Development of group-specific *nosZ* quantification method targeting active nitrous oxide reducing population in complex environmental samples

DaeHyun D. Kim[1], Doyoung Park[1,2], Hyun Yoon[1,3], Min Joon Song[1], Taeho Yun[1], and Sukhwan Yoon[1#]

[1] Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology (KAIST)

[2]Department of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

[3]Department of Civil and Environmental Engineering, Cornell University, Ithaca, New York, USA

Abstract

Despite the recent interest in *nosZ*-possessing organisms as the sole N_2O sink in soil and aquatic environments, quantification of *nosZ* has relied on undependable qPCR techniques prone to unspecific amplification and compromised accuracy. Here, we have combined culture-based methods with computational methods to develop TaqMan-based qPCR for quantification of microorganisms actively involved in N_2O consumption in activated sludge reactors. A sewage sample was enriched in a batch reactor fed continuous stream of N_2 containing 20-10,000 ppmv N_2O, where 14 genera of potential N_2O-reducers were identified. All available amino acid sequences of NosZ affiliated to these taxa were grouped into five subgroups (two clade I and three clade II groups), and primer/probe sets exclusively and comprehensively targeting each subgroups were designed and validated with *in silico* PCR. The *nosZ* profiles of four activated sludge samples determined with the qPCR assays were compared with those analyzed from shotgun metagenome. The results of the group-specific qPCR assays were generally in agreement with the results of the metagenomic analyses, with the two subgroups (*Flavobacterium*-like and *Dechloromonas*-like) of clade II *nosZ* dominating the *nosZ*-possessing population. These quantitative tools

will be immensely useful for identification of the key N_2O-reducing populations in environments with rapid nitrogen turnover.

Introduction

Nitrous oxide (N_2O) is one of the three major greenhouse gases with the largest contributions to global warming, along with CO_2 and CH_4 (ref. 1). Although the contribution of N_2O is estimated to be only ~6% of the net greenhouse gas emissions in terms of CO_2eq, which is far less than those of CO_2 and CH_4, eliminating one molecule of N_2O from the atmosphere has the same merit as removing ~300 molecule of CO_2 due to its high global warming potential. Besides, N_2O has also been the most consequential ozone depletion agent (ref. 2, 3). Thus, global efforts to curb the increase in atmospheric N_2O concentration are necessitated for sustainable future. A better understanding of the biogeochemical reactions functioning as N_2O sources and sinks in nitrogen-rich anthropogenic environments, e.g., fertilized agricultural soils and wastewater treatment plants (WWTPs), is especially important for N_2O emission mitigation, as such environments have been estimated as the hotspots of N_2O emission to the atmosphere (ref. 4, 5).

Particularly, the biological N_2O reduction mediated by Nos-type nitrous oxide reductases (NosZ) has recently attracted immense scientific attention as the sole sink of N_2O on the Earth's surface at non-elevated temperature (ref. 6, 7, 8). This reaction had been, for long, known as one of the stepwise reactions constituting the denitrification pathway (ref. 4, 9). Only recently has the N_2O-to-N_2 reduction been recognized as an independent energy-conserving reaction, as organisms possessing *nosZ*, but neither *nirK* nor *nirS*, and capable of growth with N_2O as the sole electron acceptor, were found across diverse phylogenetic groups of bacteria (ref. 8, 10, 11). The majority of the N_2O reducing organisms lacking *nirK* or *nirS* possess clade II *nosZ*, a novel clade of *nosZ* with substantial structural differences (e.g., Sec-type N-terminal signal peptide) from the previously known clade I *nosZ* (ref. 8). Several independent research groups have reported the higher affinity of the clade II *nosZ*-possessing organisms to N_2O, and more specifically, the organisms with *nosZ* closely affiliated to *Dechloromonas* spp. have been reported with particularly low whole-cell Michaelis constants

($K_{m,app}$), suggesting that this group of organisms may be involved in consumption of low-concentration N_2O produced *in situ* via diverse biotic and abiotic processes (ref. 11, 12, 13, 14). The microbial community developed in the laboratory-scale biofilter treating 100 ppmv N_2O included the clade II N_2O reducer *Flavobacterium* spp. as the most abundant *nosZ*-carrying organisms, also supporting the significance of clade II N_2O as an important N_2O sink (ref. 15, 16).

Increasing number of recent studies on environmental nitrogen cycling have reported the correlation between N_2O reduction activity or net N_2O emission and *nosZ* abundance or the *nosZ*-to-(*nirS*+*nirK*) gene abundance ratio (ref. 11, 17, 18, 19). Since the discovery of the clade II *nosZ*, the clade I-vs-clade II quantification has become immensely popular as indicators of the N_2O sink capability (ref. 14, 20, 21). Most, if not all, of these *nosZ* gene quantifications were performed using quantitative polymerase chain reactions (qPCR) using SYBR Green detection chemistry (ref. 18, 22). The positive results (coverage over 85% and unspecific amplification below 10%) from the *in silico* analyses of the available *nosZ* primer sets may give an unjustified impression that these qPCR assays may be reliable quantitative molecular tools (ref. 23); however, application of SYBR Green qPCR for quantification of target DNA sequences in complex environmental samples has been regarded problematic due to the specificity issues (ref. 24). The melting curve analysis, the only means to check the specificity of SYBR Green qPCR assays, is not applicable when the assays are used for targeting functional genes in complex microbiomes, as amplicons with different sequences yield distinct melting curves (ref. 18). The specificity of the oft used primer sets for clade I and clade II *nosZ* (1840F: 5'-CGCRACGGCAASAAGGTSMSSGT-3'/ 2090R: 5'-CAK RTGCAKSGCRTGGCAGAA-3' for clade I *nosZ*; nosZ-II-F: 5'-CTIGGICCIYTKCAYAC-3' / nosZ-II-R: 5'-SKSACCTTITTRCCITYICG-3' for clade II *nosZ*) have not yet been closely examined, although these primers have been repeatedly used in diverse fields of study (ref. 19, 25, 26, 27). Besides, the qPCR efficiencies reported for amplification with these degenerate primer sets have been repeatedly reported to be below 80%, also putting the reliability of the previous qPCR results to question (ref. 19, 28, 29).

Quantitative functional gene analysis using shotgun metagenomics approach has become increasingly popular recently (ref. 30, 31, 32, 33, 34). In this approach, short sequencing reads are either 1) mapped onto the targeted functional gene sequences in assembled contigs identified using BLAST and/or HMMER algorithms or 2) directly assigned as quality-trimmed raw reads to the targeted functional gene sequences using the same algorithms, and the read counts are used for calculating the relative abundances of the functional genes with respect to the total microbial population (determined from the read counts of a single-copy housekeeping gene, e.g., *rpoB*). A number of such functional gene analysis tools with relatively high reliability are now available on line and clade-specific quantification of *nosZ* using short-read HiSeq sequencing reads have been previously performed (ref. 30, 33); however, the high cost of the high-throughput shotgun sequencing (for any meaningful results, a sequencing depth of >5 Gb is often required for shotgun metagenome analyses) and prolonged processing time still pose a yet unsurmountable barrier for implementation in real-time monitoring of environmental microbiomes undergoing rapid changes (ref. 35). Therefore, qPCR is still regarded as the preferred choice of quantitative molecular tool for such purpose, in spite of its own imperfectness.

The specificity issues with the SYBR-Green qPCR can be resolved by implementing the TaqMan chemistry-based qPCR technique that requires an additional consensus region for probe binding (ref. 24); however successful cases of such approach has not been reported for *nosZ* genes, presumably due to the difficulty of finding three neighboring consensus regions among the prohibitively divergent *nosZ* gene sequences (ref. 8, 10, 33). Here, we have successfully developed TaqMan-based qPCR reactions targeting four groups (two clade I *nosZ* groups and two clade II *nosZ* groups) of *nosZ* with high coverage and specificity, while maintaining PCR efficiency above 90%, which is often regarded as the criterion for reliable quantification (ref. 36). The N_2O-reducing population information gathered from laboratory cultivations of activated sludge samples with varying concentrations of N_2O was used to restrict the targets to active N_2O-reducing organisms, enabling design of mutually exclusive primer and probe sets targeting the five *nosZ* groups. The designed qPCR methods were validated with different activated sludge samples by comparing the qPCR quantification results with

nosZ quantification data processed from the shotgun metagenome sequences. The PCR amplicons were also analyzed to verify the specificity and coverage of the qPCR assays.

Materials and methods

Sample collection

The wastewater inoculum for the N_2O enrichment experiments were grab-sampled from the anoxic section of the activated sludge tank at Daejeon municipal WWTP (36º23'5" N, 127º24'28" E) in September 2016 (denoted as Daejeon1). Additional wastewater samples for validation of the designed qPCR reactions were collected at the same WWTP (Daejeon2) in February 2019 and two other activated sludge WWTPs located in Gwangju and Gapyeong (35°09'22.4"N 126°49'51.6"E and 37°49'00.1"N 127°31'13.0"E, respectively) in January and February of 2019, respectively. Each wastewater sample was collected in a 2-L polyethylene bottle filled up to the brim to minimize oxygen ingress. The sample bottles were immediately placed in a cooler and transported to the laboratory. The wastewater samples were stored at 4°C until use.

Fed-batch enrichment of activated sludge samples and identification of active N_2O-reducing groups

A simple fed-batch bioreactor was constructed for enrichment of active N_2O-reducers in the Daejeon1 wastewater sample. A 500-mL glass bottle with a side port (Duran, Mainz, Germany) was fitted with a GL45 cap with three ports. Two of the ports were used as the inlet for N_2O-carrying gas and the gas outlet from the bottle and the remaining port was used for aqueous phase sampling. The side port of glass bottle was used for sampling of the gaseous phase. The modified MR-1 medium was prepared by adding per 1 L of deionized water, 0.5 g NaCl, 0.41 g sodium acetate, 0.23 g KH_2PO_4, 0.46 g K_2HPO_4, 0.026 g NH_4Cl, 1 ml of 1000X trace metal solution, and 1 mL of 1000X vitamin stock solution (ref. 37). The reactor with 200 mL MR-1 medium (40% of the total reactor volume) was flown through with N_2 gas carrying 0 ppmv, 20 ppmv, 200 ppmv or 10,000 ppmv N_2O (Samo specialty gas, Daejeon, South Korea) for 30 hours before inoculation. After inoculating the medium with 2 mL of the Daejeon1 sample, the same gas used for flushing was bubbled through the medium

at the volumetric flowrate of 20 mL min^{-1} to provide the sole electron acceptor N_2O to the microbial culture and maintain the reactor at anoxic condition. The reactor operated with >99.999% N_2 gas served as the control to confirm that the microbial consortia were enriched with N_2O as the electron acceptor and the contribution of O_2 contamination to microbial growth was minimal. The N_2O concentration in the headspace of the reactor was monitored with a HP6890 Series gas chromatography fitted with an HP-PLOT/Q column and an electron capture detector (Agilent, Palo Alto, CA), with the injector, oven and detector temperatures set to 200, 85, and 250 ºC, respectively (ref. 38). The O_2 concentration in the fed-batch reactor was monitored with a FireSting-O_2 oxygen meter (Pyroscience, Aachen, Germany); however, due to the relatively high detection limit of the sensor (the gas phase concentration of ~0.1% v/v), complete absence of O_2 was not guaranteed.

The growth of bacterial population in the reactor was monitored with quantitative polymerase chain reaction (qPCR) using TaqMan chemistry targeting the conserved region of eubacterial 16S rRNA genes. At each sampling time point, an 1.5-mL aliquot was collected from the aqueous phase of the reactor and DNA was extracted from the pellets using DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) following the protocol provided by the manufacturer. Quantitative PCR was performed with the 1055f (5'-ATGGCTGTCGTCAGCT-3') / 1392r (5'-ACGGGCGGTGTGTAC-3') / Bac1115Probe (5'-CAACGAGCGCAACCC-3') primer and probe set targeting a conserved region of bacterial 16S rRNA genes using a QuantStudio3 platform (Thermo Fisher Scientific, Waltham, MA) (ref. 39). Incubation was halted when the population of the enrichment reached a plateau, as indicated by three consecutive measurements without increased gene counts. The reactor was dismantled and the aqueous phase was collected for microbial composition analysis.

The hypervariable V6−8 region of the 16S rRNA gene was amplified with 926F: 5′-AAACTYAAAKGAATTGRCGG-3′ / 1392R: 5′-ACGGGCGGTGTGTRC-3′ primer set and MiSeq sequencing of the amplicons was outsourced to Macrogen Inc. (Seoul, Korea). The raw sequence reads have been deposited in the NCBI short reads archive (SRA) database (accession: PRJNA552413). The raw sequence data were processed using the QIIME pipeline v 1.9.1. The

sequence reads with the quality scores lower than the default cut-off (the phred score of 20) were removed. The trimmed reads were clustered to the Greengenes (v 13.8) 16S rRNA gene database with the cut-off value set to 0.97. The remaining reads, which failed to cluster with the reference database, were clustered into *de novo* OTUs with the same cut-off value (0.97). The taxonomy was assigned to each OTU using the RDP classifier against the Greengenes v13.8 database. For each enrichment sample, the genera assigned to the OTUs with the relative abundances higher than 0.3% were selected. The lists of all NosZ proteins belonging to the organisms affiliated to these genera were extracted from the Uniprot (www.uniprot.org) database (accessed in March 2017) (Table S1). The initial list obtained with the search keywords, "NosZ", "nitrous", "nitrous oxide", and "nitrous-oxide" was manually curated to retain only the NosZ proteins.

Design of degenerate primers and probes for group-specific qPCR of *nosZ* genes

The pooled *nosZ*/NosZ sequence data included, in total, 174 nucleotide sequences and the corresponding amino acid sequences from 14 distinct genera. Subsequently, a multiple sequence alignment was performed with these amino acid sequences, using MUSCLE algorithm with the parameters set to the default values (ref. 40). The NosZ phylogenetic tree was constructed using the neighbor-joining method in MEGA 7.0 with the bootstrap value of 500 (ref. 41). The *nosZ* gene sequences were clustered into five groups (NosZG1-5) according to the positions of the corresponding NosZ sequences in the phylogenetic tree, which consisted of five phylogenetically distinct subbranches.

For each *nosZ* group, a primers and probe set was designed using the PriMux software to comprehensively and exclusively target the *nosZ* gene sequences within the group (ref. 42). The parameters were modified from the default values to obtain the optimal candidate degenerate oligonucleotide sequences for qPCR (length: 18-24 bp, amplicon size: 80-400 bp, T_m: 56-62°C for primers and 68 - 72°C for probes). Several candidate primers and probe sets were generated for each *nosZ* group, implementing the ***min***, ***max***, and ***combo*** algorithms of the Primux software. The performance of each candidate primers and probe set was predicted with *in silico* PCR performed

with the simulate_PCR software against all complete genome sequences of the organisms belonging to each of the five *nosZ* groups (Table S1) (ref. 42, 43). Coverage within the target *nosZ* group and mutual exclusivity across the groups were the two major criteria for assessment of the candidate primers and probe sets. The *in silico* PCR tests were also performed against the complete genomes of 50 bacterial strains lacking *nosZ* to preclude the possibility of unspecific detection.

Experimental verification of the designed primer and probe sets on model N_2O reducer strains and construction of calibration curves

For experimental verification of each *nosZ* primer and probe set, a model organism was selected harboring the targeted *nosZ* gene(s). The selected model organisms were *Pseudomonas stutzeri* DCP-Ps1 (NosZG1), *Acidovorax soli* DSM25157 (NosZG2), *Flavobacterium aquatile* LMG4008 (NosZG3), *Ignavibacterium album* JCM16511 (NosZG4), and *Dechloromonas aromatica* RCB (NosZG5). *Acidovorax soli* DSM25157 and *F. aquatile* LMG4008 were acquired from Korean Collection for Type Cultures and *I. album* JCM16511 from Japanese Collection of Microorganisms. The axenic batch cultures of these organisms were prepared as previously described in the literature or using the media and incubation conditions specified by the distributors. The cells were harvested at OD_{600nm} = 0.1 and the *nosZ* gene(s) in the DNA extracted from the cell pellets were amplified using the designed primers. The calibration curve for each qPCR reaction was constructed using ten-fold serial dilutions of PCR2.1® vectors (Invitrogen, Carlsbad, CA) carrying the *nosZ* amplicons. The thermocycle for the qPCR of NosZG1-G5 was as follows: 95°C for 10 min and 40 cycles of 95 °C for 30 s, 58 °C for 60 s, and 72 °C for 60 s.

Group specific quantitative PCR targeting *nosZ* gene in activated sludge samples

The group-specific quantification of the *nosZ* genes in the activated sludge samples were performed with the designed primer and probe sets (Table 1) on a QuantStudio 3 real-time PCR instrument (Thermo Fisher Scientific, Waltham, MA) using TaqMan detection chemistry (FAM as the reporter and NFQ-MGB as the quencher). DNeasy Blood & Tissue Kit (QIAGEN) was used to extract DNA

from the activated sludge samples. Each 20-µL qPCR reaction mix contained 10 µL of 2X TaqMan master mix (Applied Biosystems, Foster city, CA, USA), 5 µM each of the forward and reverse primers, 0.5 µM of the probe, and 2 µL of the DNA sample. Calibration curves were used to calculate the copy numbers of the targeted *nosZ* genes from the C_t values. Eubacterial 16S rRNA genes in the extracted DNA samples were quantified using the 1055F/1392R/Bac115Probe set for quantification of total bacterial population in the activated sludge samples. The copy numbers of the *nosZ* genes in the wastewater samples were normalized with the 16S rRNA gene copy numbers to facilitate comparison with the relative abundances of the *nosZ* groups from the metagenomic analyses.

The PCR amplicons of the Daejeon1 sample amplified with NosZG1-5 primer sets were sequenced using Illumina Miseq platform (San Diego, CA) at the Center for Health Genomics and Informatics at University of Calgary. The raw sequence reads have been deposited in the SRA database (accession number: PRJNA552418). After quality trimming and merging of the paired-end sequences, the sequences without the probe-binding region were removed, and the remaining reads were clustered into OTUs with 0.97 cut-off using cd-hit-est v. 4.6 (ref. 44). The OTUs were annotated using blastx against the bacterial Refseq database downloaded in June, 2018, with the e value cut off set to 10^{-3} and word size to 3 and no seg option,

Computational quantification of *nosZ* genes from shotgun metagenomes of the activated sludge samples

The DNA samples for shotgun metagenome sequencing were extracted from 50 mL each of the four activated sludge samples using DNeasy PowerSoil Kit (Qiagen). Sequencing of the metagenomic DNA was performed at Macrogen Inc., where Hiseq X Ten sequencing platform (Illumina, San Diego, CA) was used for generating 5-10 Gb of paired-end reads data with 150-bp read length. The raw sequence reads have been deposited in the NCBI short reads archive (SRA) database (accession numbers: PRJNA552406). The raw reads were then processed using Trimmomatic v0.36 software with the parameters set to the default values (ref. 45). The trimmed reads were translated into amino

249 acid sequences using all six possible reading frames and screened for clade I and II *nosZ* sequences
250 using hidden Markov models (HMM). The two HMM algorithms for clade II *nosZ* were downloaded
251 from the Fungene database (accessed in October 2016). As the HMM for clade I *nosZ* in the database
252 had specificity issues, the HMM algorithm was constructed *de novo* using *hmmbuild* command of
253 HMMER v3.1b1. The clade I NosZ sequences used to build the HMM were manually curated from
254 the pool of NosZ sequences downloaded from the NCBI database (accessed in October 2016), to
255 represent diverse subgroups within the clade (Table S3) (ref. 8, 10). The candidate partial NosZ
256 sequences were extracted from the translated shotgun metagenome reads using the *hmmsearch*
257 command of HMMER v3.1b1 with the e value cutoff set to 10^{-5}. The nucleotide sequence reads
258 corresponding to the extracted partial NosZ sequences (in separate bins for clade I and clade II NosZ)
259 were assembled into contigs using metaSPAdes v3.12.0 with parameters set to default values (ref. 46).
260 The assembled contigs with lengths shorter than 200 bp were filtered out. The overlapping contigs
261 appearing in both clade I and clade II NosZ bins were identified by clustering the two sets of contigs
262 against each other with a nucleotide identity cutoff of 1.0. The identified overlapping contigs were
263 manually called to the correct bin according to the BLASTX results. The trimmed sequence reads
264 were then mapped onto the contigs in the *nosZ* bins using Bowtie2 v2.2.6, yielding sequence
265 alignments for the contigs and the mapped reads. The alignment files were further processed using
266 samtools v0.1.19, and the PCR duplicates were removed using MarkDuplicates function of picard-
267 tools v1.105 (ref. 47). The sequence coverages of the contigs were calculated using bedtools v2.17.0.
268 The number of reads mapped on to the contigs were normalized with the lengths of the respective
269 contigs. The contigs were assigned taxonomic classification using blastx, and those matching the taxa
270 identified from amplicon sequencing were assigned to NosZG1-5 bins, while those affiliated to
271 *Ignavibacterium nosZ* were manually assigned to NosZG4 (Table S5). The contigs without matching
272 sequence were binned as 'other *nosZ* sequences'. As no HMM method has been developed for rRNA
273 genes, the sequence coverage of *rpoB* gene, a single-copy housekeeping gene, was used for
274 normalization of the *nosZ* abundance data. The HMM algorithm for *rpoB* was downloaded from the
275 Fungene database. The attempt to use 16S rRNA sequences extracted with Meta-RNA and assembled
276 with EMIRGE as the template for mapping was not successful, as the coverage of the extracted 16S

277 rRNA sequences turned out to be an order of magnitude lower than the *rpoB* coverage (Data not
278 shown) (ref. 48, 49).
279
280 **Results**
281 **Active N_2O reducers enriched in fed-batch incubation with varying N_2O concentrations**
282 The microbial communities of the three N_2O-reducing enrichments, each prepared with different N_2O
283 concentrations, were analyzed to identify active N_2O-reducing groups of microorganisms (Table S6).
284 Screening for OTUs with >0.3% abundance in any of the three enrichments yielded 69 OTUs
285 assigned to 33 genera in total, and 14 of these genera were identified with phylogenetic subgroups
286 harboring clade I or clade II *nosZ*. The OTUs belonging to these putatively *nosZ*-harboring genera
287 amounted to 50.8% – 63.2% of the total microbial population in the enrichments. The abundances of
288 the genera putatively harboring clade II *nosZ* were observed to be greater than the genera harboring
289 clade I of *nosZ* (Table S6). In the enrichment incubated with 20 ppmv N_2O, *Cloacibacterium* (18.9%),
290 *Flavobacterium* (14.2%), and *Acidovorax* (13.5%) were identified as the dominant genera.
291 *Dechloromonas* (17.3%) and *Flavobacterium* (15.8%) were the dominant genera in the 200 ppmv
292 enrichment, and *Dechloromonas* was the predominant population in the enrichment incubated with
293 10000 ppmv of N_2O, constituting 46.0% of the total microbial population.
294
295 **Designing of the degenerate *nosZ* primer/probe sets**
296 A phylogenetic tree was constructed with 174 translated NosZ sequences affiliated to the 14 genera
297 putatively harboring active N_2O-reducing organisms (Figure 1). The NosZ phylogenetic tree
298 composed of five distinct branches (NosZG1 – NosZG5), to which each of the NosZ sequences was
299 assigned (Figure 1). NosZG1 and NosZG2 were identified as clade I NosZ and NosZG3-NosZG5 as
300 clade II NosZ. After the iterative process of designing candidate primers/probe sets and performing *in*
301 *silico* PCR tests, the final sets of degenerate primers/probe sets were designed to comprehensively and
302 exclusively target the five *nosZ* groups (Table 1). The *in-silico* PCR performed against 174 genomes
303 from which the target *nosZ* sequences were extracted and 50 genomes without *nosZ* confirmed the
304 high levels of coverage (57.3-100%) and complete exclusivity for all five degenerate primers/probe

sets (Table 1). The qPCR calibration curves were constructed with the selected model organisms (Figure, S3, Table S7), and despite the high levels of degeneracy (up to 108), amplification efficiencies above 90% were attained for all five primer and probe sets after rigorous optimization process.

Cross-validation of the group-specific qPCR quantification results with shotgun metagenome data

The *nosZ* gene abundances in the four activated sludge samples were quantified using the newly designed qPCR assays, and the *nosZ* sequences extracted from the shotgun metagenomes of these same samples were quantitatively analyzed in parallel, for cross-validation of the qPCR results (Figure 2). The qPCR results of the four activated sludge samples invariably showed the dominance of the clade II *nosZ* genes (NosZG3 and NosZG5) over clade I *nosZ* genes (NosZG1 and NosZG2), with at least five-fold higher copy numbers (Figure 3), suggesting the importance of clade II *nosZ*-possessing organisms in denitrification and N_2O reduction in activated sludge reactors. The qPCR targeting the NosZG4 group failed to amplify the clade II *nosZ* belonging to this group. The NosZG4 primers/probe set was designed from a single complete genome (*Ignavibacterium album*) and six sequences from metagenome assembled genomes (MAGs), which may have been insufficient to cover the sequence divergence of this *nosZ* subgroup.

The distribution of the *nosZ* sequences extracted from the shotgun metagenomes (contig statistics provided in Table S4) were also severely biased towards clade II, with the sequence coverages of the clade II *nosZ* groups higher than those of the clade I *nosZ* by at least 4.5 folds. The compositions of the *nosZ* genes exhibited high level of similarity across the activated sludge samples. The calculated relative abundances of the two most abundant groups, NosZG3 and NosZG5, were relatively consistent across the samples, varying by less than 1.7-fold. The contributions of NosZG1 and NosZG2 to the total *nosZ* abundances were minor in all of the samples and largely variable. The fold differences between the samples with the highest relative abundance of NosZG1 and NosZG2 and those with the lowest abundance were 3.5 and 2.3, respectively. The *nosZ* genes sorted as NosZG4

333 constituted <1.2% of the total *nosZ* genes grouped as NosZG1-NosZG5, suggesting that this *nosZ*
334 group has less significant role in N_2O reduction in activated sludge tanks than other analyzed groups.
335
336 As the qPCR data were normalized with the 16S rRNA copy numbers and the metagenome-derived
337 data were normalized with the sequence coverages of *rpoB*, direct comparison of the outcomes was
338 not possible; however, the similarity in the general trends was evident. The ratios of the *nosZ*/*rpoB*
339 (metagenome) to *nosZ*/16S (qPCR) were within a narrow range between 1.6 and 2.4 for NosZG5
340 across the four activated sludge samples. Quantification of other *nosZ* subgroups yielded less
341 consistent results between the two methods. The *nosZ*/*rpoB*-to-*nosZ*/16S ratio varied from 3.3 to 25.9
342 for NosZG3 and the results for less abundant NosZG1 and NosZG2 were less consistent. The
343 *nosZ*/*rpoB*-to-*nosZ*/16S ratios of NosZG1 and NosZG2 for the Daejeon1 sample were 20.4 and 90.1,
344 respectively, while these ratios were 1.9 and 1.8 for the Daejeon2 sample collected from the same
345 WWTP at a different time of the year. Nevertheless, in 9 of the 16 qPCR assays performed (excluding
346 NosZG4), the *nosZ*/*rpoB* and *nosZ*/16S ratios stayed within the same order of magnitude, and the
347 general agreement with the metagenome-driven data that the NosZG3 and NosZG5 were the
348 dominant *nosZ* subgroups indicated predictive accuracy of the group-specific quantification using the
349 newly-designed primers.
350
351 The TaqMan-based qPCR assays were also compared with the most frequently used qPCR assays for
352 clade I and clade II *nosZ* using SYBR green detection chemistry (Figure S1). The subpar
353 amplification efficiencies of 1840F–2090R (targeting clade I *nosZ*; 79.3%) and nosZII-F–nosZII-R
354 (targeting clade II *nosZ*; 61.9%) primer sets posed a critical accuracy problem for these SYBR Green
355 based qPCR assays (Table S2). The abundances of clade I *nosZ*, as determined with the SYBR Green
356 assays, were relatively consistent with the cumulative abundances of NosZG1 and NosZG2
357 determined with the TaqMan qPCR assays, except for Daejeon1 sample, where the SYBR Green
358 qPCR yielded 87 times lower copy numbers than the TaqMan qPCR. The SYBR Green qPCR with
359 nosZII-F–nosZII-R primer set underestimated the clade II *nosZ* copy numbers by several orders of
360 magnitudes for three of the four analyzed samples, with the largest difference (1.8×10^3-fold)

observed with the Dajeon1 sample. The importance of the clade II *nosZ* among the N_2O-reducing microbial community in these activated sludge samples would have gone unnoticed due to this underestimation, if the conventional qPCR assays were used as the sole means of *nosZ* quantification.

Analyses of the amplicons amplified with the *nosZ* qPCR primers

The *nosZ* amplicons of the Daejeon1 sample, amplified with the newly-designed qPCR primers (NosZG1, NosZG2, NosZG3, and NosZG5), were sequenced and analyzed to check whether the qPCR reactions were comprehensive and mutually exclusive (Figure 4). As the NosZG4 primers failed to amplify the targeted genes, NosZG4 was excluded from the downstream analysis. The *nosZ* taxa extracted from the shotgun metagenome data matching the taxa observed in the amplicon analysis were considered as targeted by the primer and the probe. The *nosZ* genes of putative active N_2O reducers captured by the NosZG1 and NosZG2 primer sets accounted for 54.2% of the entire pool of the clade I *nosZ* genes extracted from the shotgun metagenome. The NosZG3 and NosZG4 primer sets captured 63.4% of the clade II *nosZ* genes recovered from the metagenome (Figure S2). No amplification outside the targeted group occurred for any of the primers, confirming the complete mutual exclusivity of these primers. Each of the qPCR assays were comprehensive within its target groups, as analyzed *nosZ* amplicon sequences covered all genera originally targeted by each primer and probe set.

Overall, the OTUs identified to be abundant from sequence analyses of the PCR amplicons coincided with the dominant *nosZ* OTUs from shotgun metagenome analysis. The OTUs affiliated to *Rhodobacter capsulatus*, Rhodobacteraceae bacterium QY30, *Hyphomicrobium denitrificans*, and *Bradyrhizobium* sp. were the dominant OTUs among the OTUs NosZG1 amplicons. The two most abundant *nosZ* OTUs among the NosZG3 amplicons were affiliated to *Flavobacterium columnare* and *Niastella koreensis*, which were also the most abundant *nosZ* taxa according to the metagenome analyses. Discrepancies between the amplicon sequencing data and the shotgun metagenome data were observed for NosZG2 and NosZG5 to some degree. The OTUs assigned to the genus *Thauera* and *Ruvrivivax gelatinosus* were the dominant among NosZG2, constituting 50.4% of the amplicons,

389 but together constituted only 2.2% of the metagenome-derived *nosZ* sequences belonging to this
390 group. Instead, an OTU affiliated to *Ramlibacter tataouinensis* was recovered in high relative
391 abundance in the metagenome-derived *nosZ* pool, suggesting possible misassembly due to the
392 sequence similarity with *nosZ* of *R. gelatinosus.* The low relative abundance of *Dechloromonas*
393 *aromatica* in the NosZG5 amplicons (0.2%) may also be explained as a consequence of possible
394 misassembly, as the relative abundance of the OTU assigned to *Azospira oryzae* was much more
395 abundant (54.6%) in the qPCR amplicon than in the metagenome-derived *nosZ* pool.
396
397 **Discussion**
398 Quantification of the functional genes encoding the nitrogen cycle enzymes (e.g., bacterial/archaeal
399 *amoA*, comammox *amoA*, *nrfA*, and clade I and II *nosZ*) has been used in increasing number of
400 studies across diverse disciplines of environmental sciences and engineering, as predictors of nitrogen
401 cycling activity in diverse soil and aquatic environments (ref. 23, 50, 51, 52). Despite the frequent use
402 of qPCR in determining the potential N_2O-reducing populations, i.e., clade I and clade II *nosZ*-
403 posessing microorganisms, the predictability of qPCR quantification has rarely been scrutinized in the
404 previous studies. The new group-specific qPCR assays exhibited definite advantages over these
405 frequently used primers. Although the targets were limited to the groups of *nosZ*-possessing
406 organisms enriched with N_2O in *ex situ* cultures, the copy numbers obtained were, in many cases,
407 orders of magnitudes larger than those measured using 1840F/2090R (nosZ I) and nosZII-F/nosZII-
408 R(nosZ II) primer sets. The diversity of the sequences recovered in the amplicon sequences indicates
409 the exhaustive coverages of the qPCR assays developed in this study within the targeted groups. Most
410 of the major *nosZ* OTUs (>1% in relative abundance) recovered in metagenome-based *nosZ* profiling
411 of the examined activated sludge samples were amplified with exactly one of the four primer sets,
412 thus confirming the absolute mutual exclusiveness of the primers/probe sets. The amplicon
413 sequencing results suggest trade-offs between coverage and specificity when choosing between
414 SYBR Green and TaqMan chemistry, as the probes select against non-*nosZ* sequences but reduce the
415 diversity of covered *nosZ* sequences.
416

417 The major *nosZ*-possessing organisms identified in the qPCR and metagenomic analyses differ greatly
418 from the major *nosZ*-possessing populations of the agricultural soils analyzed previously with shotgun
419 metagenome sequencing (ref. 33). None of the six most abundant *nosZ* phylogenetic groups in the
420 Havana and Urbana agricultural research site soils (the *nosZ* genes affiliated to *Anaeromyxobacter*,
421 *Opitutus*, *Hydrogenobacter*, *Ignavibacterium*, *Dyedobacter*, and *Gemmatimonas*) was identified as a
422 major population in any of the activated sludge samples examined in this study. Of the organisms
423 affiliated to these genera, *Anaeromyxobacter dehalogenans* and *Gemmatimonas aurentiaca* have been
424 confirmed of N_2O reduction activity; however, the N_2O reduction rates measured *in vitro* were orders
425 of magnitudes lower for these organisms than other examined N_2O-reducing organisms throughout
426 the entire range of N_2O concentration, and *G. aurantiaca* lacked the capability to utilize N_2O as the
427 growth substrate (ref. 10, 12, 38). The differences in the compositions of the *nosZ*-possessing
428 populations may be attributed to the inherent difference in the rates of the biochemical turnover
429 processes in WWTP activated sludges and soils. Rapid nitrogen cycling reactions take place on a
430 constant basis in WWTPs, while the time scale of the biogeochemical processes in agricultural soils is
431 orders of magnitude longer and often, nitrogen supply provided through fertilization and plant
432 exudation is sparse and sporadic (ref. 53). Besides, upland agricultural soils are often well-aerated and
433 thus, the capability of N_2O utilization would not provide specific benefits to the *nosZ*-harboring
434 organisms at ordinary dry conditions (ref. 54). In fact, incubation of soil microbial consortium with
435 N_2O in the same fed-batch cultivation resulted in enrichment of similar organismal groups as
436 observed in the activated sludge (e.g., *Pseudomonas*, *Flavobacterium*, *Acidovorax*, and
437 *Chryseobacterium*), suggesting that under occasions of pulse stimulation of denitrification and N_2O
438 reduction, e.g., flooding events immediately following fertilization, these organisms may become the
439 relevant N_2O sinks in soils (Table S8.).
440
441 *Ignavibacterium* spp. were identified in all enrichment cultures and the activated sludge samples in
442 significant populations (>0.1% of the total 16S rRNA amplicon sequences), and *nosZ* gene sequences
443 affiliated to this genus were recovered in the shotgun metagenome data as well. The only isolated
444 strain of the genus, *Ignavibacterium album* strain JCM 16511 had identified the organismal group as a

445 strict anaerobe only capable of growth on fermentation; however the enrichment of these organisms,
446 albeit modest, in the fed-batch cultures provided only with a non-fermentable electron donor
447 suggested the possibility that these understudied organisms may be capable of growth on N_2O
448 respiration (ref. 55, 56). Presumably due to the paucity of sequence information used to design the
449 degenerate primers/probes, the NosZG4 primers/probe set designed to target the *nosZ* affiliated to
450 *Ignavibacterium* spp. was not successful in detecting these *nosZ* sequences. The development of
451 quantitative molecular tools targeting this group may need to wait until more isolates and sequenced
452 genomes are available. Nevertheless, the future investigations of this unique non-phototrophic
453 member of the *Chlorobi* phylum may bring about novel insights to unique N_2O reduction physiology.
454
455 In summary, the group-specific *nosZ* qPCR methods developed in this study substantially improved
456 the accuracy in estimating the population of *nosZ*-possessing organisms, especially those harboring
457 the novel clade II *nosZ*. Cross-checking with the *nosZ* sequences extracted from the shotgun
458 metagenome verified that the NosZG1-5 primers/probe sets are comprehensive within the groups and
459 mutually exclusive, and thus are definite improvement over the previously available clade I and clade
460 II *nosZ* primer sets in this respect. Any methods for quantification of functional genes are prone to
461 error, and the *nosZ* qPCR methods developed here also have certain deficiencies, including inability
462 to amplify *nosZ* genes affiliated to relatively abundant *Iganvibacterium* spp. In designing the group-
463 specific primers, a trade-off between coverage and mutual exclusivity was inevitable, which could
464 have been the cause of modest discrepancy with the metagenome data. Nevertheless, despite these
465 drawbacks, there is little doubt that the qPCR assays developed in this study are the most reliable
466 tools for real-time quantification of the environmentally significant *nosZ* genes available to date.
467 These tools are expected to be immensely useful not only for estimation of overall N_2O-reducing
468 capability of a microbiome, but also for identification of the key N_2O-reducing population in different
469 soil and aquatic environments.
470
471 **Acknowledgements**

This work was financially supported by "the R&D Center for reduction of Non-CO2 Greenhouse Gases (Grant No. 2017002420002)" funded by Korea Ministry of Environment (MOE).

Conflict of interest

The authors declare that they have no competing interests.

References

1. Ciais P, Sabine C, Bala G, Bopp L, Brovkin V, Canadell J *et al*. Carbon and other biogeochemical cycles. *Climate change 2013: the physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press: Cambridge, UK, 2014, pp 465-570.
2. Ravishankara A, Daniel JS, Portmann RW. Nitrous oxide (N_2O): the dominant ozone-depleting substance emitted in the 21st century. *Science* 2009; **326:** 123-125.
3. Portmann R, Daniel J, Ravishankara A. Stratospheric ozone depletion due to nitrous oxide: influences of other gases. *Phil Trans Royal Soc B.* 2012; **367:** 1256-1264.
4. Yoon S, Song B, Phillips RL, Chang J, Song MJ. Ecological and physiological implications of nitrogen oxide reduction pathways on greenhouse gas emissions in agroecosystems. *FEMS Microbiol Ecol.* 2019; **95:** fiz066.
5. Law Y, Ye L, Pan Y, Yuan Z. Nitrous oxide emissions from wastewater treatment processes. *Phil Trans Royal Soc B.* 2012; **367:** 1265-1277.
6. Thomson AJ, Giannopoulos G, Pretty J, Baggs EM, Richardson DJ. Biological sources and sinks of nitrous oxide and strategies to mitigate emissions. *Philos Trans Royal Soc B.* 2012; **367:** 1157-1168.
7. Frutos OD, Quijano G, Aizpuru A, Muñoz R. A state-of-the-art review on nitrous oxide control from waste treatment and industrial sources. *Biotechnol Adv.* 2018; **36:** 1025-1037.
8. Hallin S, Philippot L, Löffler FE, Sanford RA, Jones CM. Genomics and ecology of novel N_2O-reducing microorganisms. *Trends Microbiol.* 2017; **26:** 43-55.

9. Zumft WG. Cell biology and molecular basis of denitrification. *Microbiol Mol Biol Rev.* 1997; **61:** 533-616.
10. Sanford RA, Wagner DD, Wu Q, Chee-Sanford JC, Thomas SH, Cruz-García C *et al.* Unexpected nondenitrifier nitrous oxide reductase gene diversity and abundance in soils. *Proc Natl Acad Sci.* 2012; **109:** 19709-19714.
11. Jones CM, Spor A, Brennan FP, Breuil M-C, Bru D, Lemanceau P *et al*. Recently identified microbial guild mediates soil N_2O sink capacity. *Nat Clim Change.* 2014; **4:** 801-805.
12. Yoon S, Nissen S, Park D, Sanford RA, Löffler FE. Nitrous oxide reduction kinetics distinguish bacteria harboring clade I versus clade II NosZ. *Appl Environ Microbiol.* 2016; **82:** 3793-3800.
13. Suenaga T, Riya S, Hosomi M, Terada A. Biokinetic characterization and activities of N_2O-reducing bacteria in response to various oxygen levels. *Front Mirobiol.* 2018; **9:** 697.
14. Domeignoz-Horta LA, Spor A, Bru D, Breuil m-c, Bizouard F, Leonard J *et al*. The diversity of the N_2O reducers matters for the N_2O:N_2 denitrification end-product ratio across an annual and a perennial cropping system. *Front Microbiol* 2015; **6:** 971.
15. Yoon H, Song MJ, Yoon S. Design and feasibility analysis of a self-sustaining biofiltration system for removal of low concentration N_2O emitted from wastewater treatment plants. *Environ Sci Technol.* 2017; **51:** 10736-10745.
16. Yoon H, Song MJ, Kim DD, Sabba F, Yoon S. A serial biofiltration system for effective removal of low-concentration nitrous oxide in oxic gas streams: mathematical modeling of reactor performance and experimental validation. *Environ Sci Technol.* 2019; **53:** 2063-2074.
17. Morales SE, Cosart T, Holben WE. Bacterial gene abundances as indicators of greenhouse gas emission in soils. *ISME J.* 2010; **4:** 799.
18. Henry S, Bru D, Stres B, Hallet S, Philippot L. Quantitative detection of the *nosZ* gene, encoding nitrous oxide reductase, and comparison of the abundances of 16S rRNA, *narG*, *nirK*, and *nosZ* genes in soils. *Appl Environ Microbiol.* 2006; **72:** 5181-5189.

19. Conthe M, Wittorf L, Kuenen JG, Kleerebezem R, van Loosdrecht MCM, Hallin S. Life on N_2O: deciphering the ecophysiology of N_2O respiring bacterial communities in a continuous culture. *ISME J.* 2018; **12:** 1142-1153.

20. Juhanson J, Hallin S, Söderström M, Stenberg M, Jones CM. Spatial and phyloecological analyses of *nosZ* genes underscore niche differentiation amongst terrestrial N_2O reducing communities. *Soil Biol Biochem.* 2017; **115:** 82-91.

21. Domeignoz-Horta LA, Philippot L, Peyrard C, Bru D, Breuil M-C, Bizouard F *et al*. Peaks of in situ N_2O emissions are influenced by N_2O-producing and reducing microbial communities across arable soils. *Glob Change Biol.* 2018; **24:** 360-370.

22. Jones CM, Graf DR, Bru D, Philippot L, Hallin S. The unaccounted yet abundant nitrous oxide-reducing microbial community: a potential nitrous oxide sink *ISME J.* 2013; **7:** 417.

23. Ma Y, Zilles JL, Kent AD. An evaluation of primers for detecting denitrifiers via their functional genes. *Environ Microbiol.* 2019; **21:** 1196-1210.

24. Hatt JK, Löffler FE. Quantitative real-time PCR (qPCR) detection chemistries affect enumeration of the *Dehalococcoides* 16S rRNA gene in groundwater. *J Microbiol Methods.* 2012; **88:** 263-270.

25. Graf DRH, Zhao M, Jones CM, Hallin S. Soil type overrides plant effect on genetic and enzymatic N_2O production potential in arable soils. *Soil Biol Biochem* 2016; **100:** 125-128.

26. Paranychianakis NV, Tsiknia M, Kalogerakis N. Pathways regulating the removal of nitrogen in planted and unplanted subsurface flow constructed wetlands. *Water Res.* 2016; **102:** 321-329.

27. Waller LJ, Evanylo GK, Krometis L-AH, Strickland MS, Wynn-Thompson T, Badgley BD. Engineered and environmental controls of microbial denitrification in established bioretention cells. *Environ Sci Technol.* 2018; **52:** 5358-5366.

28. Di HJ, Cameron KC, Podolyan A, Robinson A. Effect of soil moisture status and a nitrification inhibitor, dicyandiamide, on ammonia oxidizer and denitrifier growth and nitrous oxide emissions in a grassland soil. *Soil Biol Biochem* 2014; **73:** 59-68.

552 29. Stoliker DL, Repert DA, Smith RL, Song B, LeBlanc DR, McCobb TD *et al.* Hydrologic
553 controls on nitrogen cycling processes and functional gene abundance in sediments of a
554 groundwater flow-through lake. *Environ Sci Technol.* 2016; **50:** 3649-3657.
555 30. Orellana LH, Konstantinidis KT, Rodriguez-R LM. ROCker: accurate detection and
556 quantification of target genes in short-read metagenomic data sets by modeling sliding-
557 window bitscores. *Nucleic Acids Res.* 2016; **45:** e14-e14.
558 31. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G *et al.* Structure and
559 function of the global ocean microbiome. *Science* 2015; **348:** 1261359.
560 32. Li L, Yin X, Jiang X-T, Zhang T, Chai B, Cole JR *et al.* ARGs-OAP v2.0 with an expanded
561 SARG database and Hidden Markov Models for enhancement characterization and
562 quantification of antibiotic resistance genes in environmental metagenomes. *Bioinformatics*
563 2018; **34:** 2263-2270.
564 33. Orellana LH, Rodriguez-R LM, Higgins S, Chee-Sanford JC, Sanford RA, Ritalahti KM *et al.*
565 Detecting nitrous oxide reductase (*nosZ*) genes in soil metagenomes: method development
566 and implications for the nitrogen cycle. *MBio* 2014; **5:** e01193-14.
567 34. Orellana LH, Chee-Sanford JC, Sanford RA, Löffler FE, Konstantinidis KT. Year-round
568 shotgun metagenomes reveal stable microbial communities in agricultural soils and novel
569 ammonia oxidizers responding to fertilization. *Appl. Environ Microbiol.* 2017; **84:** e01646-17.
570 35. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from
571 sampling to analysis. *Nat Biotechnol.* 2017; **35:** 833.
572 36. Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M. How good is a PCR efficiency
573 estimate: recommendations for precise and robust qPCR efficiency assessments. *Biomol*
574 *Detect Quantif.* 2015; **3:** 9-16.
575 37. Yoon S, Sanford RA, Löffler FE. *Shewanella* spp. use acetate as an electron donor for
576 denitrification but not ferric iron or fumarate reduction. *Appl Environ Microbiol.* 2013; **79:**
577 2818-2822.
578 38. Park D, Kim H, Yoon S. Nitrous oxide reduction by an obligate aerobic bacterium
579 *Gemmatimonas aurantiaca* strain T-27. *Appl Environ Microbiol.* 2017; **83:** e00502-17.

39. Ritalahti KM, Amos BK, Sung Y, Wu Q, Koenigsberg SS, Löffler FE. Quantitative PCR targeting 16S rRNA and reductive dehalogenase genes simultaneously monitors multiple *Dehalococcoides* strains. *Appl Environ Microbiol* 2006; **72:** 2765-2774.
40. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; **32:** 1792-1797.
41. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for bigger datasets. *Mol Biol Evol* 2016; **33:** 1870-1874.
42. Hysom DA, Naraghi-Arani P, Elsheikh M, Carrillo AC, Williams PL, Gardner SN. Skip the alignment: degenerate, multiplex primer and probe design using K-mer matching instead of alignments. *PloS One* 2012; **7:** e34560.
43. Gardner SN, Slezak TJBB. Simulate_PCR for amplicon prediction and annotation from multiplex, degenerate primers and probes. *BMC Bioinformatics* 2014; **15:** 237.
44. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006; **22:** 1658-1659.
45. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014; **30:** 2114-2120.
46. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* 2017; **27:** 824-834.
47. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* 2009; **25:** 2078-2079.
48. Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol.* 2011; **12:** R44.
49. Huang Y, Gilna P, Li W. Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* 2009; **25:** 1338-1340.
50. Pjevac P, Schauberger C, Poghosyan L, Herbold CW, van Kessel MAHJ, Daebeler A *et al.* *AmoA*-targeted polymerase chain reaction primers for the specific detection and quantification of comammox *Nitrospira* in the environment. *Front Microbiol.* 2017; **8:** 1508.

608 51. Meinhardt KA, Bertagnolli A, Pannu MW, Strand SE, Brown SL, Stahl DA. Evaluation of
609 revised polymerase chain reaction primers for more inclusive quantification of ammonia-
610 oxidizing archaea and bacteria. *Environ Microbiol Rep*. 2015; **7:** 354-363.
611 52. Chen X, Peltier E, Sturm BSM, Young CB. Nitrogen removal and nitrifying and denitrifying
612 bacteria quantification in a stormwater bioretention system. *Water Res* 2013; **47:** 1691-1700.
613 53. Cassman KG, Dobermann A, Walters DT. Agroecosystems, nitrogen-use efficiency, and
614 nitrogen management. *Ambio*. 2002; **31:** 132-140, 139.
615 54. Aulakh MS, Khera TS, Doran JW. Mineralization and denitrification in upland, nearly
616 saturated and flooded subtropical soil I. Effect of nitrate and ammoniacal nitrogen. *Biol Fert*
617 *Soils*. 2000; **31:** 162-167.
618 55. Iino T, Mori K, Uchino Y, Nakagawa T, Harayama S, Suzuki K-I. *Ignavibacterium album*
619 gen. nov., sp. nov., a moderately thermophilic anaerobic bacterium isolated from microbial
620 mats at a terrestrial hot spring and proposal of *Ignavibacteria* classis nov., for a novel lineage
621 at the periphery of green sulfur bacteria. *Int J Syst Evol Microbiol.* 2010; **60:** 1376-1382.
622 56. Liu Z, Frigaard N-U, Vogl K, Iino T, Ohkuma M, Overmann J *et al*. Complete genome of
623 *Ignavibacterium album*, a metabolically versatile, flagellated, facultative anaerobe from the
624 phylum *Chlorobi*. *Front Microbiol*. 2012; **3:** 185.

625

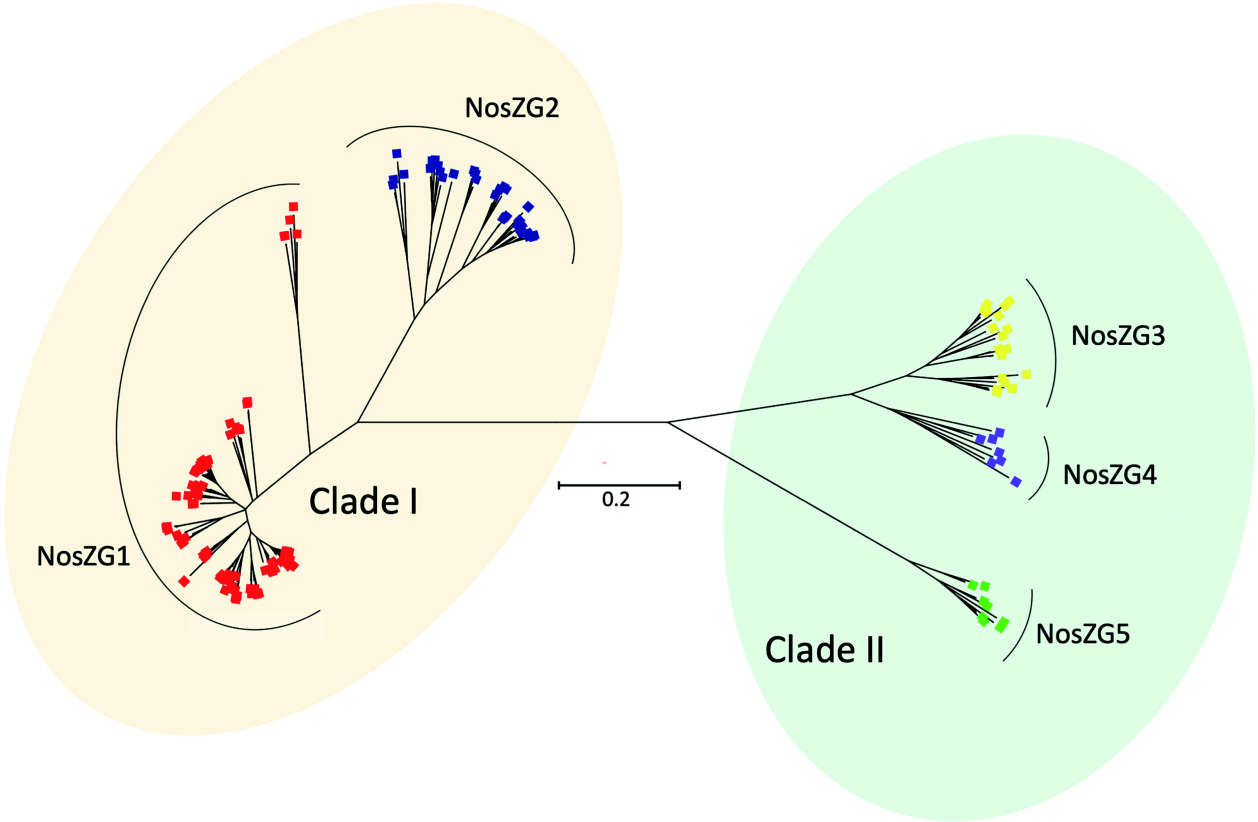
626 **Figure legend**

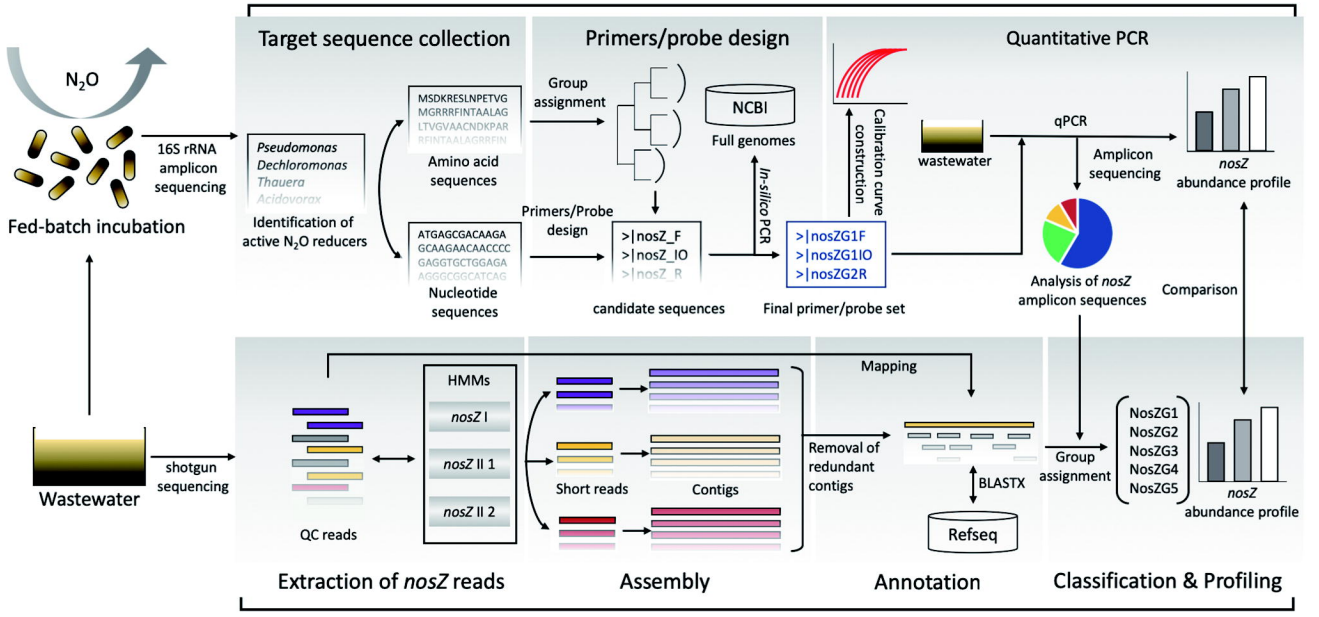
627 **Figure 1.** The phylogenetic tree of 174 *nosZ* sequences belonging to 14 genera enriched after fed-
628 batch incubation with N_2O. The *nosZ* sequences were imported from full, draft, and metagenome-
629 assembled genome sequences imported from the NCBI database and the tree was built using
630 neighbor-joining algorithm with 500 bootstrap replicates

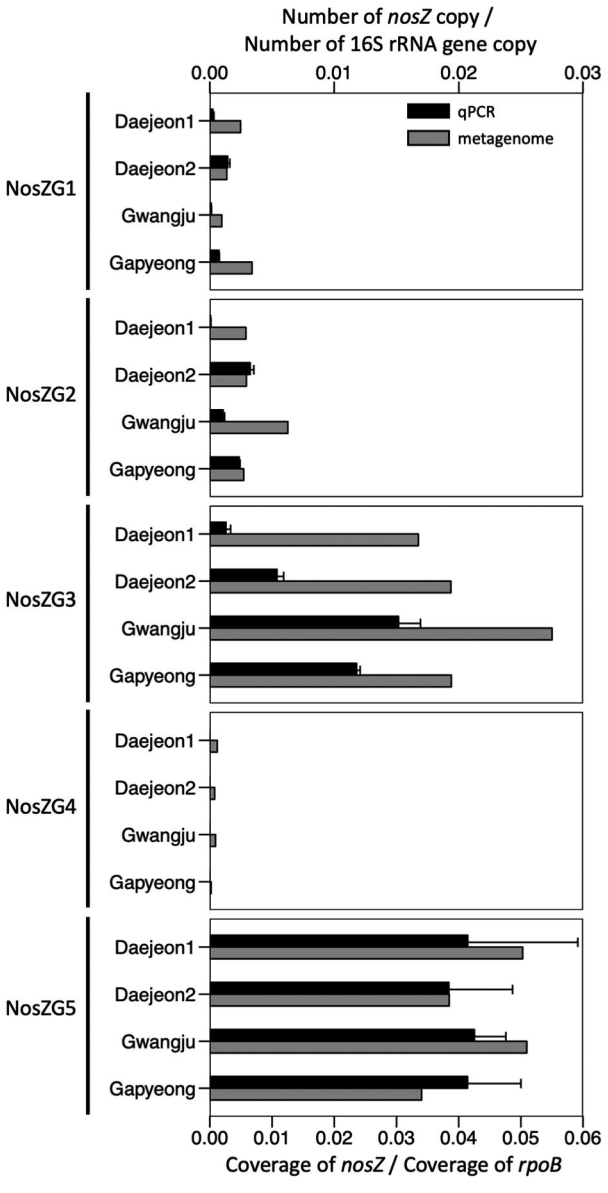
631 **Figure 2.** Flow chart for designing of the group-specific *nosZ* primer/probe sets and cross-validation
632 with *nosZ* sequence data extracted from shotgun metagenome data

633 **Figure 3.** The relative abundances of NosZG1-5 *nosZ* genes as quantified using qPCR and
634 metagenome analysis. The *nosZ* copy numbers from the qPCR assays were normalized with the copy
635 numbers of eubacterial 16S rRNA genes. The coverages of *nosZ* genes in the metagenome data were
636 normalized with the coverages of *rpoB* genes

637 **Figure 4.** The phylogenetic tree constructed with the OTUs generated from the partial *nosZ* sequences
638 amplified with NosZG1-5 primers. The phylogenetic tree was generated using neighbor-joining
639 method with 500 bootstrap replications. The heatmap shows the relative abundances of the OTUs
640 within each *nosZ* group, computed from the amplicon sequencing data and the *nosZ* sequence data
641 extracted from the metagenome


NosZG2
NosZG1
Clade I
0.2
NosZG3
NosZG4
NosZG5
Clade II


N_2O
16S rRNA amplicon sequencing
Fed-batch incubation
Wastewater
shotgun sequencing
Target sequence collection
Pseudomonas
Dechloromonas
Thauera
Acidovorax
Identification of active N_2O reducers
MSDKRESLNPETVG
MGRRRFINTAALAG
LTVGVAACNDKPAR
RFINTAALAGRRFIN
Amino acid sequences
ATGAGCGACAAGA
GCAAGAACAACCCC
GAGGTGCTGGAGA
AGGGCGGCATCAG
Nucleotide sequences
Primers/probe design
Group assignment
NCBI
Full genomes
In-silico PCR
>|nosZ_F
>|nosZ_IO
>|nosZ_R
candidate sequences
>|nosZG1F
>|nosZG1IO
>|nosZG2R
Final primer/probe set
Calibration curve construction
Quantitative PCR
wastewater
qPCR
Amplicon sequencing
nosZ abundance profile
Analysis of nosZ amplicon sequences
Comparison
HMMs
nosZ I
nosZ II 1
nosZ II 2
QC reads
Short reads
Contigs
Mapping
Removal of redundant contigs
BLASTX
Refseq
Group assignment
NosZG1
NosZG2
NosZG3
NosZG4
NosZG5
nosZ abundance profile
Extraction of nosZ reads
Assembly
Annotation
Classification & Profiling


Number of nosZ copy /
Number of 16S rRNA gene copy
0.00
0.01
0.02
0.03
qPCR
metagenome
NosZG1
Daejeon1
Daejeon2
Gwangju
Gapyeong
NosZG2
Daejeon1
Daejeon2
Gwangju
Gapyeong
NosZG3
Daejeon1
Daejeon2
Gwangju
Gapyeong
NosZG4
Daejeon1
Daejeon2
Gwangju
Gapyeong
NosZG5
Daejeon1
Daejeon2
Gwangju
Gapyeong
0.00
0.01
0.02
0.03
0.04
0.05
0.06
Coverage of nosZ / Coverage of rpoB

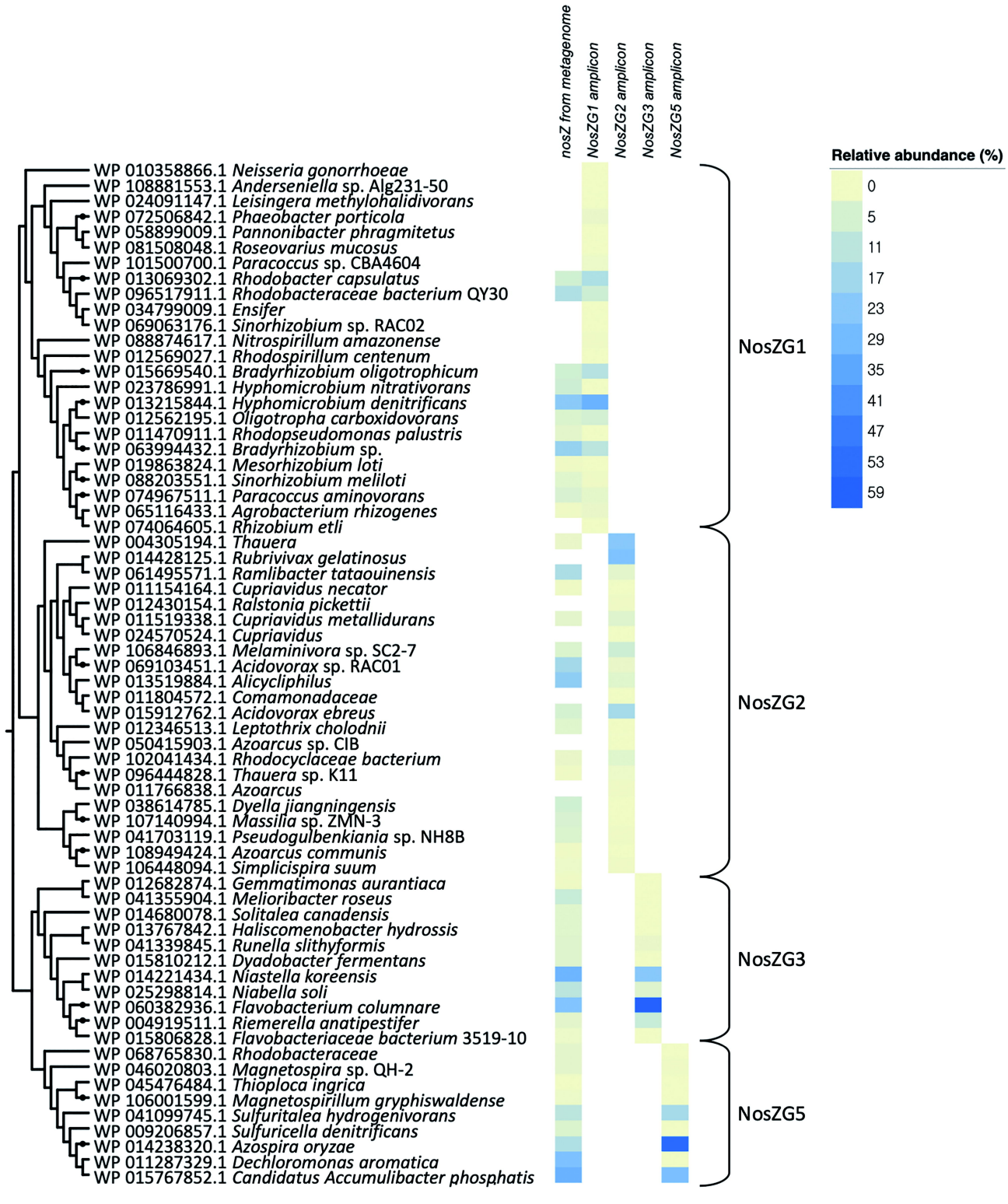

nosZ from metagenome
NosZG1 amplicon
NosZG2 amplicon
NosZG3 amplicon
NosZG5 amplicon
Relative abundance (%)
0
5
11
17
23
29
35
41
47
53
59
WP 010358866.1 Neisseria gonorrhoeae
WP 108881553.1 Anderseniella sp. Alg231-50
WP 024091147.1 Leisingera methylohalidivorans
WP 072506842.1 Phaeobacter porticola
WP 058899009.1 Pannonibacter phragmitetus
WP 081508048.1 Roseovarius mucosus
WP 101500700.1 Paracoccus sp. CBA4604
WP 013069302.1 Rhodobacter capsulatus
WP 096517911.1 Rhodobacteraceae bacterium QY30
WP 034799009.1 Ensifer
WP 069063176.1 Sinorhizobium sp. RAC02
WP 088874617.1 Nitrospirillum amazonense
WP 012569027.1 Rhodospirillum centenum
WP 015669540.1 Bradyrhizobium oligotrophicum
WP 023786991.1 Hyphomicrobium nitrativorans
WP 013215844.1 Hyphomicrobium denitrificans
WP 012562195.1 Oligotropha carboxidovorans
WP 011470911.1 Rhodopseudomonas palustris
WP 063994432.1 Bradyrhizobium sp.
WP 019863824.1 Mesorhizobium loti
WP 088203551.1 Sinorhizobium meliloti
WP 074967511.1 Paracoccus aminovorans
WP 065116433.1 Agrobacterium rhizogenes
WP 074064605.1 Rhizobium etli
WP 004305194.1 Thauera
WP 014428125.1 Rubrivivax gelatinosus
WP 061495571.1 Ramlibacter tataouinensis
WP 011154164.1 Cupriavidus necator
WP 012430154.1 Ralstonia pickettii
WP 011519338.1 Cupriavidus metallidurans
WP 024570524.1 Cupriavidus
WP 106846893.1 Melaminivora sp. SC2-7
WP 069103451.1 Acidovorax sp. RAC01
WP 013519884.1 Alicycliphilus
WP 011804572.1 Comamonadaceae
WP 015912762.1 Acidovorax ebreus
WP 012346513.1 Leptothrix cholodnii
WP 050415903.1 Azoarcus sp. CIB
WP 102041434.1 Rhodocyclaceae bacterium
WP 096444828.1 Thauera sp. K11
WP 011766838.1 Azoarcus
WP 038614785.1 Dyella jiangningensis
WP 107140994.1 Massilia sp. ZMN-3
WP 041703119.1 Pseudogulbenkiania sp. NH8B
WP 108949424.1 Azoarcus communis
WP 106448094.1 Simplicispira suum
WP 012682874.1 Gemmatimonas aurantiaca
WP 041355904.1 Melioribacter roseus
WP 014680078.1 Solitalea canadensis
WP 013767842.1 Haliscomenobacter hydrossis
WP 041339845.1 Runella slithyformis
WP 015810212.1 Dyadobacter fermentans
WP 014221434.1 Niastella koreensis
WP 025298814.1 Niabella soli
WP 060382936.1 Flavobacterium columnare
WP 004919511.1 Riemerella anatipestifer
WP 015806828.1 Flavobacteriaceae bacterium 3519-10
WP 068765830.1 Rhodobacteraceae
WP 046020803.1 Magnetospira sp. QH-2
WP 045476484.1 Thioploca ingrica
WP 106001599.1 Magnetospirillum gryphiswaldense
WP 041099745.1 Sulfuritalea hydrogenivorans
WP 009206857.1 Sulfuricella denitrificans
WP 014238320.1 Azospira oryzae
WP 011287329.1 Dechloromonas aromatica
WP 015767852.1 Candidatus Accumulibacter phosphatis
NosZG1
NosZG2
NosZG3
NosZG5

Table 1. The primers/probe sets developed in this study for group-specific quantification of *nosZ* genes

Clade	Group	Primers / Probe	Sequence (5' → 3')	Degeneracy	Coverage	Slope	Y-intercept	Efficiency (%)
Clade I	NosZG1	NosZG1F	AAG GTN CGB GTN TAC ATG	48	70% (77/110)	-3.28	36.84	91.6
		NosZG1R	CSN NCA TYT CCA TGT GCA	64				
		NosZG1P	FAM-ACT GCM VBT GGT TCT GCC AYG C-MGBNFQ	36				
	NosZG2	NosZG2F	GRC ATC WKC MMC GAC AAG	32	84.4% (27/32)	-3.49	38.22	93.4
		NosZG2R	HYC TCG RYG TTG TAC TGG	24				
		NosZG2P	FAM-ACC ACS CGC GTG TTC TGC G-MGBNFQ	2				
Clade II	NosZG3	NosZG3F	CAY TTT GCW CCD GAY AAT ATT GAA	24	95% (19/22)	-3.56	40.04	90.7
		NosZG3R	BSH WGT TTC ACC BGG CAT	108				
		NosZG3P	FAM-AAY YTR GAA CAA GAY TGG GAT GTA CCK C-MGBNFQ	32				
	NosZG4	NosZG4F	ATW GTT GCY GGH GGM AAA	24	57.1% (4/7)	-3.41	38.39	94.9
		NosZG4R	TTC CCA DGT KCC DAW TTT	36				
		NosZG4P	FAM-MGG YGA AGT KSA GAA TCC GGG WT-MGBNFQ	32				
	NosZG5	NosZG5F	AAC GAC AAG KCS AAY CCG	8	100% (5/5)	-3.36	38.87	97.7
		NosZG5R	GCG GTC GAA CTT CCA GTA	0				
		NosZG5P	FAM-GCS GTG MTC GAY CTG CGB G-MGBNFQ	24				