

# Analysis of 16S rRNA gene variability in soil nitrifying bacteria of the genus *Nitrosomonas*

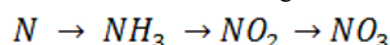
Anastasia Nechayeva\*, Konstantin Boyarshin, Olga Bespalova, Violetta Klyueva, Olesya Makanina, and Irina Batlutskaya

Belgorod State National Research University, Belgorog, Russia

**Abstract.** The main goal of the work was to assess variability of 16S rRNA gene sequence within the nitrifying bacterial genus *Nitrosomonas* to find specific sequences for its detection. To achieve it, we had to find and to assess sequences that are highly conservative on the level of the genus and to find and to assess sequences variable on the level of genus but conserved on the level of species. In the SILVA database of ribosomal RNA sequences, 231 sequences of 16S rRNAs of bacteria of the genus *Nitrosomonas* were collected, of which were sorted 132 sequences by length from 1400 to 1541 (full-sized gene) nucleotides. We conducted an analysis of the taxon-specificity of sequences conserved at the genus level. More than a hundred full matches were found by the BLAST program in the nr database with other genera of the same and other families. So, in *Nitrosomonas* 16S rRNA gene are present some highly conservative regions, but they are not genus-specific due to high coincidence with other genera. Wherein, a variable region 994-1041 is highly species-specific for the species *N. eutropha*. Generally, the sequence of 994-1041 region of *Nitrosomonas* 16S rRNA genes tends to be clustered, being very close between some species.

## 1 Introduction

Nitrification is an important step in the nitrogen cycle in the nature. The nitrification process as a whole consists in the oxidation of organic nitrogen to nitric acid residue. It is carried out in three phases: in the first phase, ammonia ( $\text{NH}_4^+$ ) is formed under the influence of various bacteria, in the second phase, ammonia is oxidized by bacteria of the genus *Nitrosomonas* etc. to nitrous acid residue ( $\text{NO}_2^-$ ), and finally, in the third phase, it is oxidized by the *Nitrobacter* genus etc. to nitric acid residue ( $\text{NO}_3^-$ ) [1]. The general scheme of these transformations has the following form:



Nitrifying organisms are chemoautotrophs, and use carbon dioxide as their carbon source for growth. *Nitrosomonas europaea*, as well as populations of soil-dwelling ammonia-oxidizing bacteria, have been shown to assimilate the carbon dioxide released by the reaction to form biomass via the Calvin Cycle, and harvest energy by oxidizing ammonia to nitrite [2, 3]. Many of such microorganisms also are capable of oxidizing urea and this feature may explain enhanced growth of ammonia-oxidizing bacteria in the presence of urea in acidic environments [4].

\*Corresponding author: [nechayeva@list.ru](mailto:nechayeva@list.ru)

Bacteria of the genus *Nitrosomonas* play a key role in the conversion of ammonium soil nitrogen to nitrite, providing the use of mineral and organic fertilizers by agricultural plants. This genus contains 12 species.

In this study, sequence variability of the 16S rRNA gene was evaluated within the nitrifying bacterial genus *Nitrosomonas* to find specific sequences for its detection, sequences that are highly conserved at the genus level, and sequences that vary at the genus level but remain conserved at the species level were found and evaluated. Analysis of taxon-specificity of sequences that are kept conserved at the genus level was carried out too.

## 2 Problem statement

The presence of nitrate nitrogen in the zone of the root system is of great importance for plants. The nitrifying ability of the soil microbiota reflects its potential in the accumulation of nitrogen usable for plants.

In soils with an increased intensity of nitrification processes, a significant amount of nitrates accumulates, thereby creating conditions favorable for denitrification as well [5]. Knowledge of the dynamics of the growth of microorganisms and the intensity of the passage of ammonification and nitrification will allow us to regulate these processes and establish the doses, dates and types of nitrogen fertilizers depending on the soil conditions.

This determines the use of the introduced fertilizers and the reduction of losses during their application.

### 3 Research question

The main question of this study is what DNA sequences can serve as molecular markers to detect the abandons of microorganisms of genus *Nitrosomonas* in soil microbiota. On this path, we sought to evaluate the sequence variability of the 16S rRNA gene within the nitrifying bacterial genus *Nitrosomonas* in order to find specific sequences for its detection. To achieve this goal, it was necessary to select a set of sequences that might serve for binding of PCR primers specific for this genus. These should be some evolutionary conservative sequences that are the same for the main representatives of the *Nitrosomonas* genus. We also needed to find and evaluate sequences that vary at the genus level, but persist at the species level.

### 4 Purpose of the study

The aim of this study was to assess variability of 16S rRNA gene sequence within the nitrifying bacterial genus *Nitrosomonas* to find specific sequences for its detection.

### 5 Methods and materials

In the SILVA database of ribosomal RNA sequences, 231 sequences of 16S rRNA of bacteria of the genus *Nitrosomonas* were collected. Of these, using the online interface, they were sorted by the length extracting 132 sequences with length from 1,400 to 1,541 (full-sized gene) nucleotides. The obtained data set contained from one to twelve sequences belonging to each of the twelve species of the genus *Nitrosomonas*, as well as 63 sequences that were identified on the genus level only.

132 sequences of 12 species and unclassified bacteria of the genus *Nitrosomonas* were aligned by the Clustal Omega program. This way three conservative sequences more than 20 nucleotides in length have been identified.

On the table I traditionally distinguished highly conserved sequences of bacterial (E) and archaeal (A) rRNAs are shown (table I, [6]). These sequences should be too conserved to be genus-specific. But they are not the same even within the same genus, as we show, aligning 132 *Nitrosomonas* – as we have only three of all these sequences which are truly conservative, and only one of them is completely conservative (the second).

The conservative sequences found at the level of the genus *Nitrosomonas* overlap to some extent with the universal conservative sequences traditionally distinguished in special literature [6] (tab. 2).

As known, group-specific PCR is PCR for related sequences within the same or between different species using conservative primer-binding sequences. It means that the selection of universal primers for 16S ribosomal genes for amplification of a species-specific part of the gene implies that the sequence of 16S gene is conserved

between species, therefore, amplification of the sequence will be performed for all studied species.

As usual, several steps are used to construct group-specific primers, including collecting data on the sequences of the selected gene from several adjacent taxa, aligning the nucleotide sequences and analyzing them, identifying regions with the optimal level of variability, selecting primer sequences, and checking that the sequences match the molecules in database. The most time-consuming steps: comparing the aligned areas of closely related groups aimed at detecting unique group features, and searching for randomly matching sequences.

**Table 1.** The conservative fragments in archael and eubacterial 16s rDNAs [6].

Bacteria	Position	Conservative fragment
E	104–120	GGCGVACGGGTGAGTAA
E	314–368	CAYTGGRACCTGAGACACGGYCCARACTCTACGGG AGGCAGCAGTRRGAATHHT
A	344–367	AYGGGGYGCAGCAGCGRGAARAC
E	505–539	GGCTAACTHCGTGCCAGCAGCCCGGTAATACGDA
A	506–547	GGYAAGDCYGGTGYCAGCCGCCGGTAHAHACCRC CDRTGGCGAA
E	683–707	GTGTAGRGGTGAATKCGYAGAKAT
E	764–806	CGAAAGYGTGGGKAKRCAGGATTAGATACCTCGGT AGTCC
A	779–806	CRAACSGGATTAGATACCCSGGTAGTCC
E	879–893	CCRCTGGGGAGTAC
A	882–936	CCTGGGRAGTACGKHCCAAGDRTGAAACTTAAAGG AATTGGCGGGGAGCAC
→ E	909–940	ACTCAAAGAATTGACGGGGGCCGCCACAAGC
A	947–973	GCSTGCGGYTAAATGGABTCAACGCC
E	949–964	ATGTGGTTAATTCGA
E	969–985	ACGCGARGAACCTTACC
A	1043–1073	GAGAGWGWTGCATGGCCGYCYCAGYTCGT
→ E	1048–1114	GTGSTGCATGGYGTCTGTCAGCTCGTYGCTGAGRT GTYGGGTTAAGTCCRYAACGAGCGCAACCC
A	1094–1111	GTCAGRYAACGARGCAGA
→ E	1177–1197	GGAAGGYGGGAYGACGTCAA
A	1225–1242	ACACGCGSGCTRCAAWGG

That is why, for a more simplified obtaining of sequences with degenerate positions, we created a special program based on a high-level general-purpose programming language, aimed at increasing developer productivity and code readability “Python 3.7”.

The script package written in “Python 3.7” is developed to facilitate these steps. It includes three main tools. Two of them are destined to build consensus sequences on the base of alignments generated by well-known “Clustal Omega” tool [7]. Unlike other common programs for building consensus sequences the scripts of the package can insert in them characters marking variable positions and indicating nucleotides that can occupy them. Also, it is possible to represent consensus in form of six lines with percentage of occurrence of every nucleotide or deletion.

To facilitate analysis of data obtained by “Nucleotide BLAST” package, the tool for comparison of two files containing results of searches in nucleotide sequence database with “BLAST” was developed. It allows fast check whether two primers in the designed pair have coinciding sequences in the same molecule that may lead to false positive results. This greatly simplifies the analysis of possible problems with the specificity of the selected primer pair. In general, the use of the described scripts demonstrates their positive role in solving problems related to the selection of species-specific and group-specific primers.

## 6 Results

As a result, three regions with a length of more than twenty nucleotides were identified, identical for 95 percent or more of the representatives of the genus (Table II). A comparison of the observed sites conservative at the genus level with a map of rRNA sites conservative at high taxonomic levels showed their incomplete overlap. Conservative plot 1169-1255 to the least extent overlaps with the traditionally allocated highly conserved sequences (28% overlap).

We also performed an analysis of the taxon-specificity of sequences stored at the genus level. More than a hundred full matches were found by the BLAST program in the nr database with other genera of the same and other families. The conserved sequences found were not specific for the genus.

Highly variable sequences are located in regions 1-104, 994-1041 and 1431-1541. For further analysis, we chose the short sequence 994-1041 in the middle of the molecule. To evaluate the intraspecific and interspecific variability of this region with the help of a computer program, we created sequences from sets of sequences for generating consensus. After making sure that 16S rRNA sequences that are highly conserved at the genus level are not genus-specific, we shifted our attention to highly variable sequences and traced their variability at the species level.

Mobilis	UUGACAUGCNUGG-AAUUUUACGGAGACAU--AAAAGUGCCCGAAAGGGAACCAAGACAC	1028
Stercoris	UUGACAUGCNUGG-AAUCUAAUGGAGACAU--AAGAGUGCCCGAAAGGAGCCAAGACAC	1016
Halophila	UUGACAUGCNUGG-AAUCCUGUGGAGACAU--GGGAGUGCCCGAAAGGAGCCAAGACAC	1044
Europaea	UUGACAUGCNUGG-AAUCUAAUGGAGACAU--AAGAGURCCCGAAAGGAGCCAAGACAC	1022
Eutropha	UUGACAUGCNUGG-AAUCUAAUGGAGACAU--AAGAGUGCCCGMAAGGGAGCCARGACAC	1034
Nitrosa	UUGACAUGCUCGAAA---UUNNGCAGAGAUGCAGAAAGUGCCCGAAAGGGAUUCGAGACAC	1035
Communis	UUGACAUGCUCGAAA---UCCACAGAGAUGUGGAAGUGCCCGAAAGGGAUUCGAGGCAC	1016
Ureae	UUGACAUGUYA-GAAAAWUUC--AGAGAUGAAUUUGUGUCGAAAGAGAUCURAACAC	1016
Oligotropha	UUGACAUGUYA-GAAAAUUUC--AGAGAUGRAUUUGUGCCCGAAAGGGAUUCUGAACAC	1008
Cryotolerans	UUGACAUGU-AUCUAAAGCUCGC--UGAGAGGUGAGUGUGCCCGAAAGGGAUUAACAC	1035
Aestuarii	UUGACAUGUAGCGAAUA-UUUU--AGAGAUAAAUAUGUGCCUU--CGGGAACCCUAACAC	1013
Marina	UUGACAUSKWRSSRAWUUUKSU--AGAGAUASMUUAGKGCUCUU--SGGRAMSSYWRMMM	1028
	***** * * * * *	

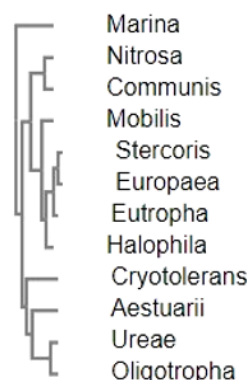
**Fig. 1.** Alignment of consensus sequences of positions 994-1041 of 16S rRNA gene of all twelve *Nitrosomonas* species, built by Clustal Omega program [7].

**Table 2.** Partial overlay of the sequences in 16S rRNA gene, conserved on the genus level, with known highly conservative sequences (from table 1).

Positions	Length, n.	% of identical sequences within the genus	% overlay with known highly conservative sequences	Conserved sequence
896-919	24	97	46	UCCGAAGAUUAAAACUCAAGGAA
1043-1115	74	98	91	CACAGGUGCUGCAUUGGUGUCGUA GCUCGUGUCGUGAGAUUUGGGUUA AGUCCCGCAACGAGCGCAACCCUU
1169-1255	74	95	28	AAACCGGAGGAAGGUGGGGAUGACG UCAAGUCCUAUGGCCUUUAGGGU AGGGUUCACACGUAUACAAGG

*Nitrosomonas* species are divided to some clusters by 994-1041 sequence. The alignment of all these twelve consensus sequences depicts high variability (Fig. 1). Based on these data, a tree of consensus sequences was built (Fig. 2).

As it turned out, the sequence 994-1041 is highly conservative at the species level, for example, for *Nitrosomonas eutropha* (Fig. 3). After aligning the consensus sequences, we got the general one with the help of our special program based on a high-level general-purpose programming language, aimed at increasing developer productivity and code readability “Python 3.7”. Then we decided to check on the basis of nr “BLAST” database whether the created consensus sequence would be detected within the species *Nitrosomonas eutropha*.



**Fig. 2.** Phylogenetic tree of consensus sequences of positions 994-1041 of 16S rRNA gene, built by Clustal Omega program [7].

FNMM01000002.145311.146851EUTROPHA	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
DI358917.4.1508	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
DI166430.4.1505	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
DL479336.4.1505	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
DI246295.4.1505	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
LY417117.118725.120241	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
LZ214934.118725.120241	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
CP000450.231716.233245	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
QICQ01000008.9712.11252	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
AY123795.1.1498	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
FNYP01000008.78106.79646	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
KF618626.1.1425	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
PZZQ01000009.89377.90917	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
FMTW01000008.89377.90917	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57
FPBL01000007.54958.56498	UUGACAUGCUUGGAAUCUAAUGGAGACAUAAAGAGUGCCCGAAAGGGAGCCAGGACAC	57

**Fig. 3.** Alignment of fifteen sequences of region 994-1041 of *N. eutropha* 16S rRNA gene

As a result, only three genera, *Azoarcus*, *Thauera* and *Burgholderia* have the same sequence in this region. So, this region is highly specific for *N. eutropha* and is promising for elaboration of species-specific PCR primers.

## 7 Conclusion

As the result of the work, highly conservative regions of *Nitrosomonas* 16S rRNA gene are present, but are not genus-specific due to high coincidence with other genera.

Conservative sequence sets are often used to generate phylogenetic trees, since it can be assumed that organisms with similar sequences are closely related [3]. The choice of sequences may vary depending on the taxonomic scope of the study [8, 9]. The most highly conserved genes such as 16S ribosomal RNA and other sequences are useful for reconstructing deep phylogenetic relationships and identifying bacterial phyla in metagenomic studies [6].

As is known, the genome of most species of the genus *Nitrosomonas* has not yet been determined [10]. But fortunately, we were able to determine that sequence of 994-1041 region of *Nitrosomonas* 16S rRNA tends to be clustered, being very close for some species. Variable region 994-1041 is highly species-specific at least for species *Nitrosomonas eutropha*. Though the genome *Nitrosomonas eutropha* has not been determined for all strains, the genome of *Nitrosomonas eutropha* C91 C71 has been sequenced and investigated extensively [9, 11].

Although the genomes of some *Nitrosomonas* species have yet to be sequenced, there are many research areas

that have benefited from sequencing the more studied representatives of the *Nitrosomonas* species *N. europaea* and *N. eutropha*, including wastewater treatment, agriculture and biogeochemistry.

## References

1. F. Beeckman, H. Motte, T. Beeckman, Curr. Opin. Biotechnol. **50**, 166 (2018)
2. R. Hatzepichler, Appl Environ Microbiol **78**, 7501 (2012)
3. U. Purkhold, A. Pommerening-Roser, S. Juretschko [et al.], Appl Environ Microbiol **66**, 5368 (2000)
4. K. L. Marsh, G. K. Sims, R. L. Mulvaney, Biol. Fert. Soil **42**, 137 (2005)
5. K. Kazeev, G. Mokrikov, U. Akimenko, M. Myasnikova, S. Kolesnikov, Environmental Assessment of the Use of No-Till Technology in the Rostov Region, Monograph, 2018.
6. Y. Wang, P.-Y. Qian, PLoS ONE **4**, e7401 (2009)
7. J. D. Thompson, T. J. Gibson, F. Plewniak, F. Jeanmougin, D. G. Higgins, Nucleic Acids Res **25**(24), 4876 (1997)
8. S. R. Gadagkar, M. S. Rosenberg, S. Kumar, J. Exp. Zool. Part B **304B**, 64 (2005)
9. L.Y. Stein, D.J. Arp, P.M. Berube [et al.], Environ. Microbiol. **12**, 2993 (2007)
10. P. Chain, J. Lamerdin, F. Larimer, W. Regala [et al.], J. Bacteriol **185**, 2759 (2003)
11. H.-P. Koops, B. Böttcher, U. Möller, A. Pommerening-Röser, G. Stehr, J Gen Microbiol, **13**, 1689(1991)