegel, 1998, Lee et al., 2001). *Magnetospirillum magnetotacticum* is a microaerophilic spirillum of the α class within the phylum Proteobacteria, and is involved in the formation of crystals of the iron mineral magnetite (Fe_3O_4). *Magnetospirillum* species have been reported as N_2 fixers as well as denitrifiers (Bazylinski et al., 2000; Shinoda et al., 2005), and some species can grow on toluene, phenol, and other aromatic compounds (Shinoda et al., 2005; Kawaguchi et al., 2006). Finally, *A. dehalogenans* is a microaerophilic δ -proteobacterium recently characterized as a member of the Myxococcales order, and uses several compounds as electron donors (e.g. acetate, hydrogen) and acceptors (e.g. *ortho*-halophenols, reduced metals and NO_3^-) (Sanford et al., 2002). *Anaeromyxobacter* spp. 16S rRNA gene sequences have been found in subsurface samples from the uranium-contaminated U.S. DOE-NABIR Field Research Center (Oak Ridge, TN) and in flooded rice field soils (Treude et al., 2003; Petrie et al., 2003).

Physiological studies confirmed the capacity of *W. succinogenes* to reduce NO_3^- to NH_4^+ , and N_2O to N_2 , without producing N_2O from NO (Yoshinari, 1980). However, these kinds of studies that measure simultaneously several N species (NO_3^- , NO_2^- , NH_4^+ , NO,

N₂O, and N₂) have not been carried out in other bacteria mentioned above.

Molecular biology techniques

Cultivation provides information about the physiological capabilities of a small fraction of microorganisms because the majority of microbes cannot be grown in the laboratory. It has been estimated that nearly 99% of naturally occurring microbes cannot be cultured (Hughes et al., 2001). An alternative to overcome the limitations of culturability is provided by molecular biology techniques, since gene sequences coding for different enzymes involved in the biological N cycling can be obtained directly from natural samples (Rösch et al., 2002).

Several oligonucleotide primers have been designed to amplify the *nosZ* genes by Polymerase Chain Reaction (PCR) technique (Scala and Kerkhof, 1998). This molecular method enables researchers to produce millions of copies of a specific DNA sequence using a thermostable DNA-copying enzyme, called DNA polymerase. The polymerase recognizes a short oligonucleotide (primer) bound to the DNA template (usually a genomic DNA), and begins to use free nucleotides available to synthesize a new DNA strand complementary to the template. Generally, in PCR a nucleic acid extraction from cells lysed by different methods can be used as a template. The accurate design of two primers -called the *forward* and *reverse* primer pair-, which are complementary to the beginning and the end of the DNA sequence to be amplified, is the key factor to achieve the amplification of only the DNA sequence that is being targeted.

A suitable genetic target should have two conserved regions for the annealing of the primer pair and, between these conserved ends, a variable region that can be used for phylogenetic differentiation (Nocker et al., 2007). Each amplicon in the pool of PCR products can be cloned individually in a cloning host like *Escherichia coli*. An individual PCR-amplified sequence is inserted in a genetic vector and introduced into a bacterial receptor, which then is grown on agar slants with a rich nutrient medium. Usually, one bacterial colony corresponds to a clonal population in which each cell harbors a vector with one PCR product. The identification of the PCR product is done by sequencing, which is the basis for the construction of phylogenetic trees (Nocker et al., 2007). Through sequencing, one can identify sequences that are similar to those already published in databases, but also novel sequence information can also be generated.

Culture-independent molecular approaches have also demonstrated the disparity between culturable and the actual microbial diversity (Nocker et al., 2007). Microbial ecologists sample natural communities to assess the “true” diversity of microbes. The term *diversity* usually means *richness*, or the number of different types, and its unit of measurement is the operational taxonomic unit (OTU), that can be understood as an *species*. Usually, 16S ribosomal rRNA gene sequences with greater than 97% sequence identity are typically assigned to the same species, but also a unique terminal restriction fragment can be regarded as an OTU. In any community, the number of types of organisms observed increases with sampling effort until all types (OTUs) are observed. The relationship between the number of types observed and sampling effort gives information about the total diversity of the sampled community. Statistical approaches have been developed to

compare and estimate microbial species richness and there are methods available to assign sequences to OTUs (Hughes et al., 2001). Schloss and Handelsman (2005) developed DOTUR (Distance-Based OTU and Richness) program, which calculates values that are used to construct rarefaction curves, diversity indexes, and richness estimators.

A rarefaction curve compares the observed richness among habitats that have been unequally sampled. An accumulation curve is constructed by plotting the cumulative number of types observed versus sampling effort. If the sampling continued, the curves would eventually reach an asymptote at the actual community richness. A rarefied curve results from averaging randomizations of the observed accumulation curve. Shannon and Simpson are diversity indexes that have been used in microbial ecology (Hughes et al., 2001). Shannon describes the average degree of uncertainty of predicting the specie of an individual picked at random from the community. If the diversity of a habitat is low, the certainty of finding the same species again is high. High diversity means high uncertainty. Simpson includes the probability that two individuals randomly selected from a sample will belong to the same species. The richness estimators Chao and ACE (abundance-based coverage estimators) are particularly useful for datasets that tend to types of low abundance, as it happens for microorganisms.

PCR products can also be used for genetic profiling methods, such as the terminal restriction fragment length polymorphism (T-RFLP) technique that analyzes many PCR fragments in a single run, providing snapshots of a community (Nocker et al., 2007). This

technique relies on the sequence variation of species-specific amplicons. The pool of PCR products will be differentially cut with a restriction enzyme [cleaves DNA molecules at a precisely defined site(s)], generating different lengths of restriction enzyme products. In T-RFLP, only the terminal fragment carrying the fluorescent label from the primer is visualized. Different species can yield identical terminal restriction patterns when only one restriction enzyme is used, therefore, the use of more than one restriction enzyme will be necessary to correctly discern different species. A community profile is obtained by the electrophoretic separation of fluorescently labeled terminal restriction fragments detected by an automated system (Nocker et al., 2007).

The temporal and spatial variability of *nosZ* genotypes have been studied by T-RFLP in the intertidal rocky biofilms and sediment sites of the Douro River estuary (Portugal) (Magalhães et al., 2007) and in the continental shelf sediments of Tuckerton (NJ) (Scala and Kerkhof, 2000). Thirteen distinct *nosZ* T-RFLP patterns were detected in the estuary, while 71 were found in the continental shelf.

PCR has many variants that offer distinct advantages. One of those variants is *In Situ* PCR (IS-PCR) that offers a direct detection of even single copy genes inside the host cell without DNA or RNA extraction (Nuovo, 1996). IS-PCR has to be carried out inside the cell and, in the case of bacteria, their cell wall has to be permeabilized in order to allow PCR reactants to enter the cell. Commonly, the IS-PCR reaction is performed on glass slides (Nuovo, 1996). Microbial ecologists use the IS-PCR technique to determine the identity, activity, and distribution of the bacterial species in a specific niche. This may be possible if the 16S rRNA and functional gene transcripts are detected simultaneously

(Chen et al., 1997). PCR amplification with only one primer can also be used to generate a larger DNA amplicon, which prevents leaking of short amplicons in permeabilized cells (Kenzaka et al., 2005). The technique can be applied to detect genetic expression when combining it with Reverse Transcription PCR (RT-PCR) (Hodson et al., 1995; Tani et al., 1998), in which an enzyme called reverse transcriptase synthesizes double stranded DNA from single stranded RNA templates. Following the initial RT step, a regular PCR can be carried out to increase the number of DNA copies. Population density measurements and cell morphology, for example, can be obtained by observing cells with fluorescently labeled RNA or DNA, which emit a signal under an epifluorescence microscope. To increase the specificity of the technique, instead of using labeled nucleotides in PCR, Fluorescent *In Situ* Hybridization (FISH) can be used. FISH employs a fluorescent probe targeted to a specific PCR amplicon, avoiding detection of non-specific PCR products that are generated when primers bind to non desired DNA sequences (Tani et al., 1998). Multiple labeled fluorescent probes can also be used reducing background signal interference (Kenzaka et al., 2005).

Phylogenetic analysis

Sequence databases have allowed the comparison of protein structures, genes and entire genomes. Proteins with significant sequence similarity and/or with similar structure and function, belong to the same protein family, and share a strong evolutionary relationship. Proteins that do not share a common ancestor are very unlikely to show statistically significant sequence similarity, making sequence alignment a powerful tool for

identifying the members of protein families. Proteins that are evolutionarily related and that perform the same function in different species are called homologous proteins. Some sequence variation among organisms is possible and it may have little or no effect on the function of the protein, but other regions essential to the function are conserved (Nelson and Cox, 2000).

With the extensive amount of data present in databases, it is now common for bioinformaticians to describe each gene by its closest match in the database (usually using the BLAST program). While this practice is useful as a first cut, a robust phylogenetic analysis is required to confidently identify related sequences (Logsdon and Faguy, 1999). Phylogenetic trees constructed using DNA or protein sequence data are very important tools to clarify phylogenetic relationships among organisms. Taxonomic relationships (based on 16S and 18S rRNA genes) have demonstrated that, in general, the number of residues that differ in homologous proteins between two species is proportional to the phylogenetic difference between those species. This information allows the construction of evolutionary trees that show the origin and sequence of appearance of different species during the course of evolution (Nelson and Cox, 2000).

In molecular phylogeny, the data usually consists of multiple gene or protein sequence alignments that can be generated using programs like clustalW. The topology of the tree represents the degree of relatedness among different sequences, and the length of the branches of the tree corresponds to the distance between them. The external nodes are the sequences we are studying, and the internal nodes are inferred ancestors. Currently, there

are three major categories of tree-building methods: distance method, maximum parsimony, and maximum likelihood method (Abascal, n.d.).

In the distance method called Neighbor-Joining (NJ), the best phylogenetic tree is the one that minimizes the total tree length (i.e. the sum of the length of all branches in the tree). A distance for each pair of sequences is calculated from the multiple alignment. This method is extremely fast and useful as a preliminary inference. The NJ method generates an unrooted tree, and an outgroup (e.g., a homologous sequence, or a closely related species but from other taxonomic group) must be included to identify a common ancestor. The use of an outgroup could be explained with an analogy. If we had a circular book and we wanted to know which is the first page, an outgroup would be the hardcover (Abascal, n.d.).

In Maximum Parsimony (MP) method, the most probable tree is the one that represents the lower number of changes. However, if we have many sequences to analyze, the number of possible trees will be greater the amount of atoms in the universe. For this reason, the method to find the most parsimonious tree is approximated and not exhaustive, and this method does not guarantee finding the optimal solution; however, the analysis can be accomplished in reasonable time periods (Abascal, n.d.).

The Maximum Likelihood (ML) method needs an evolution model to calculate the probabilities of seeing the observed data. Some factors that have been taken in account by models of protein evolution are different frequencies of amino acid substitutions,

evolution rate heterogeneity among different species, and invariant and very variable sites along the same sequence. Different bioinformatic programs are available to study evolutionary history: ModelTest, ModelGenerator, and ProtTest. ML considers all the positions (each amino acid residue position) in the multiple alignments, and finds the tree that generated the data with the highest probability (Abascal, n.d.).

While ML works with the probability of seeing the observed data given a model, the Bayesian inference works on the probability that the model is correct given the observed data. The Bayesian method starts with a random tree and parameters, and each generation it randomly proposes a new tree topology or a new model parameter value. If the new tree has a higher probability than the previous model, then it is accepted and the process repeated; if, in the other hand, the new tree has a lower probability, it is accepted but with a probability proportional to the worsening produced. After a cumulative probability has been reached, the sampled trees are sorted based on probability and the best tree is selected (Abascal, n.d.).

Phylogenetic trees usually include a support value at every node of the tree, and commonly a bootstrap value is used. The bootstrap test can be used to statistically test the reliability of the phylogenetic tree constructed. Bootstrap values can be computed by resampling the characters (DNA or protein sequence data) randomly a number of times, and computing the frequency with which a node returns in the pooled results (Liu 2005). The use of an inappropriate evolutionary model could lead to low bootstrap supports (<50%) This node could be collapsed during the computing of a consensus tree. A

collapsed branch is a branch that is not present in a number of replicate trees analyzed.

MATERIALS AND METHODS

Bacterial Cultures and Growth Conditions

Axenic cultures of *Anaeromyxobacter* spp, *Desulfitobacterium* spp., and *Wolinella succinogenes* were used in this study as reference cultures harboring *nosZ* genes. *Anaeromyxobacter* species were grown statically in anoxic R2A broth amended with acetate and pyruvate (10mM each) (Sanford et al., 2002). *Desulfitobacterium* species were grown in bicarbonate-buffered basal salts medium amended with pyruvate (30mM) and vitamin solution (Wolin et al., 1963). *W. succinogenes* was grown in *Wolinella* medium (as previously described by the German Collection of Microorganisms and Cell Cultures, DSMZ). Cultures were grown in 160ml serum bottles with boiled degassed medium and closed with butyl rubber stoppers. *Pseudomonas stutzeri* OX1 was grown in Luria Bertani (LB) broth and used as a denitrifying reference strain. Strains used in this study were provided by Dr. Frank Löffler, School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, GA. Denitrifying microorganisms were provided by the Center for Microbial Ecology (CME) at Michigan State University (East Lansing, MI), and the Tropical Microbial Ecology Laboratory (TMEL) (University of Puerto Rico at Mayagüez).

DNA and RNA extractions

Both pure and environmental samples were used in this study. For pure cultures, 1ml of each fully-grown bacterial culture (two-day-culture for *Wolinella* and *Anaeromyxobacter*, and seven-day-culture for *Desulfitobacterium*) was used for extraction and purification of total DNA using the Instagene Matrix method (BioRad). Before PCR amplification, purified DNA was centrifuged at 11,000rpm for 2min. *A. dehalogenans* strain 2CP-C DNA was extracted using QIAGEN Genomic-tip (100ml of a two-day-culture in R2A anoxic medium with fumarate). For environmental samples, total community DNA (Table 1) was extracted using the PowerSoil™ DNA Isolation Kit (MoBio Laboratories, Inc).

RNA extractions were done with RNeasy mini kit (QIAGEN) with an additional DNaseI (530kunits/mg solid) (Sigma) treatment after RW1 buffer (10µl of DNaseI [100U/ml] stock solution was added to 70µl DNase buffer [40mM Tris·HCl pH 7.9, 10mM NaCl, 6mM MgCl₂, 0.1mM CaCl₂] and mixed by pipetting only directly onto the RNeasy membrane, and incubated at room temperature for 15min). DNA and RNA concentrations were measured spectrophotometrically at 260nm.

Development of PCR Primers

Thirty two nucleotide sequences were retrieved from Pipeline Functional Gene/Repository (PFGR) (<http://flyingcloud.cme.msu.edu/fungene/>) and aligned using ClustalW (Thompson et al., 1994) (<ftp://ftp.ebi.ac.uk/pub/software/unix/clustalw/>), and

BioEdit Sequence Alignment Editor (Hall, 1999). Primer Select Program (DNASTAR, Inc, Madison, WI) was used to obtain an initial primer set. *In silico* amplifications were performed using FastPCR 4.0.27 (Kalendar, 2004) and insilico.ehu.es (<http://insilico.ehu.es/PCR/>). Nucleotide sequences and their accession number are listed in Appendix 2.

The design of *nosZ* priming sequences specific for *Anaeromyxobacter dehalogenans* strain 2CP-C was obtained using Primer Select Program. A total of 113 forward primers and 119 reverse primers were identified by this program. The specificity of each primer binding site was evaluated using basic local alignment search tool (BLAST) from Joint Genome Institute (JGI) (<http://genome.jgi-psf.org/cgi-bin/runAlignment?db=anade&advanced=1>) and National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/BLAST>). Only those primer sites showing the fewest hits against the BLAST database were selected.

The level of conservancy was evaluated for each primer with few hits in BLAST. Characteristics such as a length of 17bp or more, high melting temperature (T_m , temperature at which half of the primer population is denatured), and high GC near the 3' end were preferred. The primer pairs were chosen based on the product length (>400bp) and a T_m difference (<5°C) between both primers of a primer pair. Potential dimer formation was evaluated using FastPCR 4.0.27.

Due to the differences in the codon usage of different organisms, the design of *nosZ* priming sequences specific for *nosZ* sequences, such as the ones of *W. succinogenes*, *D. aromatica*, *A. dehalogenans*, *M. magneticum*, and *D. hafniense*, was achieved through degenerated primers. These are DNA fragments in which some base positions exhibit flexibility resulting in several primer sequence options. The universal degenerate code is: M=(A/C), R=(A/G), W=(A/T), S=(G/C), Y=(C/T), K=(G/T), V=(A/G/C), H=(A/C/T), D=(A/G/T), B=(C/G/T), N=(A/G/C/T). For example, GCTNATCSGTBT has 24 degeneracies ($4 \times 2 \times 3$). Less than 1024 degeneracies were preferred. Primer binding sites were verified manually. Numbers used for naming primers are relative to the *nosZ* sequence gene of *A. dehalogenans* strain 2CP-C (Figure 2).

Amplification of nosZ Genes

All oligonucleotide primers were synthesized by Integrated DNA Technologies (IDT, Coralville, IA). To evaluate the effectiveness of *in silico* designed primers, a temperature gradient was used to test each primer pair in a gradient thermocycler (MasterCycler, Eppendorf). The temperature gradient was established based the T_m (reported by the manufacture of the oligonucleotides, IDT), setting the limits of the gradient 3°C above and below the T_m . *A. dehalogenans* strain 2CP-C DNA was used as a model unit to optimize the annealing temperature. The following steps were employed in the PCR (Scala and Kerkhof, 1998): 1 cycle at 94°C for 5min; 35 cycles, each consisting of 95°C for 0.5min, annealing for 1.5min, and 72°C for 2min; and a final extension of 72°C for 10min.

For PCR amplifications approx. 150ng of DNA template, 1X Taq polymerase reaction buffer, 0.2pmol/μl of each primer (0.4-0.8pmol/μl for degenerate primers), and 2.5mM of each dNTP were used in a 20-25μl reaction volume. Different magnesium chloride (MgCl₂) concentrations were evaluated (1.5, 2.0, 2.5, 3.0, 3.5mM). The amplification product was visualized in 1% agarose-gel electrophoresis. Gels were run at 90V for 30min, and DNA bands were stained with ethidium bromide and observed using a ultraviolet transilluminator. After setting these parameters, degenerated primer pairs were tested with DNA from other organisms: *A. dehalogenans* 2CP-1, *A. dehalogenans* 2CP-3, *Anaeromyxobacter* strain R, *Desulfitobacterium hafniense*, *Desulfitobacterium chlororespirans* Co23, *Desulfitobacterium* strain PCE1, *Wolinella succinogenes*, and *Pseudomonas stutzeri*.

Cloning of nosZ PCR Products and Sequencing

Total community DNA of environmental samples from a fluidized bed reactor (FBR), the Mayagüez wastewater treatment plant (MWWT) and from from the petrochemical complex of the Commonwealth Refining Company, Inc. (CORCO) (Table 1) were PCR amplified with NosZ943F/NosZ1789R primer pair and cloned using pGEM (R)-T Vector System (Promega). In the same way, PCR products from genomic DNA from bacterial cultures of *A. dehalogenans* strain 2CP-C and *P. stutzeri* generated using NosZ256F/NosZ1807R, and Nos661/Nos1773 (Scala and Kerkhof, 1998) primer pairs, respectively, were cloned. For the environmental samples, JM109 competent cells

provided with the cloning kit were used, while *Escherichia coli* DH5 α was made competent by cold 50mM CaCl₂ treatment and used for cloning of PCR products from bacterial cultures.

For preparing CaCl₂ competent cell stock, one fresh colony of *E. coli* DH5 α was transferred in 5ml of LB broth and 37°C at 150rpm. After an overnight incubation, 250 μ l were transferred to 25ml of LB broth and grown for 4h at 37°C and 150rpm; centrifuged at 2,500rpm for 10min, and the supernatant was discarded. The pellet was gently resuspended in 12.5ml of cold 50mM CaCl₂ and maintained on ice for 30min, centrifuged at 2,500 rpm for 10min and the supernatant was discarded and maintaining the pellet in ice until 1.5ml of 50mM CaCl₂ plus 20% glycerol were added. Aliquots of 100 μ l were prepared and kept at -80°C.

The molar ratio of PCR product to vector was 3:1, except for the MWWT clone library, for which 10:1 and 5:1 ratios were used. Positive clones were identified by white/blue selection (clones that contain PCR products produce white colonies) in Luria Bertani agar with 100 μ g/ml ampicillin, 0.5mM IPTG, and 80 μ g/ml X-Gal (LB/Amp/IPTG/X-gal). White colonies were picked and transferred to 10 μ l double distilled water (ddH₂O), and isolated in LB/Amp/IPTG/X-gal.. The selected colonies were heated to 100°C for 10min, centrifuged at 16,000g for 5min, and used as DNA template (1 μ l) in a colony-PCR using either T7 promoter and SP6 promoter, or M13F(-20) and M13R(-27) vector primer sets. For PCR amplifications, 1X Taq polymerase reaction buffer, 1pmol of each primer, 2.5mM of MgCl₂, and 2.5mM of each dNTP were used in a 50 μ l reaction volume. The

following steps were employed in the PCR: 1 cycle at 94°C for 2min; 20 cycles each consisting of 94°C for 0.5min, 52°C for 2min, and 72°C for 2min; and a final extension of 72°C for 10min. Inserts of the correct size were examined by agarose-gel electrophoresis. Positive clones were sequenced using Nevada Genomics Center services. For sequencing, only 20 cycles were performed for each PCR amplification (to prevent accumulation of mistakes by Taq polymerase), and two replicates of 50µl were prepared. Replicates were mixed and PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega). The DNA concentration of clean PCR products was measured spectrophotometrically at 260nm.

Sequence and Phylogenetic Analysis

The identities of clone sequence data from FBR, CORCO, and MWWT libraries were compared with sequences in the Genbank using BLASTn. Sequences were edited manually using BioEdit; vector sequences were deleted and in the case of sequences that were inserted in a reverse orientation during pGEM cloning, their reverse-complement sequences were used. Clone sequences with hits within *nosZ* genes sequences were subjected to unique genotypes analysis for each clone library using the PHYLIP package (Felsenstein, 1989) and coverage analysis using ASLO program (<http://www.aslo.org/lomethods/free/2004/0114a.html>) (Kemp and Aller, 2004) under the default parameters. Each clone sequence library was aligned independently using ClustalW (Thompson et al., 1994) and each file *.aln was used to construct a distance matrix by using DNADIST (PHYLIP

<http://evolution.genetics.washington.edu/phylip.html>). Each outfile was then used in DOTUR 1.53 software (Schloss and Handelsman, 2005) applying the furthest-neighbor algorithm to assign sequences genotypes. Rarefaction curves were constructed for each distance level (<http://www.plantpath.wisc.edu/fac/joh/dotur.html>), and richness estimators and diversity indexes were calculated (Table 6). Only clones that represent a unique genotype with less than 97-99% of sequence similarity with other genotypes were subjected to a second round of sequencing in the opposite direction (Nevada Genomics Center).

Sequences were translated *in silico* using BioEdit Sequence Alignment Editor (Hall, 1999) and a protein similarity matrix was generated using MatGat Matrix Global Alignment Tool (Campanella et al., 2003). Only the region of approximately 290 amino acids flanked by primers NosZ943F and NosZ1789R was analyzed. Blosum62 was used as the similarity matrix.

Neighbor-joining (NJ) phylogenetic analysis was conducted using MEGA version 3.1 (Kumar et al., 2004) using 1,000 replicate trees based on the alignment of approx. 180 nucleotides and 60 amino acid residues of the NosZ fragments. After this preliminary NJ analysis, the dataset were subjected to a Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI) phylogenetic analysis. For MP phylogenetic analysis, MEGA 3.1 program (Kumar et al., 2004) was used with 1000 bootstrappings. To generate the ML phylogenetic tree, the optimal evolutionary model (WAG+I+G) was obtained from ProtTest 1.4 (Abascal et al. 2005), then Seqboot and

Protmlk from Phylip 3.67 package (Felsenstein, 1989) were used for bootstrapping and for ML analysis. BI analysis was done using Mr. Bayes, with WAG+I+G as evolutionary model. The following parameters were used: 500,000 generations, Nrun=4, Nchains=4, Temp=0.2, Swapfreq=1, samplefreq=100, relburnin=yes, burninfrac=0.25. Phylogenetic trees were viewed using Treeview (Roderick 2000).

Terminal Restriction Fragment Length Polymorphism of nosZ Amplicons

Reverse primer NosZ1789R was labeled with the fluorophor InfraRedDye700 (LI-COR, Lincoln, NE). Total community DNA from the aerobic and anaerobic stations of the Mayagüez wastewater treatment plant (MWWT) collected in May 2007 was PCR amplified with NosZ943F/NosZ1789R primer set as previously described. Additionally, clone MWWT34 containing an amplicon generated with primer pair NosZ943F/NosZ1789R was included as a control. Two replicates were done, and electrophoresed in 1% agarose gels. The bands were cut out of the gel, and then replicates were mixed and purified using Spin Columns and Elution Tubes (Ambion, Austin, TX).

Most suitable restriction enzymes for amplicons generated with NosZ943F/1789R were identified by an *in silico* analysis using BioEdit restriction map tool. All 4-base cutter enzymes were evaluated, and enzymes that cut all amplicons (Table 7 and 8) and generate terminal fragments between 50 and 700bp were preferred. PCR products were used for separate restriction digests with the following enzymes: *HinPI*, *NlaIII*, and *RsaI*. Each digestion reaction consisted of 0.25µl of 10X reaction buffer, 1 unit of each restriction enzyme and 2µl of PCR product added for a total reaction volume of 2.5µl. The digestions were incubated at 37°C for 4h followed by 10min at 65°C to inactivate the

enzymes. A volume of 2µl of each restriction reaction was mixed 1:1 with IR2 stop solution (LI-COR, Lincoln, NE) and electrophoresed using the services of the Sequencing and Genotyping facility (SGf) (University of Puerto Rico, Rio Piedras Campus). KB^{Plus}-LICOR (50-700bp) was used as a molecular sizing standard. The T-RFLP fingerprints were collected as a TIF image and analyzed using Gel Pro Analyser (Media Cybernetics, Silver Spring, Maryland, USA).

In Situ RT-PCR

In Situ RT-PCR (ISRT-PCR) allows the detection of gene expression inside the cell. This is accomplished by cell permeabilization, after which PCR and hybridization reactants can pass through the cell wall and enter the complex cell matrix. To achieve specificity and sensitivity this procedure has to be carefully optimized. Several key steps are: protease digestion, which can define the intensity of the signal; one-step RT-PCR systems, which allow reproducible amplification; and FISH as a downstream procedure, in which an internal, fluorescently labeled probe is specific to the PCR product generated in the previous step and this allows a more specific detection.

The ISRT-PCR protocol used in this study was previously described by Cordero (2003) with modifications to optimize the procedure. *In Situ* RT-PCR was developed in an RNase-free environment, cleaning equipment and lab bench daily with Eliminase (Decon Labs), treating all solutions with 0.1% (v/v) DEPC, oven baking glassware at 240°C overnight, and rinsing plasticware with Eliminase and DEPC-treated ddH₂O.

Bacterial cell fixation. Bacterial samples were mixed with RNAProtect Bacteria Reagent (QIAGEN) for 5min at room temperature. The cells were treated with 50ml freshly prepared 4% paraformaldehyde and incubated for 2h. After paraformaldehyde treatment, cells were harvested by centrifugation at 5,000rpm for 5min and washed with 50% ethanol in 1X phosphate-buffered saline (PBS: 120mM NaCl and 2.7mM KCl in 10mM phosphate buffer [pH 7.6]). Cells were centrifuged again at 5,000rpm for 5min, resuspended in absolute ethanol, and stored at -80°C until analysis. Fixed cells were used before one week of storage. Cells were recovered by centrifugation and resuspended in 5ml 1X PBS, and 10µl of cell suspension were spotted on IS-PCR designed glass slides (Applied Biosystems, Foster City, CA). Slides were incubated in 0.1% fresh DEPC for 12min at room temperature, and immersed for 2 to 3s in 0.5% metaphor agarose at 37°C, and placed on ice until the agarose had solidified. Excess agarose on the back side of the slides was removed, and the samples were dehydrated in 50, 80, and 96% (vol/vol) ethanol for 5min each (Daims et al., 2001).

Cell permeabilization. A range of conditions were evaluated to optimize cell permeabilization by either electroporation or enzymatic treatment with lysozyme/proteinase K. An overnight RNA- and DNA-nuclease treatment was explored to quickly test the effectiveness of these treatments on cell integrity, using 1mg/ml RNase A and 100U/ml DNase with a subsequent staining of samples with 10ng/µl of 4',6'-diamidino-2-phenylindole (DAPI) (Sigma, St. Louis, MO) for 5min (Hodson et al., 1995). Removal of nucleic acids (lack of fluorescence) caused by the entrance of

nucleases through the cell envelope was used as one criterion of permeabilization success. All samples were examined by epifluorescence microscopy (Olympus BH2-RFCA, Olympus, Japan).

Permeabilization by electroporation was evaluated as follows: a density of 10^7 fixed cells/ml was washed twice with ddH₂O and subjected to successive pulses with electroporation time constants of 10, 25, 50, 100, or 200ms at 2.5kV, 25 μ F, and 1000 Ω . Cold water and 1mm gap electroporation cuvettes (Molecular Bioproducts) were used.

Enzymatic cell permeabilization was evaluated by three parameters: (i) varying the time of exposure to paraformaldehyde between 2 and 24h; (ii) the time of exposure to 1mg/ml lysozyme from 15 to 30min at 37°C; and (iii) using proteinase K (0.5, 1, 2, 5, and 10 μ g/ml final concentration) for 10min at 37°C. A 10 μ l of cell suspension was spotted on a glass slide and incubated with the enzyme mixture. After the incubation, reagents were removed by three consecutive washes in coupling jars with 1X PBS and a final wash with absolute ethanol. Proteolytic enzymes were inactivated by heating the slides for 2min at 94°C (Hodson et al., 1995). After complete processing of samples by ISRT-PCR and FISH detection (see below), the fluorescent intensity and cell morphology were used to identify the gentlest effective treatment for further analysis.

ISRT-PCR protocol. PCR amplification of *nosZ* gene fragments was performed using the QIAGEN OneStep RT-PCR Kit and the primer pair NosZ334F/1789R. The spotted cells were then covered with 50 μ l of reaction mixture (30 μ l of RNase free water, 10 μ l of 5X RT-PCR Buffer, 2 μ l of dNTP Mix 10mM each, 0.6 μ M of each primer, and 2 μ l of

RT-PCR Enzyme Mix) and sealed using the assembly tool (Perkin Elmer, Norwalk, CT) to create a micro-humid chamber. The reactions were performed in an automated *in situ* PCR thermal cycler (Gene Amp *In Situ* PCR System 1000 Perkin Elmer, Norwalk, CT). The following temperatures and time profiles were used: reverse transcription at 50°C for 30min, a hotstart activation step at 95°C for 15min; followed by 35 cycles each consisting of melting at 94°C for 0.5min, annealing at 57°C for 1.5min and extension at 72°C for 2min; and a final extension was performed at 72°C for 10min. After amplification, the slides were washed twice with 1X PBS.

Detection of amplified gene products by fluorescent *in situ* hybridization (FISH).

After ISRT-PCR, slides were incubated in denaturation buffer (0.5M NaOH/1.5M NaCl) at room temperature for 10min. Subsequently, samples were transferred to neutralizing buffer (1M Tris-HCl [pH 8.0] 1.5M NaCl) for 5min at room temperature and then serially dehydrated with ethanol (50, 80, and 100%). The slides were prehybridized in DIG Easy Hybridization Solution (Roche Applied Science, Germany) for 30min. The cyanine dye (Cy3) labeled NosZ943 probe (probe Nos1527 [Scala and Kerkhof, 1998] was used for negative controls) was boiled for 5min and kept at 4°C. Prior to use, the probe was diluted 1:10 in DIG Easy hyb solution to achieve a final concentration of 5ng/μl. The spotted cells were then covered with 30μl of the hybridization mixture, sealed using the assembly tool (Perkin Elmer, Norwalk, CT) and incubated 24(*nirS*)-48h (*nosZ*) at 41°C. After hybridization, the slides were washed twice at 41°C with pre-warmed hybridization buffer (900mM NaCl, 20mM Tris-HCl [pH 7.2] and 0.01% SDS) at 48°C for 20min. Finally, the slides were washed with washing buffer (900mM NaCl/100mM Tris-HCl

[pH 7.2] for 5min at 41°C. Samples were counterstained with DAPI (10ng/μl for 5min) and let air dry. Slides were covered with a drop of UV mount media and a coverslip with nail polish around its periphery. Finally, they were examined under epifluorescence microscopy (Olympus BH2-RFCA, Olympus, Japan). Photographs were taken using a digital SPOT Insight Color Camera and the corresponding SPOT Advanced software V 3.2 (Diagnostic Instruments, USA). UG-1 filter (Olympus, Japan) and XF108-2 filter set (Omega Optical, Brattleboro, Vermont) were used to visualize DAPI and Cy3 signals, respectively. Samples were also examined using a Laser Confocal Microscope model Olympus FluoView 300 at the UPRM.

In order to evaluate the group specificity of each probe, two clones of *E. coli* DH5α were generated which carry the *A. dehalogenans* strain 2CP-C and the *P. stutzeri nosZ* gen fragments, respectively, generated as described in “cloning of *nosZ* PCR products and sequencing”. Hybridization conditions were examined including probe concentration (5 and 25ng/μl), hybridization time (24 and 48h; as recommended for degenerated probes) (Sambrook et al., 1989, p 11.45), and washing stringency (2min room temperature following by 2min at hybridization temperature; and 10min at 48°C).

ISRT-PCR protocol-Environmental samples. Modifications were made to apply the protocol to environmental samples. At the moment of sampling, a 500μl aliquot of the bioreactor was removed and mixed 1:1 with RNAprotect Bacteria Reagent (QIAGEN), or it was filtered through a IsoporeTM polycarbonate membrane filter (pore size, 0.2μm) (Millipore Corp., Bedford, MA) or through a black polycarbonate membrane filter (pore

size, 0.2 μ m) (Poretics Corp., Livermore, CA) (only a small volume of approx. 100 μ l passed through the filter), then the membrane filter was removed and immersed in RNAProtect Bacteria Reagent. The suspension or the membrane filter were treated with 50ml of 4% paraformaldehyde for 2h. In the case of the cell suspension, it was stirred constantly on a magnetic stirring hot plate and cells were recovered by centrifugation at 5,000rpm for 5min, they were resuspended in 10ml of 1X PBS, and 10 μ l of cell suspension was spotted on IS-PCR designed glass slides and let dry in a 55°C incubator. Slides and membrane filters were treated with 0.1% DEPC, immersed in agarose, dehydrated in ethanol, and cell permeabilized with lysozyme and proteinase K, as described previously. After cell permeabilization, slides and membrane filters were boiled in 10% acid-washed polyvinylpolypyrrolidone (PVPP) (Sigma) in distilled water for 10 min (PVPP reduces PCR inhibition) (Holben et al., 1988; Morgan et al., 1998). In the case of glass slides, they were sealed using an assembly tool (Perkin Elmer, Norwalk, CT) and they were heated in an *In Situ* PCR thermal cycler (GeneAmp In Situ PCR System 1000 Perkin Elmer, Norwalk, CT). PVPP was removed by three consecutive washes with 1X PBS, and then glass slides and membrane filters were serially dehydrated with ethanol (50, 80, and 100%) for 3min each. PCR and FISH were performed as described previously. In the case of membrane filters, PCR and FISH were performed in PCR tubes in a Perkin Elmer model 2400 thermal cycler (GenAmp PCR System). After DAPI counterstaining, membrane filters were placed onto glass slides and covered with UV mount media and a coverslip.

RESULTS

Development of PCR primers

Primers specific to *nosZ* from *Anaeromyxobacter* spp.

Identification *in silico* of specific priming sites to amplify *Anaeromyxobacter nosZ* gene sequences was accomplished by Primer Select Software (DNASTar), with the exception of primers NosZ256F and NosZ1366F which were designed manually. Among 113 suggested forward and 119 reverse primers, only six sites were selected for further analysis (Table 2). The selection of primers was based on length (17 to 27bp), high T_m, higher GC content near the 3' end, and specificity suggested by BLAST analysis. Importantly, a region of high similarity of nucleotides (Figure 3) among *Anaeromyxobacter* and other non-denitrifying strains was used as a critical parameter. The presence of unique bases to enhance the probability that the primers work for all *Anaeromyxobacter* strains but not for other species was considered as well.

For primer testing, genomic DNA was used from several *Anaeromyxobacter* species (*A. dehalogenans* strain 2CP-C, *A. dehalogenans* strain 2CP-1, *A. dehalogenans* strain 2CP-3, and *Anaeromyxobacter* strain R), as well as from other members of the non-denitrifying group (*W. succinogenes*, and *Desulfitobacterium* spp), from one organism phylogenetically related to *Anaeromyxobacter* spp. (*Myxococcus xanthus*), and from an organism lacking the *nosZ* gene (*Dehalococcoides* sp. strain BAV1). Optimal PCR conditions were achieved for seven primer pairs (Table 3): NosZ187F/1207R,

NosZ256F/718R, NosZ334F/1789R, NosZ903F/1207R, NosZ903F/1789R, NosZ903F/1807R, and NosZ943F/1789R. Figures 5 and 6 show an example of a gradient PCR product specifically for primer pair NosZ943F/1789R. Results for each primer pair will be described in decreasing order of importance to the project.

The use of primer pair NosZ903F/1789R (Figure 8) yielded the expected 913bp PCR product using any of the annealing temperatures tested (55-65°C). A temperature of 63°C worked best for this primer pair. Using this primer set, successful and specific amplification for a *nosZ* fragment from all *Anaeromyxobacter* strains occurred, and no PCR products from other non-denitrifying organisms were observed (*M. xanthus* was not tested). The optimum magnesium concentration was 2.5mM.

PCRs using primer pair NosZ256F/718R (Figure 7 and Table 3) yielded a PCR product of the expected size (483bp) at all tested annealing temperatures (55-68°C); however, higher temperatures within this range generated better amplification yields (brighter bands). For this primer pair, 68°C was selected for further applications. This primer set generated the expected amplicon with genomic DNA from all *Anaeromyxobacter* strains, although for *A. dehalogenans* strain 2CP-1 a low yield was obtained. Along other non-denitrifying organisms, *D. hafniense* and *D. chlororespirans* yielded a very weak PCR product as well. A strong amplification product was obtained for *M. xanthus* whereas amplification with *W. succinogenes* and *D. ethenogenes* DNA yielded no products. Figure 3 shows that for the forward primer NosZ256F, *D. hafniense* shares many of the

bases along the priming site, while priming site for NosZ718R differs more. The optimum magnesium concentration was 2mM.

PCR conditions for NosZ187F/1207R, NosZ903F/1207R, NosZ903F/1807R are also described in Table 3. The PCR products of the expected size were obtained for each primer pair, and were 1043, 924, and 327bp in length, respectively, with DNA from *Anaeromyxobacter* strains. Primer set NosZ903F/1807R and NosZ903F/1207R amplified *W. succinogenes* *nosZ* gene robustly. In contrast, NosZ187F/1207R generated multiple unspecific fragments in *D. hafniense*, *D. chlororespirans* Co23, *W. succinogenes*, and *D. ethenogenes*. No other primer pair generated an amplicon for *D. ethenogenes*.

Primers specific to novel *nosZ* genes

In order to retrieve *nosZ* sequences from environmental samples, degenerate primers were designed based on *A. dehalogenans* strain 2CP-C, *W. succinogenes*, *D. hafniense*, *D. aromatica* and *M. magnetotacticum* *nosZ* sequences. Only these *nosZ* sequences from non-denitrifying strains were available in public databases at the beginning of this study. A primer binding site suitable for amplification of all *nosZ* varieties, including traditional ones could not be established (sequences used in the multiple sequence alignment can be found in Appendix 2). Due to observed differences in codon usage, Primer Select program failed to identify a region of conserved nucleotides long enough to serve as a primer binding site. Therefore, target priming sites were located manually (Figure 9).

In order to identify PCR primers suitable for *nosZ* genes from a wide range of taxonomic groups, degenerate primers were designed. Primer pair NosZ334F/1789R generated a larger amplicon of approx. 1485bp, while NosZ943F/1789R produced a smaller amplicon of approx. 870bp. Figure 9 shows the priming sites for these degenerate primers in an alignment of partial *nosZ* sequences of members from the non-denitrifying *nosZ* group. Figure 9 also shows major priming differences with members of the traditional-*nosZ* group (*Achromobacter cycloclastes*, *Bradyrhizobium japonicum*, *Brucella* spp., *Burkholderia* spp., *Paracoccus pantotrophus*, *Pseudomonas* spp., *Sinorhizobium meliloti*, among others).

The amplification reaction using primer pair NosZ334F/1789R (Figure 11) generated a PCR product of the expected size (1482bp) with a wide range of annealing temperatures (57 to 65°C); 57°C was selected as optimal, while 66°C yielded no amplification. All *Anaeromyxobacter* and *Desulfitobacterium* strains yielded amplicons, although *Desulfitobacterium* PCE-1 DNA required a second round of PCR with the same primer pair for its detection. For *W. succinogenes*, a larger PCR fragment of nearly 3000bp was obtained. Using an *in silico* PCR software (<http://insilico.ehu.es/PCR/>), these primers were evaluated against the *W. succinogenes* genome (insilico.ehu.es). According to the genome sequence, NosZ334F/1789R should produce an amplicon of 2844bp in length (positions 870868-873711), which matches the size of the electrophoresis gel band of about 3000bp (Figure 11, lane 7). The same type of *in silico* amplification with *A. dehalogenans* strain 2CP-C genome generated a 1481bp long product (positions 2709144-2710624), also in agreement with the experimental PCR results.

Optimization of the annealing temperature was a critical aspect for primer pair NosZ943F/1789R, since at lower annealing temperatures the main observed PCR product was larger than the expected (Figure 5, lanes 1-4, correspond to 40, 41.2, 43.3, 46.3°C). Temperatures between 52 to 62°C (Figure 5, lanes 6-8; and Figure 6 lanes 1-7) generated a PCR product of the predicted size (873bp), with 59°C as the best annealing temperature for this primer pair. However, due to the high level of degeneracy of primer NosZ943F (144 degenerate positions), additional amplicons were always observed, and the specificity could not be improved by varying the magnesium concentration (1.5, 2, 2.5, 3, and 3.5mM). Therefore, this PCR product required clean-up steps of the desired fragment before further analysis could be done. Among the other non-denitrifying organisms, NosZ943F/1789R produced an amplicon of the correct size with DNA from *Desulfitobacterium* spp. (*W. succinogenes* was not tested).

The primer pair NosZ943F/1789R was also used to evaluate environmental DNA samples obtained from a fluidized bed reactor (FBR) treating diesel contaminated groundwater, soil samples from a petrochemical complex (CORCO), activated sludge from the largest secondary wastewater treatment unit operated in Puerto Rico (MWWT), marine communities from the Mayaguez Bay, and soil samples from the “Bosque del Pueblo” rainforest. Since it is common to find humic acids in environmental samples which are inhibitory for PCR amplifications, replicates of PCR reactions were prepared for each environmental sample, one with total environmental DNA, and the other reaction with total environmental DNA combined with *A. dehalogenans* strain 2CP-C DNA. All

samples generated amplicons of the expected size with the exception of the marine bacterioplankton community from Mayagüez Bay, thus demonstrating widespread distribution of novel *nosZ* genes in nature (Figure 12, lanes 9-20). NosZ334F/1789R primer pair was also tested with monthly samples from Mayagüez Wastewater Treatment Plant (MWWT) (Figure 12, lanes 1-8) and yielded positive amplification (Figure 13, lanes 5-5).

Finally, primer pair NosZ943F/1789R was chosen for assessing diversity of *nosZ* in environmental samples by clone sequencing and by T-RFLP. Figure 13 shows the screening of positive clones using this primer pair.

Environmental nosZ-gene libraries

The *nosZ* gene family appears to be widely distributed and highly diverse in nature (Rösch et al., 2002), but little information is available for the non-denitrifying NosZ group. Using degenerate primers, a broad range of *nosZ* sequences were sought in order to compare environmental sequences from Puerto Rico with those previously published and available in public databases. A total of 104 sequences of approx. 870bp was obtained: 12 from the FBR, 24 from CORCO, and 68 from MWWT. Based on nucleotide sequences from reference strains, a protein prediction was prepared for clone sequences using BioEdit Sequence Alignment Editor. The priming regions were deleted in order to avoid mistranslations, since only a one-base error during sequencing may result in a different protein. To prevent these mistakes, only clones that were sequenced in both

directions were included in this analysis. The second-round sequencing did not yield the complete PCR fragment and a stretch of approx. 400bp was represented by a double strand sequence. Translated *nosZ* genes revealed the presence of conserved histidine residues in all the environmental gene sequences (Simon et al., 2004) (Figure 14).

Through a Similarity Matrix analysis (MatGat) (Table 4), protein similarities were established for clone sequences and reference sequences. FBR1, FBR13, FBR31, FBR35, CORCO6, MWWT29, MWWT31, MWWT43, MWWT72, MWWT73, MWWT82, MWWT83, MWWT86, MWWT90, MWWT94, and MWWT102 had a percentage of similarity of 94, 94, 96, 94, 95, 94, 97, 84, 93, 94, 93, 92, 97, 93, 92, and 89%, respectively, with *D. aromatica*. FBR15, FBR25, FBR28, CORCO8, CORCO26, CORCO30, and MWWT24 had a protein similarity with NosZ from *Anaeromyxobacter dehalogenans* strain 2CP-C of 96, 91, 96, 88, 82, 88, and 87%, respectively. The highest protein similarity of CORCO7 was 68% with *M. gryphawaldense*. CORCO18 has a protein similarity of 71% with *C. fetus* and *M. gryphawaldense*. Clones CORCO32 and CORCO33 have 86% similarity with each other. The highest protein similarities with a reference strain were 66 and 67%, respectively, with *M. gryphawaldense*.

A Similarity Matrix Analysis was performed to reference traditional-*nosZ* sequences (Table 5). Members of the order Rhizobiales (*Sinorhizobium meliloti*, *Brucella ovis*, *Rhodopseudomonas palustris*, *Bradyrhizobium japonicum*, *Methylobacterium* sp.) shared between 89 to 96% similarity, while similarity percentages in Rhodobacterales (*Stappia aggregata*, *Silicibacter pomeroyi*, *Roseovarius* spp., *Dinoroseobacter shibae*, and

Rhodobacter sphaeroides) ranged between 84 to 92. Members of α -Proteobacteria, such as Rhizobiales (*Sinorhizobium meliloti*, *Brucella ovis*, *Rhodopseudomonas palustris*, *Bradyrhizobium japonicum*, *Methylobacterium* sp.) and Rhodobacterales (*Paracoccus pantotrophus*, *Stappia aggregata*, *Silicibacter pomeroyi*, *Roseovarius* sp., *Dinoroseobacter shibae*, and *Rhodobacter sphaeroides*) tend to share only 77 to 84% protein similarity. Members of the β -Proteobacteria including *Burkholderia pseudomallei*, *Cupriavidus necator*, *Azoarcus* sp., *Thiobacillus denitrificans* (Tm. *denitrificans*), and *Acidovorax* sp. share a protein similarity between 81-95%. Representatives of the γ -Proteobacteria such as *Alkalilimnicola ehrlichei*, *Reinekea* sp., *Psychromonas ingrahamii*, *Colwellia psychrerythraea*, *Photobacterium profundum*, *Shewanella loihica*, *Marinobacter* sp., *Hahella chejuensis*, and *Pseudomonas aeruginosa*, share 83-96% of protein similarity. Some members of different divisions within the Proteobacteria share more than 90% protein similarity. This is the case of *Paracoccus pantotrophus*, an α -proteobacterium whose NosZ is more similar to that of γ -Proteobacteria *Pseudomonas aeruginosa* (92%), *Pseudomonas fluorescens* (92%), and *Pseudomonas denitrificans* (*Ps. denitrificans*) (97%), while protein similarity ranges 70-75% relative to the Rhodobacterales previously described. This is also the case for *Achromobacter cycloclastes*, which belongs to the order Burkholderiales, and only shares 67-68% similarity to other Burkholderiales, like *Burkholderia pseudomallei*, *Cupriavidus necator*, and *Acidovorax* sp.

Phylogenetic analysis of nosZ genes

The phylogenetic analysis revealed that the new sequences from environmental samples were different from traditional-*nosZ* sequences, while exhibiting more similarity relative to the non-denitrifying NosZ group. None of the clonal sequences shares more than 62% of protein similarity with traditional-*nosZ* sequences (similarity matrix not shown). To further confirm the identity of environmental sequences, a phylogenetic analysis was conducted for the NosZ family based on 46 nearly full-length sequences available in PFGR (<http://flyingcloud.cme.msu.edu/fungene/>). The inclusion of sequences from the environmental clone libraries is especially useful, since traditional and non-denitrifying N₂ORs have distinctive traits, and larger number of sequences could give insights into their evolution. The amino acid and the nucleotide sequences were analyzed independently, and most branches were very well supported, based on bootstrap analysis (1000 replicates). First, phylogenetic neighbor-joining trees were constructed with *nosZ* full-length sequences without clone sequences; then the 46 full-length sequences of reference were truncated in the same region for which clone sequence information is available (approx. 0.9kb). Another phylogenetic tree was constructed based on the NosZ region available for the bone fish *Crenimugil crenilabis*, for which a *nosZ* conserved domain of about 90 amino acids was identified with NCBI Conserved Domain Search tool (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?>). To accomplish this goal, all sequences including clone sequences were truncated to approximately 0.2kb (corresponding to the *C. crenilabis* 90 amino acid domain) and all characters where gaps appeared were deleted (Figure 15). The topologies of 0.9 and 0.2kb based trees were congruent, and fit the topology of trees constructed with full-length sequences (not shown). The addition of clone sequences did not change the overall phylogenetic

relationships between reference sequences in consensus trees, but largely amplified the information obtained for *nosZ* genes.

A phylogenetic analysis with Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI) algorithms was followed (Figure 16, 17, and 18, respectively). Through the phylogenetic analysis, 9 clades of *nosZ* variants were identified with the 3 models: Clade I, *Dechloromonas aromatica*, FBR31, FBR35, FBR13, CORCO6, FBR1, FBR4, MWWT31, MWWT90, MWWT82, MWWT72, MWWT94, MWWT83, MWWT73, MWWT29, MWWT86; Clade II, *Magnetospirillum gryphiswaldense*, *Magnetospirillum magneticum*, *Magnetospirillum magnetotacticum*; Clade III, *Thiomicrospira denitrificans*, *Wolinella succinogenes*, *Campylobacter fetus*; Clade IV, CORCO32, CORCO33, CORCO7; Clade V, *Psychroflexus torquis*, *Robiginitalea biformata*, CORCO47, *Desulfitobacterium hafniense*, *Geobacillus thermodenitrificans*, *Gramella forsetii*; Clade VI, *Anaeromyxobacter dehalogenans* strain 2CP-C, *Anaeromyxobacter* sp. strain Fw109-5, CORCO8, CORCO26, CORCO30, FBR15, FBR25, FBR28, MWWT24; Clade VII, *Halorubrum lacusprofundi* and *Haloarcula marismortui*; Clade VIII, *Azoarcus* sp., *Thiobacillus denitrificans*, *Burkholderia pseudomallei*, *Acidovorax* sp, *Cupriavidus necator*; and Clade IX, *Achromobacter cycloclastes*, *Alkalilimnicola ehrlichei*, *Bradyrhizobium japonicum*, *Brucella ovis*, *Colwellia psychrerythraea*, *Dinoroseobacter shibae*, *Hahella chejuensis*, *Marinobacter* sp, *Methylobacterium* sp, *Paracoccus pantotrophus*, *Photobacterium profundum*, *Pseudomonas aeruginosa*, *Pseudomonas denitrificans*, *Pseudomonas stutzeri*, *Psychromonas ingrahamii*, *Reinekea* sp., *Rhodobacter sphaeroides*,

Rhodopseudomonas palustris, *Roseovarius sp*, *Shewanella loihica*, *Silicibacter pomeroyi*, *Sinorhizobium meliloti*, *Stappia aggregata*. Clones CORCO18, MWWT102, MWWT43 and *Pyrobaculum calidifontis* were not included in any particular clade as their phylogenetic relationship varies among the different models.

With the similarity matrix made using MatGat, protein similarity percentage can be analyzed for each clade in the MP phylogenetic tree (Figure 17). Protein similarity for each clade was as follows: 91-100% for clade I; 89-99% for clade II; 83-91% for clade III; 75-86% for clade IV; 59-90% for clade V; 81-96% for clade VI; 77% for clade VII; 81-95% for clade VIII; and 70-97% for clade IX. *P. calidifontis* and CORCO18, which were not included in any particular clade, show a protein similarity of 70%. CORCO47, which belongs to clade V, show its highest protein similarity with members of the clade VI (71-75%).

Traditional reference *nosZ* sequences belong to clades VIII and IX. Clade VII consisted of two archaeas, *Halorubrum lacusprofundi* and *Haloarcula marismortui*. The majority of environmental clone sequences fell into clades I, IV, V, and VI, and did not follow a geographical pattern (except for clade IV). Clade IV was composed exclusively of CORCO clones. Clone CORCO18 by ML and BI models show a close relationship with the archaea *Pyrobaculum calidifontis*, while CORCO47 was most related to Flavobacteria. Clones MWWT43 and MWWT102 fell in two distinct clades (I and II) most closely related to either *D. aromatica* or *Magnetospirillum* spp. Clades I and VI lost resolution in ML and BI models, respectively. Since these clades kept the same overall

topology in the other models, this may indicate that the support value is not high enough thus collapsing during the inquiry, and more sequence data is needed.

Diversity indexes and richness estimators of nosZ-gene libraries

Since all environmental clones were randomly sequenced without any previous selection, CORCO, FBR, and MWWT libraries could be subjected to a diversity and richness analysis using PHYLIP package. Nine (9), 14, and 30 unique genotypes were identified for FBR (α 0.03), CORCO (α 0.04), and MWWT (α 0.04), respectively. The estimated coverage was 33, 54, and 72% for FBR, CORCO, and MWWT *nosZ*-gene libraries.

Rarefaction curves (Figure 19) are used to determine differences in relative richness between libraries that would be observed if sampling were to be continued. If the slope of the curve is too steep, the sampling and sequencing efforts were incomplete. For CORCO library, it was observed that when a distance of 27% was used to define an OTU, the sampling was not completed. However, when a distance of 21% was used to define an OTU for FBR library, the rarefaction curve began to stabilize, showing that this library could harbor a lower number of OTUs.

DOTUR (Distance-based OTU and Richness) is a method used to compare libraries statistically. It defines an OTU using every possible distance between sequences, and it calculates Shannon-Weaver and Simpson diversity indexes (Magurran 1988), and the abundance-based coverage estimator (ACE) (Chao and Lee 1992; Chao and Yang 1993), bias-corrected Chao1 (Chao 1984), interpolated jackknife (Burnham and Overton 1979),

and bootstrap (Smith and van Belle 1984) richness estimators. The estimated indexes are shown in Table 6. This analysis shows that the Shannon and Simpson diversity indexes (representing the richness and evenness of a sample, respectively) are higher for CORCO than for FBR library, even when a 4% of difference between OTU's is considered in CORCO clonal library instead of 3%, as for FBR clonal library. The most exhaustively sampling effort was accomplished for the wastewater treatment unit (MWWT). The sampling disparity could influence the richness found for MWWT gene library, which has the highest richness within the three clonal libraries. It also has the highest number of genotypes based on ACE, Boot, Chao, and Jack estimators. Nevertheless, the evenness for MWWT clone library was the lowest. This suggests that even though that MWWT library has the highest number of OTU's, there are sharp differences between the relative abundances of individual OTUs, only a few predominating in the population.

Terminal Restriction Fragment Length Polymorphism of nosZ Amplicons

Figure 26, 27, and 28 show the electropherogram (a plot of results from an analysis done by electrophoresis) profiles generated with *HinPI*, *NlaIII*, and *RsaI*, respectively, for the aerobic (AES) and anaerobic (ANS) stations of Mayagüez Wastewater Treatment Plant samples in May 2007. As a restriction control, clone MWWT34 was also digested with the same restriction enzymes. The T-RF profiles show a higher number of peaks for the ANS sample. In *RsaI* digests, the AES pattern presented 16 distinct peaks, while ANS patterns showed 30 peaks. Among the 16 peaks, only one was unique for the AES, while 15 were unique for ANS. Clone MWWT34 was represented by the peak with a fragment

length of 223bp, which was present in AES and ANS, and corresponds to the fragment length expected by the *in silico* restriction (Table 7).

In *Nla*III digests, the AES pattern presented 12 distinct peaks, while ANS pattern showed 18 peaks. Only the ANS pattern showed unique peaks (Figure 27). Clone MWWT34 was represented by the peak with molecular weight 124bp, which was present in AES and ANS; but in the *in silico* restriction, the T-RF fragment length expected was 117bp (Table 7), which is 7bp shorter than the observed fragment. The reason could be a gel distortion commonly called the “smiley effect”.

In *Hin*PI digests, the AES pattern presented 13 distinct peaks, while ANS pattern showed 19 peaks (Figure 28). Among the 13 peaks, only one was unique for the AES, while 8 were unique for ANS. Clone MWWT34 was represented by the peak with molecular weight 274bp, which was presented only in ANS, and corresponds to the molecular weight expected by the *in silico* restriction (275bp) (Table 7).

In Situ-PCR

Cell Permeabilization

Enzyme permeabilization and electroporation were evaluated for their effectiveness to make stable pores in the cell wall, large enough to allow access of both RNase (approx. 13,000Da) and DNase (approx. 33,000Da) into the cells, while preserving the intact

overall morphology. Both enzymatic treatments and electroporation were performed after cells were fixed with 4% paraformaldehyde. This pre-treatment rigidly fixes cell wall proteins preventing closure of pores after permeabilization (Nuovo 1996; Hixson et al. 1981). The ability of RNases and DNases to cross the cell membranes and to destroy nucleic acids after an overnight incubation was evaluated with DAPI staining, which binds strongly to nucleic acids, and can be observed as a blue fluorescence. The electroporation treatment of 50ms (Figure 20 A and B) was insufficient to promote effective entry of enzymes, since numerous DAPI-emitting cells were observed after nuclease treatment. On the contrary, 200ms (Figure 20 E and F) resulted in no cells being detected with or without nuclease treatment, demonstrating the destruction of all cells. Treatment with 100ms (Figure 20 C and D) provided enough electroporation time to generate large pores that enable the entry of nucleases while the proper cell morphology was retained. This assay was used to test enzymatic permeabilization treatments as well.

The most effective enzymatic cell permeabilization protocol was observed with 1mg/ml lysozyme treatment during 30min of incubation. In addition, proteinase K at 0.5µg/ml was effective with cells fixed for 2h in 4% paraformaldehyde, while 2µg/ml were required for 24h-fixed cells. Nuovo (1996) documented needs for higher concentrations of pronase when digesting cells fixed for longer periods. Nonetheless, the best permeabilization condition was achieved with 2h-fixed cells incubated for 30min with a lysozyme solution and 0.5µg/ml of proteinase K. Although electroporation appears to work as well, the enzymatic permeabilization protocol was chosen for ISRT-PCR.

ISRT-PCR

Prior to *nosZ* *in situ* detection, *nirS* was tested (Figure 21). This gene has been previously used in our lab for the detection of denitrifying populations (Cordero, 2003). After evaluating several cell permeabilization parameters, specificity of the *nirS*-specific primers with pure and mixed cultures was tested. ISRT-PCR amplification was performed and successfully achieved for exponentially growing denitrifying cultures of *P. stutzeri* while background signal was not observed in RNase negative controls (Figure 21 B). When ISRT-PCR was tested with a co-culture of *P. stutzeri* and a non-denitrifying rod, only *Pseudomonas* cells were detected by fluorescence microscopy (Figure 21 A).

After *nirS* ISRT-PCR was optimized and validated, this approach was used for detecting *nosZ* gene products inside active cells. Expression of *nosZ* gene was first analyzed by RT-PCR through RNA extraction and gel-electrophoresis (Figure 22). Results showed robust amplification from *P. stutzeri* cells grown in nitrate broth, and a weak band when cells were grown in BLK (Nitrogen-free medium) with histidine (Figure 22). *E. coli* DH5 α generated a weak band as well (Figure 22, lane 6). After this preliminary assay, the *nosZ* gene of *P. stutzeri* was examined by an *in situ* technique. *P. stutzeri* was grown in nitrate broth and BLK plus histidine, fixed with 4% paraformaldehyde, spotted in glass slides, permeabilized enzymatically, and subjected to ISRT-PCR with primers Nos661F and Nos1773R (Scala and Kerkhof, 1998) with unlabeled nucleotides. After these steps, FISH was used to target an internal region of the previous PCR amplicon with Cy3-labeled Nos1527 probe (Scala and Kerkhof, 1998), and signal-emitting cells were

detected with fluorescence microscopy. A strong Cy3 signal was observed for *P. stutzeri* cells grown in nitrate broth, while a weak Cy3 signal was observed in *P. stutzeri* cells grown in BLK plus histidine as the only nitrogen source (Figure 23 B), but no Cy3 signal was observed in RNase negative control (Figure 23 F).

The detection of novel *nosZ* genes was also achieved by IS-PCR. For this, two *E. coli* DH5 α clones were constructed with pGEM Easy Vector (Promega), one carrying a partial *nosZ* gene from *P. stutzeri* and the other carrying the partial *nosZ* gene from *A. dehalogenans* strain 2CP-C. The following controls were used: (i) *E. coli* DH5 α without pGEM vector, against Probe943 after PCR with NosZ334F/NosZ1789R primer pair, to ensure that *E. coli* genome is not responsible for the detection signal; (ii) clone carrying *Pseudomonas nosZ* partial gene against Probe943 after PCR with Nos661/Nos1773 primer pair, to demonstrate specificity of Probe943; (iii) clone carrying *Anaeromyxobacter nosZ* partial gene against Probe1527 after PCR with NosZ334F/NosZ1789R primer pair, to demonstrate that Probe1527 is not suitable to detect this type of non-denitrifying sequence. IS-PCR of genomic DNA typically fails due to the highly super-coiled structure of this molecule embedded in a dense matrix of proteins and cell structures (Nuovo, 1996). Since pGEM is a high copy vector, it was expected that, through IS-PCR, the insert could be directly detected without the need of the RT step. Results showed that the clone carrying the *Anaeromyxobacter nosZ* gene emits a strong Cy3 signal when detected with Probe943, but not with Probe1527 (Figure 24 B and 25 D, respectively). On the other hand, the clone harboring the *Pseudomonas* gene did not emit a Cy3 signal when detected with Probe943 nor did *E. coli* DH5 α (Figure 24 H and 24 F, respectively).

ISRT-PCR was further assayed with environmental samples obtained at the wastewater treatment plant operated to remove nitrogen by biological means. With laboratory cultures, Cy3-emitting cells could account for all or nearly all cells detected, based on DAPI detection. However, with natural samples the detection was considerably lower-around 4% of Cy3-emitting cells account for all possible cells detected by DAPI staining in the sample (only cells that were clearly distinguished from background signal were counted). The objective was not to prove exhaustively the abundance of these genes in the environment, but to prove that the designed primers could amplify genes that are harbored by bacteria in environmental samples. Figure 25 B shows at least three rods emitting Cy3 signal, while sections 25 D and F show a higher number of cells emitting Cy3 signal. Figures 25 A and 25 B show a high fluorescence background that was observed whenever polycarbonate filters were used. Polycarbonate black filters did not eliminate the background fluorescence. Figure 25 F shows acceptable background levels obtained when glass slides were used.

DISCUSSION

Primers specific to nosZ from Anaeromyxobacter spp.

Development of specific-primers for *Anaeromyxobacter nosZ* gene sequence is useful for molecular monitoring of N₂O reduction by bacterial populations. For example, quantitative molecular studies demonstrate that *Anaeromyxobacter* 16S rRNA gene

sequences are the most abundant group in agricultural soils, rice fields, and contaminated sites (e.g., the uranium and NO₃⁻ contaminated Field Research Center site in Oak Ridge, TN). The selection of primer pairs was based on their specificity suggested by BLAST analysis and empirical results. Those that selectively amplify *Anaeromyxobacter* targets were preferred. The primer pair NosZ903F/1789R (Figure 6) provided enough specificity to amplify all *Anaeromyxobacter* strains, while not amplifying *Desulfotobacterium* spp. Since primer NosZ1789R was designed to amplify non-denitrifying *nosZ* sequences, the specificity must be given by the differences between non-denitrifying sequences and *Anaeromyxobacter* sequences along the priming site of NosZ903F. These two types of sequences indeed have marked differences (Figure 3). For example, five basepairs differentiate *A. dehalogenans* strain 2CP-C priming sequence from *D. hafniense nosZ*-sequence, the most similar known specie among the non-denitrifying *nosZ* priming region. The region in *A. dehalogenans* strain 2CP-C (PFGR), *Anaeromyxobacter* sp. Fw109-5 (PFGR), *A. dehalogenans* strain 2CP-3 (this study, see Appendix 1), and *Anaeromyxobacter* strain R (this study, see Appendix 1) *nosZ* sequences varied in only one position along the 18 nucleotide-long primer.

Differences in the amplification yields were observed for NosZ256F/718R primer pair (Figure 5) among *Anaeromyxobacter* strains. It was therefore predicted that some regions of their NosZ domain differ among these strains. Actually, it was observed through the sequencing of the partial *nosZ*-sequence of *Anaeromyxobacter* strain R and *A. dehalogenans* strain 2CP-3, and through their comparison with *A. dehalogenans* strain 2CP-C (PFGR) and *Anaeromyxobacter* sp. strain Fw109-5 (PFGR) *nosZ*-sequences, that

they differ in some bases. With these four *nosZ*-sequences now available for *Anaeromyxobacter* spp. it is possible to optimize primer design.

Primers Specific to novel nosZ Genes and environmental analysis

Discrimination of *nosZ* varieties of the traditional- and the non-denitrifying group was successfully accomplished with primer pairs NosZ334F/1789R and NosZ943F/1789R. Partial *nosZ* sequences retrieved directly from environmental samples captured new sequences while traditional *nosZ* sequences were selectively excluded. Therefore, field-based studies with emphasis on specific functional microbial groups affecting the N₂O budget are now possible by employing these primers.

Evidence that strongly supports that these *nosZ* gene fragments are involved in N₂O reduction in the environment includes: (i) the presence of the NosZ domain; (ii) the presence of conserved histidine residues, essential for function and structure of N₂OR (Simon et al., 2004) (Figure 14) in all environmental gene sequences; (iii) fragments of the expected size; (iv) high similarity index to *nosZ* sequences from cultured organisms; and (v) detection of microbial cells harboring *nosZ* mRNA by ISRT-PCR using the newly designed primers.

Horn and coworkers (2006) determined that approximately 90% of any two denitrifiers that share >97% 16S rRNA sequence similarity, will also share higher than 90% NosZ protein similarity (65% for nucleic acid similarity). In addition, they predicted that

similarity levels of a translated *nosZ* gene below 68% (or 48% nucleic acid similarity) could be associated to uncultured organisms. Since cloned sequences FBR1, FBR13, FBR31, FBR35, CORCO6, MWWT29, MWWT31, MWWT72, MWWT73, MWWT82, MWWT83, MWWT86, MWWT90 and MWWT94 had a protein similarity of 92-97% relative to *D. aromatica*, and FBR15, FBR25 and FBR28 were 91-96% similar to *Anaeromyxobacter* sp., it is also possible that they share higher than 97% 16S rRNA sequence similarity, and therefore, they could belong to the same genus or to genera closely related to *Dechloromonas* and *Anaeromyxobacter*, respectively. This is the case for some traditional-*nosZ* members of the order Rhizobiales and Rhodobacterales of the α -Proteobacteria, which share 89-96% and 84-92%, respectively, within the members of their taxonomic order. On the contrary, similarity levels for clones CORCO7, CORCO32, and CORCO33 were 68, 66, and 67%, respectively, with their most similar cultured organism. Perhaps these gene sequences are carried by unique and yet to be cultured organisms in the environment.

A protein similarity of 84% to the NosZ of *D. aromatica* was observed for clone MWWT43, while MWWT24, CORCO8, CORCO26 and CORCO30 shared between 82-88% protein similarity to NosZ of *Anaeromyxobacter* spp. At this point, the identity of the taxonomic groups carrying these novel *nosZ* genes is unclear, but these similarity percentages suggest that these clones belong to the classes β - (MWWT43) and δ -Proteobacteria (MWWT24, CORCO8, CORCO26 and CORCO30). A similar inference can be made for members within the α -Proteobacteria belonging to different orders

(Rhizobiales and Rhodobacterales), which tend to share between 77-84% of protein similarity (Table 5).

Phylogenetic analysis of nosZ genes

Through the construction of phylogenetic trees using three models (MP, ML, and BI), it was observed that traditional-*nosZ* genes were very related to each other and belong to clades VIII and IX. Unexpectedly, clade VII, composed of two Euryarcheota (*Halorubrum lacusprofundi* and *Haloarcula marismortui*), was the next most closely related clade. Additional archaeal sequences will be necessary to confirm this observed relationship. These three clades (VII, VIII, and IX) form a major distinctive cluster.

Clades V and VI were more related to each other than to any other clade. Clade V includes Flavobacteria and Firmicutes *nosZ* sequences, while clade VI includes species of the δ -proteobacterium *Anaeromyxobacter*. According to the MP model, clades I, II, III, and IV were more related to each other. Their phylogenetic relationships varied among the three models used. The collapse in fewer branches, perhaps due to little support values in BI and ML models, indicates that the most informative model was the MP. This collapsing process suggests that there are still key missing *nosZ* genes that will further clarify the phylogenetic relationships among these clades. Moreover, clade IV showed lack of relatedness to known reference sequences; six clones for the CORCO site were found (only three OTU representatives are shown in the phylogenetic tree), suggesting a newly discovered divergent NosZ variant.

Among the nine distinctive clades, there were groups which shared a low protein similarity to each other, such as clade IV and V, as well as between *P. calidifontis* and CORCO18. These clades will require additional sequence information for better resolution, as subclades within these major groups are predicted.

Sequence similarities among *nosZ* genes have been previously correlated to taxonomic relatedness (Horn et al., 2006). However, in the case of the non-denitrifying *nosZ* group, their taxonomic relatedness is unclear, with a few representatives associated to various taxonomic families. The exception to this rule is found with *W. succinogenes*, *Campylobacter fetus*, and *Tm. denitrificans* all of which are members of the ϵ -Proteobacteria (Gupta, 2006); *Anaeromyxobacter dehalogenans* strain 2CP-C and *Anaeromyxobacter* sp. strain Fw109-5, which belong to the same genera; *Desulfitobacterium* and *Geobacillus* of the phylum Firmicutes; and *Robiginitalea*, *Psychroflexus* and *Gramella*, all members of Flavobacteria class. In the phylogenetic analysis these organisms group together in the same clade as their taxonomic relative(s). Notably, those oligonucleotides specifically designed for *Anaeromyxobacter* spp. were not suitable for PCR amplification of *Desulfitobacterium* spp. and *W. succinogenes nosZ* genes. However, they were suitable for PCR amplification from *M. xanthus* genomic DNA (Figure 5), a species belonging to the same class as *A. dehalogenans* (δ -Proteobacteria).

There are genes with difficult evolutionary history due to possible lateral gene transfer. For example, *Thermotoga maritima* contains significant amounts of ‘archaeal-like’ genes, which have been associated to energy metabolism or transport of macromolecules (Nelson et al., 1999). Similarly, the *nosZ* gene could have conferred a strong selective advantage in terms of energy and therefore, lateral gene transfer has been suggested as a mechanism involved in the widespread distribution of this gene in nature. For example, different strains of *Bradyrhizobium japonicum* could be N₂O producers and N₂ producers (Sameshima-Saito, Chiba, and Minamisawa, 2006). Furthermore, *Dechloromonas aromatica* and *Dechloromonas denitrificans* are members of the non-denitrifying *nosZ* and traditional *nosZ* groups, respectively (Horn et al., 2006). This apparent conflict could be explained by means of lateral *nosZ* gene transfer from *W. succinogenes* to *M. magnetotacticum* or *D. aromatica* (Zumft and Kroneck, 2006). However, the core *nos* genes, *nosZDFYL*, share a common phylogeny and in general, these taxonomic groups employ a similar biochemical pattern with great phylogenetic correspondence according to 16S rRNA sequence analysis (Zumft and Kroneck, 2006).

Environmental clone sequences of *nosZ* genes may reflect the presence of diverse taxonomic groups of organisms according to their phylogenetic analysis. Several environmental clones were related to *A. dehalogenans* and others to *D. aromatica*. CORCO47 was more related to the Flavobacteria; however this relatedness was not supported by BI model. CORCO18 was more related to the Crenarcheota *P. calidifontis*; however, similar to the case of CORCO47, this relatedness was not supported by MP model and it will require further analysis. CORCO32, CORCO33 and CORCO7 were

overly unrelated to any reference strain in the current database, but more related to the non-denitrifying *nosZ* sequences. Furthermore, the substantial branch lengths support the idea that some of the sequences may come from distinctive genera whose *nosZ* genes are yet to be described. This divergence was observed in cases such as CORCO7, CORCO8, CORCO18, CORCO26, CORCO30, CORCO32, CORCO33, CORCO47, MWWT24, MWWT43 and MWWT102.

CORCO, MWWT, and FBR clone sequences belong to a major clade together with all known non-denitrifying *nosZ* members. However, the branch location of these environmental sequences does not follow a geographical pattern, except for clones CORCO7, CORCO32, and CORCO33, which do not have a close relative among sequences retrieved from the bioreactor samples (FBR and MWWT). BLAST tool showed that the NosZ1789R primer has a high degree of identity with some unidentified bacteria from maize planted fertilized soil (Dambreville et al., 2006), suggesting that additional non-denitrifying *nosZ* bacteria may be identified in other environmental samples. Agricultural areas may be promising areas to find novel *nosZ* sequences, since petrochemical complex CORCO has a history of sugar cane cultivation and prove to harbor many different *nosZ* genotypes. Also, 16S rRNA gene sequences related to *Anaeromyxobacter* have been found in several agricultural sample soils (F. Löffler personal communication), and in Europe 75% of the N₂O is derived from agricultural fields (Freibauer and Kaltschmitt, 2003). In the presence of substrates like NO₃⁻, NO₂⁻, NH₄⁺, abundant carbon sources in agricultural fields, physiological changes might be triggered, thus impacting the sources and sinks of N₂O in the environment. Better

understanding of communities that reduce N₂O could be very useful for agricultural applications for which fertilizers can be avoided or their application rate reduced (Avrahami et al., 2002).

Zumft and Kroneck (2006) suggest that evolution of N₂O respiration took place before the separation of the domains Bacteria and Archaea. For this reason, an Eukaryote outgroup was chosen for this phylogenetic analysis. *Crenimugil crenilabis* is a bone fish whose mitochondria harbor a *nosZ* domain. Animal mitochondrial DNA is less variable in size and organization than mtDNA from unicellular organisms. It has been proposed that mitochondria originated from symbiosis of an organism having a physiology comparable to that of modern *Paracoccus denitrificans* (*Pa. denitrificans*) (Kurland and Andersson, 2000); thus, it might resemble an ancient *nosZ* from the time before the separation of the domains Bacteria and Archaea.

The diversification of *nosZ* genes into at least 9 clades could have had its origin on early Earth when the oxygen levels began to increase and a strong selective pressure could have driven the evolution of NosZ into different structural variants. Teraguchi and Hollocher (1989) identified several differences between the N₂OR of *W. succinogenes* and *Pa. denitrificans*, including the molecular weight, copper atoms per subunit, and an iron atom and cytochrome *c* covalently associated with the enzyme. An explanation for *nosZ* sequence diversification could have its bases on the oxygen recruitments of the organisms carrying over N₂O reduction. It is known that organisms harboring non-

denitrifying *nosZ* sequences are obligate anaerobes or microaerophilic bacteria; perhaps this feature was once a driving evolutionary force selecting for different classes of N₂OR.

Diversity indexes and Richness estimators of nosZ-gene libraries

DOTUR (Distance-based OTU and Richness) defines an OTU using every possible distance between sequences. For 16S rRNA gene, a distance value of 0.03 is commonly used to differentiate at the species level; 0.05 at the genus level; 0.10 at the family/class level; and 0.20 at the phylum level (Schloss and Handelsman, 2005). For other genes of phylogenetic relatedness, such as β -subunit RNA polymerase gene (*rpoB*), a 0.20 distance value is used to differentiate between species (Schloss and Handelsman, 2005). This apparent discrepancy demonstrates a wide range of criteria to link genetic distance and phylogenetic relationships. According to Horn and coworkers (2006), approximately 0.10 may be used to differentiate *nosZ* genes at the species level. This group suggested that any two denitrifiers that shared higher than 97% 16S rRNA sequence similarity will share more than 90% NosZ protein similarity as well (which corresponds to a 0.1 distance value). Table 6 shows diversity indexes and richness estimators using a 0.03-0.04 distance value in order to compare the three *nosZ* gene libraries (the next most similar distance values among the gene libraries were 0.09, 0.1, and 0.13 for CORCO, MWWT, and FBR, respectively). Nevertheless, a higher distance value (0.27-0.29) was used for rarefaction curves comparison. At this distance value, it can be seen that CORCO *nosZ* gene library has the higher number of OTU's.

In clone libraries obtained from FBR and CORCO samples, the number of different OTUs was less than twice the square root of the total Chao1 richness. This value is indicative of a low correlation between richness and sequencing efforts. However, Schloss and Handelsman (2005) demonstrated that only after 690 16S rRNA DNA sequences, the Chao1 richness estimator at the 3% difference level began to stabilize in a Sargasso Sea collection. On the contrary, the number of clones sequenced from the MWWT library was satisfactory. Both, diversity indexes and richness estimators gave insights of the relative complexity of the communities. Perhaps, microbial communities at the CORCO site were represented by many populations with different *nosZ* genes. On the contrary, in the bioreactors some species appear to be enriched by the continuous growth conditions, especially a hydrocarbon-degrading community in the fluidized bed reactor (FBR) and a high organic matter-degrading community in the anaerobic station of Mayagüez wastewater treatment plant (MWWT). Competitive exclusion after natural enrichment could explain lower *nosZ* diversity gene indexes as compared to the CORCO soil community.

To measure how well the sample represents the whole environment, the Good Coverage Index was calculated (www.aslo.org/lomethods/free/2004/0114a.html). The low values of coverage (33% for FBR, 54% for CORCO and 72% for MWWT) show that the diversity of *nosZ* variants is higher than that recovered in this study. These values suggest that a clone library of at least 40 clones will be necessary for FBR unit, 50 for CORCO, and more than 80 clones for MWWT to scrutinize the *nosZ* gene diversity of these three environments.

In previous works, Rodríguez-Martínez and coworkers (2006) detected *nosZ* genes in a Functional Gene Microarrays (FGA) analysis. This is not surprising, since the uptake of oxygen, NO_3^- , and sulfate in this fluidized bed reactor of Vega Baja were indicative of both aerobic and anaerobic respiration. In general, they found that 19% of cultures partially sequenced were closely related to the Actinobacteria division, 37% belonged to Bacilli division, and 44% to α -, β -, and γ - proteobacteria subdivisions; no ϵ -proteobacteria were detected. In the present study did not find any *nosZ* gene related to the *nosZ* of ϵ -Proteobacteria *Wolinella*, *Thiomicrospira*, or *Campylobacter*. Given that the primers worked well for *W. succinogenes* genomic DNA, the number of ϵ -Proteobacteria *nosZ* like genes may be low or absent in these environments.

Terminal Restriction Fragment Length Polymorphism of nosZ Amplicons

T-RFLP and the phylogenetic analysis revealed that *nosZ* can not be used for setting taxonomic relationships. A band in the T-RFLP fingerprint pattern can not be understood as an OTU, but it certainly represents a group of *nosZ* genes that fragments in a similar way, based on its DNA sequence. Among the expected peaks for reference organisms carrying an non-denitrifying *nosZ* gene (Table 8), only those for *M. magneticum* and *D. aromatica* could be represented by the electropherograms obtained (Figure 26, 27, and 28). However, the *nosZ*-amplicon of these organisms will generate a *RsaI*-restriction product of 21bp that can not be analyzed, since the molecular weight standard has a lowest band of 50bp, and also these small fragments will be confused with fluorescent

unbounded NosZ1789R primers (27bp). This is in line with the results of the phylogenetic analysis, in which *D. aromatica* and *Magnetospirillum* spp. were the most related organisms to the clonal sequences.

T-RF peaks expected for each MWWT clone analyzed *in silico* (Table 7) were visualized in each electropherograms, except for the expected peaks of MWWT20 with *HinPI* (167bp), MWWT93 with *HinPI* (521bp), and MWWT94 with *RsaI* (416bp). Although only 26 MWWT clones could be analyzed by an *in silico* restriction, we can compare the information provided by the MWWT clonal library and the T-RFLP analysis. Some redundancies between these two analyses were observed. In Table 7 we can find some fragment lengths that are repeated among several clones. For *HinPI*, the fragment 134-137bp can be found in 12 clones; this fragment could be represented by the 136bp T-RF peak, which has the highest value of Optical Denisity (OD) in the AES and ANS electropherograms. For *NlaIII*, the fragments 86-87bp and 116-117bp can each be found in 7 clones, while the fragment of 397bp can be found in 5 clones; these three fragments are represented by T-RF peaks with high OD values. For *RsaI*, the fragment 516-517bp is presented in 5 clones; this fragment could be represented by the 515bp T-RF peak, which has the highest value of OD in the AES and ANS electropherograms.

Although the reference organisms that carry non-denitrifying *nosZ* genes are microaerophilic and obligate anaerobic organisms, it is possible that these gene-variants could be carried by facultative organisms, since many of the variants were found in the aerobic station (AES) of the treatment plant, which has a constant injection of oxygen. In

this plant, the wastewater treatment includes a seed consortium that is recycled continuously; therefore, it is possible that facultative organisms are being selected in this consortium, and could function either in anoxic conditions or under an oxygen atmosphere.

By applying this technique, we proved that many variants of the non-denitrifying *nosZ* gene exist that are not yet represented by cultured organisms. These variants could represent a high functional diversity for reducing N₂O in the environment.

IN SITU PCR

Cell Permeabilization

An insufficient permeabilization can result in false negatives, since PCR and FISH reactants cannot entry the cell. On the contrary, excessive permeabilization can result in either false positives, when leaking of amplicons can entry inside negatively permeabilized cells, and false negatives, when leaking of amplicons leave positive cells. Our results showed that electroporation, a faster and more economical procedure, could be employed for IS-PCR. Nonetheless, a limitation with electroporation is the low cell density requirement (10^7 cells/ml), from which 10 μ l are used to spot on *in situ* PCR slides. Since natural cell losses are expected during washes of IS-PCR, a cell concentration technique, such as centrifugation, could be necessary. This additional step

can potentially compromise the integrity of already weakened cells in the IS-PCR protocol.

ISRT-PCR

Despite controlled fixation and permeabilization, direct IS-PCR or ISRT-PCR with labeled nucleotides will yield non-specific amplification products and therefore, a high frequency of false positive results (Komminoth and Long 1993 and 1995; Chen et al., 1999). Weak primer annealing during PCR amplification can also result in a number of non-specific amplicons with detectable signal. Therefore, the detection of *in situ* PCR products could be more effectively assessed by performing a FISH step using a fluorescently labeled DNA oligonucleotide probe targeting an internal region of the amplicon.

Although it is probable that not all cells in exponentially growing cultures are actually active, false negatives could result for various reasons as previously observed by Long and co-workers (1993). Possible explanations for the false negative results include: (i) cell to cell variations in the amplification efficiency due to differences in membrane permeability; (ii) associated proteins causing non-accessibility of mRNA; (iii) cell fixation or permeabilization biases in older cultures; and (iv) loss of amplification products during washing steps in the detection procedures.

Results show a RT-PCR amplification from *P. stutzeri* cells grown either in nitrate broth or BLK (nitrogen-free medium) with histidine, as well as a weak amplification from *E. coli* cells (Figure 22, lane 6). This could be due to non-specific binding of primers Nos661F/Nos1773R (Scala and Kerkhof, 1998). In contrast, ISRT-PCR amplification showed a Cy3 signal from *P. stutzeri* cells grown in either nitrate broth or BLK with histidine, but no signal from *E. coli* cells, demonstrating a more specific detection using the internal probe Nos1527. In both cases (RT or ISRT), the amplification from *P. stutzeri* cells grown in BLK with histidine was very weak. Härtig and Zumft (1999) report that the induction of *nosZ* gene expression in *P. stutzeri* was in response to low oxygen tension, and the presence of NO_3^- and NO_2^- . This weak amplification could be due to the basal expression of the gene, since the pathway for histidine utilization in *Pseudomonas* sp. does not require the presence of NO_3^- or NO_2^- (Lessie and Neidhardt, 1967).

IS-PCR was used to prove that probe NosZ943 was specific enough to detect the *nosZ* gene from *A. dehalogenans* strain 2CP-C, but not the traditional *nosZ* gene from *P. stutzeri*. An *in silico* analysis showed that Probe943 will find 5 mismatches along its 17bp if it were to be hybridized against the *Pseudomonas nosZ* sequence. Similarly, 4 mismatches could be found between the 19bp of probe Nos1527 (Scala and Kerkhof 1998) and the ‘potential’ probing site of *A. dehalogenans* sequence. During the hybridization between Probe943 and the PCR-amplified *Pseudomonas* partial *nosZ* sequence, and Nos1527 and the PCR-amplified *A. dehalogenans* partial *nosZ* sequence,

the probes may find a partial complementary target, but they will be unbound and washed away during the washing step.

The low percentage of cells (~4%) that were observed expressing *nosZ* genes in the ISRT-PCR from natural samples could be for several reasons. Firstly, there was a high background signal; it is possible that the PCR amplicons leak from positive cells and they get trapped in a dense matrix of debris, resulting in fluorescent clumps that hide positive cells. Secondly, there may be nonspecific binding of fluorescent probes to organic matter (Thomas et al., 1997). Finally, permeabilization conditions for some microbial cells in the sample may have been suboptimal. For further optimization of this technique, multilabelled probes could be used to facilitate the distinguishing of positive cells, also the simultaneous detection of 16S rRNA genes and *nosZ* genes with two different fluorescent dyes could be used for a more unambiguous detection. Additional steps for disruption of debris clumps will favor the visualization of positive cells, since signal-emitting free cells can be easily observed. Although further studies are necessary to optimize and validate this approach at this scale, this preliminary assay documented the potential application of ISRT-PCR to study gene expression in natural systems.

LITERATURE CITED

- Abascal, F. No date. Introducción a la Filogenia Molecular.
http://www.cnb.uam.es/~fabascal/Filogenia_molecular/Filogenia-resumen.pdf
- Abascal, F., R. Zardoya, D. Posada. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21: 2104-2105.
- Avrahami S., R. Conrad, and G. Braker. 2002. Effect of Soil Ammonium Concentration on N₂O Release and on the Community Structure of Ammonia Oxidizers and Denitrifiers. *Appl. Environ. Microbiol.* 68: 5685-5692.
- Baar, C., M. Eppinger, G. Raddatz, J. Simon, C. Lanz, O. Klimmek, R. Nandakumar, R. Gross, A. Rosinus, H. Keller, P. Jagtap, B. Linke, F. Meyer, H. Lederer, and S.C. Schuster. 2003. Complete genome sequence and analysis of *Wolinella succinogenes*. *Proc. Natl. Acad. Sci. U.S.A.* 100: 11690–11695.
- Barnes, J., and N. J. P. Owens. 1998. Denitrification and Nitrous Oxide Concentrations in the Humber Estuary, UK, and Adjacent Coastal Zones. *Mar. Pollut. Bull.* 37: 247-260.
- Bazylnski, D.A., A. J. Dean, D. Schöler, E. J. P. Phillips, D.R. Lovley. 2000. N₂-dependent growth and nitrogenase activity in the metal-metabolizing bacteria, *Geobacter* and *Magnetospirillum* species. *Environ. Microbiol.* 2: 266–273.
- Bouwman, A.F., K.W. Van der Hoek, and J.G.J. Olivier. 1995. Uncertainties in the global source distribution of nitrous oxide, *J. Geophys. Res.*, 100: 2785–2800.
- Burnham, K.P., and W.S. Overton. 1979. Robust estimation of population size when capture probabilities vary among animals. *Ecology* 60: 927–936.

- Campanella, J.J., Bitincka L., and J. Smalley. 2003. MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinformatics* 4: 29.
- Chan, Y.-K. and R. Wheatcroft. 1993. Detection of a Nitrous Oxide Reductase Structural Gene in *Rhizobium meliloti* Strains and Its Location on the nod Megaplasmid of JJ1c10 and SU47t. *J. Bacteriol.* 175: 19-26.
- Chan, Y.-K., and W.A. McCormick. 2004. Experimental evidence for plasmid-borne *nor-nir* genes in *Sinorhizobium meliloti* JJ1c10. *Can. J. Microbiol.* 50: 657–667.
- Chao, A. 1984. Non-parametric estimation of the number of classes in a population. *Scand. J. Stat.* 11: 265–270.
- Chao, A., and S.M. Lee. 1992. Estimating the number of classes via sample coverage. *J. Am. Stat. Assoc.* 87: 210–217.
- Chao, A., M.C. Ma, and M.C.K. Yang. 1993. Stopping rules and estimation for recapture debugging with unequal failure rates. *Biometrika* 80: 193–201.
- Chen, F., W. A. Dustman, and R. E. Hodson. 1999. Microscopic detection of the toluene dioxygenase gene and its expression inside bacterial cells in seawater using prokaryotic *in situ* PCR. *Hydrobiologia*. 401: 131-138.
- Chen, F., J. M. Gonzalez, W. A. Dustman, M. A. Moran, and R. E. Hodson. 1997. In Situ Reverse Transcription, an Approach To Characterize Genetic Diversity and Activities of Prokaryotes. *Appl. Environ. Microbiol.* 63: 4907-4913.
- Coates, J. D., R. Chakraborty, J. G. Lack, S. M. O'Connor, K. A. Cole, K. S. Bender, and L. A. Achenbach. 2001. Anaerobic benzene oxidation coupled to nitrate reduction in pure culture by two strains of *Dechloromonas*. *Nature* 411: 1039-1043.

- Cordero-Rodríguez, C. 2003. Isolation, Characterization and Detection by *In Situ* Gene Amplification of Marine Denitrifying Bacteria. Thesis submitted for a M.S. degree at the University of Puerto Rico, Mayagüez.
- Coyle, C. L. W.G. Zumft, P. M. H. Kroneck, H. Korner, and W. Jakob. 1985. Nitrous oxide reductase from denitrifying *Pseudomonas perfectomarina* Purification and properties of a novel multicopper enzyme. *Eur. J. Biochem* 153: 459-467.
- Daims, H. N. B. Ramsing, K.-H. Schleifer, and M. Wagner. 2001. Cultivation-Independent, Semiautomatic Determination of Absolute Bacterial Cell Numbers in Environmental Samples by Fluorescence In Situ Hybridization. *Appl. Environ. Microbiol.* 67: 5810–5818.
- Dambreville, C., S. Hallet, C. Nguyen, T. Morvan, J.-C. Germon, and L. Philippot. 2006. Structure and activity of the denitrifying community in a maize-cropped field fertilized with composted pig manure or ammonium nitrate. *FEMS Microbiol. Ecol.* 56: 119-131.
- Dong, L. F., D. B. Nedwell, G. J. C. Underwood, D. C. O. Thornton, I. Rusmana. 2002. Nitrous Oxide Formation in the Colne Estuary, England: the Central Role of Nitrite. *Appl. Environ. Microbiol.* 68: 1240-1249.
- Felsenstein, J. 1989. PHYLIP - Phylogeny Inference Package Version 3.2). *Cladistics* 5: 164-166.
- Firestone, M. K. and Tiedje J. M. 1979. Temporal Change in Nitrous Oxide and Dinitrogen from Denitrification Following Onset of Anaerobiosis. *Appl. Environ. Microbiol.* 38: 673-679.

- Freibauer, A. and M. Kaltschmitt. 2003. Nitrous oxide emissions from agricultural mineral soils in Europe - controls and models. *Biogeochemistry* 63: 93-115.
- Gallagher, E., L. McGuinness, C. Phelps, L. Y. Young, and L. J. Kerkhof. 2005. ¹³C-Carrier DNA Shortens the Incubation Time Needed To Detect Benzoate-Utilizing Denitrifying Bacteria by Stable-Isotope Probing. *Appl. Environ. Microbiol.* 71: 5192-5196.
- García-Ruiz, R., S. N. Pattinson, and B. A. Whitton. 1998. Kinetic Parameters of Denitrification in a River Continuum. *Appl. Environ. Microbiol.* 64: 2533-2538.
- Gupta, R. S. 2006. Molecular signatures unique proteins and conserved Indels) that are specific for the epsilon proteobacteria Campylobacterales). *BMC Genomics* 7: 167.
- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* 41: 95-98.
- Härtig, E., and W. G. Zumft. 1999. Kinetics of *nirS* Expression (Cytochrome cd₁ Nitrite Reductase) in *Pseudomonas stutzeri* during the Transition from Aerobic Respiration to Denitrification: Evidence for a Denitrification-Specific Nitrate- and Nitrite-Responsive Regulatory System. *J. Bacteriol.* 181: 161-166.
- Hixson, D. C., J. M. Yep, J. R. Glenney, Jr., T. Hayes, and E. F. Walborg, Jr. 1981. Evaluation of Periodate/Lysine/Paraformaldehyde Fixation as a Method for Cross-linking Plasma Membrane Glycoproteins. *J. Histochem. Cytochem.* 29: 561-566.
- Hodson, R. E., W. A. Dustman, R. P. Garg, and M. A. Moran. 1995. In situ PCR for visualization of microscale distribution of specific genes and gene products in prokaryotic communities. *Appl. Environ. Microbiol.* 61: 4074-4082.

- Holben, W. E., J. K. Jansson, B. K. Chelm, and J. M. Tiedjel. 1988. DNA Probe Method for the Detection of Specific Microorganisms in the Soil Bacterial Community. *Appl. Environ. Microbiol.* 54: 703-711.
- Horn, M. A., H. L. Drake, and A. Schramm. 2006. Nitrous Oxide Reductase Genes *nosZ* of Denitrifying Microbial Populations in Soil and the Earthworm Gut Are Phylogenetically Similar. *Appl. Environ. Microbiol.* 72: 1019–1026.
- Hughes, J. B., J. J. Hellmann, T. H. Ricketts, and B. J. M. Bohannan. 2001. Counting the Uncountable: Statistical Approaches to Estimating Microbial Diversity. *Appl. Environ. Microbiol.* 67: 4399–4406.
- Kalendar, R. 2004. FastPCR, PCR primer design, DNA and protein tools, repeats and own database searches program. http://www.biocenter.helsinki.fi/bi/bare-1_html/fastpcr.htm.
- Kawaguchi, K., Y. Shinoda, H. Yurimoto, Y. Sakai, and N. Kato. 2006. Purification and characterization of benzoate-CoA ligase from *Magnetospirillum* sp. strain TS-6 capable of aerobic and anaerobic degradation of aromatic compounds. *FEMS microbiol. lett.* 257: 208-213.
- Kelso, B., R. V. Smith, R. J. Laughlin, and S. D. Lennox. 1997. Dissimilatory Nitrate Reduction in Anaerobic Sediments Leading to River Nitrite Accumulation. *Appl. Environ. Microbiol.* 63: 4679-4685.
- Kemp, P. F., and J. Y. Aller. 2004. Estimating prokaryotic diversity: When are 16S rDNA libraries large enough?. *Limnol. Oceanogr. Methods.* 2: 114-125.

- Kenzaka, T., S. Tamaki, N. Yamaguchi, K. Tani, and M. Nasu. 2005. Recognition of Individual Genes in Diverse Microorganisms by Cycling Primed In Situ Amplification. *Appl. Environ. Microbiol.* 71: 7236–7244.
- Komminoth, P., and A. A. Long. 1993. In situ polymerase chain reaction: an overview of methods, applications and limitations of a new molecular technique. *Virchows Arch. B Cell Pathol.* 64: 67-73.
- Komminoth, P. and Long, A. A. 1995. *In situ* polymerase chain reaction—methodology, applications and non-specific pathways. Boehringer Mannheim, ed. PCR application manual. Mannheim, Germany. p. 97-106.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinformatics* 5: 150-163.
- Kurland, C. G. and S. G. E. Anderson. 2000. Origin and Evolution of the Mitochondrial Proteome. *Microbiol. Mol. Biol. Rev.* 64: 786–820.
- Lee, T., T. Tokunaga, A. Suyama, K. Furukawa. 2001. Efficient dechlorination of tetrachloroethylene in soil slurry by combined use of an anaerobic *Desulfitobacterium* sp. strain Y-51 and zero-valent iron. *J. Biosci. Bioeng.* 92: 453-458.
- Lessie, T. G., and F. C. Neidhardt. 1967. Formation and Operation of the Histidine-degrading Pathway in *Pseudomonas aeruginosa*. *J. Bacteriol.* 93: 1800-1810.
- Liu, X. 2005. Do Neighbor-joining and Maximum Likelihood Methods Produce Similar Bootstrap Consensus Trees?. An internship report submitted for a M.S. degree at the Arizona State University.

- Logsdon, J. M., and D. M. Faguy. 1999. Evolutionary genomics: *Thermotoga* heats up lateral gene transfer. *Curr. Biol.* 9: 747-751.
- Long, A. A., P. Komminoth, and H. Wolfe. 1993. Comparison of indirect and direct *in situ* polymerase chain reaction in cell preparations and tissue sections. *Histochemistry* 99: 151-162.
- Ludeña-Hinojosa, Y. 2007. Cianobacterias en la Bahía de Mayagüez: Abundancia, Distribución y su Relación con las Propiedades Bio-Ópticas. Thesis submitted for a M.S. degree at the University of Puerto Rico, Mayagüez.
- Mackiewicz, M. and J. Wiegel. 1998. Comparison of Energy and Growth Yields for *Desulfitobacterium dehalogenans* during Utilization of Chlorophenol and Various Traditional Electron Acceptors. *Appl. Environ. Microbiol.* 64: 352–355.
- Magurran, A. E. 1988. Ecological diversity and its measurement. Princeton University Press, Princeton, N.J.
- Magalhães, C., N. Bano, W. J. Wiebe, A. A. Bordalo, J. T. Hollibaugh. 2007. Dynamics of Nitrous Oxide Reductase Genes (*nosZ*) in Intertidal Rocky Biofilms and Sediments of the Douro River Estuary (Portugal), and their Relation to N-biogeochemistry. *Microb. Ecol.* 55: 259-269.
- Matsubara, T., K. Frunzke, W. G. Zumft. 1982. Modulation by copper of the products of nitrite respiration in *Pseudomonas perfectomarinus*. *J Bacteriol.* 149: 816–823.
- Morgan, U. M., L. Pallant, B. W. Dwyer, D. A. Forbes, G. Rich, and R. C. A. Thompson. 1998. Comparison of PCR and Microscopy for Detection of *Cryptosporidium parvum* in Human Fecal Specimens: Clinical Trial. *J. Clin. Microbiol.* 36: 995–998.

- Müller C., M. Martin, R.J. Stevens, R.J. Laughlin, C. Kammann, J. C. G. Ottow, and H. - J. Jäger. 2002. Processes leading to N₂O emissions in grassland soil during freezing and thawing. *Soil Biol. Biochem.* 34: 1325-1331.
- Muñoz-Hincapié, M., J. M. Morell, J. E. Corredor. Increase of nitrous oxide flux to the atmosphere upon nitrogen addition to red mangroves sediments. *Mar. Pollut. Bull.* 44: 992-996.
- Nelson, D. L. and M. M. Cox. *Lehninger Principles of Biochemistry*. 3rd ed., p. 138-140, 183-189, Worth Publishers, 2000.
- Nelson, K. E. , R. A. Clayton, S. R. Gill, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, W. C. Nelson, K. A. Ketchum, L. McDonald, T. R. Utterback, J. A. Malek, K. D. Linher, M. M. Garrett, A. M. Stewart, M. D. Cotton, M. S. Pratt, C. A. Phillips, D. Richardson, J. Heidelberg, G. G. Sutton, R. D. Fleischmann, J. A. Eisen, O. White, S. L. Salzberg, H. O. Smith, J. C. Venter, and C. M. Fraser. 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature* 399: 323-329.
- Nocker, A., M. Burr, A. K. Camper. 2007. Genotypic Microbial Community Profiling: A Critical Technical Review. *Microb. Ecol.* 54: 276–289.
- Nuovo, G.J. 1996. The foundation of successful RT *In Situ* PCR. *Front. Biosci.* 1: c4-c15.
- Petrie, L., N. N. North, S. L. Dollhopf, D. L. Balkwill, and J. E. Kostka. 2003. Enumeration and Characterization of Iron(III)-Reducing Microbial Communities from Acidic Subsurface Sediments Contaminated with Uranium(VI). *Appl. Environ. Microbiol.* 69: 7467–7479.

- Rasmussen, T., B. C. Berks, J. N. Butt, and A. J. Thomson. 2002. Multiple forms of the catalytic centre, CuZ, in the enzyme nitrous oxide reductase from *Paracoccus pantotrophus*. *Biochem. J.* 364: 807-815.
- Rodríguez-Martínez, E., E. X. Pérez, C. W. Schadt, J. Zhou, and A. Massol-Deyá. 2006. Microbial Diversity and Bioremediation of a Hydrocarbon-Contaminated Aquifer in Vega Baja, Puerto Rico. *Int J Environ Res Public Health.* 3: 292-300.
- Rösch, C., A. Mergel, and H. Bothe. 2002. Biodiversity of Denitrifying and Dinitrogen-Fixing Bacteria in an Acid-Forest Soil. *Appl. Environ. Microbiol.* 68: 3818-3829.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. *Molecular Cloning: A Laboratory Manual* 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY 1989.
- Sameshima-Saito, R., K. Chiba, and K. Minamisawa. 2006. Correlation of Denitrifying Capability with the Existence of *nap*, *nir*, *nor* and *nos* Genes in Diverse Strains of Soybean *Bradyrhizobia*. *Microbes Environ.* 21: 174-184.
- Sanford, R. A., J. R. Cole, F. E. Löffler, and J. M. Tiedje. 1996. Characterization of *Desulfitobacterium chlororespirans* sp. nov., Which Grows by Coupling the Oxidation of Lactate to the Reductive Dechlorination of 3-Chloro-4-Hydroxybenzoate. *Appl. Environ. Microbiol.* 62: 3800–3808.
- Sanford, R. A., J. R. Cole, and J. M. Tiedje. 2002. Characterization and Description of *Anaeromyxobacter dehalogenans* gen. nov., sp. nov., an Aryl-Halo-respiring Facultative Anaerobic Myxobacterium. *Appl. Environ. Microbiol.* 68: 893–900.
- Scala, D. J., and L. J. Kerkhof. 2000. Horizontal Heterogeneity of Denitrifying Bacterial Communities in Marine Sediments by Terminal Restriction Fragment Length Polymorphism Analysis. *Appl. Environ. Microbiol.* 66: 1980–1986.

- Scala, D. J., and L. J. Kerkhof. 1998. Nitrous Oxide Reductase *nosZ* Gene-specific PCR primers for detection of denitrifiers and three *nosZ* genes from marine sediments. *FEMS Microbiol. Lett.* 162: 61-68.
- Schloss, P. D., and J. Handelsman. 2005. Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. *Appl. Environ. Microbiol.* 71: 1501-1506.
- Shinoda, Y., J. Akagi, Y. Uchihashi, A. Hiraishi, H. Yukawa, H. Yurimoto, Y. Sakai, N. Kato. 2005. Anaerobic degradation of aromatic compounds by *Magnetospirillum* strains: isolation and degradation genes. *Biosci. Biotechnol. Biochem.* 69: 1483-1491.
- Simon, J. 2002. Enzymology and bioenergetics of respiratory nitrite ammonification. *FEMS Microbiol. Rev.* 26: 285-309.
- Simon, J., O. Einsle, P. M. H. Kroneck, and W. G. Zumft. 2004. The unprecedented *nos* gene cluster of *Wolinella succinogenes* encodes a novel respiratory electron transfer pathway to cytochrome c nitrous oxide reductase. *FEBS Lett.* 569: 7-12.
- Smith, E. P., and G. van Belle. 1984. Nonparametric estimation of species richness. *Biometrics* 40: 119-129.
- Smith, M. S. 1982. Dissimilatory Reduction of NO_2^- to NH_4^+ and N_2O by a Soil *Citrobacter* sp. *Appl. Environ. Microbiol.* 43: 854-860.
- Smith, M.S., and K. Zimmerman. 1981. Nitrous oxide Production by Nondenitrifying Soil Nitrate Reducers. *Soil Sci. Soc. Am. J.* 45: 865-871.
- Tani, K., Kurokawa K., and M. Nasu. 1998. Development of a Direct In Situ PCR Method for Detection of Specific Bacteria in Natural Environments. *Appl. Environ. Microbiol.* 64: 1536-1540.

- Teraguchi, S., and T. C. Hollocher. 1989. Purification and Some Characteristics of a Cytochrome *c*-containing-Nitrous Oxide Reductase from *Wolinella succinogenes*. *J. Biol. Chem.* 264: 1972-1979.
- Thomas, J. C., M. Desrosiers, Y. St-Pierre, P. Lirette, J.-G. Bisailon, R. Beaudet, R. Villemur. 1997. Quantitative flow cytometric detection of specific microorganisms in soil samples using rRNA targeted fluorescent probes and ethidium bromide. *Cytometry* 27: 224–232.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucl. Acids Res.* 22: 4673-4680.
- Treude, N., D. Rosencrantz, W. Liesack, S. Schnell. 2003. Strain FAc12, a dissimilatory iron-reducing member of the *Anaeromyxobacter* subgroup of Myxococcales. *FEMS Microbiol. Ecol.* 44: 261-269.
- Trimmer, M., J. C. Nicholls, and B. Deflandre. 2003. Anaerobic Ammonium Oxidation Measured in Sediments along the Thames Estuary, United Kingdom. *Appl. Environ. Microbiol.* 69: 6447-6454.
- Trimmer, M., J. C. Nicholls, N. Morley, C. A. Davies, and J. Aldridge. 2005. Biphasic Behavior of Anammox Regulated by Nitrite and Nitrate in an Estuarine Sediment. *Appl. Environ. Microbiol.* 71: 1923-1930.
- Trogler, W. C. 1999. Physical properties and mechanisms of formation of nitrous oxide. *Coord. Chem. Rev.* 187: 303-327.

- Wolf, I., and R. Russow. 2000. Different pathways of formation of N₂O, N₂ and NO in black earth soil. *Soil Biol. Biochem.* 32: 229-239.
- Wolin, E. A., M. J. Wolin, R. S. Wolfe. 1963. Formation of methane by bacterial extracts. *J. Biol. Chem.* 238: 2882–2886.
- Yoshinari, T. 1980. N₂O Reduction by *Vibrio succinogenes*. *Appl. Environ. Microbiol.* 39: 81-84.
- Zumft, W. G. 1997. Cell Biology and Molecular Basis of Denitrification. *Microbiol. Mol. Biol. Rev.* 61: 533–616.
- Zumft, W. G., and P. M. Kroneck. 2006. Respiratory transformation of nitrous oxide (n(2)o) to dinitrogen by bacteria and archaea. *Adv. Microb. Physiol.* 52: 107-227.
- Zumft, W. G., and T. Matsubara. 1982. A novel kind of multi-copper protein as terminal oxidoreductase of nitrous oxide respiration in *Pseudomonas perfectomarinus*. *FEBS Lett.* 148: 107-112.

FIGURES

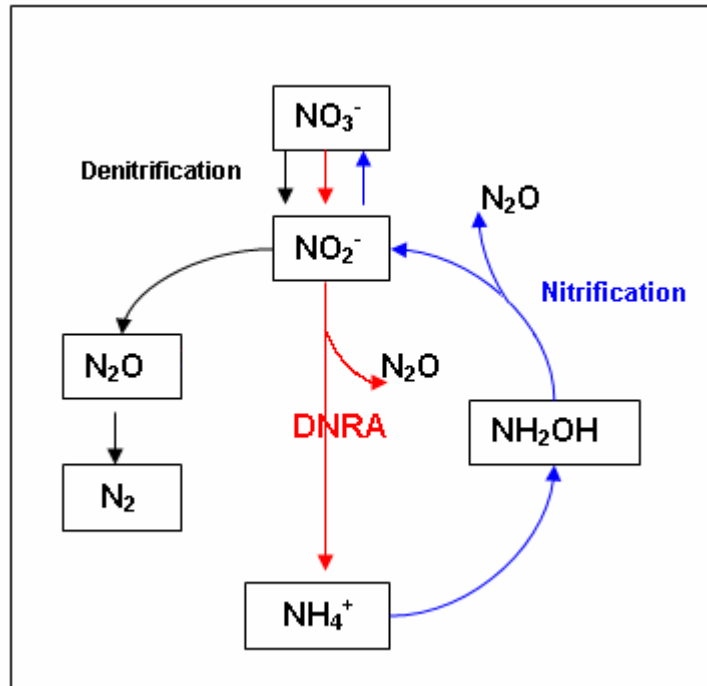


Figure 1. Denitrification, DNRA and nitrification: three known pathways involved in nitrous oxide production (modified from Kelso et al., 1997).

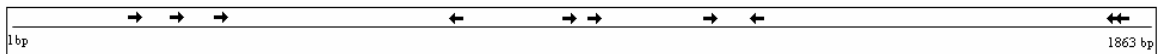


Figure 2. Graphic depiction of primer binding sites along the 1863 bp *nosZ* gene sequence of *A. dehalogenans* strain 2CP-C.

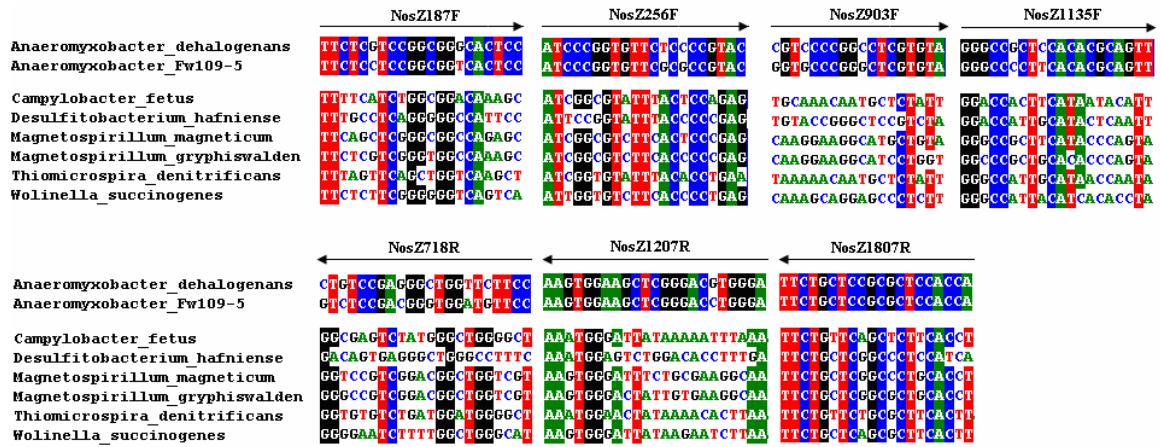


Figure 3. Partial multiple *nosZ* nucleotide sequence alignments showing priming sites for forward (upper) and reverse (lower) primers.

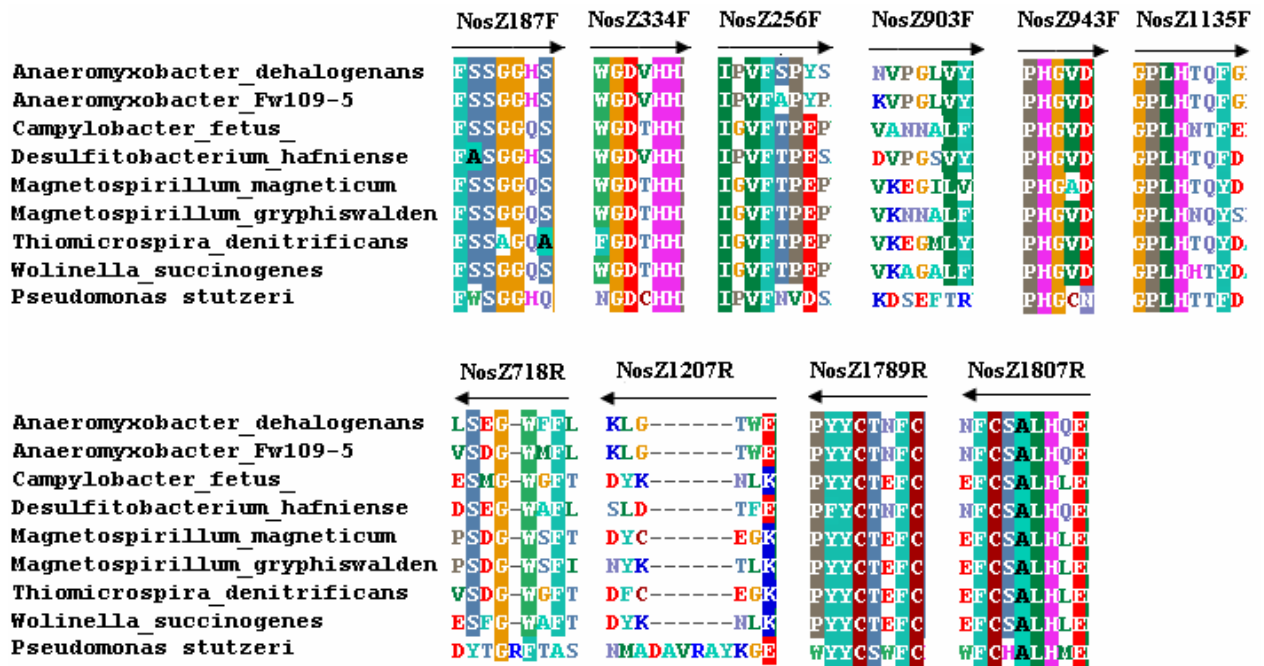


Figure 4. Partial multiple *nosZ* amino acid sequences alignments showing priming sites for forward (upper) and reverse (lower) primers.

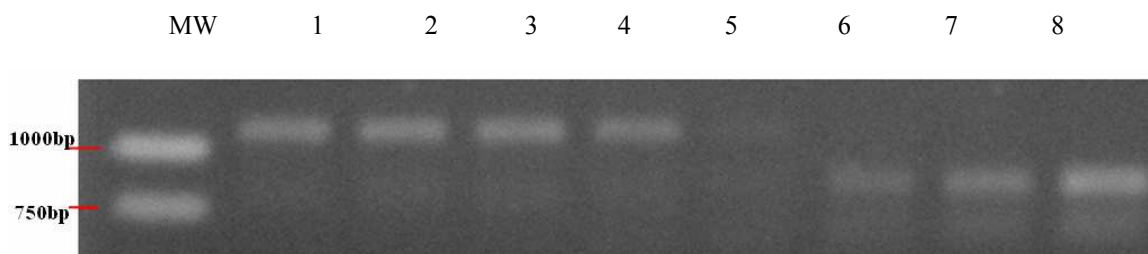


Figure 5. Optimization of PCR conditions for amplification with primer pair NosZ943F/1789R. Results from *A. dehalogenans* 2CP-C genomic DNA amplifications using PCR annealing temperatures of 40, 41.2, 43.3, 46.3, 49.8, 52.8, 54.9, and 56°C (Lanes 1-8). Five microlitres of each amplification reaction were loaded onto an agarose gel (1%). MW, 1Kb DNA ladder (Promega). NosZ943F/1789R amplicons with the expected size of 873bp were obtained at 52.8, 54.9, and 56°C (Lanes 6-8). Unspecific amplification occurred at lower temperatures (Lanes 1-5).

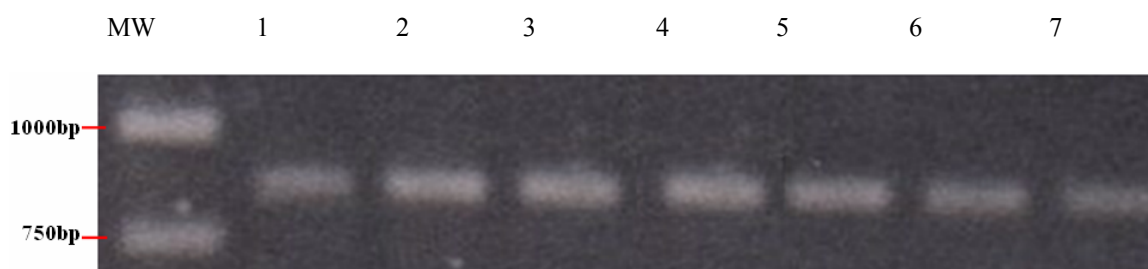


Figure 6. Optimization of PCR conditions for primer pair NosZ943F/1789R. Results from *A. dehalogenans* 2CP-C DNA amplifications using PCR annealing temperatures of 56, 57.3, 58.4, 59.7, 60.8, 61.6 and 62 °C (Lanes 1-7). Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW, 1Kb DNA ladder (Promega). Primer pair NosZ943F/1789R generated an amplicon of 873bp.

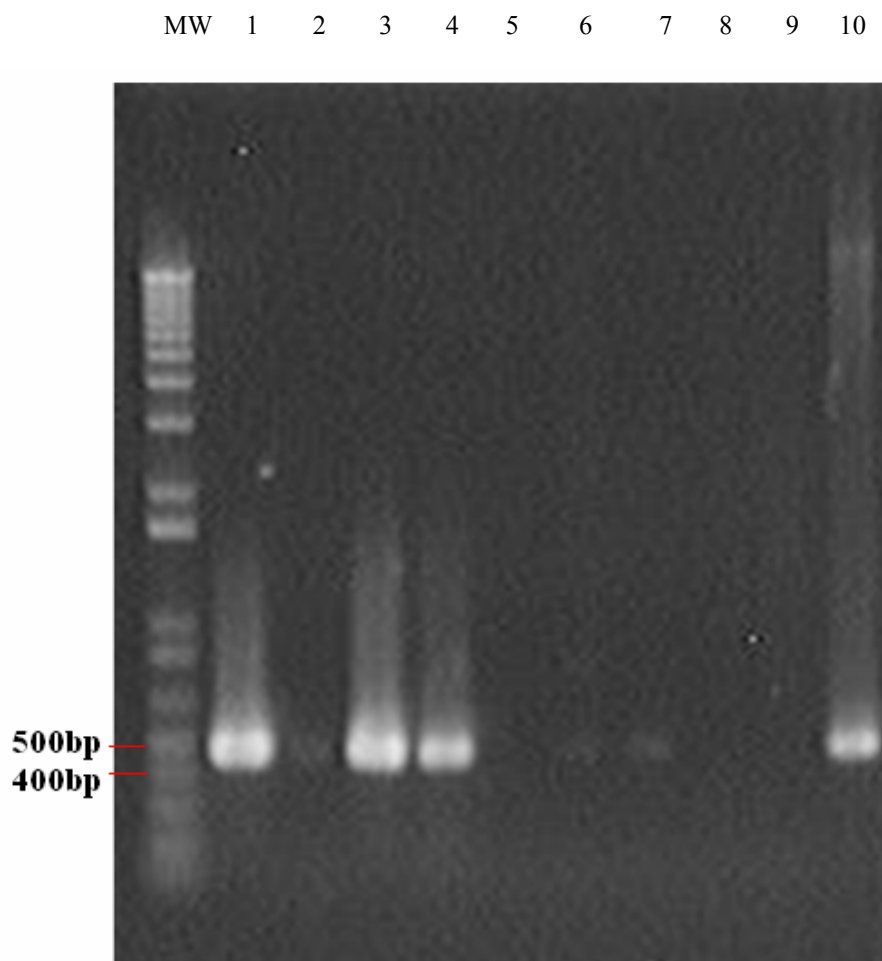


Figure 7. Analysis of PCR amplifications with primer pair NosZ256F/718R. Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW, 1Kb plus DNA ladder (Invitrogen). Primer pair NosZ256F/718R generated an amplicon of 483bp. A PCR annealing temperature of 68°C and a magnesium concentration of 2mM were used. Lanes 1-4, *Anaeromyxobacter dehalogenans* strains. Lane 1, strain 2CP-C DNA amplification. Lane 2, strain 2CP-1 DNA amplification. Lane 3, strain 2CP-3 DNA amplification. Lane 4, strain R DNA amplification. Lane

5, *Desulfitobacterium* sp. PCE-1 DNA amplification. Lane 6, *D. hafniense* DNA amplification. Lane 7, *Desulfitobacterium chlororespirans* Co23 DNA amplification. Lane 8, *W.succinogenes* DNA amplification. Lane 9, *Dehalococcoides ethenogenes* BAV1 DNA amplification. Lane 10, *Myxococcus xanthus* DNA amplification.

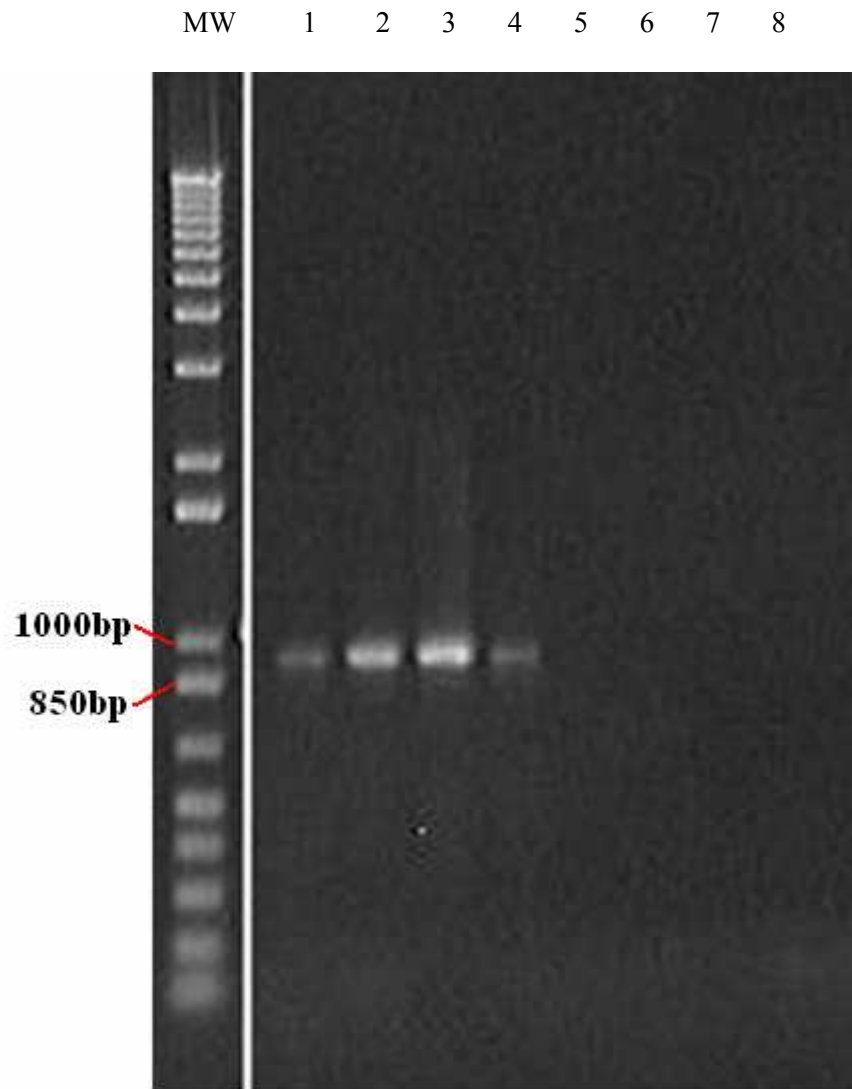


Figure 8. PCR amplifications with primer pair NosZ903F/1789R targeting the *Anaeromyxobacter* spp. *nosZ* gene and species of the non-denitrifying group. Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW, 1Kb plus DNA ladder (Invitrogen). Primer pair NosZ903F/1789R generated an amplicon of 913bp. PCR annealing temperatures of 63°C and magnesium concentration of 2.5mM were used. Lane 1, *A. dehalogenans* 2CP-C DNA amplification. Lane 2, *A. dehalogenans*

2CP-1 DNA amplification. Lane 3, *A. dehalogenans* 2CP-3 DNA amplification. Lane 4, *Anaeromyxobacter* sp. strain R DNA amplification. Lane 5, *D. hafniense* DNA amplification. Lane 6, *Desulfitobacterium chlororespirans* Co23 DNA amplification. Lane 7, *W. succinogenes* DNA amplification. Lane 8, *Dehalococcoides* sp. BAV1 DNA amplification.

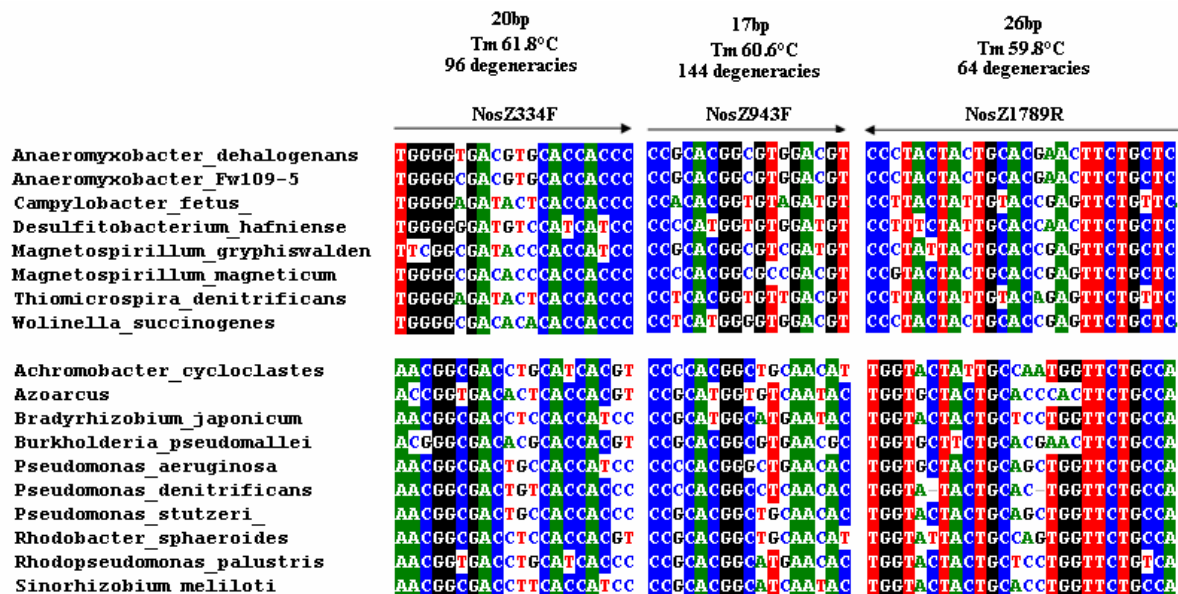


Figure 9. Partial multiple non-denitrifying (first 8 sequences) and traditional- *nosZ* (last 10 sequences) nucleotide sequences alignments showing priming sites for degenerate primers.

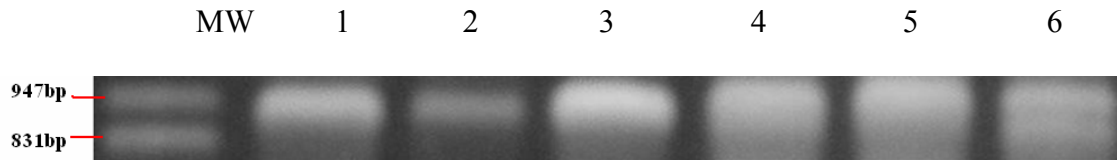


Figure 10. Analysis of PCR amplification of primer pair NosZ943F/1789R with DNA of species of the non-denitrifying group. The PCR annealing temperature was 59°C and the reactions contained 2.5mM magnesium. Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW, Lambda DNA plus *EcoRI/HindIII* molecular weight marker. Primer pair NosZ943F/1789R generated an amplicon of 873bp. Lane 1, *A. dehalogenans* 2CP-C DNA amplification. Lane 2, *A. dehalogenans* 2CP-3 DNA amplification. Lane 3, *Anaeromyxobacter* sp. strain R DNA amplification. Lane 4, *Desulfitobacterium chlororespirans* Co23 DNA amplification. Lane 5, *D. hafniense* DNA amplification. Lane 6, *Desulfitobacterium* sp. strain PCE-1 DNA amplification.

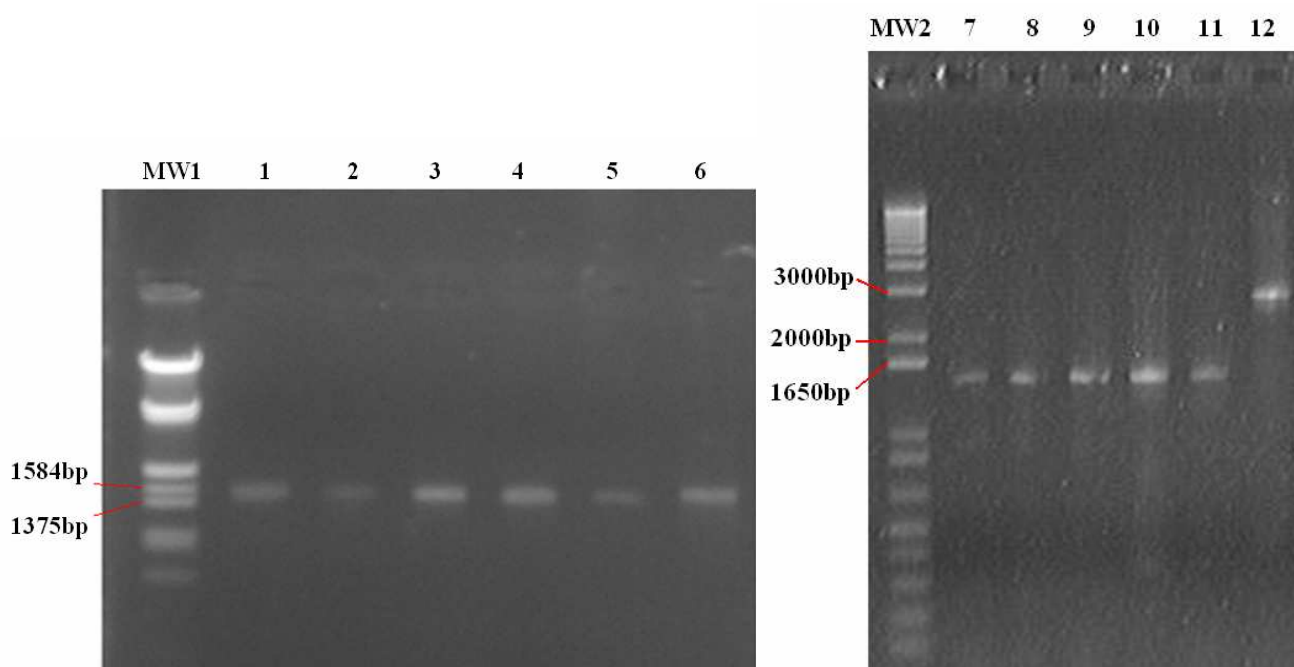


Figure 11. Optimization of PCR conditions for *nosZ* amplification with primer pair NosZ334F/1789R. Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW1, Lambda DNA plus *EcoRI/HindIII* molecular weight marker. MW2, 1Kb plus DNA ladder (Invitrogen). Primer pair NosZ334F/1789R generated an amplicon of 1482bp. A PCR annealing temperature of 57°C and a magnesium concentration of 2.5mM were used. Lane 1, *A. dehalogenans* 2CP-C DNA amplification. Lane 2, *A. dehalogenans* 2CP-3 DNA amplification. Lane 3, *Anaeromyxobacter* sp. strain R DNA amplification. Lane 4, *D. chlororespirans* Co23 DNA amplification. Lane 5, *D. hafniense* DNA amplification. Lane 6, *Desulfitobacterium* sp. PCE-1 DNA second PCR amplification using 1 µl of product from the first round of PCR amplification. Lane 7, *A. dehalogenans* 2CP-C DNA amplification. Lane 8, *A. dehalogenans* 2CP-3 DNA amplification. Lane 9, *Anaeromyxobacter* sp. strain R DNA

amplification. Lane 10, *D. chlororespirans* Co23 DNA amplification. Lane 11, *D. hafniense* DNA amplification. Lane 12, *W. succinogenes* DNA amplification.

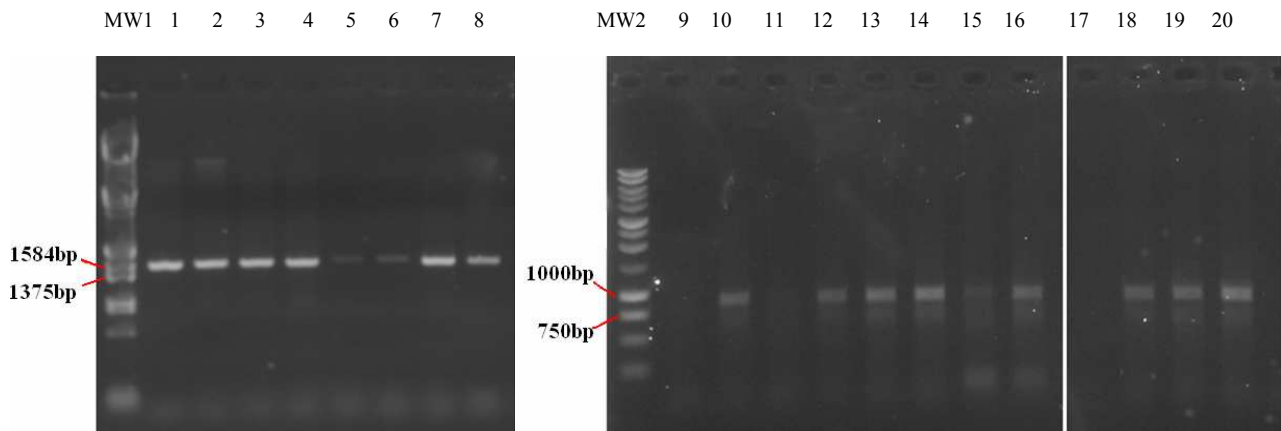


Figure 12. Analysis of PCR amplification with total community DNA of environmental samples. Five microlitres of each amplification reaction were loaded onto the agarose gel (1%). MW1, Lambda DNA plus *HindIII/EcoRI* molecular weight marker. MW2, 1Kb DNA ladder (Promega). Lanes 1-8, PCR amplification with primer pair NosZ334F/1789R Mayagüez Wastewater Treatment Plant (MWWT). Primer pair NosZ334F/1789R generated an amplicon of 1482bp. A PCR with 25 cycles and magnesium concentration of 2.5mM were used. AES, aerobic station; ANS; anaerobic station. Lane 1, AES-January. Lane 2, ANS-January. Lane 3, AES-February. Lane 4, ANS-February. Lane 5, AES-April. Lane 6, ANS-April. Lane 7, AES-May. Lane 8, ANS-May. Lanes 9-20, PCR amplification with primer pair NosZ943F/1789R using DNA of different tropical environments. Primer pair NosZ943F/1789R generated an amplicon of 873bp. A PCR annealing temperature of 57 °C and 2.5mM magnesium were used. Even lanes are replicates of the previous odd lane but with addition of *Anaeromyxobacter dehalogenans* strain 2CP-C genomic DNA. Lane 9-10, Pueblo Forest (bottom core 5-10 cm). Lane 11-12, Pueblo Forest (top core 0-5 cm). Lane 13-

14, CORCO soil. Lane 15-16, CORCO soil with addition of hydrocarbons. Lane 17-18, Mayagüez Bay; Lane 19-20, Fluidized Bed Reactor (May).

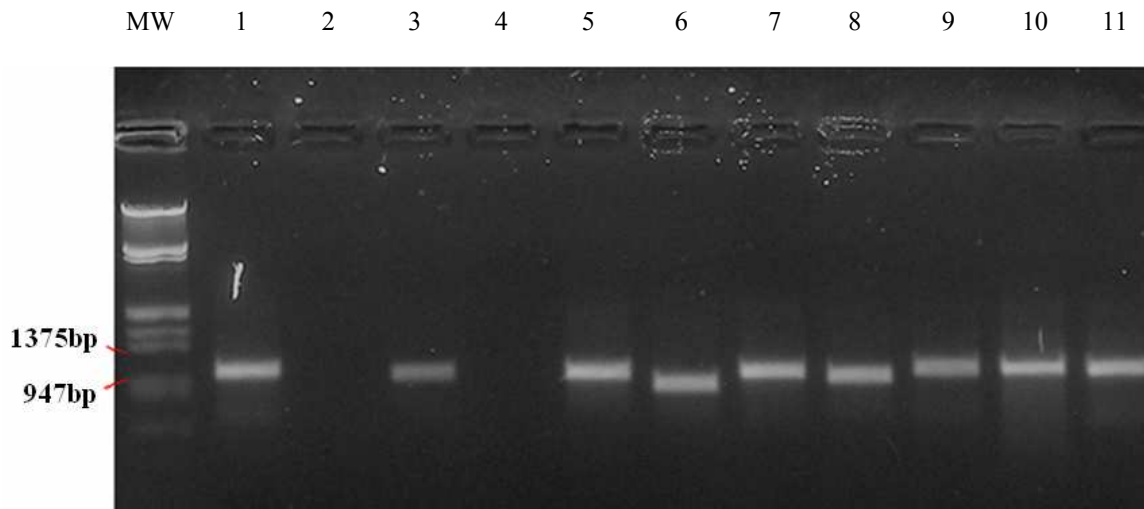


Figure 13. Analysis of colony PCR amplification of primer pair NosZ943F/1789R with environmental clones from CORCO DNA sample. PCR was performed with T7 promoter and SP6 promoter primer pair. Five μ l of each amplification reaction were loaded onto the agarose gel (1%). MW, Lambda DNA plus EcoRI/HindIII molecular weight marker. Primer pair T7 promoter/SP6 promoter generated an amplicon of approx. 1034bp. Lane 1 to 11 correspond to CORCO18, CORCO22, CORCO24, CORCO25, CORCO26, CORCO28, CORCO29, CORCO30, CORCO31, CORCO32, and CORCO33 clones.

Anaeromyxobacter 2, *Anaeromyxobacter* sp. Fw109-5; Desulfitobacterium, *D. hafniense*; Geobacillus, *G. thermodenitrificans*; Psychroflexus, *P. torquis*; Robiginitalea, *R. biformata*; Magnetospirillum 1, *M. gryphiswaldense*; Magnetospirillum 2, *M. magneticum*; Magnetospirillum 3, *M. magnetotacticum*; Campylobacter, *C. fetus*; Thiomicrospira, *Tm. denitrificans*; Wolinella, *W. succinogenes*. Symbols ‡ and * show the calcium and chloride ion ligands and the conserve histidine residues, respectively.

clone sequences from FBR, CORCO, and MWWT libraries. *Crenimugil crenilabis* served as an outgroup. Branch lines have distinct colors, in blue are traditional-*nosZ* sequences, in green are non-denitrifying *nosZ* sequences, in red are archaeal *nosZ* sequences, and in yellow are *nosZ* sequences from Flavobacteria. The bar represents 0.05 nucleotide substitutions per nucleotide.

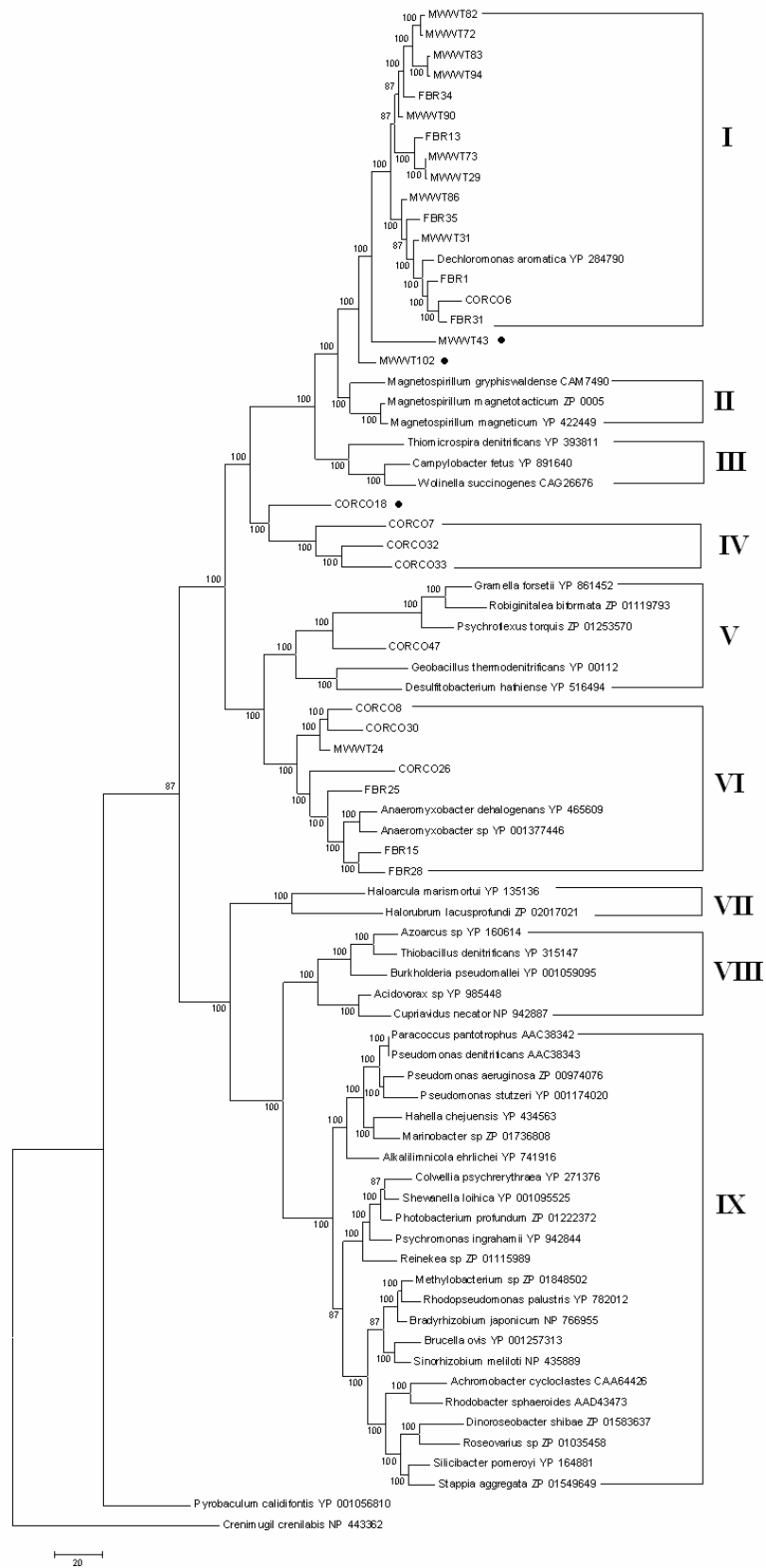
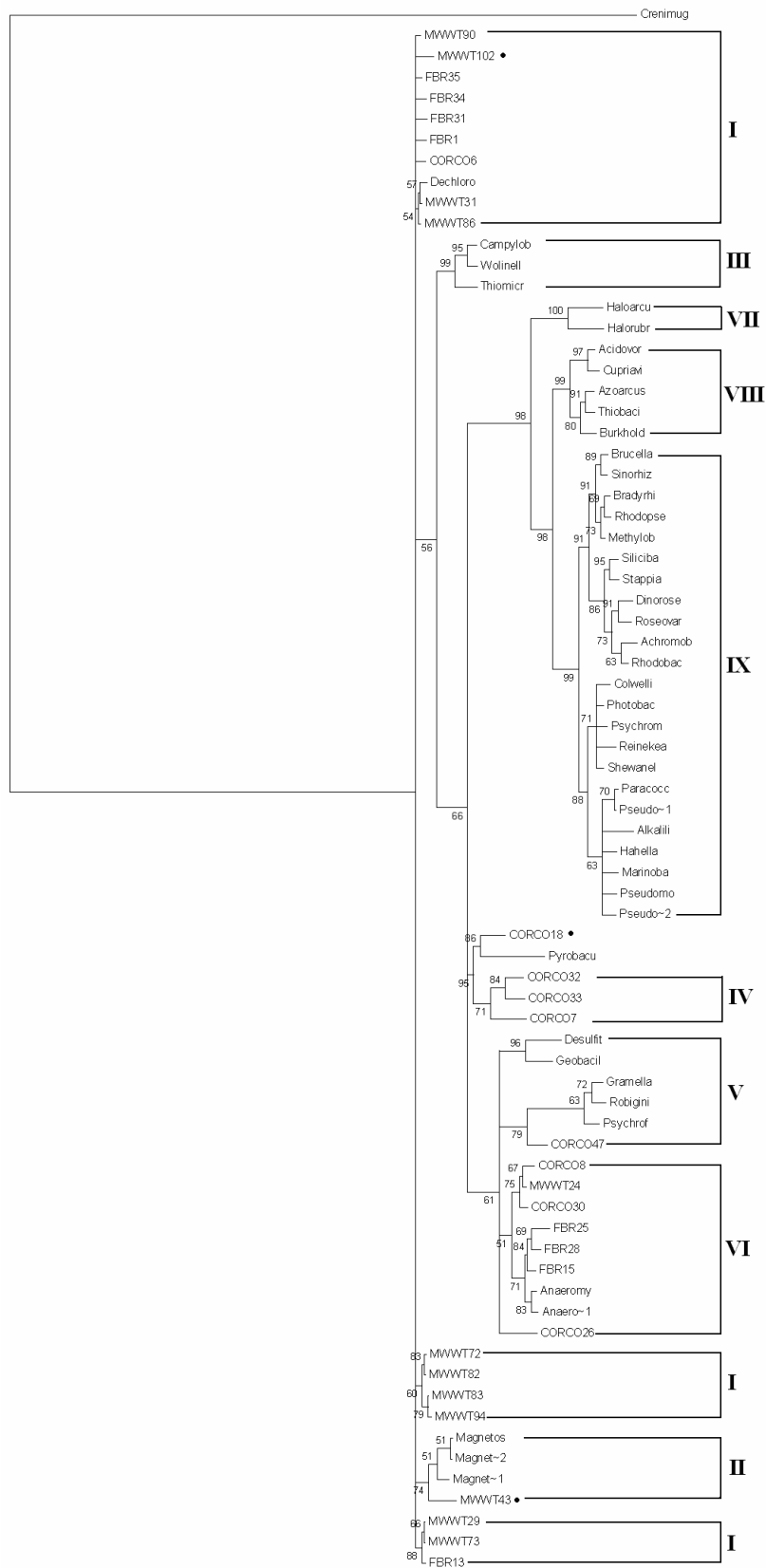


Figure 16. Strict consensus (Lenght = 1898, Consistency Index=0.48, Retention Index=0.75) of 8 most parsimonious trees (Lenght =1895, Consistency Index=0.48, Retention Index=0.75) from *nosZ* gene dataset. The sequence of *Crenimugil crenilabis* served as the root. Numbers above the branches represent the percentage of 1,000 bootstrap replications. Branch lengths correspond with numbers of amino acid replacements. The bar represents 20 replacements. Tree is based on the non-gapped multiple sequence alignment of 90 amino acids.



0.1

Figure 17. Rooted maximum likelihood tree from *nosZ* gene dataset. The sequence of *Crenimugil crenilabis* served as the root. Numbers above the branches represent the percentage of 1000 bootstrap replications, values less than 50% are not shown, and those nodes are collapsed. Scale bar represents 0.1 substitutions per amino acid position. Tree is based on the non-gapped multiple sequence alignment of 90 amino acids.

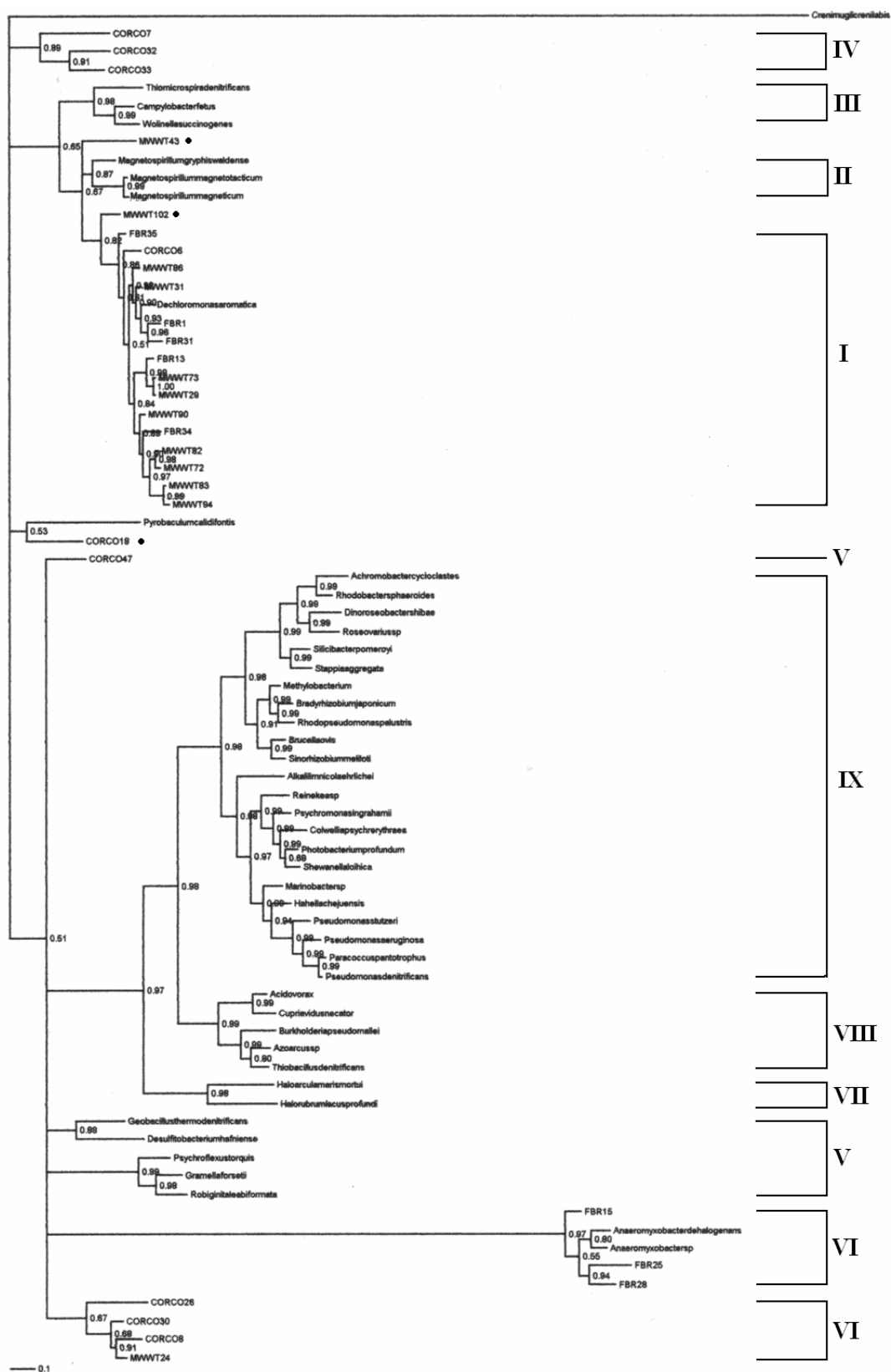


Figure 18. Bayesian phylogenetic tree from *nosZ* gene dataset after 500,000 generations of MCMCMC chains. Values to the right of the branches correspond to the posterior probabilities. The horizontal bar at the base represents 0.1 substitutions/site. Trees are based on the non-gapped multiple sequence alignment of 90 amino acids.

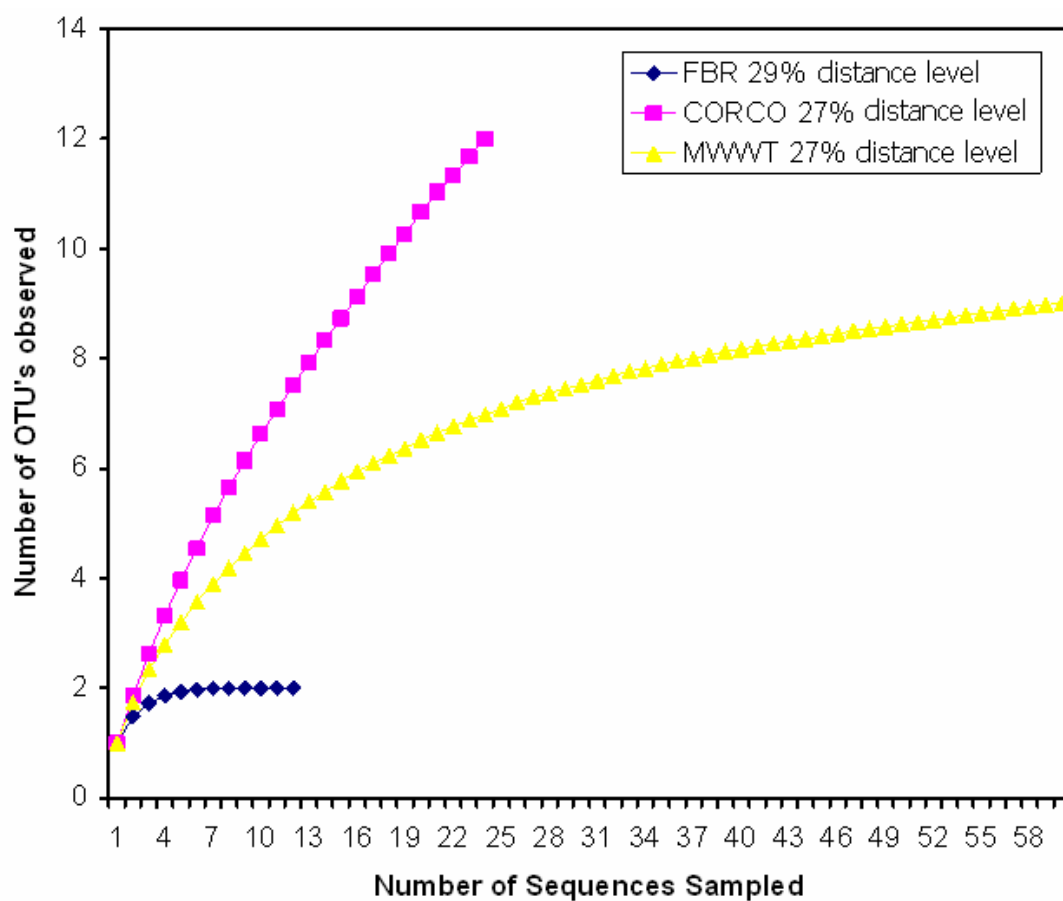


Figure 19. Rarefaction curves of *nosZ* clone sequences for the clone libraries FBR, CORCO and MWWT at 27-29% distance level.

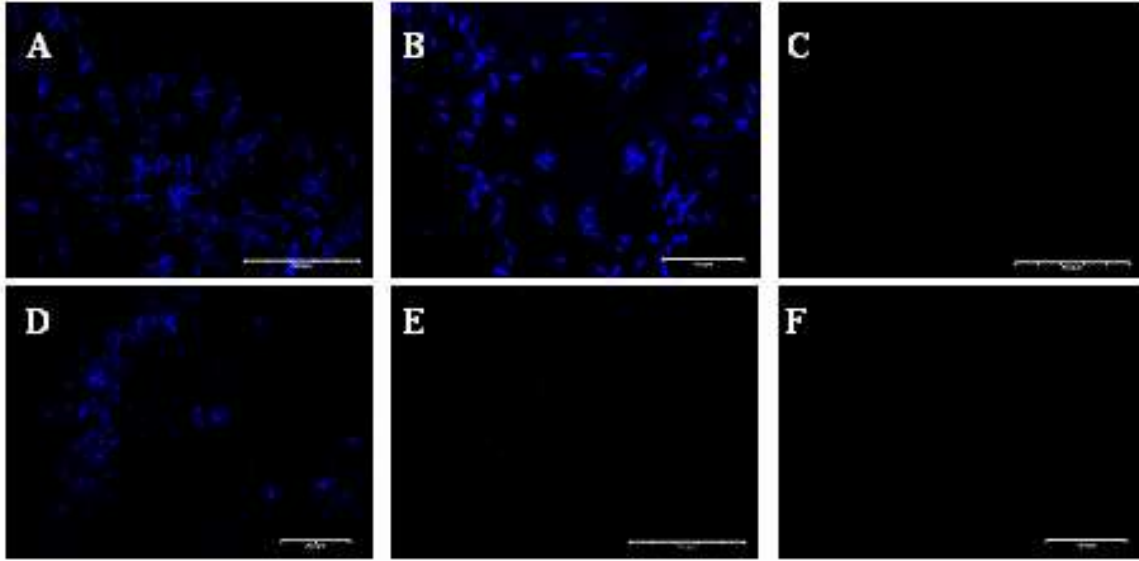


Figure 20. *Pseudomonas stutzeri* cells permeabilized by electroporation and stained with DAPI (blue) were observed by confocal microscopy. (Upper panels) Permeabilized cells showing cell integrity. (Lower panels) Permeabilized cells with RNase and DNase treatment. (A and D) Cells electroporated during 50 ms. (B and E) Cells electroporated during 100 ms. (C and F) Cells electroporated during 200 ms.

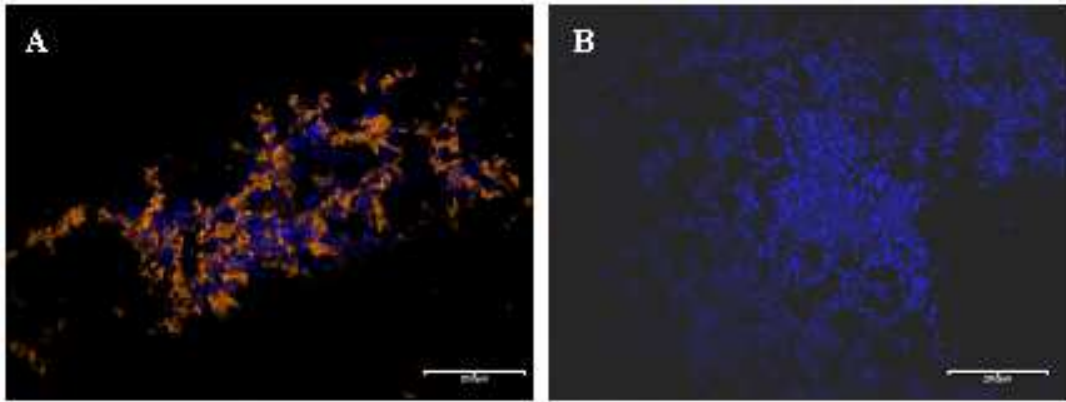


Figure 21. Confocal images of ISRT-PCR FISH of *nirS* in a co-culture of *P. stutzeri* with a non-denitrifying rod, Cy3 signal (orange) was superposed on the corresponding DAPI signal (blue). (A) Detection of *nirS* gene with a Cy3 labeled probe in a co-culture of *P. stutzeri* with a non-denitrifying rod. B, *P. stutzeri* cells treated with 1mg/ml RNase A for 1h after permeabilization.

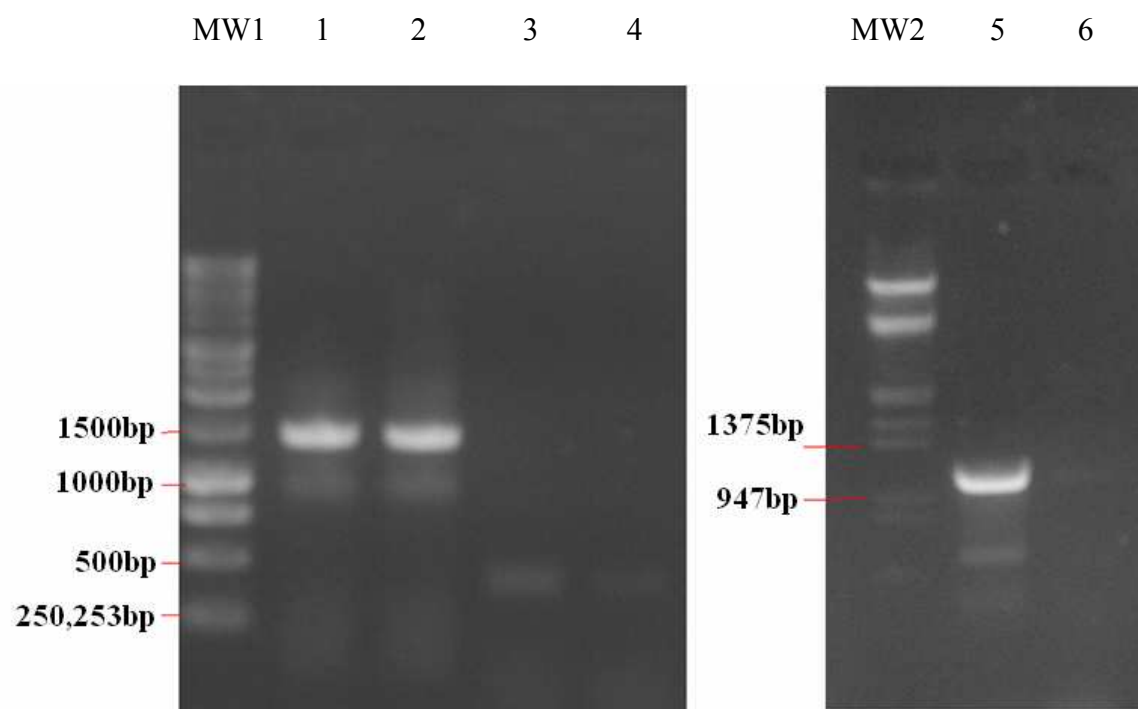


Figure 22. Analysis of RT-PCR amplification. PCR annealing temperatures of 56°C. Ten microlitres of each amplification reaction were loaded into the wells of the agarose gel (1%). MW1, 1Kb DNA ladder (Promega); MW2, Lambda DNA plus *Hind*III/*Eco*RI molecular weight marker. Primer pair NosZ1366F/1773R generated an amplicon of 407bp; primer pair Nos661F/1773R generated an amplicon of 1113bp; primer pair 27F/1392R generated an amplicon of 1394bp. Lane 1, 16S rRNA RT-PCR with primers 27F/1392R with RNA from *P. stutzeri* grown Nitrate broth; Lane 2, 16S rRNA RT-PCR with primers 27F/1392R with RNA from *P. stutzeri* grown in BLK medium with histidine; Lane 3, *nosZ* RT-PCR with primers NosZ1366F/1773R from *P. stutzeri* grown in Nitrate broth; Lane 4, *nosZ* RT-PCR with primers NosZ1366F/1773R from *P. stutzeri* grown in BLK medium with histidine; Lane 5, *nosZ* RT-PCR with primers Nos661F/1773R (Scala and Kerkhof, 1998) from *P. stutzeri* grown in Nitrate broth; Lane

6, *nosZ* RT-PCR with primers Nos661F/1773R (Scala and Kerkhof, 1998) from *E. coli* grown in Nitrate broth.

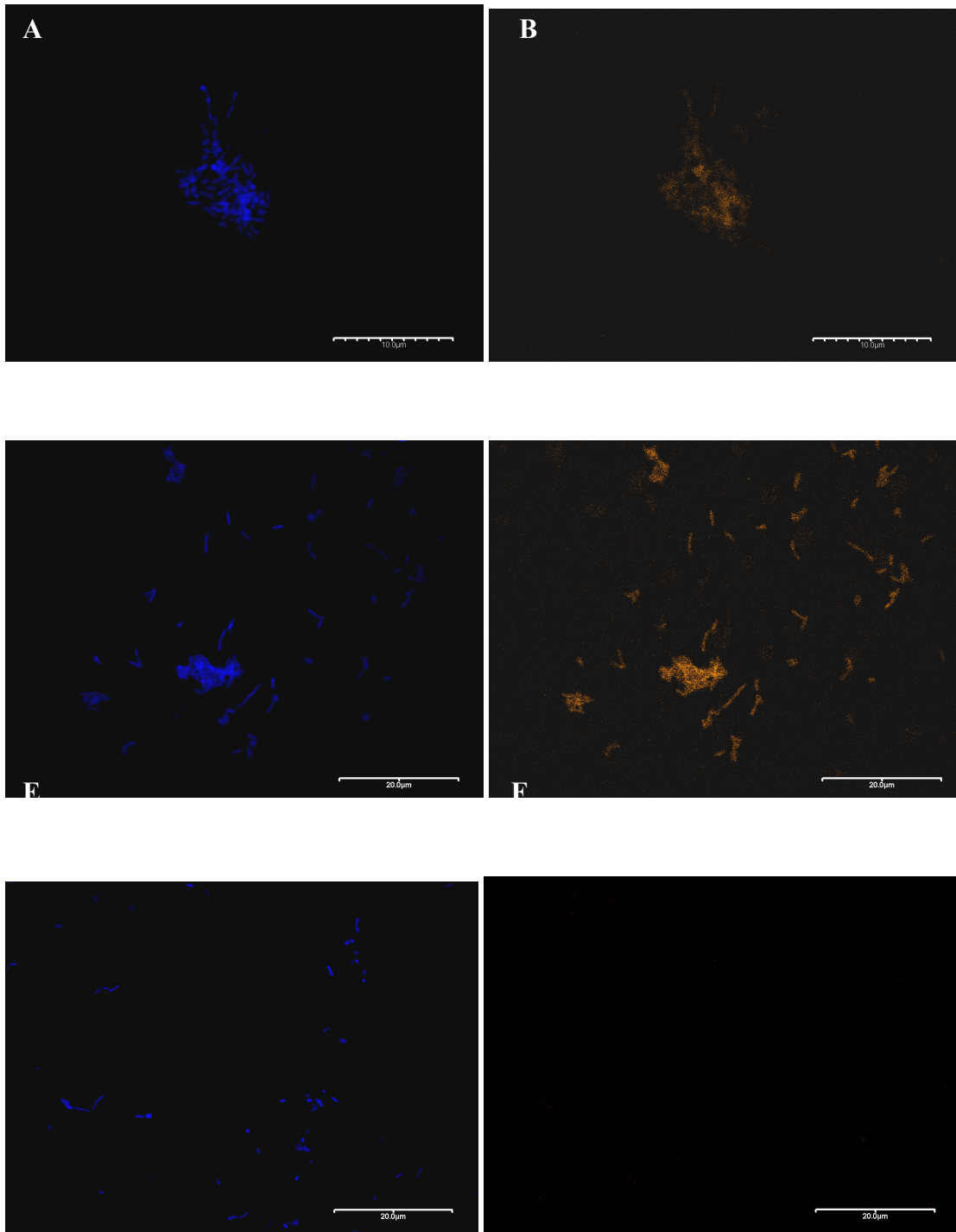
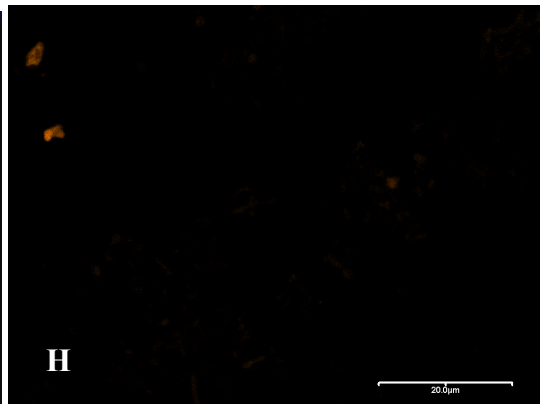
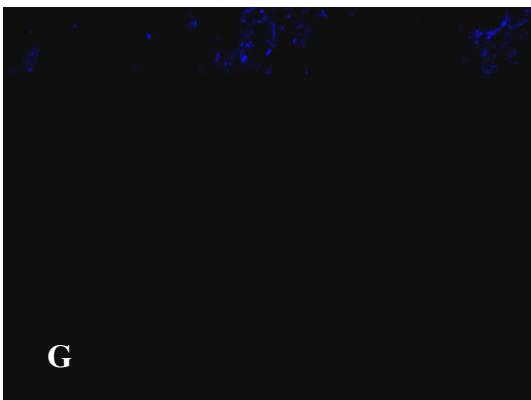
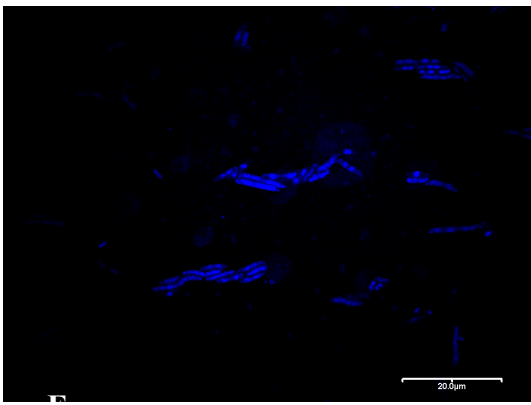
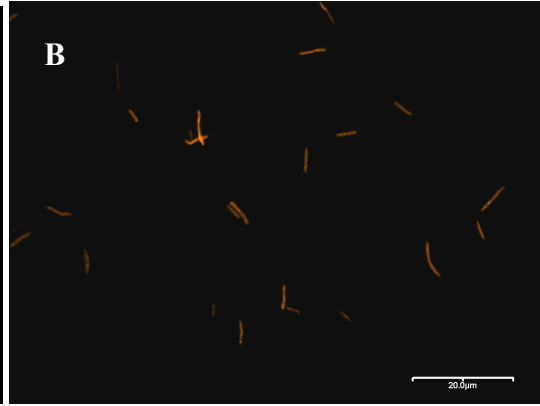
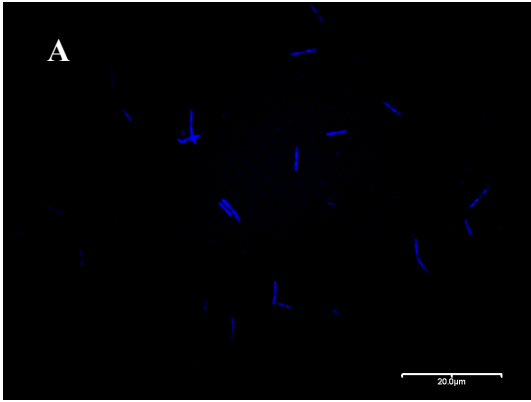


Figure 23. ISRT-PCR/FISH of *nosZ* gene of *P.stutzeri*. A-B; C-D; E-F, Confocal images corresponding to the same microscopic section. A, C, E, DAPI signal emitting-cells (blue). B, D, F, Cy3 signal emitting-cells (orange). A-B, ISRT-PCR/FISH of *nosZ* gene of *P. stutzeri* grown in BLK medium with histidine as only nitrogen source. C-D, ISRT-

PCR/FISH of *nosZ* gene of *P. stutzeri* grown in Nitrate broth. E-F, ISRT-PCR/FISH of *nosZ* gene of *P. stutzeri* grown in Nitrate broth but with a RNase treatment after the enzymatic permeabilization and before PCR reaction.



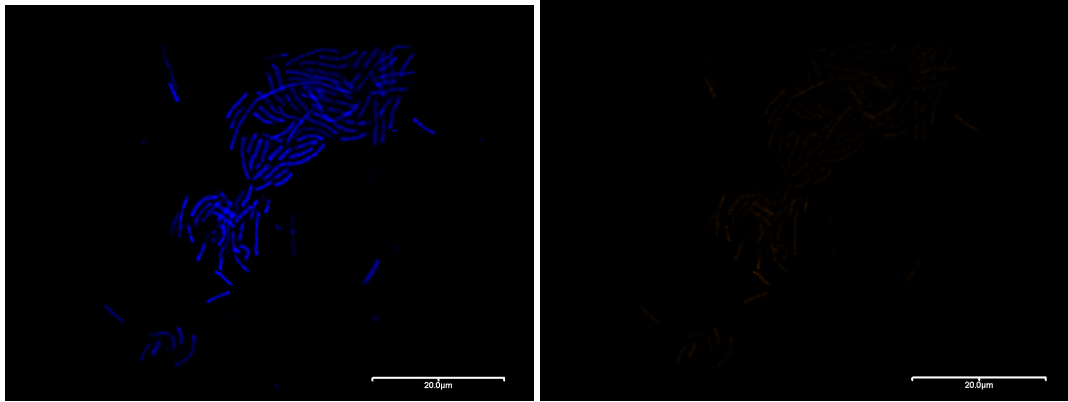


Figure 24. IS-PCR/FISH of *E.coli* DH5 α clones harboring *nosZ* genes in pGEM. A-B; C-D; E-F; G-H, Confocal images corresponding to the same microscopic section. A, C, E, G, DAPI signal emitting-cells (blue). B, D, F, H, Cy3 signal emitting-cells (orange). A-B, IS-PCR/FISH of *E.coli* DH5 α clone harboring in pGEM the partial *nosZ* gene of *A. dehalogenans* 2CP-C obtained by PCR with primers NosZ256F/1807R, Probe943 was used for FISH and primers NosZ334F/1789R were used for PCR. C-D, IS-PCR/FISH of *E.coli* DH5 α clone harboring in pGEM the partial *nosZ* gene of *A.dehalogenans* 2CP-C obtained by PCR with primers NosZ256F/1807, Probe1527 was used for FISH and primers NosZ334F/1789R were used for PCR. E-F, IS-PCR/FISH of *E.coli* DH5 α , Probe943 was used for FISH and primers NosZ334F/1789R were used for PCR. G-H, IS-PCR/FISH of *E.coli* DH5 α clone harboring in pGEM the partial *nosZ* gene of *P. stutzeri* obtained by PCR with primers Nos661/1773R, Probe943 was used for FISH and primers Nos661F/1773R were used for PCR.

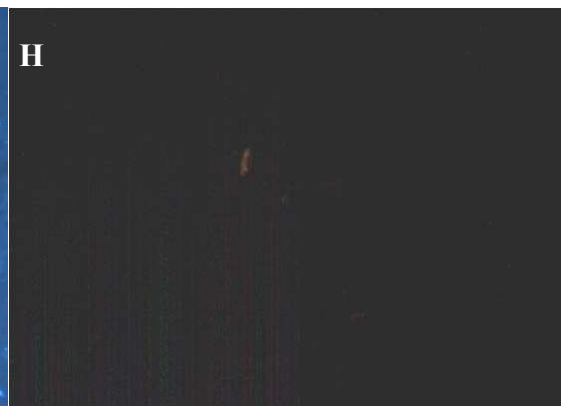
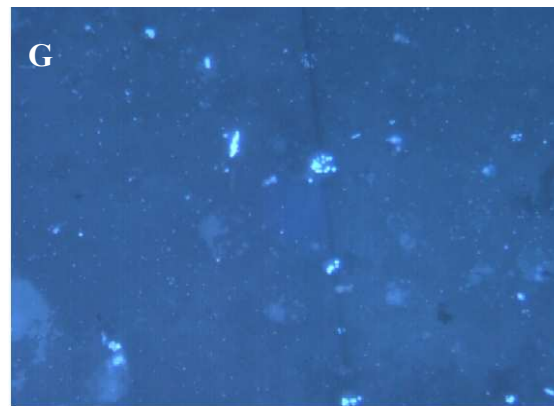
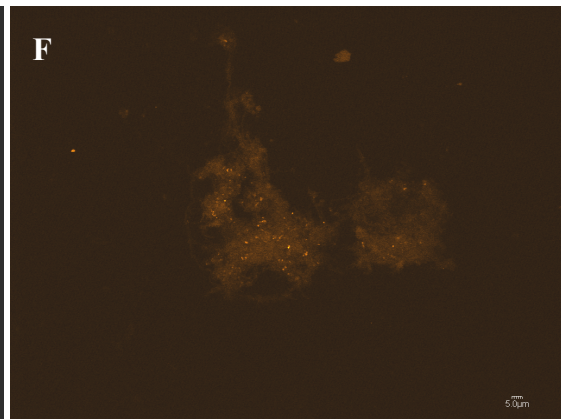
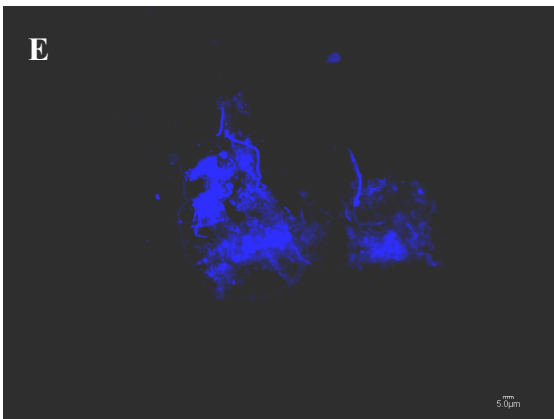
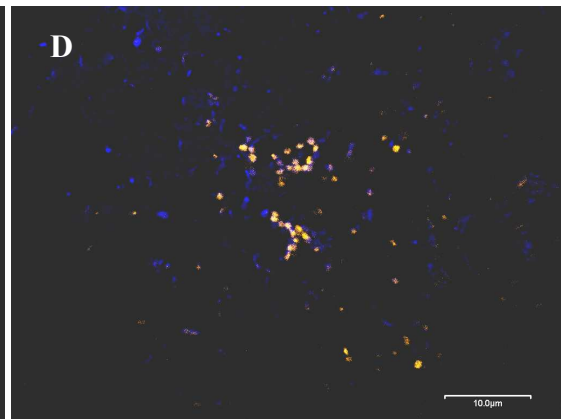
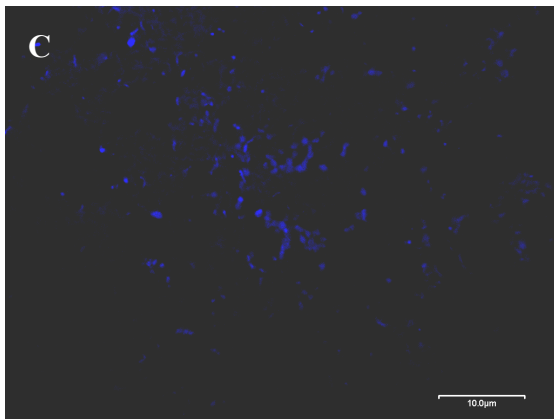
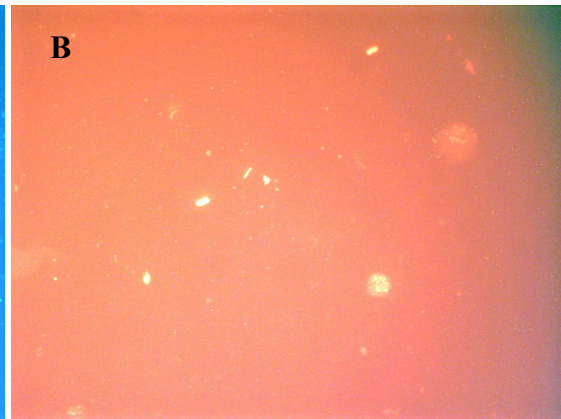
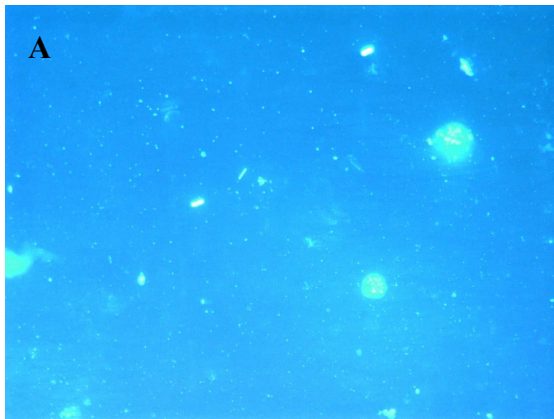


Figure 25. ISRT-PCR/FISH of *nosZ* genes in environmental samples from an anaerobic bioreactor. A-B, 1000X epifluorescence images corresponding to the same microscopic section. A and B, DAPI (blue) and Cy3 (orange) signal emitting-cells, respectively, of a ISRT-PCR/FISH of *nosZ* gene in a MWTP sample immobilized in a 0.2µm-pore-size filter. C-D, Confocal 1000X images corresponding to the same microscopic section. C section showing only DAPI signal emitting-cells (blue), while D section showing Cy3 signal (orange) superposed on the corresponding DAPI signal (blue). E-F, Confocal 600X images corresponding to the same microscopic section. E section shows DAPI signal emitting-cells (blue), while F section shows Cy3 signal (orange). G-H, 1000X epifluorescence images corresponding to the same microscopic section. G and H, DAPI (blue) and Cy3 (orange) signal, respectively, of a ISRT-PCR/FISH with a RNase treatment after the enzymatic permeabilization and before PCR reaction. C-D, E-F and G-H, ISRT-PCR/FISH of *nosZ* gene in a MWWT sample immobilized in glass slide.

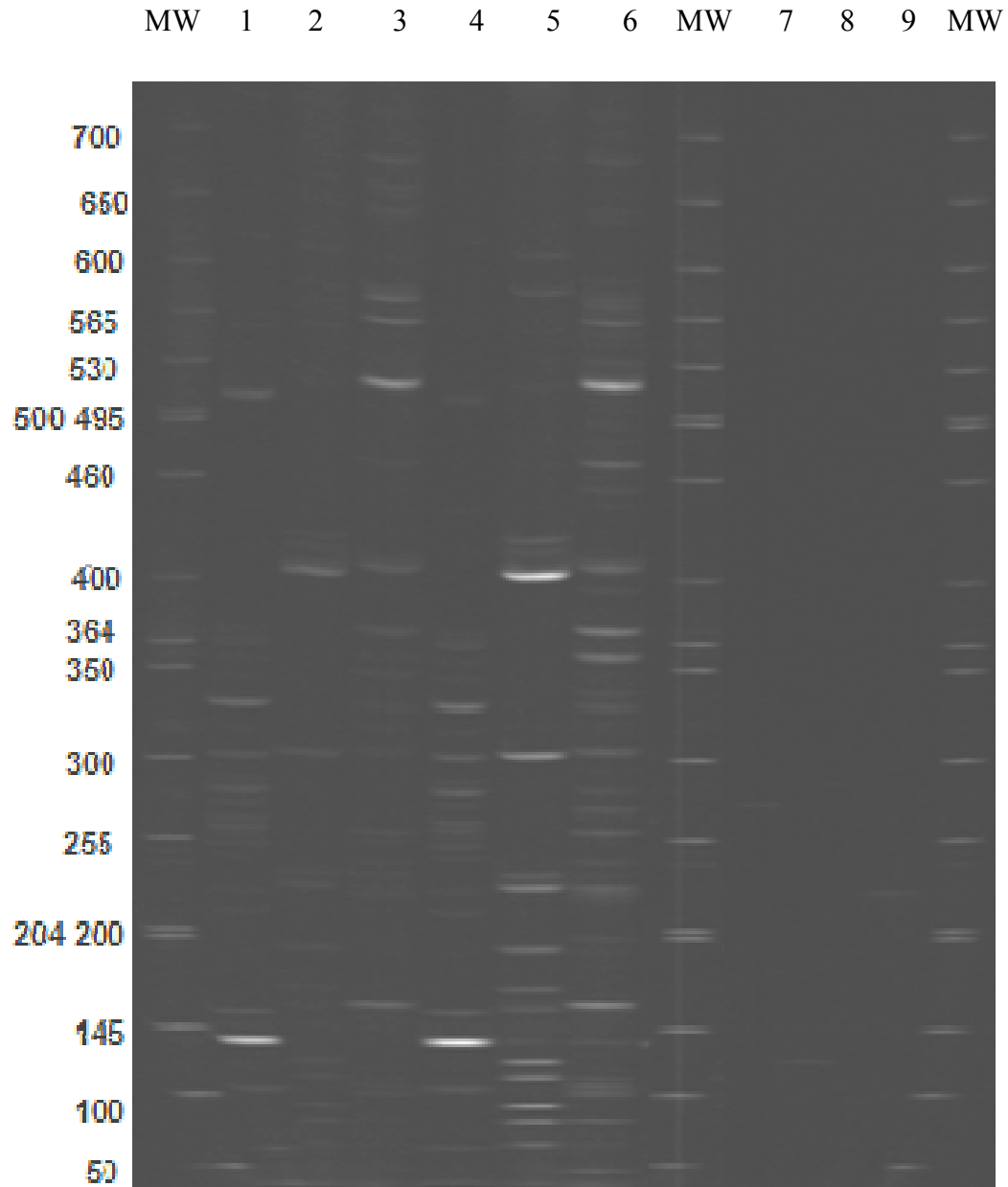


Figure 26. *nosZ*-based *HinPI* (1, 4, 7), *NlaIII* (2, 5, 8), and *RsaI* (3, 6, 9) T-RFLP fingerprint patterns of Mayaguez Wastewater Treatment Plant (MWWT) anaerobic reactor sample (1-6) and MWWT34 clone (7-9). MW, molecular sizing standard. KBPlus-LICOR (50-700bp).

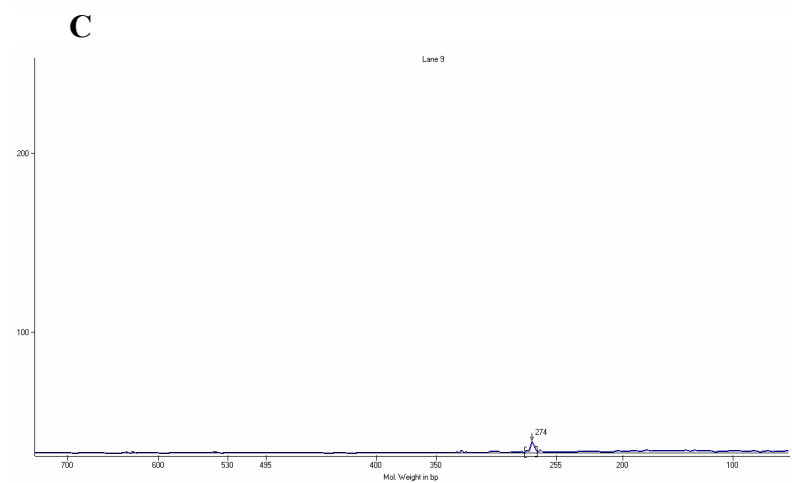
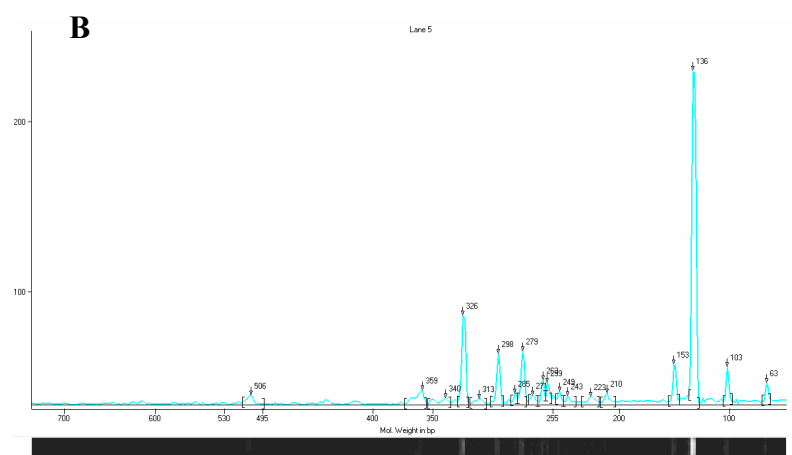
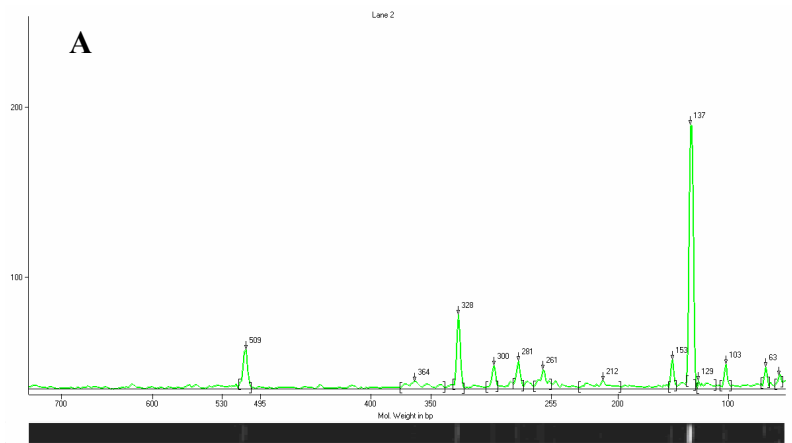


Figure 27. Comparison of electropherograms profiles generated with *HinPI* for Aerobic station (A), Anaerobic station (B) of Mayagüez Wastewater Treatment Plant samples and clone MWWT34 (C).

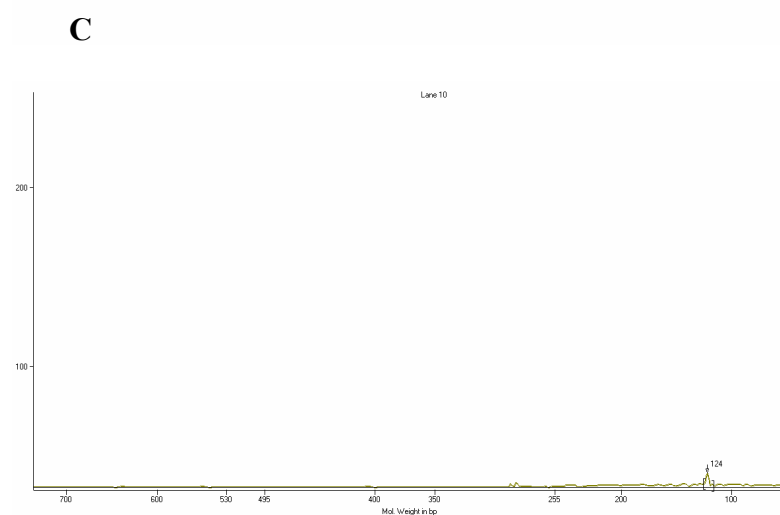
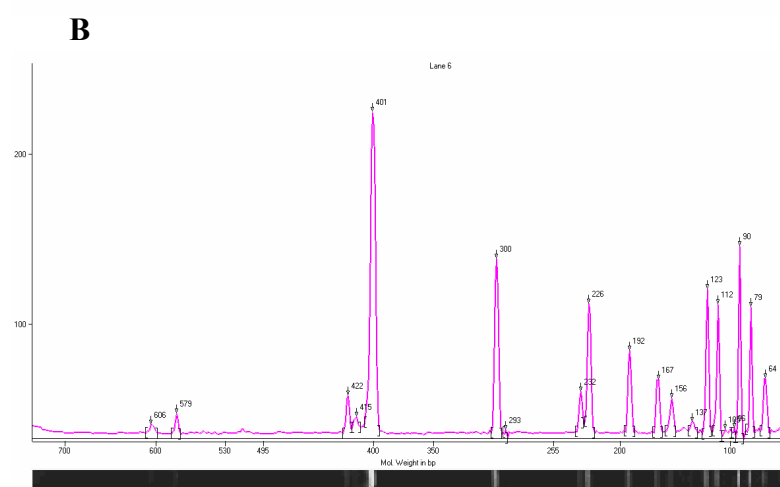
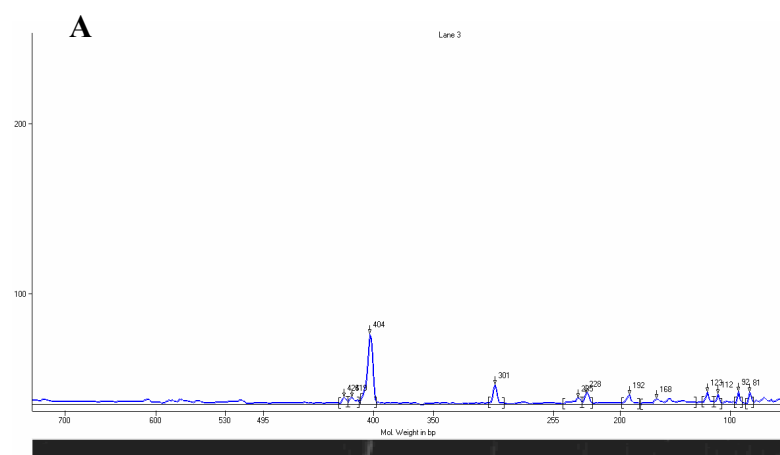


Figure 28. Comparison of electropherograms profiles generated with *NlaIII* for Aerobic station (A), Anaerobic station (B) of Mayagüez Wastewater Treatment Plant samples and clone MWWT34 (C).

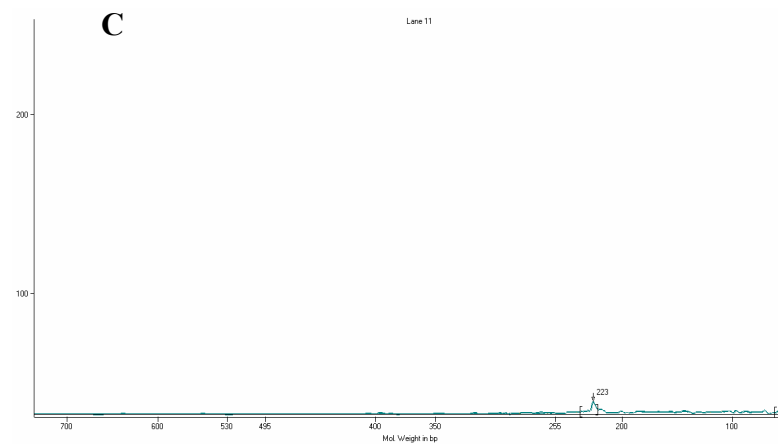
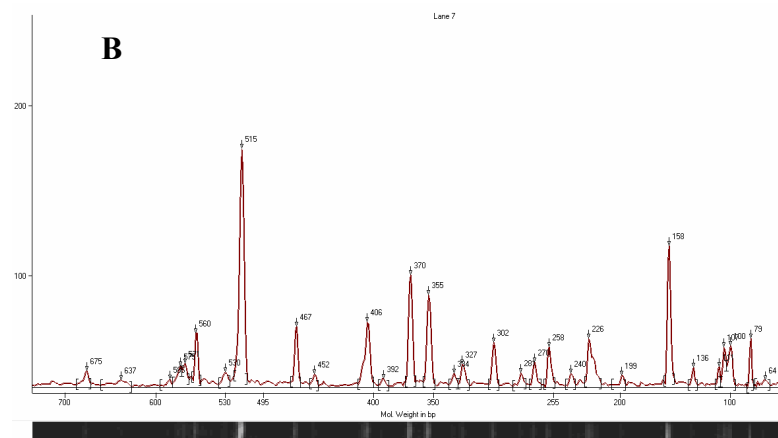
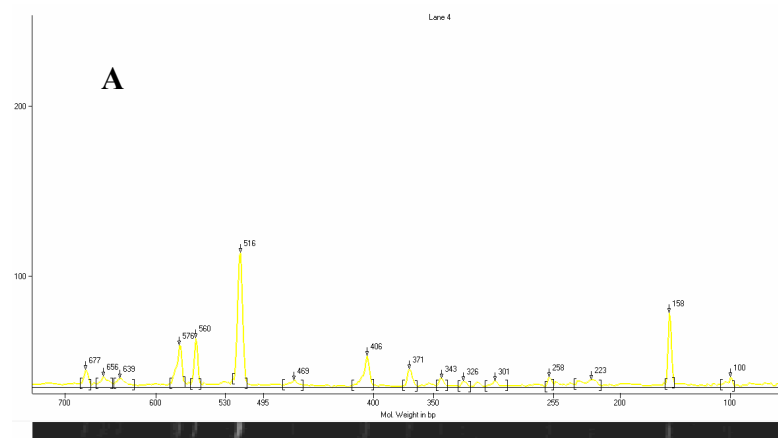


Figure 29. Comparison of electropherograms profiles generated with *RsaI* for Aerobic station (A), Anaerobic station (B) of Mayagüez Wastewater Treatment Plant samples and clone MWWT34 (C).

Table 1. Environmental samples used in this study

Sample Designation	Sample Description	Sampling Site	Reference
CORCO	0.5 gr of soil	Guayanilla	A. Massol-Deyá, TMEL
FBR	5 gr of activated carbon of a fluidized-bed bioreactor treating diesel-contaminated groundwater at Hydro Gas Station	Vega Baja	Rodríguez-Martínez et al., 2006
MWWT	activated sludge from aerobic and anaerobic units of the Mayaguez Regional Wastewater Treatment Plant	Mayagüez	A. Massol-Deyá, TMEL
MB	Coastal waters collected from Mayaguez Bay (near Guanajibo river mouth)	Mayagüez	Ludeña-Hinojosa, 2007
EBS	Soil sample from a mine deposit (0-5 cm)	Adjuntas	A. Massol-Deyá, TMEL
EBP	Soil sample from a mine deposit (5-10 cm)	Adjuntas	A. Massol-Deyá, TMEL

Table 2. Primer sequences and designation

Primer Designation	Oligonucleotide sequence
NosZ187F	ttctcgtccggcgggcactcc
NosZ256F	atcccgggtgttctccccgtac
NosZ334F	tggggbgayryvcaycaycc
NosZ903F	cgtccccggcctcgtgta
NosZ943F	ccbcayggbgysgaygt
NosZ1135F	gggccgctccacacgcagtt
NosZ1366F ^a	ggcaagctntcncc
NosZ718R	ggaagaaccagccctcggacag
NosZ1207R	tcccacgtcccagcttcactt
NosZ1789R	gagcagaantrsgtcagtagtangg
NosZ1807R	tgggtggagcgcggagcagaa

(F, forward; R, reverse)

^a for traditional-*nosZ* sequences

Table 3. Characteristics and PCR conditions for primer pairs.

Primer Pair	Product Length (bp)	T _m Difference (°C)	Optimum Mg conc. (mM)	Optimum Annealing Temp. (°C)	Amplification of <i>Anaeromyxobacter</i> spp.	Amplification of non-denitrifying species
187F / 1207R	1042	2.4	2.5	66	Y ^a	Y ^b
256F / 718R	483	2.2	1.5	69	Y	N ^c
334F / 1789R	1482	2.2	2	57	Y ^a	Y
903F / 1207R	328	0.2	2.5	66-67	Y	Y ^d
903F / 1789R	913	0.4	2.5	63	Y	N
903F / 1807R	923	4.6	2.5	67	Y	Y ^d
943F / 1789R	873	0.8	2.5	59	Y	Y

Y, yes, N, no

^a Weak amplification for *A. dehalogenans* 2CP-1

^b Unspecific amplification

^c Amplification only for *M. xanthus* and *Anaeromyxobacter* spp

^d Amplification only for *W. succionogenes* and *Anaeromyxobacter* spp.

Table 4. Similarity matrix of amino acid sequences from non-denitrifying *nosZ* group and environmental clones.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	
1. A2CP-C																																													
2. AFW109-5	95																																												
3. CFETUS	58	57																																											
4. TDENIT	58	56	83																																										
5. WSUCCI	58	57	91	83																																									
6. DAROMA	60	60	77	76	78																																								
7. DHAFNI	70	70	61	62	60	59																																							
8. GThermo	71	70	60	61	62	64	72																																						
9. MGRYPHIS	64	63	77	76	80	85	60	66																																					
10. MMAGNETI	65	64	76	74	77	83	59	65	89																																				
11. MMAGNETO	65	64	76	73	77	83	59	62	90	99																																			
12. GFORSETII	60	58	59	62	60	62	60	60	60	60																																			
13. RBIFORMA	62	62	57	55	58	62	59	62	61	59	59	90																																	
14. PTORQUIJS	62	63	57	58	59	60	59	63	60	61	61	87	88																																
15. PCALIDIFO	58	57	64	62	63	63	58	60	64	64	63	54	55	53																															
16. FBR1	62	62	78	77	79	94	62	65	86	82	82	63	61	60	62																														
17. FBR13	62	62	77	75	77	94	61	64	86	82	83	62	61	61	94	62																													
18. FBR15	96	95	57	57	59	60	70	70	64	64	64	60	62	57	63	62	62																												
19. FBR25	91	91	57	59	57	59	70	67	63	62	63	61	63	65	57	62	62	91																											
20. FBR28	96	95	57	59	58	61	69	69	64	64	64	60	62	63	58	63	63	95	91																										
21. FBR31	62	62	77	76	78	96	60	65	86	81	82	62	60	60	63	95	93	62	61	63																									
22. FBR34	62	62	78	77	78	93	61	65	86	83	83	62	60	60	63	93	93	62	64	94	94																								
23. FBR35	64	64	77	77	78	94	62	66	86	84	84	62	60	60	63	95	94	61	62	64	95	94																							
24. CORCO6	61	62	77	76	78	95	60	63	85	81	82	60	59	60	63	94	94	61	62	64	94	93	95																						
25. CORCO7	63	64	65	65	66	66	66	65	68	66	66	59	60	59	65	67	66	64	62	64	66	66	66	65																					
26. CORCO8	88	87	63	62	61	62	73	72	66	65	66	61	63	64	60	64	62	88	86	87	64	64	65	63																					
27. CORCO18	68	68	71	68	70	70	65	69	71	70	70	62	61	62	70	71	70	68	67	68	72	71	71	70	76	68																			
28. CORCO26	82	80	61	60	61	62	66	70	66	66	66	63	64	63	58	61	61	82	81	81	61	61	64	64	61	62	84	67																	
29. CORCO30	88	86	60	60	60	59	71	73	64	64	64	60	60	62	60	61	61	90	85	88	62	62	64	62	66	92	70	83																	
30. CORCO32	60	62	59	60	61	63	61	64	66	65	64	56	56	57	63	64	64	61	62	65	65	66	64	77	64	73	60	83																	
31. CORCO33	62	64	61	60	63	62	61	64	67	65	65	58	57	57	66	64	64	63	61	64	64	64	65	63	75	65	75	60	63																
32. CORCO47	71	71	63	62	64	66	70	69	66	66	66	65	68	69	59	66	67	71	75	71	66	67	68	67	64	75	66	71	75																
33. MWWT24	87	86	61	60	60	60	72	73	64	65	65	60	63	63	59	62	61	87	84	86	61	62	64	61	67	95	70	82																	
34. MWWT29	62	62	78	76	78	94	62	64	86	82	83	62	60	60	62	95	98	62	62	63	94	95	95	95	67	63	71	61	61																
35. MWWT31	61	61	78	78	79	97	61	64	86	83	83	62	61	60	63	95	95	61	60	61	96	93	96	96	66	64	71	61	61																
36. MWWT34	61	61	78	78	79	97	61	65	86	83	83	62	61	60	63	95	95	61	60	61	96	93	96	96	66	64	71	61	60																
37. MWWT43	61	60	77	75	76	84	61	62	84	81	81	60	60	59	61	83	82	60	60	61	83	83	84	82	66	62	68	61	61																
38. MWWT72	64	63	77	75	78	93	62	65	87	83	85	63	62	62	61	92	95	63	62	64	93	95	94	93	68	65	71	65																	
39. MWWT73	62	62	78	76	78	94	62	64	86	82	83	63	61	61	62	95	98	62	62	63	94	95	95	95	67	63	71	61	61																
40. MWWT82	63	63	77	75	78	93	61	65	86	82	84	62	62	62	61	92	95	62	62	64	93	95	94	93	68	65	70	65																	
41. MWWT83	62	62	78	76	78	94	62	64	86	82	83	63	61	61	62	95	98	62	62	63	94	95	95	95	67	63	71	61	61																
42. MWWT86	63	62	78	77	78	92	60	64	86	83	84	62	61	61	92	94	93	91	68	62	70	62	62	64	64	66	61	93																	
43. MWWT90	62	62	78	78	79	97	61	66	87	83	84	63	61	61	92	94	94	62	61	63	96	95	96	95	66	65	71	63																	
44. MWWT94	63	62	78	77	79	93	62	65	87	82	83	62	61	60	63	93	95	62	61	63	95	96	94	94	68	65	71	62																	

The matrix was generated using MatGat Matrix Global Alignment Tool (Campanella et al. 2003) using Blosum62. A2CP-C, *Anaeromyxobacter dehalogenans* 2CP-C; AFW109-5, *Anaeromyxobacter* sp. FW109-5; CFETUS, *Campylobacter fetus*; TDENIT, *Thiomicrospira denitrificans*; WSUCCI, *Wolinella succinogenes*; DAROMA, *Dechloromonas aromatica*; DHAFNI, *Desulfitobacterium hafniense* Y51; GTHERMO, *Geobacillus thermodenitrificans*; MGRYPHIS, *Magnetospirillum gryphiswaldense* MSR-1; MMAGNETI, *Magnetospirillum magneticum* AMB-1; MMAGNETO, *Magnetospirillum magnetotacticum* MS-1; GFORSETII, *Gramella forsetii*; RBIFORMA, *Robiginitalea biformata*; PTORQUIS, *Psychroflexus torquis*; PCALIDIFO, *Pyrobaculum calidifontis*.

Table 5. Similarity matrix of amino acid sequences from traditional-*nosZ* group.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1. <i>Paracoccus pantotrophus</i>																															
2. <i>Sinorhizobium meliloti</i>	80																														
3. <i>Brucella ovis</i>	81	94																													
4. <i>Rhodopseudomonas palustris</i>	79	90	90																												
5. <i>Bradyrhizobium japonicum</i>	80	89	89	96																											
6. <i>Methylobacterium</i> sp.	75	83	81	83	83	83																									
7. <i>Stappia aggregata</i>	74	82	81	84	84	85	92																								
8. <i>Silicibacter pomeroyi</i>	73	82	80	79	81	80	83	84																							
9. <i>Roseovarius</i> sp.	70	78	77	80	79	82	84	85	87																						
10. <i>Dinoroseobacter shibae</i>	74	82	81	83	82	83	87	86	86	86																					
11. <i>Rhodobacter sphaeroides</i>	72	79	79	79	78	79	83	83	83	83	89																				
12. <i>Achromobacter cycloclastes</i>	85	79	79	82	81	81	77	76	74	71	76	74																			
13. <i>Alkalitimmicola ehrlichei</i>	83	82	82	82	82	83	80	80	77	72	80	76	86																		
14. <i>Reinekea</i> sp.	83	83	83	82	83	82	78	78	76	74	79	75	83	89																	
15. <i>Psychromonas ingrahamii</i>	86	83	83	81	83	82	78	77	76	72	78	76	84	88	91																
16. <i>Colwellia psychrerythraea</i>	83	84	85	83	84	83	80	80	78	75	81	77	84	91	94	95															
17. <i>Photobacterium profundum</i>	83	83	84	83	83	83	79	79	76	75	79	76	85	88	92	93	96														
18. <i>Shewanella loihica</i>	86	81	82	81	82	82	78	78	75	73	76	74	86	87	86	85	84	84													
19. <i>Marinobacter</i> sp.	86	82	81	82	82	81	77	78	74	72	76	74	88	88	88	86	87	84	93												
20. <i>Marinobacter hydrocarbonoclasticus</i>	92	80	80	80	81	81	78	79	74	72	77	76	87	86	84	86	85	85	88	88	92										
21. <i>Hahella chejuensis</i>	97	81	81	80	81	82	77	77	74	71	77	75	87	87	87	88	87	86	88	88	89	91	92								
22. <i>Pseudomonas aeruginosa</i>	92	80	81	81	82	82	77	77	74	71	77	75	87	87	87	88	87	86	88	88	89	91	92	89							
23. <i>Pseudomonas denitrificans</i>	92	80	81	81	82	82	77	77	74	71	77	75	87	87	87	88	87	86	88	88	89	91	92	89	91						
24. <i>Pseudomonas fluorescens</i>	89	81	80	80	81	81	78	78	75	72	79	75	88	87	84	87	85	84	86	87	87	91	93	91	91						
25. <i>Pseudomonas stutzeri</i>	73	71	71	73	72	73	69	70	68	65	69	68	73	74	72	73	73	74	72	73	74	75	73	74	76						
26. <i>Burkholderia pseudomallei</i>	69	67	66	68	67	67	66	67	66	62	69	68	72	71	68	71	70	70	73	72	72	71	70	73	74	82					
27. <i>Cuptriavidus necator</i>	69	70	69	70	70	69	66	68	67	62	66	66	71	72	71	70	71	71	73	71	71	71	70	71	72	89	83				
28. <i>Azoarcus</i> sp.	69	67	67	70	70	68	67	68	66	61	67	66	71	72	71	69	70	70	73	70	70	70	70	71	72	90	83	93			
29. <i>Thiobacillus denitrificans</i>	69	67	66	67	68	68	67	67	66	63	67	67	71	69	68	71	69	70	73	70	70	70	70	70	72	74	83	95	82	81	
30. <i>Acidovorax</i> sp.	64	61	61	62	62	65	61	63	59	60	61	60	67	61	61	62	62	62	64	64	62	64	65	65	66	67	66	66	65	65	66
31. <i>Haloracula marismortui</i>	62	64	63	63	63	64	64	64	59	61	59	59	64	61	61	60	61	63	64	64	62	62	63	63	62	64	64	64	64	64	64
32. <i>Halorubrum lacusprofundi</i>																															77

The matrix was generated using MatGat Matrix Global Alignment Tool (Campanella et al. 2003) using Blosum62.

Table 6. Diversity indexes and Richness estimators.

Library	Total No. of sequences	No. of unique OTUs	Index of diversity			Estimated no. of genotypes			
			Shannon	Simpson	ACE	Boot	Chao	Jack	
FBR	12	9	2.13833	0.045455	18	11.4484	12.75	15	
MWWT	60	30	3.14965	0.039548	57.4095	37.1583	49.4286	48.927	
CORCO	24	14	2.49906	0.054348	22.6957	17.3461	21	22	

Shannon-Weaver and Simpson diversity indexes (Magurran 1988), the abundance-based coverage estimator (ACE) (Chao and Lee 1992; Chao and Yang 1993), and the bias-corrected Chao1 (Chao 1984), interpolated jackknife (Burnham and Overton 1979), and bootstrap (Smith and van Belle 1984) richness estimators with a 95% confidence interval (CI). Distance values were 0.03 for FBR, 0.04 for MWWT, and 0.04 for CORCO library.

Table 7. Expected *nosZ* amplicons based on *HinPI*, *NlaIII*, and *RsaI* T-RF fragment lengths for MWWT clones.

Clone	<i>HinPI</i>	<i>NlaIII</i>	<i>RsaI</i>
MWWT 20	167	117	253
MWWT 24	266	885	816
MWWT 27	135	151	27
MWWT 29	137	397	516
MWWT 31	275	117	516
MWWT 34	275	117	223
MWWT 43	352	132	36
MWWT 44	134	397	98
MWWT 46	275	117	223
MWWT 52	135	87	278
MWWT 63	137	397	517
MWWT 68	274	116	222
MWWT 70	137	87	21
MWWT 71	275	117	223
MWWT 72	135	87	516
MWWT 73	155	415	534
MWWT 82	137	397	516
MWWT 83	135	87	21
MWWT 85	134	86	26
MWWT 86	135	397	27
MWWT 87	274	116	222

MWWT 89	253	87	21
MWWT 90	137	87	642
MWWT 93	521	781	405
MWWT 94	146	98	416
MWWT 102	54	411	237

Table 8. Expected *nosZ*-amplicon based *HinPI*, *NlaIII*, and *RsaI* T-RF fragments lengths for reference organisms carrying a non-denitrifying *nosZ* gene (NosZ943F/NosZ1789R primer pair).

Organism	<i>HinPI</i>	<i>NlaIII</i>	<i>RsaI</i>
<i>A. dehalogenans</i> 2CP-C	156	202	639
<i>Anaeromyxobacter</i> sp. FW109-5	214	562	390
<i>D. hafniense</i>	431	117	560
<i>G.thermodenitrificans</i>	702	117	N.R. ^a
<i>M.gryphiswaldense</i>	307	333	642
<i>M.magneticum</i>	262	117	21
<i>M.magnetotacticum</i>	477	117	27
<i>D.aromatica</i>	137	397	21

^a N.R., no restriction product.

APPENDIX

APPENDIX 1

Partial sequences of the *nosZ*-gene of environmental clones and *A. dehalogenans*

2CP-3 and *Anaeromyxobacter* sp. strain R.

>*Anaeromyxobacter* sp. strain R

cccggtgttcgcgcgcgtactcgggcaccgggtacggcttcgacgacgagtcgaaggccatgctcggcaacctgacctgggggt
gacgtgcaccaccggcgctctccgagaccggcggcgactacgacgggcgctggctgttcgtgaacgagatgaacggggcgc
gtggcgcgcatcgacctgcgcgacttcaagacgcggcagatcatcgggccgggtgccgaacatctcgggcaaccacgggtcc
acctcatcacgcgaacagcgagttacatcctcatgtcctcgcgggttctccatccccatcccaagggggaaggcggtgtcgate
gaccgggtacgccagcgagtacaagggcgctcgcggccgggatcaaggctgacccgaagtccggccagatgtcgtgggctg
gcaggtgtcctcgcgcggttcgactgggacctcggggacgcgggcaagaagctgtcggagggtgtgttcttctcacctgtc
acaactccgagcgcgccaccgggaagctcgaggtcaccgcctcgcagcgcgaccgcgactacatcgcggccatcgactggc
gcctggccgagaaggcggcggcgagggggaaggcgagctgatcggcggcgtgaagggtgtcgtatccgaagaccgtccc
cgggctcgtgtacctgtgccctgcgggaagtcgccgcacggcgtggacgtctcgccgacggcaagtacgtggtcgggtcc
ggcaagctccaggcgctcaccaccgccttcaacttcgagaaggtcctaccgccatcaagaacaaggacttcgccggcgagg
aggacggcatcccggtctcaagtacgagtcacatcaaggacgcggaggtgccggtggggctcgggccgctgcacacgcagt
tcgggccccgacggg

>*A. dehalogenans* strain 2CP-3

[illegible]

>fbr1

ccgcacggtgtggacgtgtctccgatggcaagatgctggctgtgtcgggcaagctggacacgcacgtgtcgggttacagctt
cgagaagattcaggcggcgatcaaggccgagaagttcgaatcgaaggaccgtacggcatcccggtgatcgggatgaagga
cgcgtgcacacgcaggtgcaactgggcctgggcccgtgcacacgcaatacgaactgaagaactgcgtcgctacacgtcg

ctgtacgtggacagccaggtcgccaagtggaaacttctcgaaaggcaaggtgctggacaagatctcgggtgactacaacatcgg
tcacctgatgacgatggaaggcgactcggatccgaaggccgctacgtggtggcgtgaacaagctgtcgatcgaccgt
tcaaccgggtggggccgctgacccgcgaaccaccagctgatcgacatctgaacgacaagatgcagctctgtacgacatg
ccgtgcccgtggggaaccgcactacgtggtggcgatcgaagcgtcgaagctgaagccgggctgcgtacaaggctggc
accaacagccggacggacaagccgcacccgggctgctgcggcggaagagcggatcgagaagaagggaacaag
atcacgggtgatggcacgctgatccgttcgacatcacgccggagacgatcgaagcgggaagtcggggatgaagtgcgatcc
acctgaccaacctgaacgggcacaggacgagacgcacggcttcacgggtgcgacgtacaacgtgcatgcgtcggtcgagcc
gggcaagacgggtgacgggtgaaggtcaaggccgacaaggaaggcgtgtaaccttactactgcacccat

>fbr13

gatgtgtcgccggacggcaagtacatggctgctcggcaagctcgacacccatgtctcgggttacagcttcgagaagatccag
gcccgcacatcaaggcggggaattcgagagcaaggaccctacggcatcccgggtgatcgccctgaaggacgccctgcacgtg
caggtgccgctcggcctcggacccctgcacacgcagtatgactccaagccctgcgtcgcttacacctcgtgtacgtcgactcc
caggtcgccaagtggaaacttctcgaaaggcaaggtgctcgacaagatcagcgtgcactacaacatcgccacctgatgacct
ggagggcgactccaccaagccggcgccggaagtatctcgtcgctgaacaagctggcgatcgaccgttctgctggcggtcggc
ccgtgcatccgcgaaccaccagctcatcgacatcagcaacgacaagatgcagctgctctacgacacgccgtgccgctgg
gcgagccgcactatgccgtctccatcgccgcaccaagctgaagccgggctgcgtacaaggctggcaccgacagccgca
ccgacaaccgcacccggcgccgtgcgcggcgaggagaagaccgagaagaagggaacaaggtcacgggtcgggg
cacctgatccgctcgacatcacgccggagaccatcgaggcggaagtgggagacgagatcacctccgctgaccaacctc
gagcgcgcccaggacgagaccacggcttcacgggtgcgaccttaacacccacgcctcggtcgagccgggcaagacggct
acggtaagttcaaggccgacaaggaaggcgtctatcgtactactgcacgtacttctgctc

>fbr14

atgtggactcgccagtggcgaaagtggaaactactcgaaaggcaaggtgctggacaagatcagcgtgcactacaacatcgcca
cctgatgacctggaaggcgattcgaccaagccggcggaagtacctggtggcgtgaacaagctggcgatcgatcgcttc
gtgccgggtgggcccgtgcatccgcgaaccaccagctgatcgacatcagcggcgacaagatgcagttgctctacgacatgc
ccgtgccgctgggtgagccgcactacgtggtggccatcgatccaagaagctgaagccggcgtgcgtacaaggtcggca
ccgacagccgcaccgacaaggcgcatccggcgagtcgcgcggcgaggagaagatcgtcaagaagggaacaaggt
cgaagtcttcgccacctgatccgttcgacatcacgccggagaccatcgaagtcgatgtcggcgacgaggtgacgatcaacc
tgaccaacctgaacgggagcaggacgagacgcacggcttcacgggtcgcacctacaacgtgcatgcctcggtcgaaccggg
caagacgggtccaggtcaagttcaaggccgacaaggaaggcgtctatcctgctactgcacctaattctgctc

>fbr15

ccgcatggggtgatgtctggcccgcagggcaagtacatcctcggtccggcaagctccaggggctcaccaccgcttcaactt
cgagaaggtgatgacggcgatcaagaacaaggacttcggcgacgaggacggcatccccgtcctaagtacgagtcgat
caaggacgccgaggtccccgtcgccctcgggcccgtccacacgcagttcgggccgaacggcggggctacacgtcgtctt
cgtcgactcggcgatcggaagtggaaagctcgggacctgggaggtcagcagaagatccccgtctcctactcgtcgccacc
tgaccggcgccgaggggcactccgtcagccggacggcaagtacctcgcagcctgaacaagctctcccaggtcggcacct
ctcgggtgggcccgtcgagcccagtcgtcccagctcatcgacatcggcgaggagaagatgaagctcctctacgacgccttca
ccgagccggagccccactacgccagatgatcaaggcggacaagatcaagccgatcgaggtctacccgaaggaggagaac
aagcaccgcgtcgcggtctgggacgtcaaggacgcgcgtcccagcgcgacggggagcgaggtctcgtcaagatggtcctc
gtccggtcgacctactccgacctcatcgaggtgaaccggggcgacaccgtcaaggtcgcgatgacgaacatcgagcaga
cgacggagcagctccacggcttcgggctcctcgactacaacatcaacgtcgtcgtcgacccggcgagacgaagacgggtcac
gttcaaggccgacaagccgggctcttccgtactactgcacgtcacttctgctc

>fbr2

ggatgtctgtccagatggcaagtacgtcatcggtccggcaagctccaggggctcacgaccgcttcaacttcgagaaggtcct
caccgcgatccggaacaaggacttctcggcgaggaggacgggatccccgtcctgaagtacgagtcgatcaaggacgccga
ggccccgtcgccctcggcccgtccacacgcagttcggggccgacggatacgctacacctcgtctctcgtcgactcggcgg
tcgcaagtggaaagctcgggacctgggaggtcgtcgacaagatccccgtctcgtactccatcgccacctctgcgccgccgaa

ggggacaccgtcagccccgacgggaagtacctcgtcgccctcaacaagctctcgacggccggcacctctccgtcgggccgt
cgagcccagtgctcgcagcttatcgacatcaccgaggagaagatgaagctcctctacgacgcttcaccgagcccagcc
gcactacgccgtcatggtgaaggccgacaagctcaagccgagtgaggtctacccgaaggaggagaacaggaacccccctgc
ggtctgggacgtaaaggatccggcggtgacgcgcaaggggagcgaggtcctcgccagggtcgtcctcgtccgtccacgac
accccgagggcgatcgaggtccagcagggggacacccctgaaggtcgcgacacgaacatcgag

>fbr25

gacgtgtgaccggatgggaagtacatcatcggtcgggcaagctgcagggcgtagccacggccttcaactgggagaaggtcc
agaccgccatgcgcaacaaggacttcagcggcgaggaggacggcatccccgtgctgaagtacgagtcgatcaaggacgcc
aggtgccggtggggctgggcccgtgcacaccagttcgccccgacggctacgcctacacgtcgtgtctgtggacagcgc
catcgccaagtggaaagctcggcacctgggaggtggtcgacaaggtccccatgtcctactcgatcgccacctcagggcgcc
cacggcgacaccgtcaagccccacggcaagtacgtggtgggacctcaacaagctctcccacggcgggcacctcaacgtcggc
ccctcgcagcccagtcgagccaactggtcgacatcagcgaggagaagatgaagctgctgctcgacttcttcacggagccc
agccgcactacgccagatgtcaaggccgacgtgctcaaaccgacgaggtctacccgaaggaggtgaacaagcaccgct
ggcgatctgggacgtcaaggacgccggcggtgaccaggaacggcaacaaggtcacggtcaagatgggtggcggtgcgtcga
ccctgacccccaccgacttcgagatcaaggaaggcgacgagatcacgatcgccatcacgaacatcgagcagaccaccgacg
agctgcacggccttgcatcctcgactacaacatcaacatcgtgggtggaccggcgagaccaagacgggtcacgttcaaggcc
aggaagaagggtgtctacccgtactactgacccacttctgctc

>fbr28

cctcacggggcggtgtgacaccagatggcaagtacgtcatcggtcggcaagctccaggcgtagccacggccttcaact
cgagaaggtcctcaccgcgatccggaacaaggacttctcggcgaggaggacgggatccccgtctgaagtacgagtcgatc
gaggacgccgaggtccccgtcggcctcggcccgtccacacgcagttcggggccgacggatagcctacacctcgtcttgc
tcgactcggcggtcgcgaagtggaaagctcgggacctgggaggtcgtcgacaagatccccgtctcgtactccatcgccacctc
tgcgccgccgaaggggacaccgtcagccccgacgggaagtacctcgtcgccctcaacaagctctcgcacggccggcacctc
tccgtcggggcgtcgcagcccagtcgtcgcagtcacatcaccgaggagaagatgaagctcctctacgacgcttcac
cgagcccagccgcactacgccgtcatggtgaaggccgacaagctcaagccgatcgaggtctacccgaaggaggagaacg
ggaacccccctcgggtctgggacgtaaaggacgccggcggtgacgcgcaaggggagcgaggtcctcgccagggtcgtcctc
gtccgtccacgatccccgcaggcgatcgaggtccagcagggggacacccctgaaggtcgcgatcacgaacatcgagcag
acgacggacgagctccacgggtgggggctcctcgactacaacatcaacatcgtcgtcgacccggcgagacgaaggctcgtca
cgttcaaggtggacaagccccggcggtctacccctactactgcaccagttctgctc

>fbr31

gacgtgtctccggatggcaagttcatcaccgtttccggcaagctggacaccacgtttccgtctattccttcgagaagatccaggc
cgccatcaaggccggcaagttcgaatccaaggatccgtacggcattccggatcaacatgaaggatgccctgcacaccagg
ttcaactgggtcttggccccctccacaccagtagcactcgaagaagtgcgtcgctacacctcctgtatgctgactcgcaggt
cgtcaagtgggatactcgaaggcaaggttctcgacaagatctcgggtgactacaacatcggtcacctgatgacctggaagg
tgactccatggatccgaaggccgctatctggtcgccctgaacaagctgtcgatcgatcgcttcaacccggttggccgctgcat
ccgcagaaccaccagttgatcgacatcaccagcgacaagatgcagctgctgtacgacatgccgttggcgtgggtgagccgca
ctacgtcgtggccatcgaagttccaagctgaagccgggtgtccgtacaaggtcggtaccgattcccgtacggacaagccgc
atccggggcgccgtccgtgccggcgaagaagtaacgaccaagaagggaacaagatcgaagtaagggtacgttgatccgt
cgcatatcacgccgaaaccatcgaggcggaagtggggcgacgaaatcacctccacctgaccaacctcgagcgtgccagg
acgaaacccacggtttaccgtgtcgacctacaacgtccatcgctcggtcgagccgggcaagaccgtcacggtcaagttcaag
gctgacaaggaaggcgcttatccatactactgcaccagttctgctc

>fbr34

tcccttcataccagtagcagcaagcaatgcgtggcctacacctcgtgtatgtggactcgaagtggcgaaagtgaacta
ctcgcaaggcaaggtgctggacaagatcagcgtgactacaacatcgccacctgatgacctggaaggcgattcgaccaag
ccggccggcaagtacgtggtggcgctgaacaagctggcgatcgatcgcttcgtgccggtggggccgctgcatccgcagaacc
accagctgatcgacatcagcggcgacaagatgcagttgctctacgacatgcccggtgccggtgggtgagccgcactacgtggtg

gccatcgacgccaagaagctgaagccggcgtgcgtacaaaggtcggcaccgacagccgaccgacaaggcgcatccggg
cgagtgccgcccggcgaggagaagatcgtcaagaagggaacaaggtcgaagtctcggccaccctgatccgttcgcacatc
acgccgggagaccatcgaagtcgatgtcggcgacgaggtgacgatcaacctgaccaacctgaacggggcgaggacgagac
gcacggcttcacggctcgcacctacaacgtgatcctcggtcgaaccgggcaagacgggtccaggtcaagttcaaggccgac
aaggaaggcgtctatcccttactactgcacctaattctgctc

>fbr35

ccgcacgggtgtgcagtgagaccggatggcaagttcatcatcgtctcggcaagctcgataccacacttcgggtgtatagcttcg
agaagctgcaggccgccatcagcggcgcaagttcagtgccaaggaccgtacgggtgtgcccgtcatcgccatgaaggatgc
gctggacaagcaggtgcagctcggcctcggccccctgcataccagtagattcgaagccgtgtatcgctatacctcccttat
gtggattcccaggtgccaaagtggaaactattgcgagggcaagggtggtggacaagattccggtccattacaacatcggtcacctga
tgaccatggaaggcgattccatggatccgaagggcggttacctggtggcgctgaacaagctggccatcgaccgttcgtgccc
gtcggccccctgcacccgcagaacctcagctgatcgacattcaaacgacaagatgcagttgtgtacgacatgccgctgccg
ctgggtgagccgcactacgccgtcgccatcgaggcttcgaagctgaagccgggctccgctacaaggtgggtaccgattccc
ggaccgacaagccgcacccgggtgccgtccgtgccggcgaggagaagattgtcaagaagggttaacaagatcgaggtcttcg
ggacctgatccgttcccacatcactccgaaacctcgaggctgatgtcggcgacgaggtcacgggtgcatctgaccaacctc
gagcgtgcccaggacgagaccacgggttcaccgtctccacctacaacgtccatgcgtccatcgagccgggcaagaccgtca
ccgtcaagttcaaggctgacaaggaaggcgtctatccgtactactgcacctagtct

>fbr4

ccgcacgggtgtggatgtcagaccggatggcaagttcctaccgtcggcgcaagctcgacacgcacgtgtccgtctattccttc
gagaagatccaggccgcgatcaaggccggcaagttcgaatccaaggatccctacggcatcccgggtgatcgccatgaaggac
gcgctgcacacgcaggtatcgctgggctgggtcccttgcataccagtagacagcaagcaatgcgtggcctacacctcgct
gtatgtggactcgcaagtggcgaaagtggaaactactgcgaaggcaaggtgctggacaagatcagcgtgcactacaacatcggc
cacctgatgaccatggaaggcgattcgaccaagccggcgcaagtagctggtggcgctgaacaagctggcgatcgatcgct
cgtgccgggtgggcccgtgcacccgcagaaccaccagctgatcgacatcagcggcgacaagatgcagttgctctacgacatg
cccgtgccgtgggtgagccgcactacgtggtggccatcgacgccaagaagctgaagccggcgtgctgctacaaggtcggc
accgacagccgcaccgacaaggcgcatccggcgcgagtgccgcccggcgaggagaagatcgtaagaagggaacaag
gtcgaagtcttcgccacctgatccgttcgacatcacgccggagaccatgaagtcgatgtcggcgacgaggtgacgatcaa
cctgaccaacctcgaacggcgaggacgagacgcacggcttcacggctcgcacctacaacgtgcatgcctcggctgaaccg
ggcaagacgggtccaggtcaagttcaaggccgacaaggaaggcgtctatcccttactactgcacctaattctgctc

>fbr6

agatggcaagtacatgccgtgggcccggcaagctggacacgcacatgcgtcgggtgtatagcttcgagaagatccaggccgcgac
caggcggggcaagttcgaatcgaaggaccctacggcgtgccgggtgatcgccatgaaggaaagcgtgcacacgcaagtgcaa
ctgggcttgcccgtgcacacgcagtagcactgaagccgtgcgtggcctacacgtcgatgtacgtggacagccaggtgg
cgaagtgggactactgtcagggcaaggtgctggacaagatcagcgtgcactacaacatcggtcacctgatgacgatggaaggt
gactctgccgatccgaaggccggttacctggtggcgccgaacaagctggcgatcgaccgttctgctgccgggtgggcccgtgca
cccgagaaccaccagctgatcgacatcagcaacgacaagatgcaactgctgtacgacatgccgtgccgtgggagccg
cactacgtggttccatcgaagcgacgaagctgaagccggcgctgcgtacaaggtgggcacggacagccgtacggacaag
ccgcacccgggtgccgttcgcggcggaagagacgacgacgaaaaagggaacaagatcgaggtgaagggcacgctgat
ccgtcgcacatcacg

>fbr9

ctacacctcgctgtatgtcgactcccaggtcgccaagtggaaacttctgcgagggcaaggtgctcgacaagatcagcgtgacta
caacatcgccacctgatgacctggagggcgactccaccaagccggcgcaagtatctcgtcgcgtgaacaagctggcg
atcgaccgttctgcccgtggggccgtgcacccgcagaaccaccagctcatcgacatcagaacgacaagatgcagctgct
ctacgacatgccgtgccgtggcgagccgcactatgccgtctccatcgccgccagcaagctcaagccgggctgcgtac
aagtgccggcaccgacagccgcaccgacaagccgcacccggcgccgtgcgcgcccggcgaggagaagaccgagaagaag
ggcaacaaggtcacggctctgggtacgctgatccgtcgcacatcacgccgagacgatcaggcggaagtgggcgacgag

gtcaccatccacctgaccaacctcgagcgcgcgcaggacgagacccacggcttcacgggtgcaccttcaacacgcacgcct
cggtcgagcccggaagacgggtgacgggtcaagttcaaggccgacaaggaaggcgtctatccgtactactgcacctacttctgt
c

>corco10

ccgcatggggcggtatgcaggccagatggcaagtttggtgctgtttccgggaaactcgacccgcacgtgacgggtactccgc
cgagaagatccgcaaggcgatggccgggaagaagtattcaggatccgatgaatatgggtcccgatcctcgacttcgactccg
tcaaggaggcgcaggctgaagtcggtctgggcccgtgcacaccagtttgacgacaaggggcaggcctacacctccctgttc
ctggactcggcggtggcccgtggacgctggggggcgacttcaccaagctccacggccagcagccctggaccatggtccac
aagacgcccgtgcagtacaacatcgggcacctggtggcgccgaaggcgactcggtagccccgacggcggttacctggt
ctcgatgtccaagtggtcggtggaccgcttctcccgccgggtccgctgcttccgcagaacttcagctcctggacatctccggg
tccggggccgacatgcccgtcatctacgatcgccccatcgcggtggcgagccgcactacgcggtgatgatcaagccgaca
agctcaagccctggacggtgtatccggaggtggggtggaatccgcacaccagatgctggatccagcctcgccgcgtc

>corco11

ccgcatggggcggtatgcaggccagatggcaagtttggtgctgtttccgggaaactcgacccgcacgtgacgggtactccgc
cgagaagatccgcaaggcgatggccgggaagaagtattcaggatccgatgaatatgggtcccgatcctcgacttcgactccg
tcaaggaggcgcaggctgaagtcggtctgggcccgtgcacaccagtttgacgacaaggggcaggcctacacctccctgttc
ctggactcggcggtggcccgtggacgctggggggcgacttcaccaagctccacggccagcagccctggaccatggtccac
aagacgcccgtgcagtacaacatcgggcacctggtggcgccgaaggcgactcggtagccccgacggcggttacctggt
ctcgatgtccaagtggtcggtggaccgcttctcccgccgggtccgctgcttccgcagaacttcagctcctggacatctccggg
tccggggccgacatgcccgtcatctacgatcgccccatcgcggtggcgagccgcactacgcggtgatgatcaagccgaca
agctcaagccctggacggtgtatccggaggtggggtggaatccgcacaccagatgctgg

>corco14

cgtgactcccgatgggaaatggatagtcggcggcggttaagctggccgccgttgctgcgtacatagctttgaaaagatgcagg
aagccattagtaaaggcgattttgacggagagggttatgggaataccattctaaagtacgaaacggtcagagtcgccgaggttcc
gataggcttgggtcctctgcacaccagtttgacgaccagggtatgcttatacttcttattcctggactcagccgttgccaagtgg
aagctgggtacttgggaagtagtgataagataccggtttactataacatcgacaccttgccgcagtccatggagacaccgcta
aaccgggggcccagtagtctggccatgaataaaatcgccaagggcagatacctgccgttgaccttccatccggagagc
gcacagcttatcgacatatccgggaagaagatggaattgctgctcgtattccctacgcaggtagagcctcactacggtcaaataat
cgatgctaccttgataaagcccgtgctcgtatatccactgaagagaataaccacccttacgccattaagagcgcagaccaggct
cgtgtggaaaggaacggaaatagagtggacgtttacatgctggcaataagaagccatttcactcctcatctctttgagttt

>corco16

cgtgactcccgatgggaaatggatagtcggcggcggttaagctggccgccgttgctgcgtacatagctttgaaaagatgcagg
aagccattagtaaaggcgattttgacggagagggttatgggaataccattctaaagtacgaaacggtcagagtcgccgaggttcc
gataggcttgggtcctctgcacaccagtttgacgaccagggtatgcttatacttcttattcctggactcagccgttgccaagtgg
aagctgggtacttgggaagtagtgataagataccggtttactataacatcgacaccttgccgcagtccatggagacaccgcta
aaccgggggcccagtagtctggccatgaataaaatcgccaagggcagatacctgccgttgaccttccatccggagagc
gcacagcttatcgacatatccgggaagaagatggaattgctgctcgtattccctacgcaggtagagcctcactacggtcaaataat
cgatgctaccttgataaagcccgtgctcgtatatccactgaagagaataaccacccttacgccattaagagcgcagaccaggct
cgtgtggaaaggaacggaaatagagtggacgtttacatgctggcaataagaagccatttcactcctcatctctttgagttt

>corco17

agatggcaatacttcacgcggcggaagctgcaatcaatcaccactgcctacaactttgaaaaaatgatgacggccattcag
aacagggtattttccggcaacgaagatggcatacctgtgtgaattacgacagcgtaaggatgccgaagtcaatgtcgggtcgg
ggcgttgataccagtttgatgacaggggtacgcctacaccagcctgtttgtcgaaagcgcagtcgccaatggaaactcg
gcagtcgggaagtcgccgacaagattcccgtagttacaacatcgccacctttgcacggccgaaggcgatacggtaacccg
gacggcaattcctgatcgcgctcaataaactctcgcacggacggcatctcagcgtcgggtccctcgcaaccggaaagctcgca
attgatcaacatcggcggcgaaaagatgaagcttctgtatgacgtttaccgaaccgaaccgattatgcgcagatgatcagg

gcccacaagctgaaaccgattgaggtttatccaaaaggagaacaaaaaccgagcgccatttgggatgtcaaggacgcca
gggtcgggcgagcggaatacgggtgatcgccaaagtgtggtgtacgcagcacgattgaac
>corco18

ccgcacggcgccgacgtgagccccgatggccggttcacatcgtttccggcaagctggaccgcacgtgacgtctactcgtg
ggagaagattcagcaggccatccgcgaccggaatttcgcggaacgaccgctacggcatcccgggtgctcgacttcaacagc
gtcatcaatggccagctcgaggtcgggctcggcccgtccacacgcagttcgacgaccgcggaacgcctacacgagcatctt
cctcgacaccaaggtgacgaagcggagcctcgagccgctcgggtggtggaccagatcgacgtgcactacaacgtcggcca
cctcggcgccgagggcgacgcccgtcggcgacggcaagtacctgggtgccatgaacaagtggtcgctcgaccgatt
caatgacgtggggccgctgcacccgcagaacttcagctcatcgacatctcgggtgagaagatgcagctgctctacgacatgc
cgatcggtctcggcgagccgactaccccagatcatcaaggcggacaagctgcaccggtgcaggtgtacccccagatcgg
ctacgatcccgccacgatgagcgcgcgcggaacgccgtgcagctcggcaacgagggcatcacgcgccgtggcaacgaag
tgaccgtgcgcatgacggccatccggagccactcaagcccagctacgtccgggtgcgacggcgaccgctcacgatcc
acatcacgaacgtggagcaggcccgcgatgccacgcagtggttcgcatcgccaagtacaacgtcaacctgagcctcgagcc
cgcgagcatcagacggtagcttctgtggccgacaagcccggcggtgttccctactactgcacctagtctg
>corco2

ccgcacggggtcgtatgtgacgccagacggcaatacatcctggcagcggcaagctgcagtcgatcaccaccgccttcaatat
ggagaagatcctcaccgccatcgagaacaaggacttcaccggggatgaagacggcatccccgtctcgattacgagaaggtg
aaggatgccgaagtcgaagtcggactgggaccgctgcacacgcaattcgatgatcagggcttcgctacacgagcctcttctg
gagagcgggtggcggaagtggagctcggcagctgggaggtggtcgacaagattccgatcacctacaacatcgcccatctcg
cgacggccgaaggcgacacggtaacccggacggcaatatctggtcgcgtcaacaagctctcgacggcgacatctcgg
cgtcggcccgtcgagcccagagttcgagctgatcaacatcacggcgagaagatgaagctgctctacgatgccttaccg
agccggagccgactacgcgcagatgatcaaggccgacaaactgcagccgatcgaggtctaccgaaggaggagaacaag
aaccgcacgcgatctgggacgtcaaggacgcgggagtcgagcgcagcggcgctaattgtacgggtgaagatggtcgccgtc
cgacgacactgaaccgacttcgatcgaggtcaacc
>corco26

cctcacggggtggatgtgtcaccagacggcagttatttcgtcggcagcggcaagctggaagccgtgaccacgggtctacagcttc
gacaggatgatcgccgtatcagcaacaaggatttcagcaaggacgaagacggcattccagttctgaagtacgagagcgtggt
cgaacgcgaagtgcagatcggccttggccccctgcacacgcagtagcagcaccaggggatggcatacacaagcctgttcgtc
gacagcgaatcgctaagtgggacccgaagaccggggaagtactggacaccatccctgtcgcctatatcatcgccacctgtg
cgtcccaggcgccgacagcaccacccctacggcaagtatgtgatcgcgtcaacaagctgagccacggccggcaccacaa
cgtaggtccaagccagccggagagcagccagctcatcgacatcaccggcgagaagatgaagatcatctatgaagcctttaccg
agcccagccgactacgcccaggccatcccagccgacaagctcaagcccattgaggtttatccgaaggaagagaacacga
acccgcacgcgatctgggacattaaagattccggcgtaaccagcagtggtggaacggtagaggtaaagatggtcgcagtagt
tccacactgcacccgacaaacgcggaggtcaaccgaggcgaaaccgtgaccttccacattacaacattgagcaaaactaccga
cgagctgcacggattcgcgatctgtgaatacaacatcaacctggtcattgatccgggggaaaccaagagcgtcactttcaagct
gacaagccggcgcttatccatactactgcaccaattctgctc
>corco29

ccgctgcacacggagttcgacggccgcggaacgcgtatcagacggcggttcatanctcggagatcgtgaagtggagctgg
ggacgtgggaggtgctggaccggatccccgtctactactcgggtggccacctcatgattccgggcgggcgacagccggcagcc
gtcgaacaggtacatggtggcggtgaacaagatcacgaaggaccgctacctgccgacggggcccgagctcgcgcacgcggc
gcagctgatcgacatctcggcgacaagatgaagctgctgctggacttccgacgcagggcgagccgattacgcgcaggcg
ctccggcgagctgatccgcgacaggcaggtgcagatgttctcgtggagagcaaccggcaccgggtacggcggaagag
cgaagccgagacgaaggtggagcgaacggccgccgctggacgtgtggtgatggcgatccgcagccacttctcggcg
acaacatcgaggggatccaggtgggtgacagcgtctacttccacgtcaccaacctcgagcaggactgggatgtgccgcacgg
cttcgccgcttggcgggcgcaacgccgagctgctgatcatgccggcgacacacgcacgctgacctgggtgccgctgcg
ggtggcgctgttcccttactactgcagtaattctgctc

>corco30

gatggcaaatacttcatcggcggcggaagctgcaatcaatcaccactgcctacaactttgaaaaatgatgacggccattcaga
acagggattttccggcaacgaagatggcatacctgtgtgaattacgacagcgtcaaggatgccgaagtcaatgtcggctcggg
gccgttgcataccagtttgatgatcaggggtacgcctacaccagcctgtttgtcgaaagcgcagtcgccaaatggaaactcgg
cagtcgggaagtcgccgacaagattcccgtcagttacaacatcggccacctttgcacggccgaaggcgatacgggtcaatccgg
acggcaaattcctgatcgcgctcaataaactctcgcacggacggcatctcagcgtcggtcctcgcgaaccggaaagctcgaat
tgatcaacatcggcggcgaaaagatgaagcttctgtatgacgcttcaccgaaccgaaccgcattatgctcagatgatcaggg
ccgacaagctgaaaccgattgaggtttatccaaaaggagagaacaaaaacccgagcggccatttgggatgtcaaggacgccag
ggcggggcgacgggaatacgggtgatcgccaaagtgtggtgtacgcagcacgattgaaccgacaaaaatcgaagtcaat
cagggagacaaggtgatcatttcatcaccgacatgaacagacgaccgatgaactgcacggattcgggtgaacgaatacaa
catcaatgtggtggtgatccggcgaaaccaagacaattgaatttgcgcggataaacccggcgttttccttactactgcaccta
tttctgctc

>corco32

ccccacggcgtggatgtcagaccgatggccagtacatgggtggtcggggcaaacctcgatccacacgtgacggtttattcgatc
gacaagatcaaggccgccatcgcgcgaagaagtactgacatccgacgagtagcggcgtcccgggtgctcgactttgacgcgg
tgaaggagagcaggtcgagctgggccttgggcctctgcacacgcagttcgacaacaagggtatgcgtacacatcgctgttc
cttgattcggcagtgggccgctggctcctgggcggggattacgccaagctccacgccgagcagccttggaaagctggtgcaca
gacgccgtgcaatacaacatcggccacttgtgctcggccgagggcgacacggtcagccccgacggcaacttctgatctcg
atgtccaagtgggtccgtggaccgttcttcccaccgggccccttcttcgcagaacttccagttctcgacatcgctcggggcg
gcacgaacatgccgtgatctatgacagccccatcggcgtcggcgagccccactactgccagatgatcaaggccgacaagct
gcagacctggaaggtctaccccagctcggtggaacccgcgacgcagcagctggaccccaacgccccgaagaaaggcg
atgaaggcgtcacgcgaacgggaacaaggtctcgtcaagatgaccgccgtccgagccacttctccccggagcatgtcatg
ctcgaggaaggcgatgaggtcacctggcgcatctgcgtaccgaaaccgcgcaggtatccacgcacggcttcgcacatcggc
gggtacaacatcaacctctcacttgagccgggtgagtacagcaggtcaagttaccgccgaccgcgccggcacctaccgta
ctactgcacccaattct

>corco33

ccgcatggggcggtatgtcaggccagatggcaagtttgggtcgtttccgggaaactcgacccgcacgtgacggctactccgc
cgagaagatccgcaaggcgatggccgggaagaagtattcaggatccgatgaatatgggggtcccgatcctcgacttcgactccg
tcaaggaggcgaggtgaagtcggtcgggcccgtgcacaccagtttgacgacaaggggcaggcctacacctcctgttc
ctggactcggcggtggcccgtggacgctggggggcgacttcaccaagctccacggccagcagccctggaccatggtccac
aagacgccggtgcagtacaacatcgggcacctggtggcgccgaaggcgactcggtagccccgacggggcggtacctggt
ctcgatgtccaagtggtcggtggaccgttcttcccggcggttcggtcgttccgcagaacttccagctctggacatctccggg
tccggggccgacatgcccgtcatctacgatgcgccatcggcgtgggcgagccgactacgcggtgatgatcaaggccgaca
agctcaagccctggacgggtgtatccggaggtggggtggaatccgcacaccagatgctggatccagcctcgccgcgtcgtgg
acacgaaggcgtggtgcgcgaggggaacaaggtcacatcaacatgaccgcgatccgcagccacttcacggccgagcacgt
ggaagtccaggaggcgatcacgcgacgtggcgcatcaccacacctggagaccagcatcgacgccaccacggcttctgcat
cggcgggtacaacatcaacctgagcatcgagccgggtgaatatgaggaggcgacattcacggccgacaagccgggaacgta
cccatactactgcacccaattgtgctc

>corco35

tgcagtacaacatcggccacttgtgttcggccgagggcgacacggtcagccccgacggcaacttctcatctcgatgtccaagt
ggcgtggtgacgccttcttcccaccgggccccttcttcgcagaacttccagcttctcgacatcgctcggggcgacgaacat
gcccgtgatctatgacagccccattggcgtcggcgagccccactactgccagatgatcaaggccgacaagctgcagacctgga
aggtctaccccaggtcgggtggaacccaccacgcagcaactggatcccaatgccccgaagaaggcgatgaaggcgta
cccgaacgggaacaaggtcctggtcaagatgaccggctcgcagccacttctccccgaacacgtcatgctcaggaagg
cgatgaagtcacctggcgcatctgcgtaccgaaaccgcacaggacgccacgcacggcttctgcatcggcggttacaacatc
aacctctcacttgagccgggtgagtactgcgagttcaagttaccgccgaccgcgccggcacctacccttactactgacccac

>corco41

ccccacggggtggatgtctgccccgacggcaatacatcatcggcagcggcaagctgcagtcgatcaccaccgcttcaatat
ggagaagatcctcaccgccatcgagaacaaggacttcaccggggatgaagacggcatccccgtcctcgattacgagaaggtg
aaggatgccgaagtcgaagtcggactgggaccgctgcacacgcaattcgatgatcagggcttcgctacacgagcctcttgc
gagagcgcgggtggcgaagtggaaagctcggcagctgggaggtggtcgacaagattccgatcacctacaacatcgcccatctcg
cgacggccgagggcgacacggtaatccggacggcaatatctggctgcgctcaacaagctcccgacggcgacatctcg
gcgtcggcccgtcgagcccgagagttcgagctgatcaacatcacggcgagaagatgaagctgctctacgatgccttcacc
gagccggagccgactacgcgcagatgatcaaggccgacaaactgcagccgatcagggtctaccggaaggaggagaacaa
gaaccgcacgcgatctgggacgtcagggacgcgggagacgagcgcagcggcgctaattgtacggtgaagatggtcgccg
tccgcagcacactcgaaccgacttcgatcg

>corco46

cgtgtgcccagacggcaatacatcatcggcagcggcaagctgcagtcgatcaccaccgcttcaacatggagaagatcctga
ccgcgatcgggaacaaggacttcaccggcgatgaggagggcatccccgtcctcgactacgagaaggtgaaggatgccgagg
tcgaagtcgggctcgggcccgtgcacaccagttcgatgaccagggttcgcctacaccagcctcttgcgagagcgcgggtg
gcgaagtggaaagctcggcacctgggaggtggtcgacaagattccgatcacctacaacatcgccacctggcgacagccgag
ggagacacgggtcaatccggacggcaataactggctgcgctcaacaagctctcgacggcgccatctcagtgctggcccgt
cgcagcccagagctcgagctgatcaacatcacggacgagaagatgaaactgctctatgacgccttcaccgagccggagcc
ccattacgcgcagatgatcaaggccgacaaactcaaaccgatcaggtctatccgaaggaggagaacaagaaccgaaacgc
gatctgggatgtcaaggacgcgggagtcgagcgcagcggcgagtgctactgtaaagatggtcgccgtccgcagcacgcctc
gaaccgacctcgttcgaggtcaaccaggcgaccgggtgacgatctat

>corco47

ccgcatggcgtggacgtgagaccggacgggaaatggatagtcggcggcggttaagctggccgcccgttcctgcgtacatagct
ttgaaaagatgcaggaagccattagtaaaggcgattttgacggagaggttatgggaataccattctaaagtacgaaacggtcag
agtcgccgaggttcgataggttgggtcctctgcacaccagtttgacgaccagggatatgcttatacttcttattcctggactca
gccgttgccaagtggaaagctgggtacttgggaagtagtgataagataccggttactataacatcgggacaccttgccgcagtc
atggagacaccgtaaacccccggggccagtagctcctggccatgaataaaatcgccaaggcgagatacctgccggttgacct
tcccatccggagagcgcacagcttatcgacatatccgggaagaagatggaattgctgctcatttccctacgcaggtagagcctc
accacggtcaaataatcgatgctaccttgataaagcccgtgctcgtatataccacttgaagagaataaccacccttacgccattaag
agcgcagaccaggctcgtgtgaaaggaaacggaaatagagtggacgtttacatgctggcaataagaagccatttctcctcat
ccctttagtttcaggaaggggatgaagtttatttccatgtaccaacctggagcaggacgtggatattactcacggattcgggat
caacttctcaacctggatgatgaaatcgaaccggcgaaacaaagacggtgaaatgggtagccgacaaaccggggatatac
cttactactgcacccagttgtgctc

>corco54

cagacggcaatacatcatcggcagcggcaagctgcagtcgatcaccaccgcttcaatatggagaagatcctcaccgccatc
gagaacaaggacttcaccggggatgaagacggcatccccgtcctcgattacgagaaggtgaaggatgccgaagtcgaagtcg
gactgggaccgctgcacacgcaattcgatgatcagggcttcgcctacacgagcctcttgcgagagcgcgggtggcgaagtgg
aagtcggcagctgggaggtggtcgacaagattccgatcacctacaacatcgccatctcgcgacggccgagggcgacacg
gtcaatccggacggcaaatatctggtcgcgctcaacaagctctcgcacggcgacatctcggcgtcggcccgtcgagcccg
agagttcgagctgatcaacatcacggcgagaagatgaagctgctctacgatgccttcaccgagccggagccgcactacgc
gcagatgatcaaggccgacaaactgcagccgatcagggtcaaccggaaggaggagaacaagaaccgcacgcgatctggg
acgtcaaggacgcgggagtcgagcgcagcggcgctaattgtacggtgaagatggtcgccgtccgcagcacactcgaaccga
cttcgatcaggtcaaccaggcgaccgggtgacgatctacatcaccaacatcgagcagacgaccgatgagct

>corco55

ccgcacgggtgtggatgtctgtccggacgggagttcctggtagccggtggcaagctggccaccgtgatcccggtgcacgcct
tctcgaagatgcagaaggccatcgccgaccaggccttcgacggcgagaaggacggcatccccatcctcaagtacgaagccgt
caacttctcgagatccagaaccgggcccctcgccccctgcataccgagttcgacgggcaaggcaacgcctacacgtcgggtct

tcattctccagcgagatcgtgaagtacacgctgccgggctgcgaggtggtggaccgctgccgacctactactcgatcgccac
ctgatgggtcccgggcgacacgcgcaagccgtacggcaagtatgtcgtggccctgaacaagatcacgaaggaccgctac
ctgccaacgggccccgagctcacgcagtcggcgagctctacgacatcacgggcgacgagatgcgctgtctcgactcc
cgaccatcgggcgacccgcaactat

>corco57

acgccgtggccatcgccgccgacaagctgaagccgatggtgcgctacaagttcggcaccgacagccgcaccggcaagccg
catccggacgccaccggccgggcaaggagcgcacgcgagcgtgagggcaccacgtcaagggtgttcggcaccctgatccgc
tcgcacatcacccccgagatcgtcgaggtcgaggagggcgacaccgtcaccttcggttcaccaacaccgagcgcgccgag
gacgagaccacggcttcgccgtgtcgggccacaacctcaacctgtcgtcgagcccgcaagaccgcctcggccaccttgg
tggccagcaagcccggcgtctaccctactactgcacctacttctgctc

>corco6

gatggcaagtacatcgccgtggggcggaagctggacacgcatgcgctcggtgtatagcttcgagaagatccaggcccgatcc
aggcgggcaagttcgaatcgaaggaccctacggcgtgccggtgatcgccatgaagggaagcgtgcacacgcaagtgaac
tgggccttggccgctgcacacgcagtcgactcgaagccgtgcgtggcctacacgtcgatgtacgtggacagccaggtggc
gaagtgggactactgtcagggaaggtgtggacaagatcagcgtgcactacaacatcggtcacctgatgacgatggaaggtg
actctgccgatccgaaggccgttacctggtggcgccgaacaagctggcgatcgaccgttcgtgccgggtgggccccgctgcac
ccgcagaaccaccagctgatcgacatcagcaacgacaagatgcaactgctgtacgacatgccgtcgccgtgggcgagccgc
actacgtggttgccatcgaagcgacgaagctgaagccgggctgcgctacaaggtgggcacggacagccgtacggacaagc
cgaccccggtgccgttcgcgccggcggaagagacgacgacgaagaagggaacaagatcgaggtgaagggcacgctgat
ccgctcgacatcacgcctgagacgatcgaggcggaagtgggcgacgaaatcacggtgcacctgacgaacctggagcgtgc
gcaggacgagaccacggctttacggtgtcgacctacaacgtgcatgcgtcgatcgagcctggcaagacggtgacggtgaag
ttcaaggccgacaaggaaggcgtgtatccctactactgcacgtacttctgctc

>corco7

ccgcacggcgccgacgtcaggccggacggcaagtacatggctcgtggccggcaagcttgatccgcatgtcaccatctactcgtt
cgacaagatgatgcaggcgatcgaaaacgaggcctacaccgccgacgaattcggcattccgatcctcaacttcgacgacgtca
tgctggcgaggtagaactgggattggggccgctgcacacgcaatttgacgacaagggaacgcctacaccagcctcttctta
gacagtgcggttcgacgctggagcctgggtggtgacgaccggccggctgtcagctgctcgagaagctgccggtacagtaca
acatcgcccatattgcgtcgccgaaggcgacacagtcagcccggacggcaactacctggttcgctcaacaagtggctcggtc
gaccgttctctcgcagagaccgctcaagccgcagaatgagcagctgctcgacatcggcagcgacacgatgagcgtgattta
cgacatgccgatgggcattggcggaaccgcattatgcgcagatgatcaggccgacaggatcacggcctgggaatcctacccc
caggtgggttgggatccggaacccaggcgccatctgagtttccaccatgcccgggcaggagcgctcgagcgcgacggc
aatcatgtggaagtcttatgaccgtgtgcgctcgacctgaaccgggaacacgtcaatctcaagaagggtgaccaagtacgt
ggcacattacgaacatcgagacggccccggatgccacgcatggctttgaccttgggtgggcaaacatcagcctgagcatcgaa
ccgggcgaaacgacaaccttcgagtttgcgccgaccagagcgggtacgtttgccgtactactgcaccaattctgctc

>corco8

ccgcacggcgctcgatgtgaggccagacggcaatacatcatcggcagcggcaagctgctgtcgatcaccaccaccttcaatat
ggagaagatcctaccgccatcgagaacaaggacttcaccggggatgaagacggcatccccgtctcgattacgagaaggtg
aaggatgccgaagtcgaagtcggactgggaccgctgcacacgcaattcgatgatcagggttcgctacacgagcctcttctg
gagagcgcgggtggcggaagtggagctggcagctgggaggtggtcgacaagattccgatcacctacaacatcgcccatctcg
cgacggccgagggcgacacggtaatccggacggcaatatctggctgcgctcaacaagctctcgacgggcgacatctcg
gcgtcgccccgtcgagcccagagttcgacgtgatcaacatcacgggcgagaagatgaggctgctctacgatgccttacc
gagccggagccgcactacgcgcagatgatcaaggccgacaaactgcagccgatcgaggtctacccgaaggaggagaacaa
gaacccgcacgcgatctgggacgtcaaggacggggagtcgagcgcagcggcgtaatgtgacgggtgaagatggtcgccgt
ccgcagcacactcgaaccgacttcgatcgaggtcaaccaggcgaccgggtgacgatctacatccaacatcgagcagacg
accgatgagctgcacgggttcgggctcaatgaacacaatatcaatgtggtgatcgatccgggcgagaccaagacgattcagttt
gtcgccgacaagccccggcgtcttccctactactgcacctacttctgctc

>mwwt10

ccgcacgggtgtggatgtcagggccagacggcaagttcatggctgtggccggcaagctcgacacccacgtgtcgggttacagctt
cgagaagatccaggccgccatcaaggccggcaagttcgagagcaaggacccgtacggcatcccgggtatcgcgatgaagga
cgcgctgcacacgcaggtgtcgtcggcctcggcccgtgcacacgcagtagcagcaagccttgcttggcctacacgtcg
ctctacgtcgactcgcaggtcgcgaagtggaaacttctgcgaaggcaaggtcgtcgacaagatcagcgtgcactacaacatcgg
ccacctgatgacgatggaggcgcactccaccaagccccggcggaagtacgtgtggcgctgagcaagctggcgatcgaccg
cttcgtgccggtggggcccgtgcacccgcagaaccaccagctgac

>mwwt100

cggcaagttcatggctgtggccgacaagctcgacacccacgtgtcgggttacagcttcgagaagatccaggccgccatcaagg
ccggcaagttcgagagcaaggacccgtacggcatcccgggtatcgcgatgaaggacgcgtgcacacgcaggtgtcgtcgc
gcctcggcccgtgcacacgcagtagcagcaagccttgcttggcctacacgtcgtctacgtcgactcgcaggtcgcgaag
cggaacttctgcgaaggcaaggtcgtcgacaagatcagcgtgcactacaacatcggccacctgatgacgatggggggcgact
ccaccaagccccggcggaagtacgtgtggcgctgaacaagctggcgatcgaccgttcgtccggtggggcccgtgcacc
cgcagaaccaccagctgacgacatcagcaacgacaagatgcagctgctgtacgacatgccgtgccgtgggtgagccgca
ctatcggtgtcgatcgcgccaacaagctcaagccggccgttcgtacaaggtcgggaccgacagccggaccgatgagcc

>mwwt102

cgtgagaccggacggcaagttcatcctgtgtcgggcaagctcgacacgcacaccatgggtgtactcgttcgacaagattcagga
ggccatcaaggcgaacaagttcgaggggaaggacccgtacggcatcccacatcgggaatgaaggacgcgtccacaagca
ggtcgagctcggcctcggggcgtgcacacgcagtagcacaacagccgtgcacggctacacctcgatgtacgtcgactcg
caggtggcggaagtgggactactgcgagggcaaggtgtcgcgaagctgtcgggtgcactacaacatcggccacctcatgacga
tggaaggcgactcgggtgaagccgaagggtcgtacctgggtcgcgtcaacaagctggccatcgaccgttcaacccggtcgg
cccactgcacccgcagaaccaccagctcatcgacatctgaacgacaagatgcagctgctctacgacatgccgtgccgtcg
gtgagccgcactacgcggtcgccatcagcggccaccctgaagcctggcggttcgtacaaggtcggcaccacagccgca
ccgacgagaagcaccggggcgccgtgcgcgcaggcgaagagaagatcgttcgttcggggcaaggtcgaggtgttcggta
cgctgatccgtcgcacatcaccggagatcatcgaggtcgagaaggcgacgaggtgaccatcaacctcaccaacctcga
gcgcgcgggaagacgagaccacggcttcaccatcagcgcgatgaacgtgcacgcgtcggtcgagcccggaagacgggtc
ggtgaagttcaaggccgagcgcgagggcggtgtcccttactactcacccaattctgctc

>mwwt103

cgatgtgaggccagatggcaagtacatggctgtctccggcaagctcgacacccacgtctcgggttacagcttcgagaagatcca
ggccgccatcaaggccgggaaattcgagagcaaggacccctacggcatcccgggtatcgccatgaaagaggccctgcacgt
gcaggtgccgtcggcctggggcccgtgcacacccagtagcactgaaggcctgcgtcgcctacacctcgtttacgtggact
cgcaggtcgccaagtggaaacttctgcgaaggcaaggtgtcgcgaagatcagcgtgcactacaacatcggccacctgatgacc
atggagggcgactccaccaagccggcggaagtacctcgtcgcgtgaacaagctggccatcgaccgttcgtgccggtgg
ggccgtgcacccgcagaaccaccagctgacgacatctccggcgacaagatgcagctgctctacaacatgccgtgccgtg
ggcgagccgcactacgtcgtcgccatcgatgccaagaagctcaagccgggggtgcgtacaaggtcggcaccgacagccgc
accgacaagccgcatccggcgccgtgcgcgcgcg

>mwwt104

ccgcatggggccgacgtcactcccgatggcaagtacatcctggctcgggcaagctccaggcggtcacgaccgccttcaactt
cgagaagatccagaccgccatcaagaacaaggacttcacggcgacgaggacggcatccccatcctaagtacgagtcgac
aaggacgccgaggtgccgtcggcctcggggcgtgcacacacagttcggaccgaacggcgaagcctacacgtcgtcttc
gtcgacagcgccatcgccaagtggaaagctcggcacgtgggaagtcacgacaagatcccgatgtctactccatcggtacct
caccgcgcggaaggcgacaccgtcaatccgatggcaagtacctgatcggcctcaacaagctctcgacggacggcatctg
aacgtcggctctctcagcccagtcctctcaactcgtcgacaccacggaggagaagatgaagctcctctatgacgcc

>mwwt106

gacgtctgaccggatggcaagttcatggctgtggccggcaagctcgacacccacgtgtcgggttacagcttcgagaagatcca
ggccgccatcaaggccggcaagttcgagagcaaggacccgtacggcatcctgggtatcgcgatgaaggacgcgtgcacac

gcaggtgtcgtcggcctcgccccgtgcacacgcagtagcacagcaagccttgccctggcctacacgtcgtctacgtcgact
cgcaggtcgcgaagtggaaacttctgcaaaggcaaggtcgtcgacaagatcagcgtgcactacaacatcggccacctgatgac
gatggagggcgactccaccaagccccggcggaagtacgtggtggcgtaacaagct

>mwwt11

ccgcatggggccgacgtgtgtccggatggcaagttcatgaccgtttccggcaagctggacacccacgtttcgggtctattccttcg
agaagatccaggctgcgatcaaggccggcaagttcgaatccaaggatccgtacggcattccgggtgatcgggatgaaggatgc
gtcgcataccacaggtccagctcgggtcgtcccttgacacccagtagcattccaagccctgtatcgctacacctcgtgtatg
tcgattctcangtcgtcaaattggaatcactgcgaaggcaggtactcg

>mwwt13

cctcacggggtggatgtgtgaccagacggcaagtagcatcgtcggcggaagctggacacccacgcgtccgtgtatagctt
cgacaagatcatggccgccatccaggccggcaagttctagtccaaagatc

>mwwt14

ccgcatggggccgacgtcactcccgatggcaagtagcatcgtcggcggcaagctccaggggcgtcacgaccgccttcaactt
cgagaagatccagaccgccctcaagaacaaggacttcacggggcagcaggacggcatcccatctcaagtagcagtcgatc
aaggacgccaaggtgccggtcggcctcggggcgctccacacactgttaggaccgaaaggcaaacctacacgtcgttttcgt
cgacagcg

>mwwt17

ccgcacgggtgtggatgtcagaccagacggcaagttcatggtcgtggccggcaagctcgacacccacgtgtcgggtctacagctt
cgagaagatccaggccgccatcaaggccggcaagttcgagagcaaggacccgtacggcatcccgggtgatcgcgatgaagga
cgcgtgcacacgcaggtgtcgtcggcctcggcccgtgcac

>mwwt18

ccccatggcgtggacgtgtgaccagacggcaagttcatggtcgtggccggcaagctcgacacccacgtgtcgggtctacagctt
cgagaagatccaggccgccatcaaggccggcaagttcgagagcaaggacccgtacggcatcccgggtgatcgcgatgaagga
cgcgtgcacacgcaggtgtcgtcggcctcggcccgtgcacacgcagtagcacagcaagccttgcttggcctacacgtcg
ctctacgtcgactcgaggtcgcgaagtggaaacttctgcgaaggcaaggtcgtcgacaagatcagcgtgcactacaacatcgg
ccacctgatgacgatggaggcgactccaccaagccccggcggaagtagcgtggtggcgctgaacaagctggcgatcgaccg
cttcatgccgtggggccgctgcacccgcagaaccaccagctga

>mwwt20

ggatgtgagtcgggacggcaagttcctgaccgtcgccggcaagctcgacacccatgtctccgtttattccttcgagaagatccag
gtcgccatcaaggccggcaagttcgaatccaaggatccgtacggcattccgggtgatcggcatgaaggatgcgtgcatacca
ggtcagactcggcctcggcccgtgcataccagtagcacagcaagccttgatcgcctacacctccctttatgtcgattcga
gtcgtcaaatggaatcactgtgaaggcaaggttctcgacaagatctccgtgcactacaacatcggtcacctgatgacctggaag
gtgactcggcgatccgaagggtcgttatctggtggcactgaacaagctgtcgatcgatcgtttcaatccggttggcccgctgc
ccgcagaaccaccagttgatcgacatcagcaacgacaagatgcaactgtctacgacatccgggtaccgctcggtgagccgca
ctatgtggtcgccatcgaagcggccaagctgaagccggcgctccgttacaaggtcggcaccacagccgtaccgacaaccg
aatccggggcgagtcggggccggtgaggaacataccgacaagaaggcaacaagatcgaagtcttcgggtacgtgatccgtt
cgacatcacccggaaacgatcgaggctgaagtcggcgacgaaatcaccatccattgaccaacctcgaaacgcgccagga
tgaaacccacggtttcgcggtatcgaccttaacgtgcatgcttcggttgagccgggcaagacgggtcaccgtcaagttcaaggc
cgacaaggaaggcgtctatcc

>mwwt21

ccccatggcgtggatgtgagcccgacggcaagttcatcgtcgtcgggcaagctcgacacgcacacccatggtgtactcgttc
gacaagattcaggaggccatcaaggcgaacaagttcgagggcaaggacccgtacggcattcccatcatcggaatgaaggacg
cgtccacaagcaggtcgagctcggcctcggggcgctgcacacgcagtagcacaccaagccgtgcacgggtacacctcgat
gtacgtcgactcgaggtggcgaagtgggactactcgaggggcaaggtgctcgacaagctgtcgggtgcactacaacatcggc
cacctcatgacgatggaaggcgactcgggtgaagccgaagggtcgctacctggtcgcgctcagcaagctggccatcgaccgctt
caaccgggtcggcccactgcacccgcagaaccaccagctcatcgacatctgaacgacaagatgcagctgctctacgacatgc

cgctgccgctcggtgagccgcactacgcggtcgccatcagcgccgccaccctgaagcctggcgttcgctacaaggtcggcac
caacagccgcaccgacgagaagcaccggcgcngtg

>mwwt22

ccgcacggtgtggatgtgagaccagacgggaagtacatcgctgttcgggcaagctcgatccgcatgtgaccgtctacagcttc
gagaagatccagaaagcgatcgctgacgggaattacgagccagatcggttcggtgttcgggtgctcgacttcaagtcctgtgtc
gaagctcaggtggagctcggtcttggggccctccacacgcagttcgacgacaagggttcgcttacaccagcctcttctcgac
agcggatcgcgcggtggaaggtcggcggcgaaggagtcgaagacggctggacgctggtcgacatgattccgatgcaatac
aacaccggccacatctgcgtcgagagggtgacaccgttccccggcgggcattacatgatcggcataaacaagtgggcgct
cgatcgctccaccgggtgggaccgctgctgccccagaacttccatttgatcgacatcagtggcgagcagctggaactgatctac
gacatgccgattggcatcggcgagccgcactacgtccanatgatc

>mwwt23

ccgcacggtgtggatgtgagaccagacgggaagtacatcgctgttcgggcaagctcgatccgcatgtgaccgtctacagcttc
gagaagatccagaaagcgatcgctgacgggaattacgagccagatcggttcggtgttcgggtgctcgacttcaagtcctgtgtc
gaagctcaggtggagctcggtcttggggccctccacacgcagttcgacgacaagggttcgcttacaccagcctcttctcgac
agcggatcgcgcggtggaaggtcggcggcgaaggagtcgaagacggctggacgctggtcgacatgattccgatgcaatac
aacaccggccacatctgcgtcgagagggtgacaccgttccccggcgggcattacatgatcggcataaacaagtgggcgct
cgatcgctccaccgggtgggaccgctgctgccccagaacttccatttgatcgacatcagtggcgagcagctggaactgatcta
cgacatgccgattggcatcggcgagccgcactacgtccagatgatcaagtngagcgctcgagaacatctggggaggtctat
cccggaggt

>mwwt24

ccgcacggtgtggatgtgtggccggacgggaatacttcacgtgggagcggcaaaactgcaatcgatcacgacggcgtaaacgt
cgagaagatgctgaccgcgattcagaagaaagatttcagcggcaacgaagacggcattccggtcttgaactacgacagcgta
aggatgcggaggtcaacgtcggtctggggccgctgcacacgcagttcgacgatcagggttatgctatacaagcctcttctcg
aaagcgcggtcggaagtgggaagctcggcagttgggaggtcgccgacaagattccggtcagctacaacatcgccacctggc
cacggctgaaggcgatacggtaaatccggacggcaaatatctggtagcgtcaacaagctctcgacggcgccatcttgggg
ttggtccgtcgagcccagagctcgagctgatcaacatcagcggcgagaagatgaagctgctctacgacgcttaccgaa
cctgagccgattacgccagatcatcaaagccgacaagctgaaacccttcgaggtctatccgaaggagcagaacaaaaatcc
gcacgcgatctgggacgtcaaggacgccaaggtcgaacgcaacggcgccaatgtgacggccaagctggtcgccgtccgga
gcagcttgaaccgaccaagcttgaggtcaaccagggagacaaggtgacgattcacatccaatatgaacagacgaccgat
gagctgcacggattcgattgaacgaatacaacatcaacgtgtgatcgatccggcgagacaaaacgatcgagttcactgc
ggataaaccggcgctcttcccttactactgcacctaattctgctc

>mwwt25

ccgcacggggtggatgtgtggccggatggcaagaacatcatcggtggccggcaagctcgacaccacgtgtcggtctacagttt
cgagaagatccaggcggcgatcaaggccggcaagttcgagttcaaggaccggtatggcatcccgggtatcgcgctgaaggat
gccctgcacaccaggtgtcgctgggcttggggccctgcacacgcagttcgacagcaagtcctgcaccgcgtacacctcg
tgtacgtggactcgaggtggccaagtggaaactattgcaaggcaaggtgctcgacaagatcagcgtgactacaacatcggt
cacctgatgacgatggaggggtgactccaccaagccggccggaagtacctggtggcgctgaacaagctggccatcgaccgtt
cgtgccggtgggcccgtgcacccgcagaaccaccagctgatcgacatctcngcgacaagatgcagttgctgtacgacatg
ccgctgccgctgggtgagccgcactatcggtggccatcgcgccaacaagctcaagccggcagtgcgctacaaggtcggc
ac

>mwwt28

ccgcacggggtggatgtgtggccggatggcaagaacatcatcggtggccggcaagctcgacaccacgtgtcggtctacagttt
cgagaagatccaggcggcgatcaaggccggcaagttcgagttcaaggaccggtatggcatcccgggtatcgcgctgaaggat
gccctgcacaccaggtgtcgctgggcttggggccctgcacacgcagttcgacagcaagtcctgcaccgcgtacacctcg
tgtacgtggactcgaggtggccaagtggaaactattgcaaggcaaggtgctcgacaagatcagcgtgactacaacatcggt
cacctgatgacgatggaggggtgactccaccaagccggccggaagtacctggtggcgctgaacaagctggccatcgaccgtt
cactgatgacgatggaggggtgactccaccaagccggccggaagtacctggtggcgctgaacaagctggccatcgaccgtt

cggtgccggtgggcccgtgcacccgcagaaccaccagctgatcgacatctccggcgacaagatgcagttgctgtacgacatg
ccgctgccgctgggtgagccgactatgcggtggccatcgcgccaacaagctcaagccggcagtgcgctacaaggctcgca
ccgacgccgcaccgataanaaacaccccg

>mwwt29

ccccatggtgtggatgtgagccggatggcaagtacatggctgctcggcaagctcgacacccacgtctcggtctacagcttc
gagaagatccaggccgccatcaaggccgggaaattcgagagcaaggacccctacggcatcccggatcgccatgaaggag
gccctgcacgtgcaggtgccgctcggcctgggcccgtgcacacccagtagactcgaaggcctgcgtcgctacacctgc
tttacgtggactcgcggtcgccaagtggaaacttctgcgaaggcaaggtgctcgacaagatcagcgtgcactacaacatcggtc
acctgatgacctggaggggcactccaccaagccggccggcaagtacctcgtcgctgaacaagctggccatcgaccgctt
cgtgccggtgggcccgtgcacccgcagaaccaccagctgatcgacatcagcaacgacaagatgcagttgctctacgacatgc
cgctgccgctgggagccgcactacgtcgtcgccatcgatgccaagaagctcaagccgggggtgcgctacaaggctcgga
ccgacagccgcaccgacaagccgcacccggcgccgtgcgcgcccggcgaagagaagaccgagagggaagggaacaagg
tcacggtctggggcacgtgatccgctgcacatcacgcccagaccatcgaggcggaagtgggacgaggtcaccatcc
acctgaccaacctgagcgcgcccaggacgagaccacggcttcacggtgctgacctacaacctgcacgttcggtcgagcc
gggcaagacggtgacggtgaagttcaaggccgacaaggaaggcgtctaccctactactgcacgtattctgctc

>mwwt31

ccccatggtgcggatgtgagccggatggcaagttcatgaccgttccggcaagctggacacccacgttccggtctattccttcg
agaagatccaggctgcgatcaaggccggcaagttcgaatccaaggatccgtacggcattccggatcgccgatgaaggatgc
gctgcataccaggtccagctcggtctcggtccctgcacacccagtagattccaagccctgtatcgctacacctcgctgtatg
tcgattctcaggtcgtcaaatggaatcactgcgaaggcaaggtactcgacaagatctcggtgcattacaacatcggtcacctgat
gacctggaaggcgattcggtgatccgaagggtcgctacctgggtggcgtgaacaagctggcgatcgaccgcttctgcccgg
ttggcccgtgcacccgaaaatcaccagctgatcgacatttcaacgacaagatgcaattgtgtacgacatcccgctgccgttg
ggtgagccgcactacgctgtggccatcgaaagccagcaagctgaagccggcgctgcgctacaaggctggcaccgattccgc
acggacagccgcaccccgcatggttcgcggcggaagaacgcaccgagaagaagggaacaagatcacgggtctacg
gtacgttgatccgttccacatcaccccgaaacgatcgaggctgatgtcggcgacgaagtcacctccacctgaccaacctcg
aacgtgcccaggacgaaacctatggttcacggtgtccacgatgaacgtccacgcttcggtgaaccgggcaagacagtgacc
gtcaagttcaaggctgacaaggaaggcgtctatccgtactactgcacccagttctgctc

>mwwt32

cctcacggtgtggacgtgagacccgacggcaagtacatggctgctcggcaagctcgacacccacgtctcggtctacagcttc
gagaagatccaggccgccatcaaggccgggaaattcgagagcaaggacccctacggcatcccggatcgccatgaaggag
gccctgcacgtgcaggtgccgctcggcctgggcccgtgcacacccagtagactcgaaggcctgcgtcgctacacctgc
tttacgtggactcgaggtgccaaagtggaaacttctgcgaaggcaaggtgctcgacaagatcagcgtgcactacaacatcggcc
acctgatgacctggaggggcactccaccaagccggccggcaagtacctcgtcgctgaacaagctggccatcgaccgctt
cgtgccggtgggcccgtgcacccgcagaaccaccagctgatcgacatcagcaacgacaagatgcagttgctctacg

>mwwt33

cgatgtgtgccggatggcaagtacatcatcggtcgggcaagctccaggcgctcacgaccgccttcaacttcgagaagatcca
gaccgccatcaagaacaaggacttcacggcgacgaggacggcatccccatcctcaagtacgagtcgatcaaggacgccga
ggtgccggtcgccctcgggcccgtgcacacacagttcggaccgaacggcgaagcctacacgtcgctcttcgtcgacagcgcc
atcgccaagtggaaagctcggcacgtgggaagtcacgacaagatcccgatgtcctactccatcggtcacctcaccgcccgga
aggcgacaccgtcaatccggatggcaagtacctgatcgccctcaacaagctctcgacggacggcatctgaacgtcggtcctt
ctcagcccagtccttcaactcgtcgacatcacggaggagaagatgaagtcctctatgacgcttcaccgagcccagccg
cactacgccagatgatccaccgtgacaaggtcaagccgatcgaggtctacccgaaggaaanagaacaagcaccgctcg
atctgggacgtcaaggacggcggtgacctgcaagggggaacgangttctcgccaaggtcgtcc

>mwwt34

cctcacggggtcgacgtcagaccggatggcaagttcatgaccgttccggcaagctggacacccacgttccggtctattccttcg
agaagatccaggctgcgatcaaggccggcaagttcgaatccaaggatccgtacggcattccggatcgccgatgaaggatgc

gctgcataccaggtccagctcggctcggctcccttgacacccagtagcattccaagccctgtatcgctacacctcgtgtatg
tcgattctcaggtcgtcaaatggaatcactgcgaaggcaaggtagctcgacaagatctcgggtgcattacaacatcggtcacctgat
gaccatggaaggcgattcggctgatccgaagggtcgtacctgggtggcgtgaacaagctggcgatcgaccgctttgtgccgg
ttggcccgtgcacccgaaaatcaccagctgatcgacatttctaacgacaagatgcaattgtgtacgacatgccgctgccgttg
ggtgagccgcactacgctgtggccatcgaagccagcaagctgaagccgggcgtgcgctacaaggtcggtagcattcccga
cggacaagccgcaccccgcatggctcgcgccggcgaagaacgcaccgagaagaagggaacaagatcacggcttacggt
acgttgatccgttcccatcaccgccgaaacgatcgaggctgatgtcggcgacgaagtcacatccacctgaccaacctcgaa
cgtgccaggacgaaacccatggttcacgggtgccacgatgaacgtccacgcttcgggtgaaccgggcaagacagtgaccgt
caagttcaaggctgacaaggaaggcgtctatccatactactgcacccatttctgctc

>mwwt35

ccccatggcgtggatgtgagcccgacggcaagttcatcactgtgtcgggcaagctcgacacgcacacatgggtgactcgttc
gacaagattcaggaggccatcaaggcgaacaagttcgagggaaggaccgtacggcattcccatcctggaatgaaggacg
cgctccacaagcaggtcgagctcggcctcggccgctgcacacgcagtacgacaccaagccgtgcacgtgacacctcgat
gtacgtcgactcgaggtggcgaagtgggactactcgagggaaggtgctcgacaagctgtcgggtgactacaacatcggc
cacctcatgacgatggaaggcgactcgggtgaagccgaagggtcgctacctggctcgcgctcagcaagctggccatcgaccgctt
caaccggctcggccactgcaccgcgagaaccaccagctcatcgacatctgaacgacaagatgcagctgctctacgacatgc
cgctgccgctcggtagccgcactacgcggctcgccatcagcgccgccacctgaagcctggcgttcgctacaaggtcggcac
caacagccgcaccgacaagaagcaccnngggcgccgtgcgcgcaggc

>mwwt37

ccccatggcgtggatgtgacaccggacggcaagttcatcaccgtttccggcaagctcgacacgcacgtttcgggtctattcgttcg
acaagatccaggccgcaatcaaggccggcaagttcgagtcgaaggacgcctacggcatcccggtgatcgcatgaaggacg
cgttgacgcgcaggtccagctcggcctcggcccgtgcatacgcagtatgacgccaagccttgcgttgctacacctcgtct
acgtcgattcgcaggttgccaagtggaaacttctgcgaaggcaaggtgctcgacaagatcagcgtgactacaacatcggtcacc
tggtcacgatggaaggcgactcggctgatccgaaaggcggttatctgatcgcgctgaacaagctggcgatcgaccgcttcgttc
cggttggcccgtgcaccgcgagaaccaccagttgatcgacatcagcaacgacaagatgcagctcctgtacgacatgccgctg
ccgctcggcgagccgcactacgtcgtcgccatcgaggccggcaagctgaagccgggcgtgcgctacaaggtcggcaccga
cagccggaccgacaagccgcaccgggagccgtgcgcgccggcgcaananaan

>mwwt42

cctcacggggcgagcgtgacccagatggcaagttcatcaccgtttccggcaagctcgacacgcacgtttcgggtctattcgttcg
acaagatccaggccgcaatcaaggccggcaagttcgagtcgaaggacgcctacggcatcccggtgatcgcatgaaggacg
cgttgacgcgcaggtccagctcggcctcggcccgtgcatacgcagtatgacgccaagccttgcgttgctacacctcgtct
acgtcgattcgcaggttgccaagtggaaacttctgcgaaggcaaggtgctcgacaagatcagcgtgactacaacatcggtcacc
tggtcacgatggaaggcgactcggctgatccgaaaggcggttatctgatcgcgctgaacaagctggcgatcgaccgcttcgttc
cggttggcccgtgcaccgcgagaaccaccagttgatcgacatcagcaacgacaagatgcagctcctgtacgacatgccgctg
ccgctcggcgagccgcactacgtcgtcgccatcgaggccggcaagctgaagccgggcgtgcgctacaaggtcggcaccga
cagccggaccgacaagccgcaccgggagccgtgcgcgccggcgcaananaan

>mwwt43

gtgtgaccggacggcaagtagcatcgtcgttccggcaagctcgattcgcacaccttcgtctacagcttcgagaagatgcaggcc
ctgatccaggcgggcaattcgcgggtaaggatccgtacggcattcccatcctggatctgcagaagtcgctgcacacgcagac
cgagctggggctgggccccctgcacacgcagtacgatcgacctcctgcgttgctacacctcctctactgtgactcgatggt
ggccaagtggaaactattcgagggaagctgctggatcggtcctcactacaacatcgggcacctgatgacgatgcacg
gcgacaccgtggctccggaaggccattatctgggtggcgctcgacaagctggcgatcgaccgcttcagcccgtgggcccgt
gcatccgcagaaccaccagctcatcgacatcagcgccgagaagatggagctgctctatgacatgccgctcggcgtcggcgag
ccgactacgcggtggcgatcaacgccagcctgtcaagccgcgcaaccagtacaagctgggcaccaacagcaggacggg
ggaggactccgcttcaagaccaggggccggccgcgagcacgtcgacaagacgcggggcaaggtggaggtgttcggcacca
cgatccgatcgacatcccccgagatcatcgaggtggaagaaggcgatcaggtgatccacctgacgaaccaggagcg

ggccgaggatcagacgcatggattcacggaggacacctacaacgtgaacgcctcgatcgagccgggcaagacgatcacggt
gagcttcgtggcgagcaaggcggcggttatccgtactactgcaccaattctgctc

>mwwt44

ccgcatgggggtggatgtctcgccggacggcaagtacatcgctcgccgggcaagctcgacagccacgcattgggtctacagctt
cgagaagatccaggcggcgatcgccgcccggccgcttcgagtcgaaggatccgtacgggatcccggtgatcgcgctcaaggga
cgcatccacgtccaggtgccgctcgccctcgaccgctgcacacgcagtcgactcgccgcccgtgcatcgcttacacgtcg
ctgtacgtcgattcgaggtcggaagtggaaactactcgagggcaaggtcgctgacaagctgctgggtccactacaacatcggt
cacctgatgacgatggaggggcgattcgacgaaaccggcgccgcaagtacgtgctcgctcaacaagctcgcgatcgaccgct
tcgtgccggtggggccgctgcatccgcagaaccaccagctgatcgacatcagcaacgacaagatgcagcttctctacgacatg
ccgctgccgctcggtgagccgcactacgccgtctcgatcgccgctcgaaagctcaagcccgggggtcgctaccggggccgg
caccgacagccgcaccgacaagccgcacccgggagcgggtgaaggccggacaggagaagatcgtaagaaggggcaacaa
ggtcgaggtgttcgggacgctgatccgctcgacatcacgccggagaccatcgaggtcgacgtcggtgacgaggtcacgac
cacctgtcgaacctggagagcgcaggacgagacgcaggcttcaccgtgctgacgtacaacgtgcacgctcgatcgag
ccgggcaagacggtcagcgctgtcttaaggccgacaaggaaggcgcttatccctactactgcacgtaattctgct

>mwwt47

ccgcacgggtggatgtgtgtccagacggcaagttcatgaccgtttccggcaagctggacaccacgtttcggtctattccttcga
gaagatccaggctcgatcaaggccggcaagttcgaatccaaggatccgtacggcattccggtgatcgggatgaaggatcg
ctgcataccagggtccagctcggtctcggtcccttgacacccagtcgattccaagccctgtatcgctacacctcgctgtatgt
cgattctcaggtcgtaaatggaatcactgcgaaggcaaggtactcgacaagatctcggtgcattacaacatcggtcacctgatg
accatggaaggcgattcggtgatccgaagggtcgctacgttggtggcgctgaacaagctggcgatcgaccgctttgtgccggtt
ggcccgtgcatccgcaaaatcaccagctgatcgacatttctaacgacaagatgcaattgtgtacgacatgccgctgccgttgg
ggtgagccgcactacgctgtggccatcgaagccagcaa

>mwwt48

ccccatgggtcgatgtgtgtccggatggcaagttcatgaccgtttccggcaagctggacaccacgtttcggtctattccttcgag
aagatccaggctcgatcaaggccggcaagttcgaatccaaggatccgtacggcattccggtgatcgggatgaaggatcgct
gcataccagggtccagctcggtctcggtcccttgacacccagtcgattccaagccctgtatcgctacacctcgctgtatgtcg
attctcaggtgtcaaatggaatcactgcaaaggcaaggtactcgacaagatctcggtgcattacaacatcggtcacctgatgacc
atggaaggcgattcggtgatccgaagggtcgctacgttggtggcgctgaacaagctggcgatcgaccgctttgtgccggttgg
cccgtgcatccgcaaaatcaccagctgatcgacatttctaacgacaagatgcaattgtgtacgacatgccgctgccgttgggtg
agccgcactacg

>mwwt52

gattcggtcgatcccaaggccggttatctgggtggccctgaacaagctggcgatcgatcgctttgtcccgggtgggtcccttcgacc
cgcagaaccatcagctgatcgatcagcgaagacaaaatgcagcttctttacgacatgccgctgcctctgggtgagccgcact
acgtggttgccatcgaggccggcaagctgaagccggcggttcgtacaaggtcgggaccgacagccggaccgacaagccg
caccgggtgccgtacgggcccggcgaggaaaagacgggtcaagaagggaacaagatcgaggtcttcggcacctgatccgt
tcgcacatcacgccgaaacgatcgaggtgaggtcggcgacgaagtacgatccacctcaccaacctgaacgcgcgacg
gacgagacgcagggcttcacgatttcgacctacaacgtcatgcgtcggtcgagccgggcaagacggtcagcctcaagttcaa
ggcggaacaaggagggtgtctatccttactactgcacctacttctgctc

>mwwt53

ccgcacggcggtggatgtgaggccagacggcaagttcatgaccgtttccggcaagctggacaccacgtttcggtctattccttcg
agaagatccaggctcgatcaaggccggcaagttcgaatccaaggatccgtacggcattccggtgatcgggatgaaggatgc
gctgcataccagggtccagctcggtctcggtcccttgacacccagtcgattccaagccccgtatcgctacacctcgctgtat
gtcgattctcaggtcgtaaatggaatcactgcaaaggcaaggtactcgacaagatctcggtgcattacaacatcggtcacctgat
gaccatggaaggcgattcggtgatccaaagggtcgctacgttggtggcgctgaacaagctggcgatcgaccgctttgtgccgg
ttggcccgtgcatccgcaaaatcaccagctgatcgacatttctaacgacaagatgcaattgtgtacgacatgccgctgccgtt
gg

>mwwt55

cctcacggtgtcgacgtgtgtccggatggcaagttcatcatcgtcgccggcaagctcgacacgcacgtttccgtgtattccttga
gaagatccaggccgcatcaaggccggcaagttcgaatccaaggatccctacggcatcccgggtgatcgccctcaaggacgcg
gtgcagacgcaggtttcgtgggctggggccgctgcatacccagtagcagacgaagccctgacctgacctacacctgctcta
cgtcgactcgaggtcgccaagtggaaactactgcaaaggcaaggtggtggacaagatttcggtccattacaacatcgggccacct
gatgaccatggaatgcgattcgaccaagccggccggcaagtagctcgtcgctgaacaagctgtcgatcgaccgttcgtcc
cggtcggcccgctgcagccgcagagccaccagctgatcgacatcagcaacgacaagatgcgtctgctctacgacatgccgttg
ccgtggggcgagccgcattacgccgtctcgatcg

>mwwt57

ccccatggggtgatgtgtggcccgcagggcaagttcatggtcgtggccggcaagctcgacacccacgtgtcggtctacagcttc
gagaagatccaggccgcatcaaggccggcaagttcgagagcaaggacccgtacggcatcccgggtgatcgcatgaaggac
gcgtcgacacgcaggtgtcgtcggcctcgcccgctgcacacgcagtagcagacgaagcctgacctggcctacacgtcgc
tctacgtcgactcgaggtcgcgaagtggaaacttctgcgaaggcaaggtcgtcgacaagatcagcgtgactacaacatcggc
cacctgatgacgatggagggcgactccaccaagccggcggaagtacctggtggcgctgaacaagctggcgatcgaccgc
ttcgtgccgggtggggccgctgcacccgcagaaccaccagctgatc

>mwwt60

gtggatgtgagggccgcagggcaagaacatcatcgtggccggcaagctcgacacccacgtgtcggtctgcagtttcgagaagat
ccaggccggcgatcaaggccggcaggttcgagtcgaaggacccgtacggcatcccgggtgatcgcgctgaaggatgccctgca
caccaggtgtcgtgggctggggccctgcacacgcagtagcagacgaagtcctgcaccgcgtacacctcgctgtacgtg
gactcgaggtggccaagtggaaactactgcaaaggcaaggtgctcgacaagatcagcgtgactacaacatcggtcacctgat
gacgatggagggtagtccaccaagccggcggaagtacctggtggcgctgaacaag

>mwwt66

ggatgtgtgcccggatggcaagtacatcatcggtcgggcaagctccaggcgtagcaccgccttcaacttcgagaagatcc
agaccgccatcaagaacaaggacttcagggcgacaaggacggcatccccatcctcaagtacgagtcgatcaaggacgccg
aggtgccggctcgccctcgggccgctgcacacacagttcggaccgaacggcgaagcctacacgtcgtcttctcgacagcgc
catcgccaagtggaaagctcgccacgtgggaagtcatcgacaagatcccgatgctctactccatcggtcacctcaccgccggg
aaggcgacaccgtcaatccggatggcaagtacctgatcgccctcaacaagctctcgacggacggcatctgaacgtcggtcct
tctcagcccagtccttcaactcgtcgacatcacggaggagaagatgaa

>mwwt67

agaccagacggcaagtacatggctcgtcggcaagctcgacacccacgtcgtcggtctacagcttcgagaagatccaggccgc
catcaaggccgggaaattcgagagcaaggacccctacggcatcccgggtgatcgccatgaaggaggccctgcacgtgcaggt
gccgctcgccctggggccgctgcacacccagtagcactcgaaggcctgcgtcgctacacctcgctttacgtggactcgaggt
tcgccaagtggaaacttctgcaaaggcaaggtgctcgacaagatcagcgtgactacaacatcggtcacctgatgacctggag
ggcgactccaccaagccggcggaagtacctcgtcgctgaacaagctggccatcgaccgttcgtgcccgtggggccg
ctgatccgcagaaccaccagctgatcgacatcagcaacgacaagatgcagttgctctacaacatgccgctgccgctggggcg
agccgcactacgtctcnc

>mwwt68

ccgcatggggtggacgtcacgccggacggcaagttcatgaccgtttccggcaagctggacacccacgtttcggtctattccttcg
agaagatccaggctcgatcaaggccggcaagttcgaatccaaggatccgtacggcattccgggtgatcgggatgaaggatgc
gctgcatacccaggtccagctcggtcgtcgtccttgcacacccagtagcattccaagccctgatcgctacacctgctgtatg
tcgattctcaggtcgtcaaatggaactactgcaaaggcaaggtactcgacaagatcgtggtgcattacaacatcggtcacctgatg
accatggaaggcgattcggtgatccgaagggtcgtacctggtggcgctgaacaagctggcgatcgaccgctttgtgccggtt
ggcccgtgcacccgcaaaatcaccagctgatcgacatttcaacgacaagatgcaattgtgtacgacatgccgctgccgttg
gtgagccgcactacgtgtggccatcgaaagccagcaagctgaagccggggcggtcgctacaaggtcggtaccgattcccga
cggacaagccgcacccggcatggttcgcccggcggaagaacgcaccgagaagaagggaacaagatcacggtctacggt
acgttgatccgttcccacatccccggaaacgatcgaggctgatgtcgccgacgaagtcacatccacctgaccaacctcgaa

cggtcccaggacgaaacccatgggttcacgggtgccacgatgaacgtccacgcttcgggtgaaccgggcaagacagtaccgt
caagttcaaggctgacaaggaaggcgctctatccctactactgcacctatttctgct

>mwwt71

tgccggttgccccgctgcatcccgcaaaatcaccagctgatcgacatttctaacgacaagatgcaattgtgtacgacatgccg
ctgccgttggtgagccgcantacgctgtggccatcgaagccagcaagctgaagccggcgctgcgctacaaggtcgggtaccg
attcccgacggacaagtcgacccccggcatgggttcgcgcggcggaagaacgcaccgagaagaagggaacaagatcacg
gtctacgggtacgttgatccgttccacatcaccgggaacgatcgaggctgatgtcagcgacgaagtcaccatccacctgacc
aacctcgaacgtgccaggacgaaacccatgggttcacgggtgccacgatgaacgtccacgcttcgggtgaaccgggcaagac
agtaccgtcaagttcaaggctgacaaggaaggcgctctatccctactactgcaccacttctgctc

>mwwt72

cctcatgggtgtggatgtgacaccggatggcaagttcatcatcgttgccggcaagctcgacacgcacgtctcgggtgtacagcttcg
agaagattcaggccgcgatcaaggccggcaagttcagtcgaaggacgcctacggcatcccggtgctgagcctgaaggattc
gtcgcatacgcaggtgtcgttgggctcggccccctgcacaccagtagacagcaagccctgcctggcctacacctcgtgt
atgtggattcgcaggtcgcaagtggaaactactgcaaggcaaggtggtggacaagatttcgggtccactacaacatcgccacc
tgatgacctggaaggcgattcgaccaagccccggcgcaagtagcttggtggcctgaacaagctctccatcgatcgtttcgtgc
cggtcggccccgctgcaaccgcagaccagctgatcgacatcagcaacgacaagatgcagctgctgtatgacatgccgct
gccgctggggcgagccgcactacgccgtgtcgattgcccaacaagctgaagcccccggtgcgctacaaggtcggcaccga
cagccgcaccgacaagccgcatccggcgcggtgctgcccggcgaggagaagacggtcaagaagggaacaaggtcgag
gtttcggccacctgatccgtcgcacatcacgccggaaccatcgaggtcgatgtcggcgacgaggtgacgatccacctgac
caacctcgagcgcgcgaggacgagaccacgggttcaccgtctcgacctacaacgcgcatgcctcgatcgagccgggcaa
gaccgtcagcgtcaagttcaaggccgacaaggaaggtgtctatccctactactgcacctacttctgctc

>mwwt73

agaccagacggcaagtacatggctgtctccggcaagctcgacaccacgtctcgggtctacagcttcgagaagatccaggccgc
catcaaggccgggaaattcgagagcaaggaccctacggcatcccggtgatcggcatgaaggaggccctgcacgtgcaggt
gccgctcggcctggggccgctgcacaccagtagactcgaaggcctgcgtcgcctacacctcgtttacgtggactcgcagg
tcgcaagtggaaacttctgcgaaggcaaggtgtcgcacaagatcagcgtgcactacaacatcggtcacctgatgacctggag
ggcgactccaccaagccggcggaagtagctcgcgtgaacaagctggccatcgaccgcttcgtgccggtggggccg
ctgcatccgcagaaccaccagctgatcgacatcagcaacgacaagatgcagttgctctacgacatgccgctgccgctggggca
ggcgactacgtcgtcgccatcgatgccaagaagctcaagccgggggtgcgctacaaggtcggcaccgacagccgcaccga
caagccgatccggcgccgtgcgcggcggaagagaagaccgagaggaagggaacaaggtcacgggtcggggcac
gtgatccgctcgcacatcacgcccagaccatcgaggcggaagtggcgacgaggtcacctccacctgaccaacctcga
gcgcgcccaggacgagaccacgggttcacgggtcgcacctacaacctgcacgcttcgggtcgagccgggcaagacgggtgac
ggtgaagttcaaggccgacaaggaaggcgcttaccatactactgcacgtatttctgctc

>mwwt75

ccgcacggcggtggatgtgagcccagacggcaagttcatgaccgtttccggcaagctggacaccacgtttcgggtctattccttcg
agaagatccaggccgcgatcaaggccggcaagttcgaatccaaggatccgtacggcattccgggtgatcgggatgaaggatgc
gtcgatacccagggtccagctcgggtcgtctcttgacaccagtagattccaagccctgatcgcctacacctcgtgtatgt
cgatttctcaggtcgtcaaatggaatcactgcaaaaggcaaggtactcgacaagatctcgggtgcattacaacatcggtcacctgatg
accatggaaggcgattcgggtgatccaaagggtcgtacctggtggcgctgaacaagctggcgatcgaccgctttgtgccggtt
ggccactgcatccgcaaaatcaccagctgatcgacatttctaacgacaagatgcaattgtgtacgacatgccgctgccgttgg
ggtgagccgcact

>mwwt76

cctcatggcgctgatgtgaggccagatggcaagttcatgaccgtttccggcaagctggacaccacgtttcgggtctattccttcga
gaagatccagggtcgcgatcaaggccggcaagttcgaatccaaggatccgtacggcattccgggtgatcgggatgaaggatgcg
ctgcatacccagggtccagctcgggtcgtcttgacaccagtagattccaagccctgatcgcctacacctcgtgtatgt
cgatttctcaggtcgtcaaatggaatcactgcgaaggcaaggtactcgacaagatctcgggtgcattacaacatcggtcacctgatg

accatggaaggcgattcggctgatccgaagggtcgctacctgggtggcgctgaacaagctggcgatcgaccgctttgtgccggtt
ggccccgtgcatccgcaaaatcaccagctgatcgacatttctaacgacaagatgcaattgtgtacgacatgccgctgccgttgg
gtgagccgcactacgccgtggccatcgaagcngcaagctganccgggctgcgctacaaggctcggtaccgattcccc
>mwwt77

ggatgtgtgccccggatggcaagtacatcatcggtcgggcaagctccagggtcgacgaccgccttcaacttcgagaagatcc
agaccgccatcaagaacaaggacttcacgggacgagggacggcatccccatcctcaagtacgagtcgatcaaggacgccg
aggtgccggctggcctcgggcccgtgcacacacagttcggaccgaacggcaaagcctacagtcgctcttcgacagcgc
catcgccaagtggaaagctggcacgtgggaagtcacgacaagatcccgatgtcctactccatcggtcacctcaccgccgcgg
aaggcgacaccgtcaatccggatggcaagtacatcggtcctcaacaagctctcgacggacggcatctgaacgtcggtcct
tctcagcccagtcctctcaactcgtcgacatcacgggaggagaagatgaagctcctctatgacgccttcaccgagcccagcc
gcactacgcccagatgatccaccgtgacaaggtcaagcccgatcgaggtctacccgaagga
>mwwt79

agaccagacggcaagtacatggctcgtctccggcaagctcgacaccacgtctcggtctacagcttcgagaagatccaggccgc
catcaaggccgggaaattcgagagcaaggaccctacggcatcccggtgatcggtatgaaggaggccctgcacgtgcaggt
gccgctcggcctgggcccgtgcacaccagtcgactcgaaggcctgcgtcgctacacctcgctttacgtggactcgcagg
tcgccaagtggaaattctgcaaaggcaaggtgctcgacaagatcagcgtgcactacaacatcggtcacctgatgaccatggag
ggcgactccaccaagccggccggcaagtacctcgtcgcgctgaacaagctggccatcgaccgcttcgtgccggtgggccc
ctgcatccgcagaaccaccagctgatcgacatcagcaacgacaagatgcagttgctctacgacatgccgctgccgctgggcga
gccgcactacgtcgtcgccatcgatgccaagaagctcaagccgggggtgcgcta
>mwwt80

agaccagacggcaagtacatggctcgtctccggcaagctcgacaccacgtctcggtctacagcttcgagaagatccaggccgc
catcaaggccgggaaattcgagagcaaggaccctacggcatcccggtgatcggtatgaaggaggccctgcacgtgcaggt
gccgctcggcctgggcccgtgcacaccagtcgactcgaaggcctgcgtcgctacacctcgctttacgtggactcgcagg
tcgccaagtggaaattctgcaaaggcaaggtgctcgacaagatcagcgtgcactacaacatcggtcacctgatgaccatggag
ggcgactccaccaagccggccggcaagtacctcgtcgcgctgaacaagctggccatcgaccgcttcgtgccggtgggccc
ctgcatccgcagaaccaccagctgatcgacatcagcaacgacaagatgcagttgctctacgacatgccgctgccgctgggcga
gccgcactacgtcgtcgccatcgatgccaagaagctcaagccgggggtgcgctacaaggtggcaccgacagccgcacc
gacaagccgatccggcgccgtgcgcgccggc
>mwwt81

gtggatgtgagggccgacggcaagaacatcatcgtggccggcaagctcgacaccacgtgtcggtctgcagtttcgagaagat
ccaggccggcatcaaggccggcaggttcgagtcgaaggaccgtacggcatcccggtgatcgcgctgaaggatgccctgca
caccaggtgtcgtgggcttggggccccctgcacacgcagtcgacagcaagtcctgcaccgcgtacacctcgctgtacgtg
gactcgcaggtggcaagtggaaactactgcaaaggcaaggtgctcgacaagatcagcgtgcactacaacatcggtcacctgat
gacgatggagggtgactccaccaagccggccggcaagtacctgggtggcgctgaacaagctggccatcgaccgcttcgtgcc
gtggggcccgtgcacccgcagaaccaccagctgatcgacatctccggcgacaagatgcagttgctgtacaacatgccgctgcc
gctgggtgagccgcactatgcgggtgtcaatcgcgccaacaagctcaagccggccngtgcgctacaagg
>mwwt82

ccgcacgggtgtggatgtgacggcgacggcaagttcatcatcgtcgccggcaagctcgacacgcacgtttccgtgtattccttcg
agaagatccaggccgccatcaaggccggcaagttcgaatccaaggatccctacggcatcccggtgatcgccctcaaggacgc
ggtgcagacgcaggtttcgtgggcttggggcccgtgcataccagtcgacagcaagccctgcctggcctacacctcgtct
acgtcgactcgcaggtcgccaagtggaaactactgcaaaggcaaggtgggtggacaagattcgggtccattacaacatcgccac
ctgatgaccatggaaggcgattcgaccaagccggccggcaagtacctcgtcgcgctgaacaagctgtcgatcgaccgcttcgt
cccggctggcccgtgcagccgcagaccaccagctgatcgacatcagcaacgacaagatgcactgtctctacgacatgccg
ttgccgtgggagccgcattacccgtctcgatcgccccaacaagctgaagcccggctgcgtacaaggtggcaccg
acagccgcaccgacaagccgatccggcgcggtgcgtgccggcgaggagaagaccgtcaagaagggaacaaggtcga
agtcttcgccacgtgatccgttcgcacatcacgcctgaaccatcgaggtcgatgtcggcgacgaagtgacgatccacctgac

caatctcgagcgcgccaggacgagaccacggcttcacggtgtcgacctacaacacgcacgcttcggtcgagccgggcaa
gaccgtcagcgtcaagttcaaggccgacaaggaaggtgtctatccatactactgcaccaattctgctc

>mwwt83

cctcatggtgaggatgtcacaccagacggcaagttcatggtcgtggccggcaagctcgacacccacgtgtcagtctacagcttc
gagaagatccaggccgcatcaaggccggcaagctcgagagcaaggacccgtacggcatcccgggtatcgcgatgaagga
cgcgctgcacacgcaggtgtcgtcggcctcggcccgctgcacacgcagtagacagcaagccttgctggcctacacgtcg
ctctacgtcgactcgcaggtcgcgaagtggaaacttctgcaaaggcaaggtcgtcgacaagatcagcgtgcactacaacatcgg
ccacctgatgacgatggaggggcggtccaccaagccccggcggaagtacgtggtggcgtgaacaagctggcgatcgaccg
cttcgtgccggtggggcgctgcacccgcagaaccaccagctgatcgacatcagcaacgacaagatgcagctgctgtacgac
atgccgtgccgctgggtgagccgactatgcggtgtcgatcgcgccaacaagctcaagccggcggttcgtacaaggtcg
ggaccgacagccggaccgatgagccgcacccgttcgccaccaaggcgggcgaggagaagatcgtcaagaaggggaacaa
ggtcgacgtctatgcgacgtgatccgctgcacatcacgcccagacgatcgaggtcgacgtgggtgacgacgtggcgatc
cacctgaccaacctcgagcgcgcgaggacgagaccacggcttcaccgtgtcgacctacaacacccatgcgtcgggtcgaa
cgggcaagacggtgtcggtgaagtcaaggccgacaaggaaggcgtctaccgtactactgcaccaattctgctc

>mwwt84

cctcatggggcgacgtcactccggacggcaagttcatcatcgtgtcgggcaagctcgacacgcacaccatgatgtactcgttc
gacaagattcaggaggccatcaaggcgaacaagttcgagggcaaggacccgtacggcattcccatcatcggaatgaaggacg
cgctccacaagcaggtcgagctcggcctcggggcgctgcacacgcagtagacaccaagccgtgcatcggttacacctcgat
gtacgtcgactcgcaggtggcgaagtgggactactcgaggggcaaggtgctcgacaagctgtcggtgcactacaacatcggc
cacctcatgacgatggaaggcgactcgggtgaagccgaagggtcgctacctggtcgcgctcaacaagctggccatcgaccgctt
caacccggtcggccactgcacccgcagaaccaccagctcatcgacatctnaacgacaagatgcagctgctctacgacatgc
cgctgccgctcggtagccgcactacgcggctgccatcagcgccgccacctgaagcctggcggttcgtacaaggctcggcac
caac

>mwwt85

ccccacggtgccgatgtctcaccggacggcaagttcatcatcgtctcgggcaagctcgacacccatgtgtcggtctacagcttcg
agaagatccaggccgcatcaaggccggcaagttcgagagcaaggacccgtacggcatcccgggtatcgcgataaaggacg
cgctggcgacgcaggtgtcgtcggcctggggcgctgcacacgcagttcgacagcaggaaactcgtcgcctacacctcgtc
gtacgtcgactcgcaggtcgccaagtgggacttctcgaaaggcaaggtgctcgacaagatcagcgtgcactacaacatcggcc
acctgatgacgatggagggtgacagcaccaagccccaggggcaagtacgtggtggcgctgaacaagctggccatcgaccgctt
cgtgccggtggggcgctgcacccgcagaaccaccagctgatcgacatctccggcgacaagatgcagctgctctacgacat
ggcgctgccgctggggcgagccgcactacgcggtgtcgatcgcgggccaacaagctcaagcccggtgcgctacaaggtcgg
caccgacagccgtacggacgagccgcacccggggcgccaccaaggcgggcgaggagaaggtcgtcaagaaggggaacaa
ggtcgaggtgttcgcgacctgatccgctcgacatcacgcccagacgatcgaggtcgacgtcggcgacgacgtgaccatc
cacctgaccaacctcgagcgcgcgaggacgagacgcacggcttcacggtcagcacctacaacgtgcatgcgtccatcgagc
cgggcaagacggtcagcgtcaagtcaaggccgaccgcgaaggcgtgtaccatactactgcacctattctgct

>mwwt86

cctcatggggcgaggatgtgagcccgacggcaagttcatgaccgttcggcaagctggacacccacgttcgggtctattccttcg
agaagatccaggctgcgatcaaggccggcaagttcgaatccaaggatccgtacggcattccgggtgatcgggatgaaggatgc
gtcgataccacaggtccagctcggctcgtccttcgacacccagtagattccaagccctgtatcgctacacctcgtgtatg
tcgattctcaggtcgtcaaatggaatcactgcaaaggcaaggtactcgacaagatcctgggtgcattacaacatcggtcacctgatg
accatggaaggcgattcggctgatccgaagggtcgtacctggtggcgctgaacaagctggcgatcgaccgcttctgtccggtt
ggcccgctgcatccgcaaaatcaccagctgatcgacatttctaacgacaagatgcaattgtgtacgacatgccgctgccgttgg
gtgagccgactacgctgtggccattgaagccagcaagctggagccggcagtgcgctacaaggctggcaccgacagccgca
ccgatgagaagcaccggcgagtgcgcgcggtgaagaaaagacggtaagaagggaacaaggtcgaggtgttcggc
acctgatccgctcgacatcacacccgagacgatcgaaagtggatgtggcgacgaggtcacctccacctgaccaacctcga

gcgtgcgcaggacgagacccacggcttcaccgtgtcgacctacaacacccacgcctcggtcgagcccggaagaccgtgac
ggtgaagttcaaggccgacaaggaaggtgtgtaccatactactgcacccagttctgctc

>mwwt87

ccgcacggcggtggacgtctgaccagacggcaagttcatgaccgtttccggcaagctggacacccacgtttcggctctattccttcg
agaagatccaggctgcgatcaaggccggcaagttcgaatccaaggatccgtacggcattccgggtgatcgggatgaaggatgc
gtgcataccacagggtccagctcgggtccttcgacacccagtagattccaagccctgtatcgctacacctcgtgtatg
tcgattctcaggctcgtcaaatggaatcactgcaaaggcaaggtactcgacaagatctcggtgcattacaacatcggtcacctgatg
accatggaaggcgattcggctgatccgaagggtcgtacctgggtggcgtgaacaagctggcgatcgaccgctttgtgccggtt
ggcccgtgcatccgtaaaatcaccagctgatcgacatttctaacgacaagatgcaattgtgtacgacatgccgctgccgttgg
gtgagccgcactacgtgtggccatcgaagccagcaagctgaagccggcggtgcgttacaaggtcgggtaccgattcccga
cggacaagccgcaccccgcatggttcgcgccggcggaagaacgcaccgagaagaagggaacaagatcacgggtctacggt
acgttgatccgttcccacatcaccccgaaacgatcgaggctgatgtcggcgacgaagtcacatccacctgaccaacctcgaa
cgtgccaggacgaaacccatggtttcacggtgtccacgatgaacgtccacgcttcggtgaaccgggcaagacagtaccgt
caagttcaaggctgacaaggaaggcgctctatccatactactgcacccatttctgct

>mwwt88

cgatgtgaggccagatggcaagtacatggctgctcggcaagctcgacacccacgtctcgggttacagcttcgagaagatcca
ggccgcatcaaggccgggaaattcgagagcaaggacccctacggcatcccgggtgatcggcatgaaagaggccctgcacgt
gcagggtccgctcggcctgggcccgtgcacacccagtagactcgaaggcctgcgtcgctacacctcgtttacgtggact
cgcaggtcgccaagtggaaattctcgaaaggcaaggtgctcgacaagatcagcgtgcactacaacatcggccacctgatgacc
atggaggggcactccaccaagccggccggcaagtagctcgtcgcgctgaacaagctggccatcgaccgcttcgtgccggtgg
ggccgctgcacccgcagaaccaccagctgatcgacatctcggcgacaagatgcagctgctctacgacatgccgctgccgt
ggcgagccgcactacgtcgtcgccatcgatgccaagaagctcaagccgggggtgcgctacaaggtcggca

>mwwt90

ccgcacgggtgatgtcacgccggacggcaaatcatcaccgtttccggcaagctggacacccacgtttccgtctattccttcg
agaagatccaggccgcatcaaggccggcaagttcgaatccaaggatccgtacggcattccgggtcatcaacatgaaggatgc
ctgcacacccaggttcaactgggtcttggccccctgcacacccagtagactccaaggcctgcgtcgctacacctcgtgtac
gtggatagccaggtcgccaagtgggactactgcgcgggcaaggtgctggacaagctttccgtgcattacaacatcgccacct
gatgaccatggaaggcgactccaccaagccggccggcaagtatctcgtcgcgctgaacaagctggccatcgatcgcttcgtgc
cggtcgggtccctgcacccgcagaaccaccagctgatcgacatcagcaacgacaagatgcagctcctctacgacatgccgctg
ccgctgggcgagccgcactacgccgtcgccatcgccgccgaagctgaagccggcgctgcgttacaaggtcggcaccga
cagccgcaccgacaagccgcatccggcgcggtgctgcccggcgaggagaagaccgtcaagaagggaacaaggtcgaa
gtcttcgccacgctgatccgttcgacatcacgcctgaaaccatcgaggctgatgtcggcgacgaagtgcacgatccacctgacc
aacctcgagcgcgccaggacgagacccacggcttcacggtgtcgacctacaacgtgcatgcctcggctgagccgggcaag
accgtgcaggtgaagttcaaggccgacaaggaaggcgctctatccctactactgcacccacttctgctc

>mwwt92

ccgcacgggtgtggatgtcaggccagacggcaagttcatggtcgtggccggcaagctcgacacccacgtgtcgggttacagctt
cgagaagatccaggccgcatcaaggccggcaagttcgagagcaaggacccgtacggcatcccgggtgatcgcgatgaagga
cgcgctgcacacgcaggtgtcgtcggcctcggcccgtgcacacgcagtagacagcaagccttgctggcctacacgtcg
ctctacgtcgactcgaggtcgcgaagtggaaacttctgaaaggcaaggtcgtcgacaagatcagcgtgcactacaacatcg
ccacctgatgacgatggaggcgactccaccaagccggcggaagtagctggtggcgctgagcaagctggcgatcgaccg
cttcgtgccggtgggcccgtgcaccncagaaccaccanctgatcgacatcagcaacaac

>mwwt94

ggatgtgtgaccggacggcaagttcatggtcgtggccggcaagctcgacacccacgtgtcgggttacagcttcgagaagatcc
aggccgcatcaaggccggcaagttcgagagcaaggacccgtacggcatcccgggtgatcgcgatgaaggacgcgctgcaca
cgcaggtgtcgtcggcctcggcccgtgcacacgcagtagacagcgagccttgctggcctacacgtcgctctacgtcgac
tcgcaggtcgcgaagtggaaacttctgcaaggcaaggtcgtcgacaagatcagcgtgcactacaacatcggccacctgatgac

APPENDIX 2.

Nucleotide accession numbers for reference organisms used in this study (protein and nucleotide sequences were obtained from the Pipeline Functional Gene/Repository [PFGR] [<http://flyingcloud.cme.msu.edu/fungene/>]).

Achromobacter cycloclastes X94977

Acidovorax sp. NC_008782

Alkalilimnicola ehrlichei NC_008340

Anaeromyxobacter dehalogenans NC_007760

Anaeromyxobacter sp. Fw109-5 DQ504302

Azoarcus NC_006513

Bradyrhizobium japonicum NC_004463

Brucella ovis NC_009504

Burkholderia pseudomallei NC_009074

Campylobacter fetus NC_008599

Colwellia psychrerythraea NC_003910

Crenimugil crenilabis NC_003170

Cupriavidus necator NC_005241

Dechloromonas aromatica NC_007298

Desulfitobacterium hafniense NC_007907

Dinoroseobacter shibae NZ AAVE01000001

Geobacillus thermodenitrificans NC_009328

Gramella forsetii NC_008571

Hahella chejuensis NC_007645

Haloarcula marismortui NC_006396

Halorubrum lacusprofundi NZ_ABEB01000008

Magnetospirillum gryphiswaldense CU459003

Magnetospirillum magneticum NC_007626

Magnetospirillum magnetotacticum NZ_AAAP01003630

Marinobacter hydrocarbonoclasticus DQ504302

Marinobacter sp. NZ_AAXY01000004

Methylobacterium NZ_ABAY01000093

Paracoccus pantotrophus AF016058

Photobacterium profundum NZ_AAPH01000041

Pseudomonas aeruginosa NZ_AAKW01000028

Pseudomonas denitrificans AF016059

Pseudomonas fluorescens AF056319

Pseudomonas stutzeri NC_009434

Psychroflexus torquis NZ_AAPR01000008

Psychromonas ingrahamii NC_008709

Pyrobaculum calidifontis NC_009073

Reinekea sp. NZ_AAOE01000026

Rhodobacter sphaeroides AF125260

Rhodopseudomonas palustris NC_008435

Robiginitalea biformata NZ_AAOI01000001

Roseovarius sp. NZ_AAMV01000002

Shewanella loihica NC_009092

Silicibacter pomeroyi NC_006569

Sinorhizobium meliloti NC_003037

Stappia aggregata NZ_AAUW01000018

Thiobacillus denitrificans NC_007404

Thiomicrospira denitrificans NC_007575

Wolinella succinogenes AJ640086