environmental microbiology

Environmental Microbiology (2014) **16**(10), 3055–3071 doi:10.1111/1462-2920.12300

Erschienen in: Environmental Microbiology ; 16 (2014), 10. - S. 3055-3071

NxrB encoding the beta subunit of nitrite oxidoreductase as functional and phylogenetic marker for nitrite-oxidizing *Nitrospira*

Michael Pester,[1†§] Frank Maixner,[1‡§] David Berry,[1] Thomas Rattei,[2] Hanna Koch,[1] Sebastian Lücker,[1] Boris Nowka,[4] Andreas Richter,[3] Eva Spieck,[4] Elena Lebedeva,[5] Alexander Loy,[1] Michael Wagner[1] and Holger Daims[1*]

Divisions of [1]Microbial Ecology, [2]Computational Systems Biology and [3]Terrestrial Ecosystem Research, Department of Microbiology and Ecosystem Science, Ecology Centre, University of Vienna, Althanstrasse 14, A-1090 Vienna, Austria.

[4]Biozentrum Klein Flottbek, Microbiology and Biotechnology, University of Hamburg, Ohnhorststr. 18,D-22609 Hamburg, Germany.

[5]Winogradsky Institute of Microbiology, Russian Academy of Sciences, Prospekt 60-let Oktyabrya 7/2, Moscow, 117312 Russia.

Summary

Nitrospira **are the most widespread and diverse known nitrite-oxidizing bacteria and key nitrifiers in natural and engineered ecosystems. Nevertheless, their ecophysiology and environmental distribution are understudied because of the recalcitrance of** ***Nitrospira*** **to cultivation and the lack of a molecular functional marker, which would allow the detection of** ***Nitrospira*** **in the environment. Here we introduce** ***nxrB*****, the gene encoding subunit beta of nitrite oxidoreductase, as a functional and phylogenetic marker for** ***Nitrospira*****. Phylogenetic trees based on** ***nxrB*** **of** ***Nitrospira*** **were largely congruent to 16S ribosomal RNA-based phylogenies. By using new** ***nxrB*****-selective polymerase chain reaction primers, we obtained almost full-length** ***nxrB*** **sequences from** ***Nitrospira*** **cultures, two activated sludge samples, and several geographically and climatically distinct soils. Amplicon pyrosequencing of** ***nxrB*** **fragments from 16 soils revealed a previously unrecognized diversity of terrestrial** ***Nitrospira*** **with 1801 detected species-level operational taxonomic units (OTUs) (using an inferred species threshold of 95%** ***nxrB*** **identity). Richness estimates ranged from 10 to 946 coexisting** ***Nitrospira*** **species per soil. Comparison with an archaeal** ***amoA*** **dataset obtained from the same soils [Environ. Microbiol. 14: 525–539 (2012)] uncovered that ammonia-oxidizing archaea and** ***Nitrospira*** **communities were highly correlated across the soil samples, possibly indicating shared habitat preferences or specific biological interactions among members of these nitrifier groups.**

Received 4 July, 2013; accepted 26 September, 2013. *For correspondence. E-mail daims@microbial-ecology.net; Tel. (+43) 1 4277 76604; Fax (+43) 1 4277 876604. Present addresses: [†]Chair for Microbial Ecology, University of Konstanz, Universitätsstrasse 10, D-78457 Konstanz, Germany; [‡]Institute for Mummies and the Iceman, EURAC research, Viale Druso 1, I-39100 Bolzano, Italy. [§]Contributed equally.

Introduction

Nitrification, the microbially catalysed oxidation of ammonia to nitrate, is a key process of the biogeochemical nitrogen cycle in virtually all aerobic ecosystems. Aside from its crucial role in nature, nitrification is essential in biological wastewater treatment for the removal of excess nitrogen (Daims and Wagner, 2010) but causes problems in agriculture by mobilizing nitrogen in fertilized soils (Prosser, 2011). Nitric oxide and nitrous oxide are gaseous by-products of nitrification and belong to the most potent greenhouse gases (Stein, 2011). In-depth biological knowledge of nitrification will thus be required to achieve a sustainable agriculture and reliable sewage treatment and to better assess the impact anthropogenic nitrogen deposition has on the nitrogen cycle (Gruber and Galloway, 2008).

The first step of nitrification (ammonia oxidation to nitrite) is catalysed by chemolithoautotrophic ammonia-oxidizing bacteria (AOB) and archaea (AOA), whereas the second step (nitrite oxidation to nitrate) is carried out by chemolithoautotrophic nitrite-oxidizing bacteria (NOB). Because nitrifiers are generally recalcitrant to cultivation, molecular methods have been the tools of choice to detect and quantify these organisms in most studies on nitrification in natural or engineered systems. Widely used approaches to directly detect nitrifiers are 16S ribosomal RNA (rRNA) sequencing and fluorescence *in situ* hybridization (FISH) with rRNA-targeted oligonucleotide probes

(reviewed by Daims *et al.*, 2009). In addition, the gene coding for the alpha subunit of ammonia monooxygenase (*amoA*), which is present in all known AOB and AOA (Rotthauwe *et al.*, 1997; Purkhold *et al.*, 2000; Pester *et al.*, 2011), has found wide application as a functional marker gene for ammonia oxidizers (e.g. Juretschko *et al.*, 1998; Francis *et al.*, 2005; Leininger *et al.*, 2006; Tourna *et al.*, 2008; Gubry-Rangin *et al.*, 2011; Pester *et al.*, 2012). A functional marker gene, whose evolution was not heavily influenced by lateral gene transfer, is also a useful phylogenetic marker. Such marker genes enable the concomitant detection and identification of uncultured members of a microbial guild based on their placement in phylogenetic trees. This criterion is met by *amoA*, which furthermore provides a higher phylogenetic resolution than 16S rRNA for the differentiation of closely related ammonia oxidizers (Purkhold *et al.*, 2000; Pester *et al.*, 2012).

In sharp contrast to the ammonia oxidizers, the ecology of NOB has surprisingly received little attention in nitrification research. The resulting knowledge gap needs to be filled because nitrite oxidation is the major known biological source of nitrate, a key electron acceptor for anaerobic respiration and one of the most abundant forms of fixed nitrogen in terrestrial and aquatic ecosystems (e.g. Gruber, 2004). The genus *Nitrospira*, which belongs to the distinct phylum *Nitrospirae* (Ehrich *et al.*, 1995), comprises the most diverse and environmentally widespread known NOB that are also the dominant nitrite oxidizers in most wastewater treatment plants (e.g. Watson *et al.*, 1986; Juretschko *et al.*, 1998; Daims *et al.*, 2001; Lebedeva *et al.*, 2005). This genus consists of at least six phylogenetic lineages (Daims *et al.*, 2001; Lebedeva *et al.*, 2008; 2011), whose mainly uncultured members show pronounced divergence at the 16S rRNA level (sequence similarities between lineages are ≤ 94%). Although an encompassing set of *Nitrospira*-specific 16S rRNA-targeted FISH probes exists (Juretschko *et al.*, 1998; Schramm *et al.*, 1998; Daims *et al.*, 2001; Maixner *et al.*, 2006; Foesel *et al.*, 2008), the diversity within this genus hampers the design of *Nitrospira*-specific rRNA-targeted polymerase chain reaction (PCR) primers that would cover all known lineages. Published PCR primers, some based on modifications of FISH probes, target only particular *Nitrospira* lineages and/or do not reliably discriminate against non-target organisms (e.g. Dionisi *et al.*, 2002; Freitag *et al.*, 2005; Maixner *et al.*, 2006). No functional gene has been established yet as marker for detecting and identifying *Nitrospira* or other NOB except *Nitrobacter* (see below). Thus, the knowledge of NOB ecology is very limited partly because of the lack of specific molecular tools for detecting uncultured *Nitrospira* (and other NOB) in environments that are not easily amenable to FISH, like most soils, and for monitoring the metabolic activity of NOB based on the expression of specific functional genes.

The best candidate to become a specific functional marker for NOB is the key enzyme of nitrite oxidation, nitrite oxidoreductase (NXR). This membrane-associated enzyme occurs in two phylogenetically distinct forms, one cytoplasmic type found in the NOB *Nitrobacter*, *Nitrococcus* and *Nitrolancetus* (Sorokin *et al.*, 2012), and one periplasmic type found in *Nitrospira* and *Nitrospina* (Spieck and Bock, 2005; Lücker *et al.*, 2010; 2013) (Fig. 1). Both NXR forms belong to the complex iron–sulphur molybdoenzyme (CISM) family (Rothery *et al.*, 2008) and consist of an alpha (NxrA), beta (NxrB) and gamma (NxrC) subunit (Spieck *et al.*, 1998; Lücker *et al.*, 2010). While NxrA contains the substrate-binding site, the

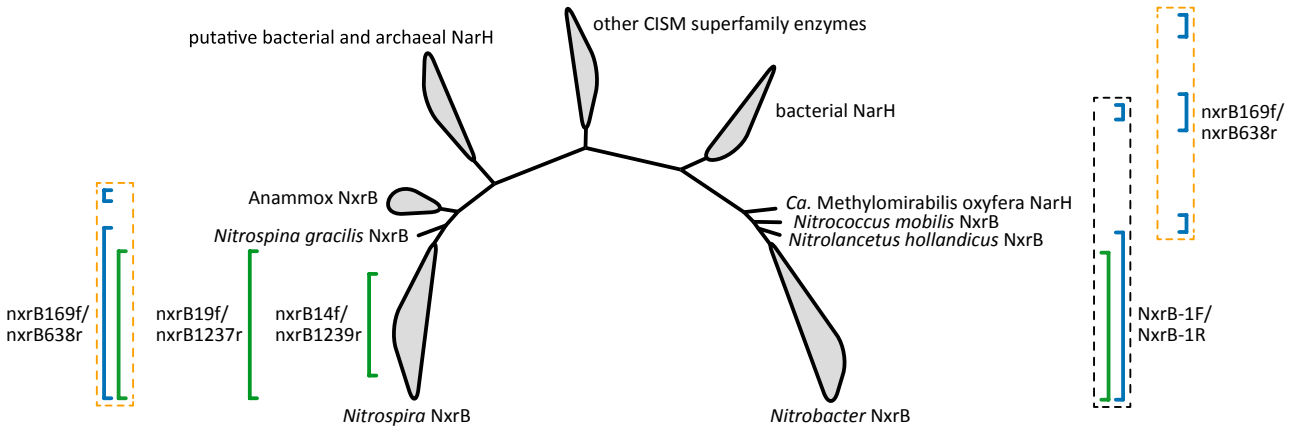


Fig. 1. Schematic illustration, based on NxrB phylogeny, of the known NXR lineages and their affiliation to nitrate reductases (NarH). The coverage and specificity of the *Nitrospira nxrB*-targeted PCR primers designed in this study and of *Nitrobacter nxrB*-targeted primers (NxrB-1F/R) developed previously (Vanparys *et al.*, 2007) are indicated by brackets. Green brackets indicate the reported primer pair specificity. Blue brackets indicate potential coverage of non-target organisms of primer pairs due to missing or weak base mismatches in at least one of the primer target regions of the respective genes. Brackets for the same primer pair are grouped by boxes. For a detailed phylogenetic analysis of NxrB and related proteins, please refer to Lücker and colleagues (2013).

Table 1. Newly designed PCR primers targeting the *nxrB* genes of *Nitrospira*.

Primer	Sequence (5′-3′)	Target site[a]	Amplicon size (bp)	Specificity
nxrB14f	ATA ACT GGC AAC TGG GAC GG	14–33	1245	*Nitrospira* lineages I, II, and V[b]
nxrB1239r	TGT AGA TCG GCT CTT CGA CC	1239–1258		
nxrB19f	TGG CAA CTG GGA CGG AAG ATG	19–39	1239	All *Nitrospira* lineages[b,c]
nxrB1237r	GTA GAT CGG CTC TTC GAC CTG	1237–1257		
nxrB169f	TAC ATG TGG TGG AAC A	169–184	485	All *Nitrospira* lineages
nxrB638r	CGG TTC TGG TCR ATC A	638–653		

a. Positions relative to the *nxrB* (NIDE3256) gene of *N. defluvii* (Lücker *et al*., 2010).
b. Different annealing temperatures were required for *nxrB* amplification from the tested *Nitrospira* cultures and environmental samples.
c. Did not amplify *nxrB* of *Hyrtios proteos* (marine sponge) symbionts.

NxrB and NxrC subunits channel electrons derived from nitrite towards the membrane-bound respiratory chain. In addition, NxrC functions as membrane anchor of the holoenzyme. In pioneering studies, PCR primers targeting the *nxrA* or *nxrB* genes of *Nitrobacter* were developed and tested (Vanparys *et al*., 2007; Poly *et al*., 2008; Wertz *et al*., 2008). Both genes were useful to differentiate closely related *Nitrobacter* strains in pure cultures and soil samples. However, the primers do not cover the *nxr* genes of other NOB such as *Nitrospira* (Wertz *et al*., 2008) (Fig. 1).

Until recently, *nxrA* or *nxrB*-targeted primers for *Nitrospira* could not be developed because of lacking genome sequences from this genus. This situation has improved since one *Nitrospira* genome has been fully sequenced (Lücker *et al*., 2010) and a second draft genome sequence has been obtained (H. Koch, A. Galushko, M. Albertsen, C. Dorninger, A. Schintlmeister, S. Lücker, A. Richter, E. Spieck, P.H. Nielsen, M. Wagner and H. Daims, unpubl. data). In this study, we used these genome sequences to design novel PCR primers that target the *nxrB* gene of *Nitrospira* and demonstrate that they cover all known lineages of this genus. Phylogenetic analyses of nitrospiral *nxrB* demonstrated consistent topologies to 16S rRNA trees, showing that *nxrB* is a useful functional and phylogenetic marker for the most diverse and widespread group of NOB. Moreover, a protocol and analysis pipeline for amplicon pyrosequencing of partial *nxrB* genes of *Nitrospira* were established and used to analyse the *Nitrospira* communities in geographically distant soils.

Results

nxrB is a functional and phylogenetic marker for Nitrospira

The major goal of this study was to establish a functional marker gene assay that targets NOB of the genus *Nitrospira*. *In silico* screening of the three NXR subunits in all published genomes of NOB revealed that for all subunits, paralogs exist in at least some species. As sequence homology of NxrB paralogs within a given strain was highest (Table S1), *nxrB* was selected for primer design and evaluation. Based on the alignment of the two nearly identical *nxrB* copies found in the *Ca*. Nitrospira defluvii genome (Lücker *et al*., 2010) and several partial *nxrB* sequences from the *Nitrospira moscoviensis* draft genome (Koch *et al*., unpubl. data), a first primer pair covering almost the entire *nxrB* gene was designed (nxrB14f/nxrB1239r, Table 1). By using this primer pair, almost full-length *nxrB* genes (1245 of 1289 bp in *Ca*. N. defluvii) were amplified from *Ca*. N. defluvii (*Nitrospira* lineage I), *N. moscoviensis* (lineage II), *Ca*. Nitrospira bockiana (lineage V), activated sludge from two full-scale wastewater treatment plants in Vienna, and Austrian beech and primeval forest soils. No amplicons were obtained from biomass of the other cultured *Nitrospira* [*Nitrospira marina* and *Nitrospira* strain Ecomares 2.1 (both lineage IV) and *Nitrospira calida* (lineage VI)]. To achieve a broader coverage of the genus *Nitrospira*, a second primer pair (nxrB19f/nxrB1237r, Table 1) was designed based on the extended dataset, which included the new *nxrB* sequences obtained by using the first primer pair. Use of these primers resulted in the amplification of 1239 bp-long *nxrB* fragments from *N. marina*, *Nitrospira* strain Ecomares 2.1, *N. calida* and from enrichments of NOB from two Russian hot springs (Gorjachinsk and Uzon). Thus, this primer pair covered the *nxrB* genes of the tested representatives from all *Nitrospira* lineages except lineage III, for which no culture has been described and an environmental sample containing these NOB was not available.

Neither of the two primer pairs allowed the successful and specific *nxrB* amplification from all tested *Nitrospira* cultures and environmental samples by using a single annealing temperature in PCR. Instead, the success of PCR varied with the cultures or samples and the applied annealing temperatures (58–68°C for primers nxrB14f/nxrB1239r and 48–59°C for primers nxrB19f/nxrB1237r). These results indicate that the primers likely have unknown base mismatches to their respective binding sites in the *nxrB* genes of different *Nitrospira*. Therefore, these primers should be applied with care in environmental studies,

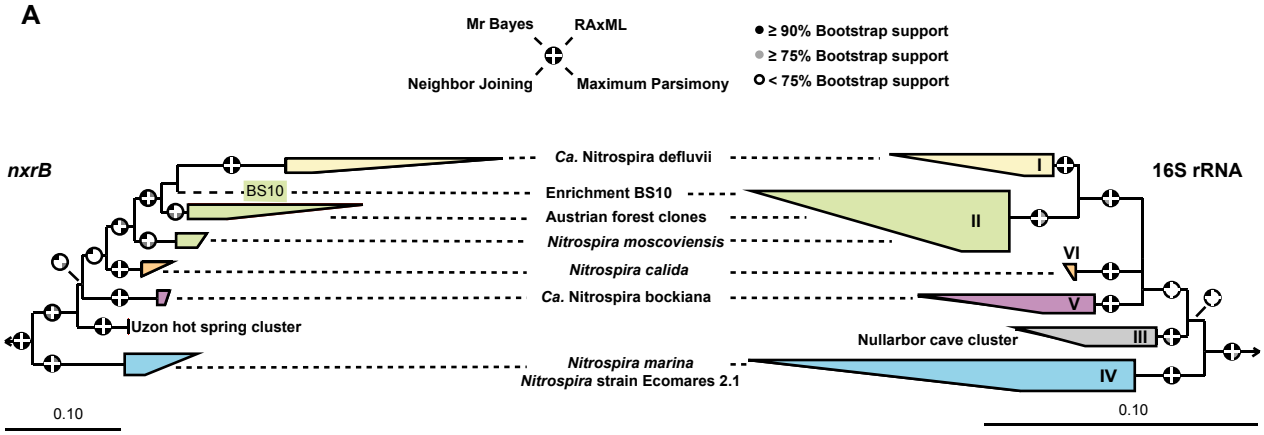
although they proved highly useful to extend our *nxrB* reference database by almost full-length sequences. Based on this extended dataset, we finally designed a third primer pair (nxrB169f/nxrB638r, Table 1) that targets internal conserved regions of all retrieved *Nitrospira nxrB* sequences. With this primer pair, 485 bp-long *nxrB* fragments were successfully amplified from all tested *Nitrospira* spp. and from a pooled sample of ten Namibian soils by using the same PCR conditions (annealing at 56.2°C) in all reactions. In addition, it was the only primer combination that yielded *nxrB* amplicons from the marine sponge *Hyrtios proteus*. Based on its broadest coverage and applicability for next-generation amplicon sequencing (see below), we propose to use this primer pair for environmental surveys. However, the forward primer (nxrB169f) has no base mismatches to the *nxrB* of *Nitrospina*, *Nitrococcus* and *Nitrolancetus* and to the *nxrB*-like gene of *Ca.* Methylomirabilis oxyfera, as well as to some nitrate reductase subunit beta (*narH*) genes. The reverse primer (nxrB638r) has three mismatches to the *nxrB* of *Nitrospina* and to some *narH* towards the 5'-end of the primer, possibly not preventing extension from the 3'-end during PCR. Therefore, use of this primer pair may lead to the unspecific amplification of non-*Nitrospira* genes if the PCR stringency is too low. The *in silico* specificities of the new *nxrB*-targeted primers are illustrated in Fig. 1.

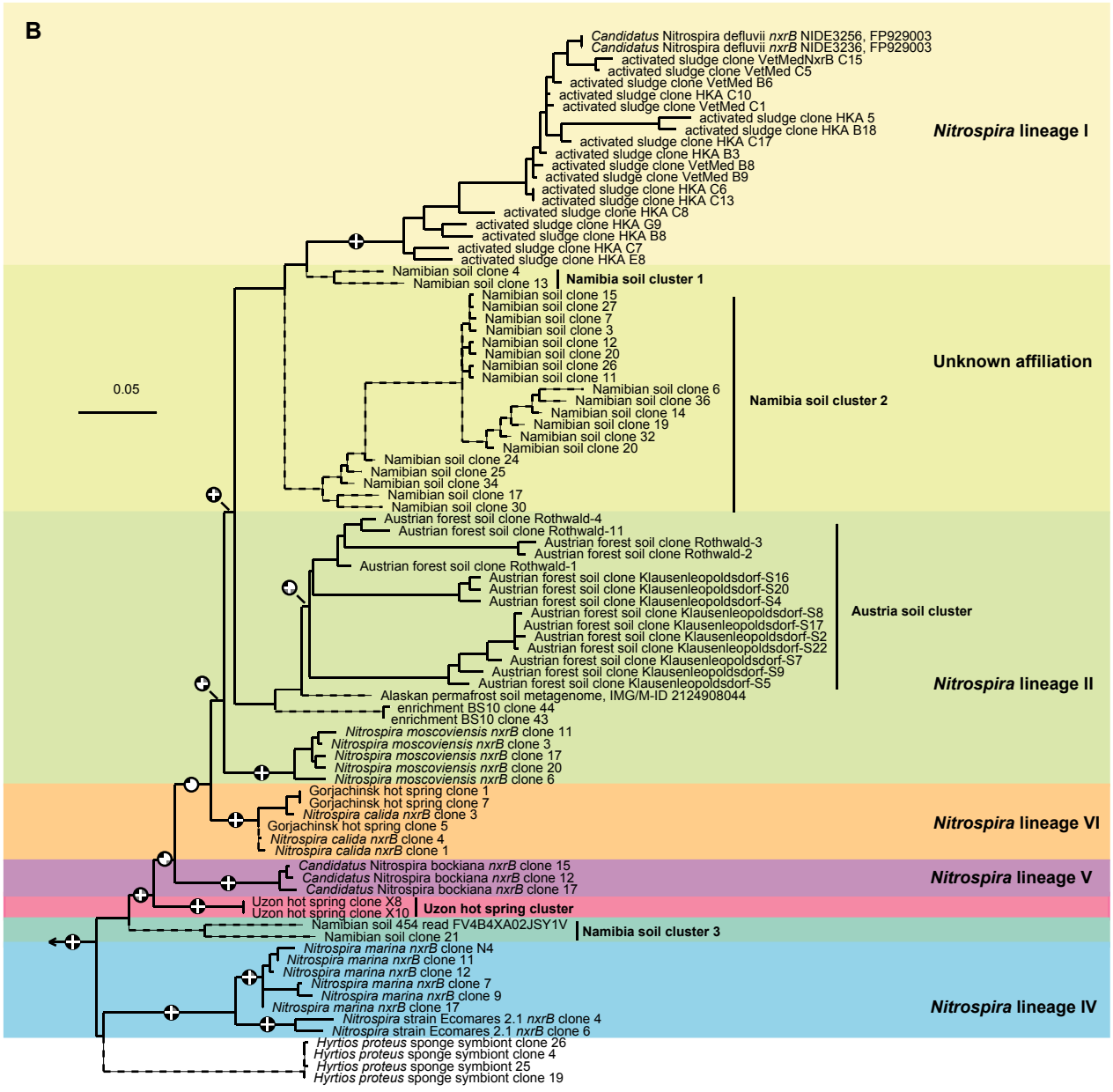
To make our phylogenetic analysis of *Nitrospira nxrB* as comprehensive as possible, we screened the NCBI, IMG/M, CAMERA and MG-RAST databases for additional publicly available entries of this gene, which cover at least the length of the shortest *nxrB* fragments as retrieved with the nxrB169f/nxrB638r primer pair. One *nxrB* sequence from an Alaskan permafrost soil metagenome (IMG/M-ID 2124908044) was the only hit with an identity below 97% to any *nxrB* sequence in our seed database and was included in our database.

A prerequisite for a functional and phylogenetic marker is a congruent evolutionary history with the 16S rRNA of the respective organisms. This condition is based on the assumption that the 16S rRNA phylogeny closely mirrors the true organismal phylogeny (Woese, 1987). To obtain a comprehensive and up-to-date 16S rRNA-based phylogeny of *Nitrospira*, we used all *Nitrospira*-related entries in the non-redundant SILVA 111 database (Pruesse *et al*., 2007) that fulfilled the SILVA quality criteria and were at least 1440 bp in length. Based on the *nxrB* and 16S rRNA databases, we calculated neighbor joining, maximum parsimony, maximum likelihood and Bayesian inference phylogenetic trees, which were combined in one consensus tree for each marker. The six 16S rRNA-based, previously proposed phylogenetic lineages within the genus *Nitrospira* (Daims *et al*., 2001; Lebedeva *et al*., 2008; 2011) were corroborated by our 16S rRNA-based analysis using the updated sequence database (Fig. S1). According to the criteria used earlier (Daims *et al*., 2001), a *Nitrospira* lineage should be monophyletic in all 16S rRNA-based trees calculated by different algorithms and should have high bootstrap support in the majority of trees, while the 16S rRNA sequence identity between the members of different lineages should be below 94%. Remarkably, no additional lineages could be defined based on these criteria and the updated dataset. However, several sequences fell between the six stable lineages and formed clusters with weak bootstrap support (Fig. S1). These clusters may fulfill the criteria for stable lineages in future analyses if more high-quality 16S rRNA gene sequences related to *Nitrospira* become available.

A comparison of the *nxrB* and 16S rRNA-based consensus trees revealed congruent topologies with high bootstrap support for all lineages that contained characterized nitrite-oxidizing *Nitrospira* as anchor points (Fig. 2A). Only *Nitrospira* lineage II, represented by one monophyletic group in the 16S rRNA-based tree, was split in three distinct but closely related lineages in the *nxrB* tree with high bootstrap support for most of these branches (Fig. 2A). Lineage II encompassed *N. moscoviensis*, *Nitrospira* clones from two Austrian forest soils, the nitrite-oxidizing enrichment BS10 and the *nxrB* sequence from an Alaskan permafrost soil metagenome (Fig. 2A and B). The more complex topology of lineage II in the *nxrB* tree may result from a higher phylogenetic resolution of *nxrB* and indicate that this lineage consists of several subgroups, which were not clearly resolved by the 16S rRNA-based phylogenetic reconstruction (Figs 2A and S1). Moreover, according to the *nxrB* analysis lineage II is paraphyletic, suggesting that lineage I evolved from an ancestor in lineage II (Fig. 2A). Alternatively, the branching pattern of the *nxrB* tree might reflect a functional diversification of *nxrB* within lineage II, which splits this clade because of functionally rather than phylogenetically caused sequence dissimilarities and groups some members of lineage II with lineage I. We can also not exclude lateral gene transfer of *nxrB* between

Fig. 2. Comparison of *nxrB*- and 16S rRNA-based phylogenies of (A) nitrite-oxidizing *Nitrospira* and (B) detailed view of the *Nitrospira nxrB* phylogeny. A detailed view of the *Nitrospira* 16S rRNA-based phylogeny is provided in Fig. S1. The Uzon hot spring and Nullarbor cave clusters (panel A) lack sequences in the respective other dataset. Both panels show majority rule consensus trees. Bootstrap support is indicated at individual branches. In total, 1197 and 1427 unambiguously aligned nucleotide positions were used in the *nxrB* and 16S rRNA analysis, respectively, and branches leading to shorter sequences are indicated by dotted lines. The scale bar indicates 10% (panel A) or 5% (panel B) estimated sequence divergence.





linages II and I. Two *nxrB* sequences from thermophilic *Nitrospira* enrichments from the Uzon Caldera, Kamchatka (Russia), formed a distinct cluster (Fig. 2A and B), indicating that they might represent a novel *Nitrospira* lineage. Indeed, a 16S rRNA sequence retrieved from one of these enrichments (Kamchatka clone Ns4a) fell into the genus *Nitrospira*, but did not cluster with significant bootstrap support with any established *Nitrospira* lineage in the phylogenetic analyses carried out here (Fig. S1) and previously (Lebedeva *et al*., 2011). However, the enrichments may contain more than one *Nitrospira* strain, and it remains unclear whether the *nxrB* and 16S rRNA sequences originated from the same strain. A comparison of the 16S rRNA and *nxrB* branching patterns was not possible for *Nitrospira* lineage III because no *nxrB* sequence from this lineage was available (Fig. 2A). Aside from these uncertainties, the reconstructed phylogenies of the *Nitrospira nxrB* and 16S rRNA genes were overall highly congruent, and thus we conclude that *nxrB* is a useful functional and phylogenetic marker for nitrite-oxidizing *Nitrospira*.

The topology of the *nxrB* consensus tree was based on near full-length *nxrB* sequences. Subsequently, this tree was extended by adding shorter sequences, which had been obtained by PCR using the primer pair nxrB169f/nxrB638r, cloning and Sanger sequencing. One short *nxrB* sequence from Alaskan permafrost soil, retrieved during the aforementioned metagenomic database search, was also added. This sequence from permafrost soil, and the *nxrB* of enrichment BS10, clustered with *nxrB* sequences from two different Austrian forest soils (Fig. 2B). Because enrichment BS10 and 16S rRNA gene sequences retrieved from the same two Austrian forest soils fell into *Nitrospira* lineage II in the 16S rRNA-based analysis (Fig. S1), this whole cluster was assigned to lineage II in the *nxrB* tree, too (Fig. 2B). In contrast, *nxrB* sequences obtained from Namibian soils formed three new clusters, two of which branched off between *Nitrospira* lineages I and II and the third one close to the root of the *Nitrospira nxrB* tree (Fig. 2B). The *nxrB* sequences retrieved from the marine sponge *H. proteus* clustered together with the *nxrB* sequences of *Nitrospira* lineage IV, which reportedly contains sponge-associated *Nitrospira* (Taylor *et al*., 2007; Off *et al*., 2010). However, because the Namibian soil and the sponge *nxrB* sequences cannot be linked to any 16S rRNA gene sequences at present, it remains unclear whether the respective *Nitrospira* represent novel lineages of this genus or should be assigned to lineages defined previously based on 16S rRNA phylogeny.

Genomic copy numbers of *nxrB* varied from two copies found in the sequenced genome of *Ca*. N. defluvii to six copies found by PCR screening in *N. marina*, and all copies from the same species where phylogenetically closely related (Fig. 2B). The ranges of nucleic acid sequence identity among the *nxrB* copies of the same species were considerably different, with a maximum identity of 99.9% observed in *Ca*. N. defluvii and a minimum identity of 96.7% observed in *N. moscoviensis* and in strain Ecomares 2.1. However, *nxrB* genes were always more divergent between different species than within the same genome (Table S2). This pattern generally applied also at the amino acid level, although few NxrB copies had a slightly higher amino acid identity across species than within the same organism (Table S3). Pairwise comparison of the *nxrB* and 16S rRNA gene identities between different *Nitrospira* revealed that *nxrB* is less conserved than the 16S rRNA (Fig. 3). For both genes, *N. marina* and *Nitrospira* strain Ecomares 2.1 were most closely related with 99.1% sequence identity of the 16S rRNA gene and ~95% nucleic acid identity of *nxrB* (Table S2). The high 16S rRNA identity and similar fatty acid profiles of these two organisms (Keuter *et al*., 2011) are contrasted by the stronger tendency of strain Ecomares 2.1 to produce extracellular polymeric substances and form biofilm and by the much higher nitrite tolerance of this culture compared with the type strain of *N. marina* (Watson *et al*., 1986; Keuter *et al*., 2011). In absence of DNA-DNA hybridization data or sequenced genomes, which would allow the calculation of the average nucleotide identity (Richter and Rossello-Mora, 2009), it is currently not possible to assess whether *N. marina* and strain Ecomares 2.1 represent the same or different *Nitrospira* species. Taking 99% sequence identity of 16S rRNA as an approximate threshold below which microbes can be assigned with a high probability to different species (Stackebrandt and Ebers, 2006), we propose based on the comparison depicted in Fig. 3 that *nxrB* sequences with less than 95% nucleic acid sequence identity likely represent different *Nitrospira* species. Considering 95% *nxrB* identity as species-level cut-off, our *Nitrospira nxrB* reference database contained 25 species-level OTUs.

High Nitrospira *richness in a selection of geographically distant soils*

To assess the applicability of *nxrB* as functional and phylogenetic marker for *Nitrospira* in ecological studies, we investigated the diversity of this gene in a selection of geographically and climatically distinct soils by using the primer pair nxrB169f/nxrB638r and 454 amplicon sequencing. In particular, we aimed to test whether the primers would mainly retrieve *Nitrospira nxrB* genes and no other, unspecific amplicons and whether the approach could detect novel *Nitrospira* representatives. In addition, our goal was to obtain first insights into the diversity of terrestrial *Nitrospira* as these organisms have barely been

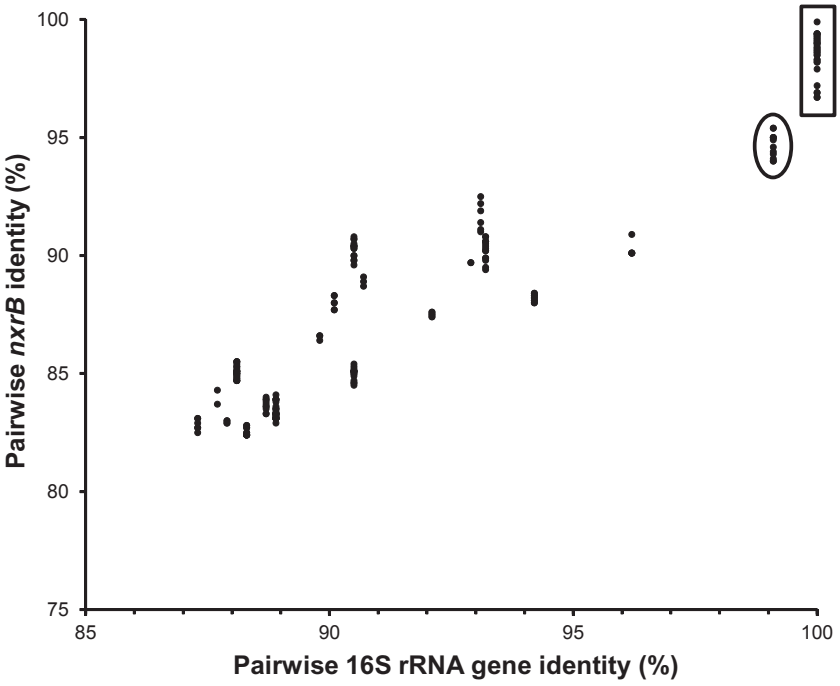


Fig. 3. Pairwise sequence identity comparison of the *nxrB* copies and 16S rRNA genes from enriched or isolated *Nitrospira*. Data points representing pairwise comparisons of gene copies from the same organism are enclosed by a rectangle. Data points representing pairwise comparisons of gene copies from *N. marina* and *Nitrospira* strain Ecomares 2.1, whose classification as the same or different species is unclear, are encircled by an ellipsis. Sequences with less than 99% identity at the 16S rRNA level are considered to belong to different species (Stackebrandt and Ebers, 2006). The organisms included in this analysis are listed in Table S2.

studied in soils and not yet by a specific deep-sequencing approach. The analysed soils encompass locations in Southern Africa (Namibia), Central America (Costa Rica), Central Europe (Austria) and the Arctic (Greenland) (Table S4). Our quality screening resulted in 101 844 high-quality reads with a median sequence length of 452 bp (read lengths ranged from 390 to 490 bp) and identified 34 chimeras at the 97% nucleic acid identity clustering level, which represented 213 *nxrB* reads in total. Phylogenetic classification of these reads using the Bayesian classifier of the mothur software and the *nxrB*/*narH* dataset (Fig. 1) revealed that for each soil, on average 94% of the reads were affiliated to *Nitrospira nxrB* (Table 2). This large fraction clearly documents the high selectivity of the primer pair nxrB169f/nxrB638r for *nxrB* of *Nitrospira* under the applied PCR conditions. Merely in two soils from Costa Rica only 53 and 66% of the sequence reads were affiliated to *Nitrospira nxrB*. The obtained non-*Nitrospira nxrB* reads were related to other groups within the CISM enzyme family. For example,

Table 2. Pyrosequencing results and observed or estimated numbers of OTUs based on reads that were affiliated to *Nitrospira nxrB*.

Geographic location	Sample	Number of unaffiliated reads	Number of *Nitrospira nxrB* reads	Good's coverage[a]	95/90/80% *nxrB* identity		
					Number of observed OTUs	Chao1 richness estimator	ACE richness estimator
Namibia	Dry woodland #06	37	3316	0.978	212/107/24	287/125/26	369/134/28
	Dry woodland #11	85	800	0.899	184/97/23	279/126/24	285/119/27
	Dry woodland #16	64	9023	0.982	576/191/27	739/216/28	733/207/29
	Arable soil #04	119	15 137	0.987	645/222/29	907/241/31	857/247/35
	Arable soil #08	814	9463	0.984	511/196/28	658/217/28	658/222/28
	Arable soil #14	722	5173	0.976	422/171/29	558/186/29	534/190/30
	Fallow soil #01	74	3621	0.975	295/123/22	384/133/25	378/139/24
	Fallow soil #10	43	492	0.888	109/67/14	192/92/14	185/114/14
	Fallow soil #17	1191	21 813	0.991	764/253/34	946/284/35	937/280/35
	Fallow soil #23	360	3479	0.972	312/143/25	465/198/26	399/164/26
Costa Rica	Rain forest	1037	1156	0.997	8/6/4	10/7/4	14/9/4
	Arable soil	769	1480	0.990	44/26/13	59/32/15	80/43/18
Austria	Riparian forest	148	4454	0.991	149/52/11	190/63/11	192/61/11
	Spruce forest	587	6630	0.998	63/28/12	74/33/13	72/32/14
	Arable soil	50	4629	0.995	86/44/17	129/74/27	128/116/77
Greenland	Tundra soil	54	5024	0.998	53/19/7	58/19/7	62/20/9

a. Calculated from the number of OTUs (at 95% *nxrB* identity level) represented by only one quality-controlled 454 read (N_1) and the total number of quality-controlled 454 reads (N) as $1 - (N_1/N)$.

several of these reads fell between the *nxrB* forms of *Nitrospira* and *Nitrospina gracilis* (but clearly outside the known genus *Nitrospira*) or were affiliated with putative bacterial and archaeal *narH* genes (data not shown).

As suggested by the Good's coverage parameter (Good, 1953), which was above 0.97 for most of the analysed soils (Table 2), the sequencing depth was sufficient to cover a large fraction of the *Nitrospira* richness at the approximate species level (95% *nxrB* nucleic acid identity) in these soils. However, in several cases, non-parametric richness estimators predicted numbers of species-level OTUs that were substantially higher than the observed OTU numbers (Table 2). Intriguingly, a high *Nitrospira* species-level OTU richness was detected in most of the soil samples (Table 2). For example, for the two Namibian soils with more than 15 000 sequence reads the non-parametric richness estimators predicted up to 946 *Nitrospira* species-level OTUs (Table 2). With 90% nucleic acid sequence identity as OTU threshold, the average numbers of estimated OTUs in the Namibian soils were still as high as 181 (Chao1) or 182 (ACE).

Novel Nitrospira *species are present in the analysed soils*

At a nucleic acid sequence identity level of 95% the *nxrB* sequence reads obtained from the different soils represented 1801 different species-level OTUs, which increased the known diversity of *Nitrospira* species by two orders of magnitude compared with the *nxrB* seed database established by Sanger sequencing. A large fraction (68%) of these *nxrB* sequence reads fell into one of the previously defined *Nitrospira* lineages and soil clusters already represented in the *nxrB* database (Fig. 2B). Most of these sequence reads belonged to the Austrian forest soil cluster within *Nitrospira* lineage II (sequences from the Austrian riparian and spruce forest soils as well as the Greenland tundra soil) or belonged to Namibia soil clusters 1 and 2 (sequences from all Namibian soils and the Austrian arable soil) (Table S5). The remaining 32% of the obtained *nxrB* sequence reads fell outside the already established *nxrB* lineages/soil clusters. At a nucleic acid sequence identity level of 95%, these *nxrB* sequence reads represented 908 species-level OTUs. These novel *nxrB* were also clustered at 80% sequence identity, a value chosen arbitrarily to get a better overview of this large sequence diversity. This grouping resulted in 31 novel '454 clusters' that were dispersed throughout the *Nitrospira nxrB* consensus tree, with the exception of lineage IV (Fig. 4). Many of these 454 clusters (13 of 31) shared a common ancestor with *Nitrospira* lineage V, whereas the second largest group (12 of 31) branched off between *Nitrospira* lineages I and II (Fig. 4). Interestingly, the Costa Rican rain forest and arable soils were both dominated by this novel *nxrB* diversity, with 454 cluster 1 (related to Namibian soil cluster 2) and 454 cluster 3 (related to *N. moscoviensis*) being the most abundant detected *Nitrospira nxrB* in these samples. All Namibian soils and the tundra soil from Greenland also showed a considerable relative abundance of novel *nxrB* types, in particular 454 cluster 8 in Namibian soils and 454 clusters 12 and 16 in the tundra soil (Table S5). These novel 454 clusters shared a common ancestor with *Nitrospira* lineage V (Fig. 4).

Discussion

Despite their ecological and biotechnological importance, *Nitrospira* are barely studied nitrifiers. Whereas most research on these organisms has focused on the representatives living in wastewater treatment plants, only few studies started to explore the ecology of *Nitrospira* in natural ecosystems (Freitag *et al*., 2005; Lebedeva *et al*., 2005; Attard *et al*., 2010; Haaijer *et al*., 2013). Here we introduce the *nxrB* gene as a functional marker for the detection and identification of *Nitrospira* in environmental samples. Aside from being a specific functional marker, *nxrB* provides a robust phylogenetic framework to distinguish uncultured *Nitrospira* even at the sublineage level and to discover novel diversity within this genus. Based on the reference sequences from cultured *Nitrospira*, we could define a nucleotide sequence identity cut-off (95%), which is useful to distinguish *Nitrospira* species by using *nxrB* as marker gene. However, because of the sequence divergence among the paralogous *nxrB* copies (Table S2), OTU formation based on *nxrB* identity thresholds above 95% could lead to an overestimation of *Nitrospira* diversity at the subspecies level. Furthermore, the range of *nxrB* copy numbers in different *Nitrospira* genomes (Fig. 2B) should be taken into account when using *nxrB* as target gene for quantitative PCR (qPCR) analyses of *Nitrospira*. This issue of qPCR is not restricted to *nxrB* but applies to all markers with varying gene copy numbers, such as *amoA* (Norton *et al*., 2002) and the 16S rRNA gene (Farrelly *et al*., 1995). Only a very limited set of *nxrB* sequences had been available as basis for the design of new *nxrB*-targeted PCR primers, but by using an iterative primer design process two primer pairs were developed for which excellent coverage of *Nitrospira nxrB* genes could be experimentally demonstrated. With these primers, numerous novel representatives outside the previously known *Nitrospira* clades from various soils and a marine sponge were recovered (Figs 2 and 4), and an next generation sequencing-based pipeline for the analysis of environmental *nxrB* was established.

Nitrospira are generally considered to be obligate autotrophic nitrite oxidizers with only limited mixotrophic capabilities (Watson *et al*., 1986; Ehrich *et al*., 1995;

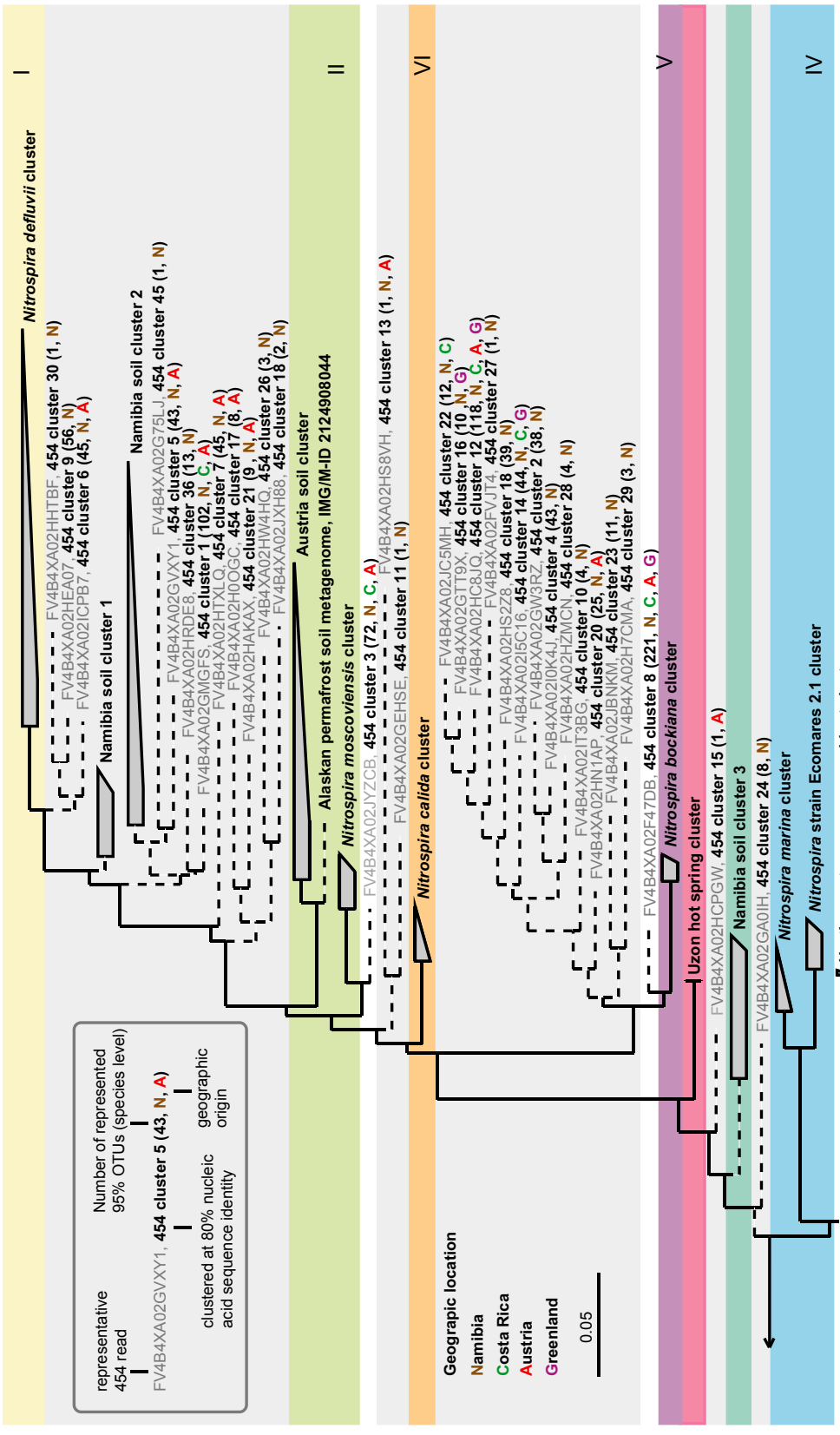


Fig. 4. Phylogenetic analysis of 454 amplicons that were classified as *Nitrospira nxrB* but did not belong to any of the previously defined *Nitrospira* lineages or environmental clusters in the *nxrB* consensus tree (Fig. 2). The '454 clusters' were formed from the 454 sequence reads based on a threshold of 80% sequence identity. Representative sequences from these clusters were then added, without changing the overall tree topology, to the *nxrB* consensus tree by using the Quick-Add Parsimony tool of the ARB software (Ludwig *et al.*, 2004). Branches leading to 454 cluster members are indicated by dotted lines to highlight their unsupported affiliation, which results from the short sequences. The scale bar indicates 5% estimated sequence divergence.

Daims *et al*., 2001). Although this restricted lifestyle appears to limit the number of ecological niches available to these organisms, a surprisingly high diversity of coexisting *Nitrospira* was detected in most of the soils investigated by *nxrB* amplicon pyrosequencing (Table 2). This interesting phenomenon should be addressed in future studies aimed to elucidate the physiological versatility of uncultured terrestrial *Nitrospira* and the factors that shape their community structure. In this context, an interesting perspective arises from the mutualistic symbiosis between ammonia and nitrite oxidizers (Stein and Arp, 1998). One may speculate that the community composition of ammonia oxidizers partly determines that of the co-occurring nitrite oxidizers (and vice versa) if the interactions between members of these guilds are specific enough to have such selective power. For example, different *Nitrospira* are known to prefer different nitrite concentrations in their microenvironment (Maixner *et al*., 2006). This property may favor co-occurrence with those AOB or AOA whose ammonia oxidation kinetics ensures that *Nitrospira* are locally provided with nitrite at their respective concentration optima. Specific interactions between nitrifiers may also include the exchange of organic compounds or growth factors or the acquisition of inorganic nutrients, especially if one of the partners lacks important biosynthetic pathways. For instance, the genome of the AOB *Nitrosomonas europaea* does not code for siderophore synthesis, whereas it contains multiple genes of siderophore receptors (Chain *et al*., 2003). These receptors may bind siderophores produced by NOB such as *Nitrobacter* and *Nitrospira* (Starkenburg *et al*., 2006; Lücker *et al*., 2010). If specific interactions partly define the ecological niches of symbiotic nitrifiers, this might increase the total number of niches and could be one factor supporting a high nitrifier diversity. Moreover, ecophysiological or biogeographical constraints on the distribution of an ammonia or nitrite oxidizer clade may influence the distribution of the preferred symbiotic partners from the other functional group. This possible scenario adds complexity to community structure analyses of nitrifiers, as it implies that surveys of only one nitrifier guild (for example, only AOA) may not provide all data needed to explain the observed distribution and abundance patterns. The aforementioned interactions between nitrifiers have not been demonstrated in terrestrial ecosystems yet. However, to illustrate the potential of analysing both functional groups, we compared the *nxrB* dataset obtained in this study to the thaumarchaeal *amoA* dataset that was derived previously (Pester *et al*., 2012) from exactly the same soil samples and DNA extractions. Procrustes analysis uncovered that AOA and *Nitrospira* communities were highly correlated across the soil samples for both OTU presence/absence [$r = 0.95$ ($P = 0.001$)] as well as OTU relative abundance [$r = 0.93$ ($P = 0.001$)] (Fig. S2). We then examined correlations in abundance of species-level OTUs within and between AOA and *Nitrospira* in the Namibian soils, which represent different soil management regimes in the same geographic area and climate (Fig. 5). As expected from the high correlation observed in the Procrustes analysis, several OTUs were highly correlated between the two groups. Positive correlations between single *amoA* and *nxrB* OTUs (e.g. *amoA* 043 and *nxrB* 0014) might indicate either shared habitat preference or a mutualistic interaction between the two organisms, and negative correlations (e.g. *nxrB* 0480/0027) might result from either divergent habitat preference or strong antagonism. Positive correlations between multiple *amoA* and *nxrB* OTUs (e.g. *amoA* 011/029 and *nxrB* 1732/0428/1269/128) might indicate a more complex interaction involving a number of factors such as weak competition, and niche partitioning either due to spatial structuring, due to different preferred ammonium or nitrite levels (Maixner *et al*., 2006) or due to heterogeneous alternative metabolisms. Positive correlations between *nxrB* OTUs but with no thaumarchaeal *amoA* OTUs (e.g. *nxrB* 0548/0055, or *nxrB* 0029/0012/0003) might indicate that these OTUs are involved in specific interactions with AOB, which were not analysed in this study. Alternatively, these OTUs might even be engaged in alternative metabolisms other than nitrite oxidation. Correlation network analysis such as presented here raises a number of fascinating alternative interpretations that cannot be resolved without more biological information (Stecher *et al*., 2013).

The *Nitrospira* richness detected in most Namibian soil samples was much higher than the richness in the non-Namibian soils (Table 2). As the Namibian soils were sampled at different sites in dry woodland, arable land and fallows, no specific habitat type or soil treatment regime could be linked to the high *Nitrospira* diversity detected in these samples. The apparent difference in richness might be caused by unknown environmental factors, biological mechanisms and the different sample size. However, interpretations of microbial community surveys should also consider possible methodical biases. Typical 454-pyrosequencing errors like homopolymer miscounts (Margulies *et al*., 2005) can lead to overestimation of microbial diversity if OTUs are defined using high sequence identity thresholds (> 97%) (Kunin *et al*., 2010). Here we can largely exclude this bias because high *nxrB* diversity was found even with a relatively low OTU clustering threshold of 90%, where the influence of sequencing errors should be small. Moreover, our analysis pipeline contains quality screening and trimming of sequences (Chou and Holmes, 2001), filtering by frameshift detection, and manual sequence curation and chimera exclusion steps, which further decrease the impact of common 454 pyrosequencing errors (Pester

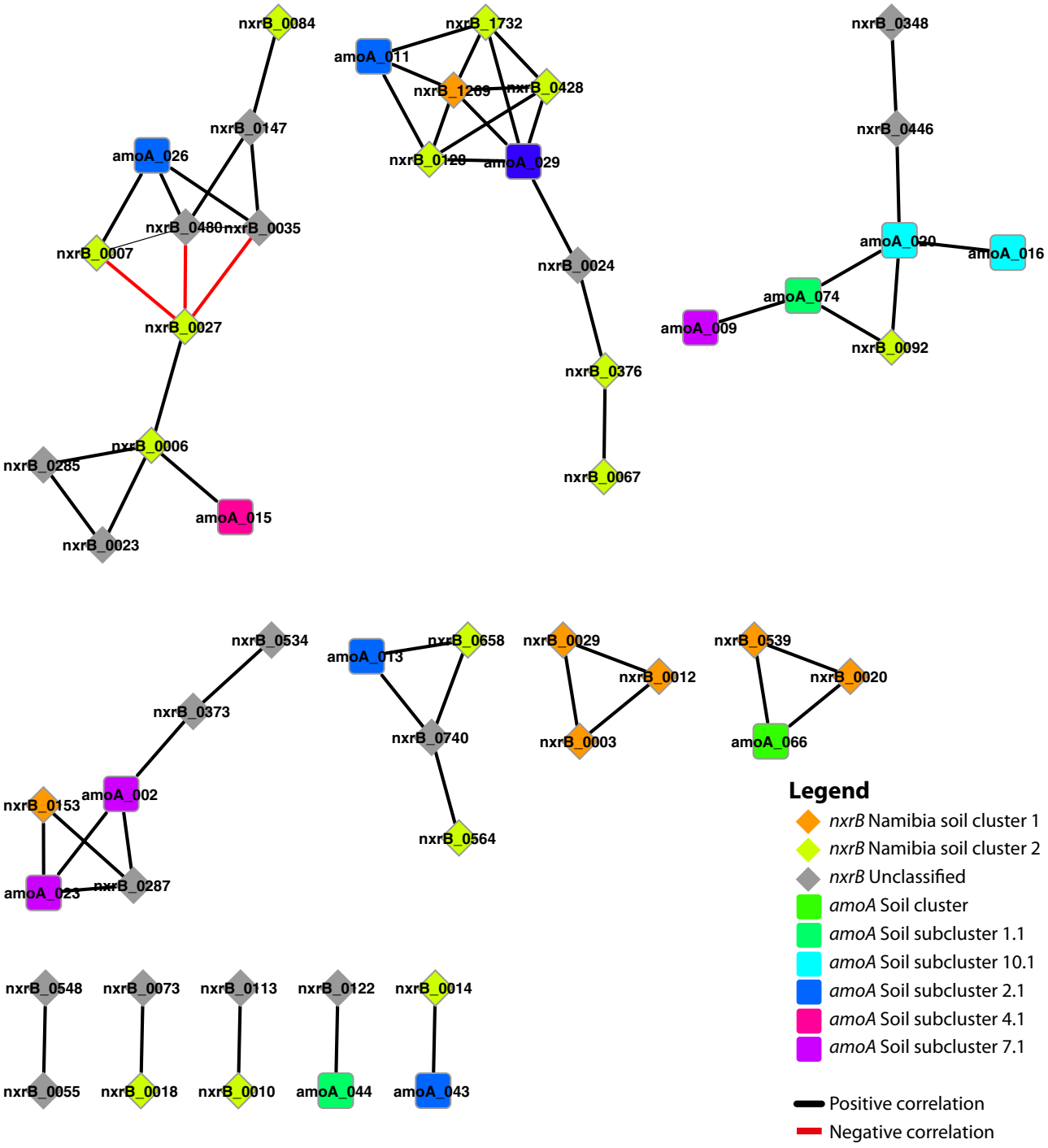


Fig. 5. Correlation network of *amoA* and *nxrB* OTUs in Namibian soils. Pairwise correlation coefficients were calculated and positive correlations (≥ 0.8) are indicated as black edges and negative correlations (≤ –0.8) as red edges. Nodes are OTUs of *amoA* (square) or *nxrB* (diamond) and are colored according to phylogeny presented here and in Pester and colleagues (2012).

et al., 2012). Consistently, the low diversity found in the Costa Rican rain forest soil (Table 2) shows that the analysis was not generally biased towards detecting inflationary diversities. Most of the OTUs, which were represented by only one or two quality-controlled 454 reads (Table S6), thus likely represent rare *Nitrospira* species, although we cannot exclude that some of these rare OTUs may result from undetected sequencing errors in single reads. Another possible, but also unlikely cause of the observed richness differences could be DNA

extraction biases. Because all characterized *Nitrospira* have a Gram-negative cell wall with a wide periplasmic space and a thin peptidoglycan layer (e.g. Watson *et al*., 1986), and the same DNA extraction protocol was applied to all soils for pyrosequencing (Table S4), *Nitrospira* cell lysis efficiencies should have been similar for all samples. The possibility that extracellular polymeric substances, which often enclose *Nitrospira* microcolonies in aquatic biofilms (Spieck *et al*., 2006), selectively impeded DNA extraction from some strains cannot be evaluated at present because nothing is known about the tendency of terrestrial *Nitrospira* to form extracellular polymers. A more likely source of bias could be extracellular DNA, whose persistence is increased by adsorption to soil particles and varies with soil mineralogy and water content (Blum *et al*., 1997). If DNA degradation was retarded in the Namibian soils, for example, due to lower water content, the detected richness may comprise both current and previous *Nitrospira* community members, whereas this 'archiving effect' would be less pronounced in other soils with higher DNA degradation rates. Short DNA fragments persist for extended periods rather than long DNA molecules in soil (Blum *et al*., 1997). Thus, the issue of extracellular DNA may be particularly significant for currently widely used amplicon pyrosequencing techniques that handle only relatively short PCR products (which can be obtained from short template molecules). Our analysis could have been affected also by a still-incomplete primer coverage of the extant *Nitrospira nxrB* diversity or by other biases of PCR such as differential amplification (reviewed by von Wintzingerode *et al*., 1997). Such problems are not specific to *nxrB* and can never be completely excluded in PCR-based microbial community analyses using functional or phylogenetic marker genes.

Because nitrifiers maintain a high cellular ribosome content even during prolonged starvation periods or when chemically inhibited (Wagner *et al*., 1995; Morgenroth *et al*., 2000), quantified amounts of rRNA or probe-conferred fluorescence intensity after rRNA-targeted FISH do not correlate with the metabolic activity of these organisms. Past research showed that *nxrB* transcription increased markedly after addition of nitrite to a starved culture of *Ca*. N. defluvii, whereas the rRNA levels were high during and after prolonged starvation periods (Lücker *et al*., 2010). These data suggest that *nxrB* mRNA is a better indicator of *Nitrospira* activity than rRNA, and thus the use of *nxrB* as genetic marker opens interesting new perspectives for future studies on the metabolic activity of uncultured nitrite-oxidizing *Nitrospira*. The *Nitrospira nxrB*-targeted primers ideally complement previously published *nxrB* primers specific for *Nitrobacter* (Vanparys *et al*., 2007) (Fig. 1), so that these two important NOB groups can be studied concomitantly in soil and other ecosystems. A crucial task will now be the design and evaluation of additional *nxrB* primers targeting the other known NOB lineages (Fig. 1). We are confident that the use of *nxrB* as marker will eventually accelerate research on NOB in a similar manner as *amoA* has boosted a large number of studies on ammonia oxidizers, leading to a more complete picture of the nitrifiers than available today.

Experimental procedures

Cultured Nitrospira *and environmental samples*

All cultures and environmental samples, which were screened for *Nitrospira nxrB* genes, are listed together with sampling details and DNA extraction methods in Table S4. Briefly, *nxrB* sequences were retrieved from pure cultures of *N. moscoviensis*, *N. marina*, *N. calida*, from highly enriched cultures of *Ca*. N. defluvii and *Ca*. N. bockiana, from nitrite-oxidizing enrichment cultures from hot springs in the Uzon Caldera, Kamchatka, Russia (Lebedeva *et al*., 2011), and from an additional *Nitrospira* culture (enrichment BS10, B. Nowka, S. Off, H. Daims and E. Spieck, unpublished). Furthermore, *nxrB* sequences of uncultured *Nitrospira* were obtained from tissue of the marine sponge *H. proteus*, two full-scale wastewater treatment plants in Vienna, Austria, and from a selection of geographically and climatically distinct pristine and agricultural soils.

PCR amplification and cloning

PCR amplification of partial 16S rRNA genes of *Nitrospira* spp. was carried out according to Maixner and colleagues (2006). Three PCR protocols based on different primer pairs (Table 1) were used to amplify *nxrB* fragments of various lengths from 5 to 100 ng of genomic DNA, which was extracted from environmental samples, or from heated cell extracts of *Nitrospira* pure or enrichment cultures. The following PCR reaction mixture was applied: 1 μM of each primer, 2 mM $MgCl_2$, 0.2 mM of each dNTP, 10 ng μl^{-1} bovine serum albumin (optional), 1.25 U Taq DNA polymerase and 1 × reaction buffer [all reagents were obtained from Fermentas, St Leon-Rot, Germany, except primers (Thermo Scientific, Ulm, Germany)]. Template DNA was initially denatured at 95°C for 5 min followed by 35 cycles of denaturation (95°C, 40 s), primer annealing (temperature depending on the primer pair, see below, 40 s) and elongation (72°C, 90 s), with 10 min of final extension at 72°C. With primer pair nxrB14f/nxrB1239r, annealing temperatures between 58 and 68°C were applied in different PCR reactions. With primer pair nxrB19f/nxrB1237r, annealing temperatures between 48 and 59°C were applied in different PCR reactions. For primer pair nxrB169f/nxrB638r, the annealing temperature was 56.2°C in all reactions except the PCRs for 454 amplicon sequencing (50°C, see below). The size of the PCR products was checked by agarose gel electrophoresis. For cloning, PCR products were loaded on a low-melting preparative agarose gel (1.5%, Biozym Sieve GeneticPure Agarose, Biozym, Hess. Oldendorf, Germany), and bands of the expected size were cut out using a micro-haematocrit tube (Brand, Wertheim, Germany). Gel pieces were melted at

70°C, diluted with 100 µl of ultra-pure water, and the retrieved PCR products were cloned in *Escherichia coli* by Topo TA cloning (Life Technologies, Paisley, UK). Cloned inserts of the right size were Sanger-sequenced by using the BigDye Terminator Cycle Sequencing Kit v3.1 and an ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions.

Screening of public databases for Nitrospira *nxrB*

Based on near full-length *Nitrospira nxrB* sequences that were retrieved by PCR in this study, we screened public repositories for additional *Nitrospira nxrB* sequences by tblastx analysis (Camacho *et al*., 2009). To define a bit score threshold for the specific retrieval of *Nitrospira nxrB* sequences, each entry of the *Nitrospira nxrB* in-house 'seed' database was used as Basic Local Alignment Search Tool query against all other in-house *Nitrospira nxrB* sequences and a set of outgroup genes. This outgroup comprised the most closely related *narH*-like genes from the marine metagenome HOTS (NCBI acc. ABEF01052189), from *Natronomonas pharaonis* DSM 2160 (NCBI acc. NC_007426.1), from *Hydrogenobaculum* strain Y04AAS1 (NCBI acc. NC_011126) and the *nxrB* (formerly annotated as *narH*) of *Candidatus* Kuenenia stuttgartiensis (NCBI acc. CT573072). The highest bit score of the outgroup entries + 10% (to make the search more conservative) was then used as the bit score threshold for the tblastx search. To account for sequence divergence among the known *Nitrospira nxrB* sequences, the bit score threshold was determined separately for each entry of the seed *nxrB* database. Thereafter, each *Nitrospira nxrB* entry was used as query, one by one and with its own bit score threshold, for screening the NCBI non-redundant and environmental databases (Benson *et al*., 2011), the IMG/M database (Markowitz *et al*., 2008), the CAMERA database (Sun *et al*., 2011) and the MG-RAST database (Meyer *et al*., 2008) (status June 2012). All sequences showing ≥ 97% nucleic acid sequence identity with a seed database entry were not considered further to reduce sequence redundancy.

Phylogenetic analysis of nxrB and 16S rRNA genes

Phylogenetic analysis of *nxrB* and 16S rRNA gene sequences was done using 1197 and 1427 unambiguously aligned nucleotide positions respectively. For reconstruction of the *nxrB* phylogeny, nucleic acid sequences were preferred over deduced amino acid sequences because of their higher phylogenetic resolution (Tables S2 and S3). No *nxrB* or 16S rRNA sequence conservation filters were applied to achieve the best possible phylogenetic resolution within the genus *Nitrospira*. Phylogenetic trees were reconstructed using (i) neighbor joining with Jukes–Cantor distance correction within the PHYLIP package (Felsenstein, 1989), (ii) maximum parsimony within PHYLIP, (iii) maximum likelihood (RAXML v7.2.8) (Stamatakis, 2006) and (iv) Bayesian phylogeny inference (MrBayes 3.1 and 3.2) (Ronquist *et al*., 2012). Bayesian inference was run with 8.0 Mio and 1.2 Mio generations for the *nxrB* and 16S rRNA gene tree and resulted in standard deviations of 0.047 and 0.007 respectively. A consensus tree was constructed from the output of the different treeing methods according to the majority rule, which defines that a cluster must occur in at least two of the four inferred trees to appear in the consensus tree (PHYLIP). Branch lengths of the consensus tree were inferred by the Fitch algorithm using a Jukes–Cantor-corrected distance matrix (PHYLIP). Bootstrap support for neighbor joining, maximum likelihood and maximum parsimony trees was determined using 1000, 1000 and 100 re-samplings respectively. Sequences that covered a smaller portion of the *nxrB* or 16S rRNA gene were added to the consensus trees, without changing the overall tree topology, using the parsimony interactive tool of ARB (Ludwig *et al*., 2004). An ARB database containing all obtained *nxrB* sequences, and the *nxrB* consensus tree is provided in File S1.

Amplicon pyrosequencing and data analysis

Amplicons of *nxrB* for pyrosequencing were generated using the primer pair nxrB169f/nxrB638r according to the procedure described above. The only modifications included a smaller cycle number (30) to reduce potential PCR biases (von Wintzingerode *et al*., 1997) and a lower annealing temperature (50°C) to account for possible base mismatches between the primers and unknown *nxrB* in environmental samples. The lower annealing temperature still resulted in a single PCR product of the expected size. Replicate *nxrB* amplicons from at least two independent DNA extractions from each soil (0.25 g of soil per extraction) were pooled before sequencing to diminish within soil heterogeneity. Barcoding of amplicons by oligonucleotide ligation at both ends, preparation of sequencing libraries and bioinformatics analysis followed the procedures described in detail previously (Pester *et al*., 2012). Modifications were (i) that only 454 reads ≥ 390 nt were analysed (97% of the obtained high quality reads) and (ii) that taxonomic classification was done using the Bayesian classifier provided by the mothur software package (Schloss *et al*., 2009). 454 reads sequenced from both ends were pooled for all analyses because they overlapped by at least 61% (more than 295 overlapping bp out of 485 possible bp). To discriminate against errors introduced by pyrosequencing, high-quality sequences were initially clustered at 97% identity level (Kunin *et al*., 2010), and cluster representatives were further screened for insertion/deletion errors using frameshift detection by an in-house adaptation of FrameD (Schiex *et al*., 2003; Pester *et al*., 2012). The FrameD-based approach corrects each read individually and does not rely on the 'majority rule' like the standard denoising programs. Furthermore, the latter were recently shown to change reads in a manner inconsistent with the known spectrum of pyrosequencing errors and resulted in addition of sequence information to shorter reads that was often dissimilar from what had been removed by quality filtering steps (Gaspar and Thomas, 2013). Detection of chimeras was performed using Uchime (Edgar *et al*., 2011) and querying against the *nxrB* reference database established in this study. Candidate chimeras were added to the reference database and checked by phylogenetic analysis of independent 5′- and 3′-end sections of the respective sequences using the parsimony interactive tool in ARB (Ludwig *et al*., 2004). As a

final quality check, remaining sequences were further clustered at the approximate species level of 95% *nxrB* sequence identity, and cluster representatives were aligned against the *nxrB* reference database. Alignments were manually curated to reduce artificial diversity caused by alignment errors.

Statistical analyses

For Procrustes analyses and correlation networks, the *amoA* (Pester *et al*., 2012) and *nxrB* (this study) datasets were subsampled without replacement at 400 reads (below the size of the smallest library), and either OTU presence/absence or relative abundance was used to compute square-root transformed Bray–Curtis distances between samples. Procrustes analysis was performed using the vegan package in R (Peres-Neto and Jackson, 2001; Oksanen *et al*., 2012; R-Core-Team, 2013). In order to generate a correlation network for Namibian soil samples, all *amoA* and *nxrB* OTUs that were detected in at least half of the Namibian soil samples (5 of 10) and had > 1% relative abundance in at least one sample were selected, and Pearson correlation coefficients were computed for all pairwise combinations. Statistical significance of each correlation was determined by randomization (1000 permutations), and resulting *P*-values were corrected for multiple comparisons using the false discovery rate method in R (Benjamini and Hochberg, 1995). Correlation coefficients with at least 0.8 or less than –0.8 and with a $P < 0.05$ were used to construct a force-directed correlation network in Cytoscape (Smoot *et al*., 2011).

Accession numbers

16S rRNA gene and *nxrB* sequences obtained by Sanger sequencing have been deposited at GenBank under accession numbers KC836093-KC836105 and KC884854-KC884938 respectively. The *nxrB* sequences obtained by 454 sequencing were submitted to the Sequence Read Archive (SRA) at GenBank under accession number SRA047303.

Acknowledgements

We thank Michael Hannig, Jörg Overmann, Barbara Reinhold-Hurek and Florian Hofhansl for sampling Namibian and Costa Rican soils, and Stefan Flechl, Sandra Hauzmayer, Brigitte Lechthaler, Jasmin Schwarz and Faris Behnam for technical support. We further thank the Norwegian High-Throughput Sequencing Centre at the University of Oslo for providing sequencing services. Soils in Namibia were sampled under the collection permit 1245/2008 and exported under export permit ES 23855. This research was supported by the Austrian Science Fund, FWF (projects P23117-B17, M. P.; P25231-B21, H. D.; P20185-B17 and P25111-B22, A. L.; P24101-B22, S. L.), the Vienna Science and Technology Fund, WWTF (projects LS216 and LS09-40, H. D.), the German Federal Ministry of Science and Education (BIOLOG/BIOTA project 01LC0621D, A. L. and M. W.) and the European Research Council [Advanced Grant Nitrification Reloaded (NITRICARE) 294343, M. W.].

References

Attard, E., Poly, F., Commeaux, C., Laurent, F., Terada, A., Smets, B.F., *et al*. (2010) Shifts between *Nitrospira*- and *Nitrobacter*-like nitrite oxidizers underlie the response of soil potential nitrite oxidation to changes in tillage practices. *Environ Microbiol* **12:** 315–326.

Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B* **57:** 289–300.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2011) GenBank. *Nucleic Acids Res* **39:** D32–D37.

Blum, S.A.E., Lorenz, M.G., and Wackernagel, W. (1997) Mechanism of retarded DNA degradation and prokaryotic origin of DNases in nonsterile soils. *Syst Appl Microbiol* **20:** 513–521.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10:** 421.

Chain, P., Lamerdin, J., Larimer, F., Regala, W., Lao, V., Land, M., *et al*. (2003) Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph *Nitrosomonas europaea*. *J Bacteriol* **185:** 2759–2773.

Chou, H.H., and Holmes, M.H. (2001) DNA sequence quality trimming and vector removal. *Bioinformatics* **17:** 1093–1104.

Daims, H., and Wagner, M. (2010) The microbiology of nitrogen removal. In *The Microbiology of Activated Sludge*. Seviour, R.J., and Nielsen, P.H. (eds). London, UK: IWA Publishing, pp. 259–280.

Daims, H., Nielsen, J.L., Nielsen, P.H., Schleifer, K.H., and Wagner, M. (2001) In situ characterization of *Nitrospira*-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl Environ Microbiol* **67:** 5273–5284.

Daims, H., Maixner, F., and Schmid, M.C. (2009) The nitrifying microbes: ammonia oxidizers, nitrite oxidizers, and anaerobic ammonium oxidizers. In *FISH Handbook for Biological Wastewater Treatment*. Nielsen, P.H., Daims, H., and Lemmer, H. (eds). London, UK: IWA Publishing, pp. 9–17.

Dionisi, H.M., Layton, A.C., Harms, G., Gregory, I.R., Robinson, K.G., and Sayler, G.S. (2002) Quantification of *Nitrosomonas oligotropha*-like ammonia-oxidizing bacteria and *Nitrospira* spp. from full-scale wastewater treatment plants by competitive PCR. *Appl Environ Microbiol* **68:** 245–253.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27:** 2194–2200.

Ehrich, S., Behrens, D., Lebedeva, E., Ludwig, W., and Bock, E. (1995) A new obligately chemolithoautotrophic, nitrite-oxidizing bacterium, *Nitrospira moscoviensis* sp. nov. and its phylogenetic relationship. *Arch Microbiol* **164:** 16–23.

Farrelly, V., Rainey, F.A., and Stackebrandt, E. (1995) Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. *Appl Environ Microbiol* **61:** 2798–2801.

Felsenstein, J. (1989) PHYLIP – Phylogeny inference package (version 3.2). *Cladistics* **5:** 164–166.

Foesel, B.U., Gieseke, A., Schwermer, C., Stief, P., Koch, L., Cytryn, E., *et al*. (2008) *Nitrosomonas* Nm143-like ammonia oxidizers and *Nitrospira marina*-like nitrite oxidizers dominate the nitrifier community in a marine aquaculture biofilm. *FEMS Microbiol Ecol* **63:** 192–204.

Francis, C.A., Roberts, K.J., Beman, J.M., Santoro, A.E., and Oakley, B.B. (2005) Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc Natl Acad Sci USA* **102:** 14683–14688.

Freitag, T.E., Chang, L., Clegg, C.D., and Prosser, J.I. (2005) Influence of inorganic nitrogen management regime on the diversity of nitrite-oxidizing bacteria in agricultural grassland soils. *Appl Environ Microbiol* **71:** 8323–8334.

Gaspar, J.M., and Thomas, W.K. (2013) Assessing the consequences of denoising marker-based metagenomic data. *PLoS ONE* **8:** e60458.

Good, I.J. (1953) The population frequencies of species and the estimation of population parameters. *Biometrika* **40:** 237–264.

Gruber, N. (2004) The dynamics of the marine nitrogen cycle and its influence on atmospheric CO_2 variations. In *NATO ASI Series*. Follows, M., and Oguz, T. (eds). Dordrecht, the Netherlands: Kluwer Academic, pp. 97–148.

Gruber, N., and Galloway, J.N. (2008) An Earth-system perspective of the global nitrogen cycle. *Nature* **451:** 293–296.

Gubry-Rangin, C., Hai, B., Quince, C., Engel, M., Thomson, B.C., James, P., *et al*. (2011) Niche specialization of terrestrial archaeal ammonia oxidizers. *Proc Natl Acad Sci USA* **108:** 21206–21211.

Haaijer, S.C., Ji, K., van Niftrik, L., Hoischen, A., Speth, D., Jetten, M.S., *et al*. (2013) A novel marine nitrite-oxidizing *Nitrospira* species from Dutch coastal North Sea water. *Front Microbiol* **4:** 60.

Juretschko, S., Timmermann, G., Schmid, M., Schleifer, K.-H., Pommerening-Röser, A., Koops, H.-P., and Wagner, M. (1998) Combined molecular and conventional analyses of nitrifying bacterium diversity in activated sludge: *Nitrosococcus mobilis* and *Nitrospira*-like bacteria as dominant populations. *Appl Environ Microbiol* **64:** 3042–3051.

Keuter, S., Kruse, M., Lipski, A., and Spieck, E. (2011) Relevance of *Nitrospira* for nitrite oxidation in a marine recirculation aquaculture system and physiological features of a *Nitrospira marina*-like isolate. *Environ Microbiol* **13:** 2536–2547.

Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. (2010) Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* **12:** 118–123.

Lebedeva, E.V., Alawi, M., Fiencke, C., Namsaraev, B., Bock, E., and Spieck, E. (2005) Moderately thermophilic nitrifying bacteria from a hot spring of the Baikal rift zone. *FEMS Microbiol Ecol* **54:** 297–306.

Lebedeva, E.V., Alawi, M., Maixner, F., Jozsa, P.G., Daims, H., and Spieck, E. (2008) Physiological and phylogenetical characterization of a new lithoautotrophic nitrite-oxidizing bacterium '*Candidatus Nitrospira bockiana*' sp. nov. *Int J Syst Evol Microbiol* **58:** 242–250.

Lebedeva, E.V., Off, S., Zumbragel, S., Kruse, M., Shagzhina, A., Lücker, S., *et al*. (2011) Isolation and characterization of a moderately thermophilic nitrite-oxidizing bacterium from a geothermal spring. *FEMS Microbiol Ecol* **75:** 195–204.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., *et al*. (2006) *Archaea* predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442:** 806–809.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhu, K., *et al*. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32:** 1363–1371.

Lücker, S., Wagner, M., Maixner, F., Pelletier, E., Koch, H., Vacherie, B., *et al*. (2010) A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proc Natl Acad Sci USA* **107:** 13479–13484.

Lücker, S., Nowka, B., Rattei, T., Spieck, E., and Daims, H. (2013) The genome of *Nitrospina gracilis* illuminates the metabolism and evolution of the major marine nitrite oxidizer. *Front Microbiol* **4:** 27.

Maixner, F., Noguera, D.R., Anneser, B., Stoecker, K., Wegl, G., Wagner, M., and Daims, H. (2006) Nitrite concentration influences the population structure of *Nitrospira*-like bacteria. *Environ Microbiol* **8:** 1487–1495.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., *et al*. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376–380.

Markowitz, V.M., Ivanova, N.N., Szeto, E., Palaniappan, K., Chu, K., Dalevi, D., *et al*. (2008) IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res* **36:** D534–D538.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., *et al*. (2008) The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9:** 386.

Morgenroth, E., Obermayer, A., Arnold, E., Brühl, A., Wagner, M., and Wilderer, P.A. (2000) Effect of long-term idle periods on the performance of sequencing batch reactors. *Water Sci Technol* **41:** 105–113.

Norton, J.M., Alzerreca, J.J., Suwa, Y., and Klotz, M.G. (2002) Diversity of ammonia monooxygenase operon in autotrophic ammonia-oxidizing bacteria. *Arch Microbiol* **177:** 139–149.

Off, S., Alawi, M., and Spieck, E. (2010) Enrichment and physiological characterization of a novel *Nitrospira*-like bacterium obtained from a marine sponge. *Appl Environ Microbiol* **76:** 4640–4646.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., *et al*. (2012) *vegan: community ecology package*. R package version 2.0-3 [WWW document]. URL http://CRAN.R-project.org/package=vegan

Peres-Neto, P.R., and Jackson, D.A. (2001) How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia* **129:** 169–178.

Pester, M., Schleper, C., and Wagner, M. (2011) The *Thaumarchaeota*: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol* **14:** 300–306.

Pester, M., Rattei, T., Flechl, S., Grongroft, A., Richter, A., Overmann, J., *et al*. (2012) *amoA*-based consensus

phylogeny of ammonia-oxidizing archaea and deep sequencing of *amoA* genes from soils of four different geographic regions. *Environ Microbiol* **14:** 525–539.

Poly, F., Wertz, S., Brothier, E., and Degrange, V. (2008) First exploration of *Nitrobacter* diversity in soils by a PCR cloning-sequencing approach targeting functional gene *nxrA*. *FEMS Microbiol Ecol* **63:** 132–140.

Prosser, J.I. (2011) Soil nitrifiers and nitrification. In *Nitrification*. Ward, B.B., Arp, D.J., and Klotz, M.G. (eds). Washington, DC, USA: ASM Press, pp. 347–383.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.

Purkhold, U., Pommering-Röser, A., Juretschko, S., Schmid, M.C., Koops, H.-P., and Wagner, M. (2000) Phylogeny of all recognized species of ammonia oxidizers based on comparative 16S rRNA and amoA sequence analysis: implications for molecular diversity surveys. *Appl Environ Microbiol* **66:** 5368–5382.

R-Core-Team (2013) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Richter, M., and Rossello-Mora, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* **106:** 19126–19131.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., *et al*. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* **61:** 539–542.

Rothery, R.A., Workun, G.J., and Weiner, J.H. (2008) The prokaryotic complex iron-sulfur molybdoenzyme family. *Biochim Biophys Acta* **1778:** 1897–1929.

Rotthauwe, J.-H., Witzel, K.-P., and Liesack, W. (1997) The ammonia monooxygenase structural gene *amoA* as a functional marker: molecular fine-scale analysis of natural ammonia-oxidizing populations. *Appl Environ Microbiol* **63:** 4704–4712.

Schiex, T., Gouzy, J., Moisan, A., and de Oliveira, Y. (2003) FrameD: a flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. *Nucleic Acids Res* **31:** 3738–3741.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., *et al*. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537–7541.

Schramm, A., de Beer, D., Wagner, M., and Amann, R. (1998) Identification and activities in situ of *Nitrosospira* and *Nitrospira* spp. as dominant populations in a nitrifying fluidized bed reactor. *Appl Environ Microbiol* **64:** 3480–3485.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., and Ideker, T. (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27:** 431–432.

Sorokin, D., Lücker, S., Vejmelkova, D., Kleerbezem, R., Muyzer, G., Sinninghe-Damsté, J., *et al*. (2012) Nitrification expanded: discovery, physiology, and genomics of a nitrite-oxidizing bacterium from the phylum *Chloroflexi*. *ISME J* **6:** 2245–2256.

Spieck, E., and Bock, E. (2005) The lithoautotrophic nitrite-oxidizing bacteria. In *Bergey's Manual of Systematic Bacteriology*. Staley, J.T., Boone, D.R., Brenner, D.J., de Vos, P., Garrity, G.M., Goodfellow, M., *et al*. (eds). New York, NY, USA: Springer Science+Business Media, pp. 149–153.

Spieck, E., Ehrich, S., Aamand, J., and Bock, E. (1998) Isolation and immunocytochemical location of the nitrite-oxidizing system in *Nitrospira moscoviensis*. *Arch Microbiol* **169:** 225–230.

Spieck, E., Hartwig, C., McCormack, I., Maixner, F., Wagner, M., Lipski, A., and Daims, H. (2006) Selective enrichment and molecular characterization of a previously uncultured *Nitrospira*-like bacterium from activated sludge. *Environ Microbiol* **8:** 405–415.

Stackebrandt, E., and Ebers, J. (2006) Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* **33:** 152–155.

Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688–2690.

Starkenburg, S.R., Chain, P.S., Sayavedra-Soto, L.A., Hauser, L., Land, M.L., Larimer, F.W., *et al*. (2006) Genome sequence of the chemolithoautotrophic nitrite-oxidizing bacterium *Nitrobacter winogradskyi* Nb-255. *Appl Environ Microbiol* **72:** 2050–2063.

Stecher, B., Berry, D., and Loy, A. (2013) Colonization resistance and microbial ecophysiology: using gnotobiotic mouse models and single-cell technology to explore the intestinal jungle. *FEMS Microbiol Rev* **37:** 793–829.

Stein, L.Y. (2011) Heterotrophic nitrification and nitrifier denitrification. In *Nitrification*. Ward, B.B., Arp, D.J., and Klotz, M.G. (eds). Washington, DC, USA: ASM Press, pp. 95–114.

Stein, L.Y., and Arp, D.J. (1998) Loss of ammonia monooxygenase activity in *Nitrosomonas europaea* upon exposure to nitrite. *Appl Environ Microbiol* **64:** 4098–4102.

Sun, S.L., Chen, J., Li, W.Z., Altintas, I., Lin, A., Peltier, S., *et al*. (2011) Community cyberinfrastructure for Advanced Microbial Ecology Research and Analysis: the CAMERA resource. *Nucleic Acids Res* **39:** D546–D551.

Taylor, M.W., Radax, R., Steger, D., and Wagner, M. (2007) Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* **71:** 295–347.

Tourna, M., Freitag, T.E., Nicol, G.W., and Prosser, J.I. (2008) Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environ Microbiol* **10:** 1357–1364.

Vanparys, B., Spieck, E., Heylen, K., Wittebolle, L., Geets, J., Boon, N., and De Vos, P. (2007) The phylogeny of the genus *Nitrobacter* based on comparative rep-PCR, 16S rRNA and nitrite oxidoreductase gene sequence analysis. *Syst Appl Microbiol* **30:** 297–308.

Wagner, M., Rath, G., Amann, R., Koops, H.-P., and Schleifer, K.-H. (1995) In situ identification of ammonia-oxidizing bacteria. *Syst Appl Microbiol* **18:** 251–264.

Watson, S.W., Bock, E., Valois, F.W., Waterbury, J.B., and Schlosser, U. (1986) *Nitrospira marina* gen. nov. sp. nov.: a chemolithotrophic nitrite-oxidizing bacterium. *Arch Microbiol* **144:** 1–7.

Wertz, S., Poly, F., Le Roux, X., and Degrange, V. (2008) Development and application of a PCR-denaturing gradient gel electrophoresis tool to study the diversity of *Nitrobacter*-like *nxrA* sequences in soil. *FEMS Microbiol Ecol* **63:** 261–271.

von Wintzingerode, F., Göbel, U.B., and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21:** 213–229.

Woese, C.R. (1987) Bacterial evolution. *Microbiol Rev* **51:** 221–271.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. 16S rRNA gene-based phylogeny of the genus *Nitrospira* including cultured and uncultured representatives. Cultured *Nitrospira* and environmental sequences that are indicated in boldface served as anchor points in the comparison of *nxrB* and 16S rRNA gene-based *Nitrospira* phylogenies (Fig. 2 in the main text). A majority rule consensus tree is shown. Bootstrap support is indicated at individual branches. In total, 1427 unambiguously aligned nucleotide positions were used in this analysis and branches leading to shorter sequences are indicated by dotted lines. The scale bar indicates 5% estimated sequence divergence.

Fig. S2. Procrustes analysis. Comparison of ordinations of *amoA* (red) and *nxrB* (black) datasets was determined using either (A) presence/absence or (B) relative abundance based Bray–Curtis distance. Black lines connect the same sample in the two ordinations. The small distances between the same samples indicate a high correlation between the two datasets [$r = 0.95$ ($P = 0.001$) for presence/absence, $r = 0.93$ ($P = 0.001$) for relative abundance].

Table S1. NxrA, NxrB and NxrC copy numbers and amino acid identities of nitrite-oxidizing bacteria with a sequenced genome.

Table S2. Range of nucleic acid sequence identities between the different *nxrB* copies within and between the analysed *Nitrospira* enrichments and pure cultures.

Table S3. Range of amino acid sequence identities between the different NxrB copies within and between the analysed *Nitrospira* enrichments and pure cultures.

Table S4. Environmental samples and cultured *Nitrospira* analysed in this study.

Table S5. Distribution and relative abundances of *Nitrospira*-related *nxrB* amplicons in various soils. Clusters with ≥ 10% relative abundance are printed in bold. The phylogenetic positions of the novel '454 clusters' are shown in Fig. 4 in the main text.

Table S6. Fraction of rare *Nitrospira nxrB* OTUs, which are represented by only one or two quality-controlled 454 reads, in relation to the total number of obtained OTUs at 95% *nxrB* sequence identity in the individual analysed soils.

File S1. *Nitrospira nxrB* ARB database encompassing the consensus tree and the source alignment of the *nxrB* reference sequences.