Degree project

UNIVERSITY OF SKÖVDE

SURVEY OF NO-PRODUCING NITRITE REDUCTASES IN POTENTIALLY MOBILE GENETIC ELEMENTS

Master Degree Project (120 credits) in Systems Biology

Second Cycle 45 credits

Autumn term

Student: Md Shakil Hasan

Supervisor: Grace Pold, Swedish University of Agricultural Sciences (SLU), Uppsala.

Examiner: Sanja Jurcevic, University of Skövde.

Abstract

Denitrification is a key microbial process that completes the nitrogen cycle by reducing nitrate to dinitrogen gas. A critical intermediate step in this pathway is the reduction of nitrite to nitric oxide, catalyzed by two non-homologous enzymes: copper nitrite reductase (NirK) and cytochrome cd_1 nitrite reductase (NirS). This process enables microbial respiration under oxygen-limited conditions and plays a vital role in both natural and engineered ecosystems. Given the ecological importance of denitrification, understanding the mobility of nir genes is essential, especially in environments subject to anthropogenic influence or rapid microbial turnover. To explore the potential for horizontal gene transfer, public databases of mobile genetic elements (MGEs)—including plasmids (IMG-PR), viruses (IMG-VR; PhageScope), and integrative/conjugative elements (ICEberg 3.0)—were systematically screened. The analysis focused on identifying instances of *nirK* and *nirS* within these MGEs, determining their host taxa, and evaluating their distribution across various biomes. Enrichment analysis was performed to detect overrepresentation in specific ecological contexts, particularly aquatic and engineered environments. Across all MGEs, *nirK* was consistently more prevalent than *nirS*, although both genes remained rare overall. Enrichment analysis revealed that *nirK* is disproportionately found in aquatic and engineered sub-biomes such as bioreactors, built environments, and industrial production systems. In contrast, *nirS* was largely confined to engineered niches. Example rates included approximately 1–3% nirK among plasmids from wastewater-related contexts, while *nirS* was detected at ≤0.1% overall, with occasional spikes in industrial settings. Viral *nirK* was observed sporadically in Caudoviricetes infecting Gammaproteobacteria, whereas *nirS* was not detected in any viral datasets surveyed. These findings support a model in which plasmids serve as the dominant vectors for *nirK* dissemination across taxonomically broad hosts and ecologically engineered niches, while *nirS* shows restricted mobility. This asymmetry in gene transfer potential has important implications for microbial community structure, process stability, and nitrous oxide (N_2O) emissions in managed ecosystems. Understanding the differential mobility of these genes can inform strategies to optimize nitrogen removal and mitigate greenhouse gas emissions in wastewater treatment and other engineered environments.

Keywords: denitrification, nirK, nirS, plasmids, bacteriophages, ICEs, horizontal gene transfer, enrichment analysis, bioreactors, wastewater.

Popular scientific summary

Nitrogen is recognized as an essential element for life, constituting a fundamental component of nucleic acids (DNA and RNA), amino acids that form proteins, and atmospheric gases. It plays a critical role in cellular structure, metabolic processes, and genetic information transfer. But nitrogen also has a "dark side." Human activities such as agriculture and industry release huge amounts of nitrogen compounds like nitrate into the environment. When too much nitrogen builds up, it pollutes rivers and lakes, causes algal blooms, and contributes to climate change through nitrous oxide (N_2O) emissions — a greenhouse gas almost 300 times more powerful than carbon dioxide.

Fortunately, microbes have a natural way of cleaning up excess nitrogen: a process called denitrification. In this process, microbes "breathe" nitrate instead of oxygen, step by step converting it back into harmless nitrogen gas (N_2) that returns to the atmosphere. A crucial step in this chain is carried out by special enzymes called nitrite reductases. These enzymes exist in two forms, known as NirK and NirS. Both do the same job, but they are built differently and are found in different groups of microbes. Through a process called horizontal gene transfer (HGT), genes can "jump" between microbes, not only from parent to offspring. This is made possible by mobile genetic elements such as plasmids (small DNA circles), viruses that infect bacteria (phages), and genomic islands (chunks of DNA integrated into chromosomes). These act like shuttles, carrying genes from one microbe to another, potentially spreading useful traits such as antibiotic resistance — or in this case, the ability to perform denitrification.

In this thesis, the presence of nitrite reductase genes (*nirK* and *nirS*) on mobile genetic elements was explored. Large public databases of plasmids, viruses, and genomic islands were used to search for and analyze the occurrence of these genes. The results were clear: plasmids play a much bigger role than viruses or genomic islands in carrying nitrite reductase genes. Even so, the genes were rare. Less than 1% of all plasmids contained *nirK* and *nirS* was even scarcer. But where they did appear, patterns emerged. *nirK* was found in a wide range of microbes and was especially common in aquatic and engineered environments such as wastewater treatment plants. *nirS*, in contrast, was almost entirely limited to engineered environments and very rare elsewhere.

Wastewater treatment plants, bioreactors, and other engineered systems are designed to reduce pollution, but they are also hotspots where microbes exchange genes. The discovery that plasmids carry denitrification genes suggests that microbial communities in these systems may adapt quickly to changing conditions, maintaining nitrogen removal even when species composition shifts. Importantly, plasmid-mediated transfer appears to favor *nirK* over *nirS*, showing that the two strategies for nitrite reduction are not equally mobile. This uneven mobility matters because it could make *nirK*-based denitrification more resilient and flexible, while *nirS*-based pathways remain more constrained. By understanding which genes are mobile, engineers can better predict how treatment systems perform and even design conditions that enhance stable nitrogen removal while minimizing nitrous oxide emissions. By improving understanding of how nitrogen-related genes are shared and spread by microbes, insights are gained that may assist society in addressing pressing environmental challenges. This knowledge could guide the design of wastewater treatment and bioreactor processes — for example, by fostering conditions that support plasmid-mediated *nirK* transfer, thereby stabilizing denitrification performance and limiting nitrous oxide emissions. In a world facing climate change and growing agricultural pressures, even the smallest players — microbes and their mobile DNA — may help secure cleaner water and a more stable nitrogen cycle.

Abbreviations

ARB – Integrated software environment for sequence data and phylogeny

FDR – False discovery rate (Benjamini–Hochberg)

GTDB – Genome Taxonomy Database

HGT – Horizontal gene transfer

HMM – Hidden Markov Model

ICE – Integrative and conjugative element

IMG-PR/VR – Integrated Microbial Genomes (Plasmids/Viruses)

MGE – Mobile genetic element

OMZ – Oxygen minimum zone

Table of Contents

Introduction

The nitrogen (N) cycle is one of the most important biogeochemical processes sustaining life on Earth, mediating the transformation of nitrogen species between atmospheric, terrestrial, and aquatic reservoirs (Galloway et al., 2004). Disruption of this cycle through anthropogenic nitrogen inputs has profound environmental consequences, including eutrophication, biodiversity loss, and the emission of nitrous oxide (N_2O), a potent greenhouse gas with a global warming potential nearly 300 times greater than carbon dioxide (CO_2) on a 100-year timescale (Change, 2023). Global assessments show that anthropogenic fixation of reactive nitrogen now exceeds natural terrestrial biological fixation (Galloway et al., 2008), creating unprecedented pressure on microbial processes that regulate nitrogen removal and transformation.

Within this cycle, microbial denitrification plays a central role by reducing nitrate (NO_3^-) to nitrogen gas (N_2), thereby closing the nitrogen loop and mitigating the accumulation of nitrogen in ecosystems (Zumft, 1997). For the cell, denitrification provides a vital respiratory strategy under oxygen-limited conditions, where nitrate and nitrite serve as alternative electron acceptors to sustain energy conservation and growth (Philippot, 2002). This process is carried out by a phylogenetically diverse group of bacteria and archaea and involves multiple enzymes encoded by distinct functional genes. Among these, nitrite reductases (Nir) catalyze the key step of reducing nitrite (NO_2^-) to nitric oxide (NO), a short-lived intermediate that links upper and lower steps of the denitrification pathway (Averill, 1996) (Figure 1).

Two structurally and evolutionarily distinct nitrite reductases are known: the copper-containing NirK and the cytochrome cd_1-containing NirS. While functionally equivalent, these enzymes have different evolutionary origins and are often distributed in a mutually exclusive manner within microbial genomes (Zumft, 1997; Jones et al., 2008). Structural differences are striking: NirK requires copper cofactors, while NirS requires the complex heme d_1, whose biosynthesis depends on auxiliary proteins such as NirN, NirF, and NirQ (Richardson et al., 2009; Einsle & Kroneck, 2004). The presence of *nirK* or *nirS* genes is therefore an important molecular marker for studying the ecology, evolution, and functional diversity of microorganisms capable of nitrite reduction (Graf et al., 2014; Varin et al., 2011). Importantly, because nitrite reduction contributes directly to cellular energy metabolism, acquisition of nir genes via horizontal gene transfer could confer a strong selective advantage in fluctuating redox environments.

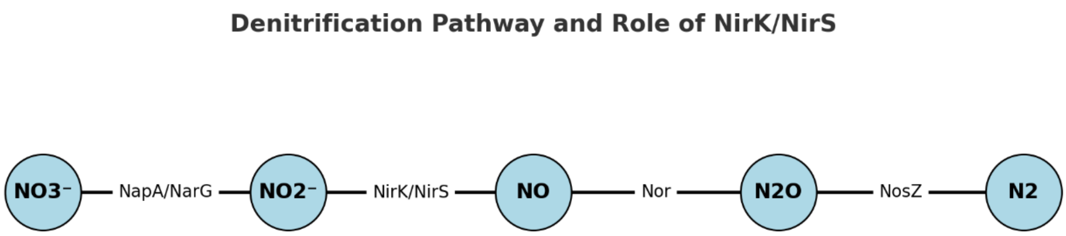

Figure 1: Denitrification pathway highlighting the role of NirK and NirS. Nitrate (NO_3^-) is sequentially reduced to dinitrogen (N_2) gas through nitrite (NO_2^-), nitric oxide (NO), and nitrous oxide (N_2O). The key enzymes are NapA/NarG (nitrate reductases), NirK/NirS (nitrite reductases), Nor (nitric oxide reductase),

and NosZ (nitrous oxide reductase). Circles denote chemical species, while enzyme names are shown on arrows to indicate the protein catalysts mediating each transformation.

Mobile Genetic Elements and Horizontal Gene Transfer

Microbial genomes are dynamic entities shaped not only by vertical inheritance but also by horizontal gene transfer (HGT), the movement of genetic material between organisms outside of parent–offspring relationships. HGT accelerates microbial adaptation, facilitates the spread of ecologically important traits, and plays a crucial role in the evolution of metabolic pathways (Frost et al., 2005; Soucy et al., 2015).

Mobile genetic elements (MGEs) such as plasmids, bacteriophages, and genomic islands are the major vectors of HGT, differing in the size of DNA they mobilize and their host range. (Figure 2)

Plasmids are extrachromosomal DNA molecules capable of autonomous replication. They can range from a few kilobases to several hundred kilobases in size, and thus can carry anywhere from a single accessory gene to entire metabolic operons. Some plasmids have a narrow host range (confined to strains of a single species), while broad-host-range plasmids can transfer across genera and even across higher taxonomic boundaries (Norman et al., 2009). This makes plasmids particularly important candidates for spreading nitrite reductase genes. Moreover, plasmids frequently co-localize with resistance genes or other metabolic functions, enabling the co-transfer of multiple adaptive traits (Smalla et al., 2015; Shintani et al., 2015).

Bacteriophages (viruses infecting bacteria) can mediate gene transfer through transduction. Specialized transduction events usually transfer one or a few host genes, while generalized transduction can occasionally move larger genomic regions (BrüSsow et al., 2004). Phage host ranges are often narrow, typically restricted to strains within a species or genus, although some phages can infect across broader taxonomic groups (Touchon et al., 2017). Phages carrying metabolic accessory genes, including those linked to biogeochemical cycling, have been increasingly documented (Roux et al., 2018). Notably, viral auxiliary metabolic genes have been implicated in nitrogen cycling in oxygen-depleted oceanic waters, where phages may mediate denitrification steps (Gazitúa et al., 2020a).

Genomic islands are chromosomally integrated DNA segments, often tens to hundreds of kilobases in size, that are acquired via HGT. They can carry multiple genes or operons, frequently including mobility genes such as integrases and transposases (Juhas et al., 2009). Some are mobilized passively, while integrative and conjugative elements (ICEs) encode their own transfer machinery, enabling direct cell-to-cell transfer. ICEs typically move between strains of the same or closely related species but in some cases cross genus-level boundaries (Johnson & Grossman, 2015). ICEs have been shown to disseminate genes involved in symbiosis and pathogenesis, underscoring their ecological and evolutionary significance (Bellanger et al., 2013).

These distinctions are critical when considering the mobility of denitrification genes: a plasmid or genomic island could, in principle, transfer a full nir operon together with its maturation factors (*nirS* with *nirN*/*nirF*), whereas phage transduction events are more likely to transfer individual genes or gene fragments.

Figure 2. Mechanisms of horizontal transfer of nir genes

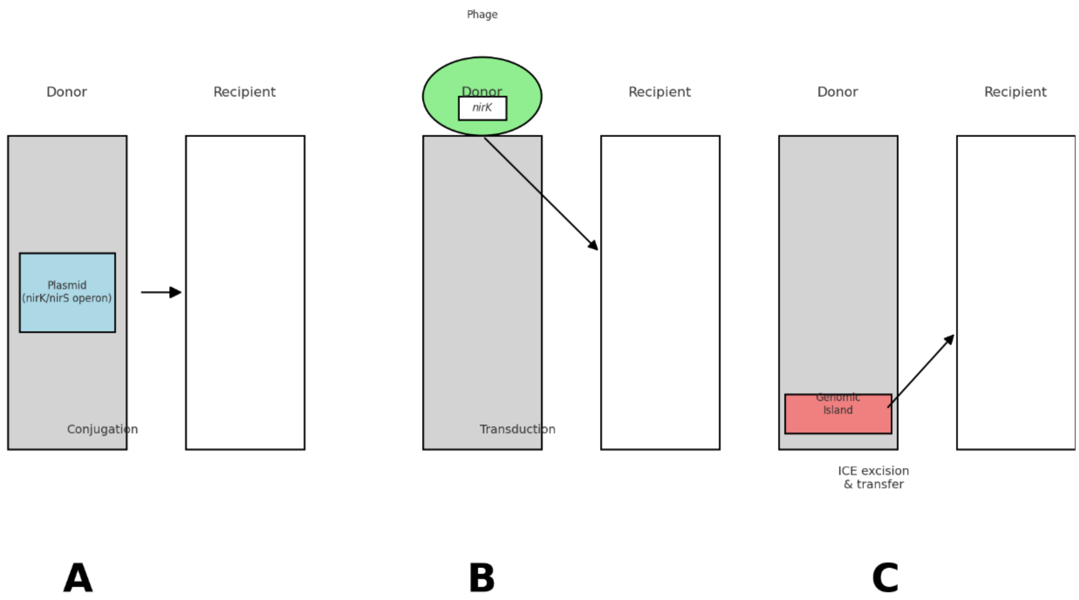

Figure 2. Mechanisms of horizontal transfer of *nir* genes. (A) Plasmids can mobilize *nirK* or *nirS* operons by conjugation, transferring DNA directly between cells. (B) Bacteriophages may capture and package *nirK* fragments during viral replication and deliver them into new hosts via transduction. (C) Integrative and conjugative elements (ICEs) are genomic islands that can excise from the donor chromosome, transfer through conjugation machinery, and integrate into recipient genomes. These processes illustrate how nir genes may spread across diverse taxa and environments, with implications for denitrification capacity under fluctuating oxygen and nitrogen conditions.

In agricultural soils, organisms encoding nitrite reductase genes (*nirK* or *nirS*) often gain a selective advantage under nitrate-rich, oxygen-limited conditions (Wang et al., 2020). This advantage reflects their capacity to reduce nitrite when other microbes cannot, and it can influence local patterns of N_2O accumulation. Beyond crop soils, similar dynamics have been observed in fertilized grasslands and rice paddies, where fluctuations in water saturation and oxygen availability create niches favoring microbes with these genes. The presence of nir genes thus represents not only a metabolic advantage for individual organisms but also a driver of greenhouse gas emissions from agriculture, linking microbial genetics to global climate processes.

In wastewater treatment plants, *nirK*- or *nirS*-carrying organisms persist through low-oxygen phases because they can continue energy generation via nitrite reduction, whereas microbes lacking these genes decline (Hallin et al., 2006). Wastewater bioreactors are among the most studied engineered ecosystems for denitrification, where microbial communities experience tightly managed oxic–anoxic cycles. Organisms harboring nir genes contribute to stable nitrogen removal performance and are thought to buffer the system against fluctuations in influent nitrate loads. The enrichment of *nirK* and *nirS* in these contexts illustrates how functional genes can be selectively maintained under strong anthropogenic pressure, with direct consequences for treatment efficiency and the release of N_2O as a byproduct.

In marine oxygen minimum zones, diverse lineages encoding nitrite reductases are key contributors to nitrogen loss processes, highlighting the ecological importance of maintaining functional *nirK* and *nirS* variants under sustained hypoxia (Ward et al., 2009). These regions account for a large fraction of global fixed nitrogen loss and are hotspots for microbial innovation.

Lineages with *nirK* or *nirS* not only control local N_2O production but also shape the balance between denitrification and competing processes such as anammox. The persistence of nitrite reductase genes in these low-oxygen marine systems demonstrates how gene-level traits scale up to influence global nitrogen budgets.

Importantly, nitrite reductases are not exclusive to classical denitrifiers: ammonia-oxidizing bacteria and archaea also harbor *nirK*, and these clades may differ in their evolutionary origins from those in canonical denitrifiers. This overlap complicates ecological interpretation, since *nirK* sequences detected in metagenomes may derive from organisms involved primarily in nitrification rather than denitrification. Because this study does not resolve the source lineages of each sequence, it is most appropriate to take a gene-based perspective. By focusing on *nirK* and *nirS* as genetic markers, rather than inferring the entire denitrification pathway, it becomes possible to describe broader patterns of ecological distribution and potential mobility without assuming functional equivalence across clades.

If *nirK* and *nirS* are mobilized via plasmids, phages, or genomic islands, their ecological distribution may be far more dynamic than previously assumed, with consequences for community adaptation and nitrogen fluxes at multiple scales. Such mobility could accelerate the spread of nitrite reduction capacity in environments undergoing rapid change, such as agricultural soils receiving heavy fertilizer inputs, wastewater reactors facing variable inflows, or expanding marine OMZs under climate-driven deoxygenation.

If these genes are mobilized via plasmids, phages, or genomic islands, their ecological distribution may be far more dynamic than previously assumed. Such mobility could:

1. Enable rapid adaptation of microbial communities to oxygen limitation and nitrate/nitrite enrichment, particularly in disturbed or engineered environments where redox conditions fluctuate strongly.

2. Facilitate the spread of denitrification capacity to phylogenetically unrelated taxa, potentially creating novel community interactions and altering competitive dynamics.

3. Alter biogeochemical nitrogen fluxes on local and global scales by influencing rates of N_2O and N_2 production, thereby coupling microbial gene flow to greenhouse gas emissions and nitrogen availability.

Moreover, examining *nirK* and *nirS* in the context of MGEs may help elucidate patterns of phylogenetic incongruence between functional gene trees and host species trees — a hallmark of HGT events. Tracking such incongruence not only provides evidence for past horizontal transfers but also highlights the ongoing role of gene mobility in shaping microbial contributions to the nitrogen cycle. In this way, *nir* genes serve as a powerful model for understanding how ecological selection, metabolic function, and horizontal transfer intersect to influence microbial community evolution and ecosystem processes.

Knowledge Gaps and Rationale for This Study

Although *nirK* and *nirS* have been extensively studied in the context of denitrification (Zumft, 1997; Graf et al., 2014), most research has neglected the potential for redistribution of these genes on ecological timescales. Few studies have systematically examined their prevalence and phylogenetic patterns when associated with MGEs (Pold et al., 2024).

Distribution — Which bacterial and archaeal taxa carry *nirK*/*nirS* on plasmids, phages, or genomic islands?

Biome specificity — Are certain environments (e.g., soil, marine, engineered) more likely to harbor MGE-associated *nirK*/*nirS*?

Addressing these questions will advance our understanding of the mechanisms driving nitrogen cycle gene dissemination, with implications for both environmental microbiology and microbial evolution.

Aim

This thesis investigates the horizontal transfer of the copper nitrite reductase (*nirK*) and cytochrome cd_1 nitrite reductase (*nirS*) genes via mobile genetic elements (MGEs), with the overarching goal of clarifying how these genes move across microbial lineages and environments that are central to the nitrogen cycle. Horizontal gene transfer plays a crucial role in microbial adaptation, yet the relative contribution of different MGEs to the spread of nitrite reductases remains poorly resolved. Addressing this gap is important, because the mobility of *nir* genes has the potential to alter ecosystem nitrogen fluxes and reshape microbial community functions under changing environmental conditions.

To achieve this goal, the study first aims to quantify the prevalence of *nirK* and *nirS* across major classes of MGEs—including plasmids, bacteriophages, and integrative conjugative elements—by systematically screening large, publicly available genome and metagenome databases. This will provide a baseline understanding of how frequently these genes occur on different vectors of horizontal transfer.

The second aim is to map the distribution of *nir* genes across microbial taxa and environmental biomes. By linking gene presence to host phylogeny and ecological context, this analysis will identify which lineages are most likely to carry transferable nitrite reductases, and in which environments these mobile copies are enriched. Special attention will be given to ecosystems with strong nitrogen turnover, such as agricultural soils, wastewater treatment plants, and marine oxygen minimum zones, where nitrite reduction plays a particularly important ecological role.

Finally, the study seeks to evaluate ecological patterns in the mobility of *nirK* and *nirS*, with emphasis on how horizontal transfer may contribute to the resilience or vulnerability of denitrification under environmental change. This includes testing whether the two genes differ in their likelihood of being mobilized, whether mobility is associated with specific host groups or environmental conditions, and how these differences may influence the stability of nitrogen removal processes and greenhouse gas emissions.

Together, these objectives will provide the first systematic comparison of *nirK* and *nirS* mobility across multiple MGE classes, offering new insights into the genetic mechanisms that underlie microbial contributions to the nitrogen cycle.

Materials and Methods

Reference HMM construction

Two curated reference alignments were used as starting points, representing manually verified *nirK* and *nirS* sequences. These alignments were originally compiled and published by Pold et al. (2024), and each sequence retains its GenBank accession number, ensuring traceability to the source database.

These alignments included diverse, high-quality representatives of each gene family and served as the basis for building Hidden Markov Models (HMMs) with the hmmbuild program from HMMER v3.3.2 (Eddy, 2011). The resulting reference HMMs were then used as query models for all subsequent database searches.

Protein search strategy

Predicted protein sequences from plasmid, phage, and genomic island datasets were used as search targets. Searches were carried out with hmmsearch (HMMER v3.3.2) using the NirK and NirS HMM profiles as queries. To scale the statistical calibration to the size of each database, the -Z parameter was set to approximate the number of protein-coding genes in the dataset. For whole genomes, this was estimated as 1000 × number of genomes, reflecting ~1000 genes per bacterial genome. For mobile genetic elements (MGEs), which are much smaller, the total number of expected genes was estimated as database size in megabytes ÷ 3500, since average prokaryotic genes are ~3.5 kb.

The --cpu 0 flag was used to parallelize across all available cores, and --tblout was specified for compact tabular output. Importantly, no initial E-value filter was applied at this stage. Instead, a moderate bit-score cutoff of -T 50 was used to retain potential hits for further evaluation.

Secondary filtering and codon anchoring

Candidate protein hits were processed with a custom Python 3.12.2 pipeline. This script retrieves the corresponding nucleotide segments for each protein hit using the sequence headers and genomic coordinates. The script then:

Realigns protein hits to the curated reference alignment using hmmalign --mapali, ensuring consistent positional homology across hits and references. Applies a stringent E-value threshold (-e 1e-20) to remove weak matches. Flags and removes problematic sequences, including those containing internal stop codons, frameshifts. Produces two synchronized outputs: an aligned protein FASTA (prot_aligned.fasta) used for tree building and phylogenetic validation, and a codon-aligned nucleotide FASTA (*_aln.fna), anchored to the protein alignment. The nucleotide alignment was preserved to facilitate codon-aware downstream analyses and metadata mapping, but not used for tree inference.

Alignment curation and phylogenetic reconstruction

The amino-acid sequences recovered from the secondary filtering step were imported into ARB v7.0 (Ludwig et al., 2004) for manual curation. Within ARB, sequences were first translated and codon-linked to ensure consistency between nucleotide and protein records, but all phylogenetic analyses were carried out on the protein alignments.

Low-quality sequences were systematically removed. This included:

● Truncated sequences, identified by incomplete coverage at the N- or C-terminus relative to the reference were removed.

● Frameshifted sequences were already excluded during secondary filtering.

● Ambiguously aligned sequences, identified in ARB as containing long stretches of gaps or poor positional homology.

Alignment refinement

To refine alignments, highly gappy or poorly aligned columns were masked using ARB, and the conserved core of each protein alignment was retained. Inconsistent terminal regions were trimmed, and the curated amino-acid alignments were exported in FASTA format for tree inference. Curation decisions were recorded in ARB metadata fields (e.g., tagging "short" for truncated entries).

Phylogenetic validation

Maximum-likelihood gene trees were reconstructed from the curated protein alignments using FastTreeMP v2.1.11. Trees were used as a quality-control step to verify that candidate sequences clustered within the *nirK* or *nirS* ingroups and remained separate from paralogues (e.g., NirN/NirF). Sequences with outlier placements triggered re-inspection of alignment quality and motifs and were removed if inconsistent. Trees were visualized in ARB and linked to metadata for auditing; they were not intended to infer species relationships, but to validate gene identity prior to ecological analyses.

Motif-based gene validation was performed to separate nirK and nirS from closely related paralogues such as NirN and NirF. Diagnostic amino-acid motifs were examined within ARB's alignment viewer:

● NirS candidates typically contained the CaGCHg heme-binding motif, -LHD- active-site residues, and the PHPGpG repeat.

● NirN and NirF sequences were distinguished from NirS by motif substitutions (e.g., -PyD- or -LDD-) and absence of the PHPGpG pattern.

● NirK candidates were validated by the presence of copper-binding motifs specific to the ingroup (e.g., HniDfH-, -YHCap-, -trpHvIG-), whereas outgroup sequences contained extra histidine insertions disrupting these patterns.

Only sequences passing both alignment-based trimming and motif validation were retained as high-confidence *nirK* or *nirS* for ecological and comparative analysis.

Taxonomic and biome classification

Taxonomic and biome metadata were linked to curated protein sequences according to their source database:

• Plasmids: Host taxonomy and biome annotations were retrieved from the Integrated Microbial Genomes & Plasmids (IMG-PR) resource (Chen et al., 2021).

• Phages: Host taxonomy, lifestyle, and biome information were obtained from IMG-VR and PhageScope metadata (Roux et al., 2020; R. H. Wang et al., 2023).

• Genomic islands (ICEs): Taxonomy and ecological metadata were extracted from the ICEberg 3.0 dataset (Liu et al., 2019).

Host taxonomy was harmonized to the phylum level using the Genome Taxonomy Database (GTDB R89) (Parks et al., 2020). Biomes were grouped into three broad categories — environmental, engineered, and host-associated — based on curated metadata fields provided by the source databases.

Data processing and visualization

All downstream analyses were performed in R v4.3.3 (R Core Team, 2023). Protein hit tables (including host taxonomy, biome, and curated gene annotations) were merged into a single metadata table. Data summaries followed a reproducible pipeline built around the tidyverse framework (Wickham et al., 2016):

• dplyr was used for filtering, grouping, and summarizing gene counts across taxonomic and biome categories (e.g., group_by(phylum, biome) %>% summarise(count = n())) (Wickham et al., 2023a).

• tidyr was used to reshape metadata into long or wide formats for downstream plotting and statistical analysis (Wickham & Girlich, 2023).

• ggplot2 v3.5.1 was used for visualization, including: Stacked proportional bar plots (geom_bar(position = "fill")) to show the relative distribution of *nirK* and *nirS* across phyla and biomes. Absolute count bar plots to visualize total numbers of hits per phylum or biome. Heatmaps (geom_tile() with scale_fill_gradient()) for combined phylum × biome distributions, allowing detection of niche-specific enrichments (Wickham, 2016).

High-confidence calls were ensured by the HMM thresholds, alignment/motif validation, and trimming; visualizations were used post-QC to contextualize results and flag rare anomalies for manual review.

Statistical analysis

All analyses were performed in R v4.3.3 (R Core Team, 2023). To reduce non-independence from closely related records, results were aggregated at the operational taxonomic unit (OTU) level wherever available: plasmid OTUs (pOTUs) in IMG-PR (Chen et al., 2021) and viral OTUs (vOTUs) in IMG-VR (Roux et al., 2020). For ICEs, records were deduplicated using the dataset's unique element identifiers from ICEberg 3.0 (Liu et al., 2019), with one observation per unique ICE. For each OTU (or deduplicated element), *nirK*/*nirS* status was set to "present" if any member of the cluster carried the gene. A sensitivity check confirmed that OTU-level results were consistent in direction with element-level summaries.

The primary quantities reported are counts and proportions of OTUs carrying *nirK* or *nirS* across biomes and phyla. Because these databases are treated as the full populations under study, only point estimates (counts and proportions) are reported, and sampling-based confidence intervals for prevalence are not computed.

To assess within-database enrichment or depletion of *nirK* or *nirS* in specific categories, 2×2 tables were constructed contrasting presence vs. absence inside vs. outside each focal biome or phylum. Association strength was summarized with odds ratios (OR), and significance was evaluated with Fisher's exact test (fisher.test in base R). When multiple categories were tested, p-values were adjusted using the Benjamini–Hochberg procedure to control the false discovery rate (FDR), with $q < 0.05$ treated as significant (Benjamini & Hochberg, 1995). Categories with very small denominators were flagged and interpreted cautiously.

For visualization, absolute counts and relative proportions (stacked bars and heatmaps) were shown, derived from the OTU-level tables. Code for data wrangling and plotting used tidyverse packages: dplyr (Wickham et al., 2023a) for filtering and aggregation, tidyr (Wickham & Girlich, 2023) for reshaping metadata, and ggplot2 (Wickham, 2016) for graphics. Enrichment tests were performed with base R functions (fisher.test, p.adjust).

Results

Screening overview across MGEs

Across the IMG-VR (177,361 records) and PhageScope (4,637 records) viral datasets, 11 *nirK*-positive viral entries were detected, while no *nirS* was identified. In the ICEberg 3.0 curated set (1,119 integrative and conjugative elements), one *nirK*-positive element was found, and no *nirS* was observed. In IMG-PR, the entire plasmid collection (n = 699,973) was searched, resulting in the identification of 5,332 *nirK*-positive plasmids (0.762%) and 92 *nirS*-positive plasmids (0.013%). Thus, *nirK* was found to be approximately 58× more frequent than *nirS* among plasmids (5,332 vs 92). Consistently, the overwhelming majority (>99%) of plasmids were not found to encode either nitrite reductase gene.

Distribution of plasmid-borne *nirK* and *nirS* across biomes

For biome-resolved summaries, plasmids with biome annotations (n = 129,282). In these data, *nirK* remained consistently more prevalent than *nirS*, but absolute proportions were low across all contexts. Engineered sub-biomes (bioreactors, wastewater, industrial production) showed the highest *nirK* frequencies (~1–3% within sub-biomes), whereas *nirS* was rare and typically <0.1%. Environmental aquatic plasmids showed lower but detectable *nirK* (~1% scale), environmental terrestrial were lower still (≈0.2%), and host-associated plasmids were near background (≤0.02% for *nirK*; *nirS* essentially absent). These patterns indicate that engineered and some aquatic contexts harbor the clearest signal of plasmid-borne *nirK*, whereas *nirS* is largely confined to isolated engineered niches. Category sizes and prevalences are shown in Figure 3.

Enrichment analysis with a fixed baseline

To avoid a shifting comparator, enrichment was assessed with a binomial logistic model (gene presence ~ biome) using a single predefined baseline and Benjamini–Hochberg adjustment across biome contrasts. *nirK* showed higher odds in several engineered sub-biomes and modestly elevated odds in environmental aquatic relative to the baseline, while host-associated categories showed lower odds. For *nirS*, only a subset of engineered categories showed detectable increases; most other biomes were not distinguishable from baseline after adjustment. Patterns are visualized in Figure 3.

Distributions across phyla

For taxonomic summaries, plasmids with phylum annotations (n = 75,253). *nirK* occurrences were concentrated in Proteobacteria (the most abundant host group by plasmid count), with smaller contributions from other phyla; *nirS* remained scarce across phyla. Counts and within-phylum proportions are reported in Figure 4, and the same fixed-baseline logistic framework was applied to qualitatively assess enrichment in the text.

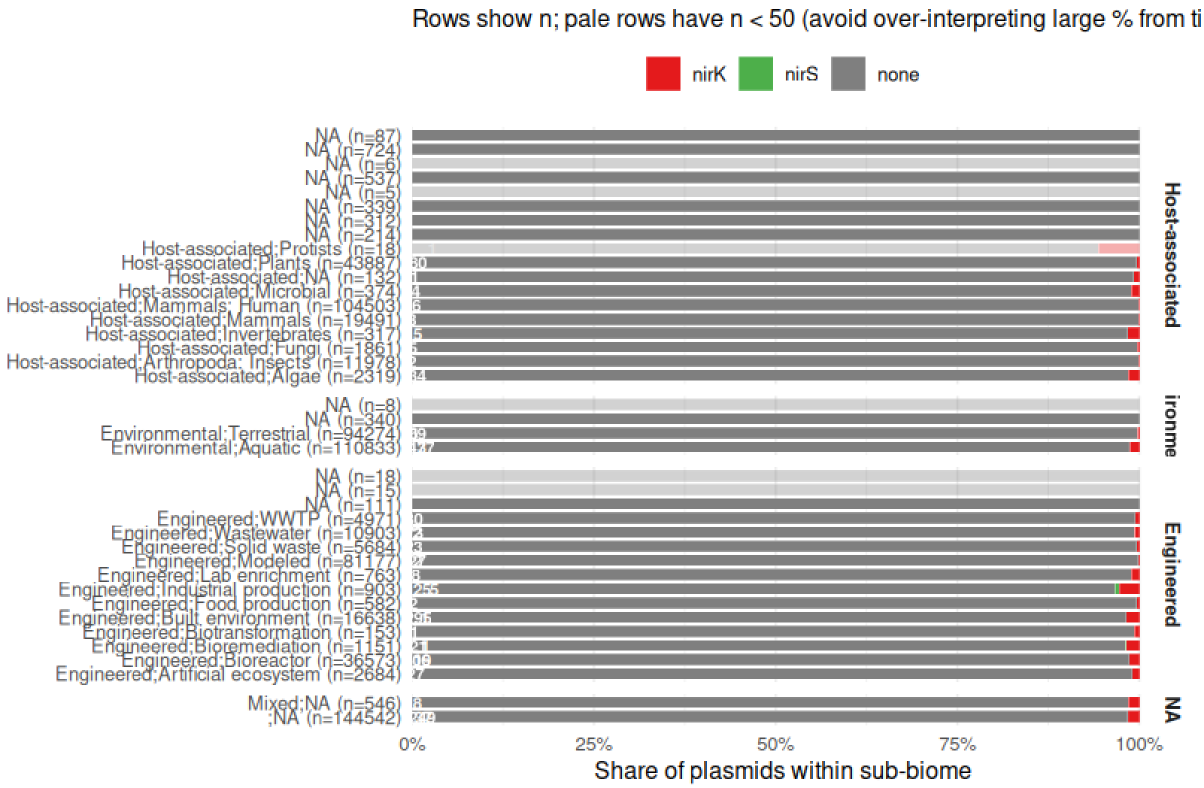

Figure 3. Normalized distribution of plasmid-borne *nirK* and *nirS* across collapsed biomes. Each horizontal bar represents a sub-biome; segments sum to 100% (gray = neither gene), with total plasmid counts shown in the row labels. Across all macro-biomes, >98% of plasmids lacked either gene. *nirK* was consistently more common than *nirS*, with the highest relative frequencies in engineered settings—bioreactors (1.4%), wastewater (0.7%), bioremediation (1.8%), and industrial production (2.8%). By contrast, *nirS* was rare, reaching only 0.05% in bioreactors and 0.6% in industrial production. In environmental plasmids, *nirK* occurred at ~1.3% (aquatic) and 0.2% (terrestrial), while *nirS* was nearly absent (<0.01%). Host-associated plasmids carried *nirK* at very low levels (<0.02% in mammals and humans) and *nirS* was undetected. Statistical comparisons confirmed enrichment of *nirK* in environmental plasmids and modest enrichment of *nirS* in engineered plasmids, with strong depletion in host-associated plasmids. Note: Rows with n < 50 are shown with reduced opacity to avoid over-interpreting large percentages from very small denominators.

Host distribution of potentially mobile elements encoding *nirK* and *nirS*

Plasmids

Phylum-level analysis (Figure 4) showed clear differences in how rarely nitrite reductase genes occur across lineages. In Proteobacteria, which dominate the plasmid dataset, only a small minority of plasmids carried any nir gene. Among the nir-positive Proteobacteria plasmids, *nirK* overwhelmingly outnumbered *nirS* (≈98% vs ≈2%). Expressed across all Proteobacteria plasmids, this corresponds to only a very small fraction carrying *nirK* (on the order of ~1%) and a vanishingly small fraction carrying *nirS* (<1%).

In Deinococcota, nir-positive plasmids were rare; among those few positives, *nirK* comprised roughly four-fifths and *nirS* one-fifth. These percentages should be interpreted with caution because the number of nir-positive plasmids is small (driving instability in the 80/20 split) and the total number of Deinococcota plasmids is also small (making within-phylum prevalence sensitive to single entries).

A handful of other phyla—including Acidobacteriota, Actinobacteriota, Bacteroidota, Cyanobacteria, Firmicutes (including Firmicutes_A), Halobacteriota, and OLB16—contained 1–3 nir-positive plasmids each, so "100% *nirK*" in these groups simply reflects tiny counts rather than true exclusivity. Plasmids without a phylum assignment were overwhelmingly nir-negative, with only ~0.01% carrying *nirK* and a similar proportion carrying *nirS*.

Viral occurrence of nir genes (IMGVR and Phagescope datasets)

In addition to plasmids, *nirK*-bearing mobile elements were also detected in viral datasets.In IMGVR, a small number of viral contigs encoding *nirK* were found, including Caudoviricetes predicted to infect Pseudomonadaceae and Glaesserella australis (both Gammaproteobacteria), as well as an Inoviridae contig predicted to infect Neisseria gonorrhoeae (Betaproteobacteria). Phagescope similarly revealed *nirK* exclusively in Caudoviricetes infecting Pseudomonas aeruginosa, all within the Gammaproteobacteria. *nirS* was not detected in either viral resource.

While these findings are consistent with earlier reports of *nirK* in Caudoviricetes linked to Gammaproteobacteria hosts in oxygen-depleted marine environments (Roux et al., 2018), they represent only a handful of viral contigs and should be interpreted cautiously. Moreover, both IMGVR and Phagescope are taxonomically enriched for Pseudomonas and related Gammaproteobacteria, so the apparent linkage may partly reflect database composition rather than a universal biological preference. Within this context, the data suggest that *nirK* has limited viral mobility and is currently observed most often in association with phages infecting Gammaproteobacteria, particularly Pseudomonads.

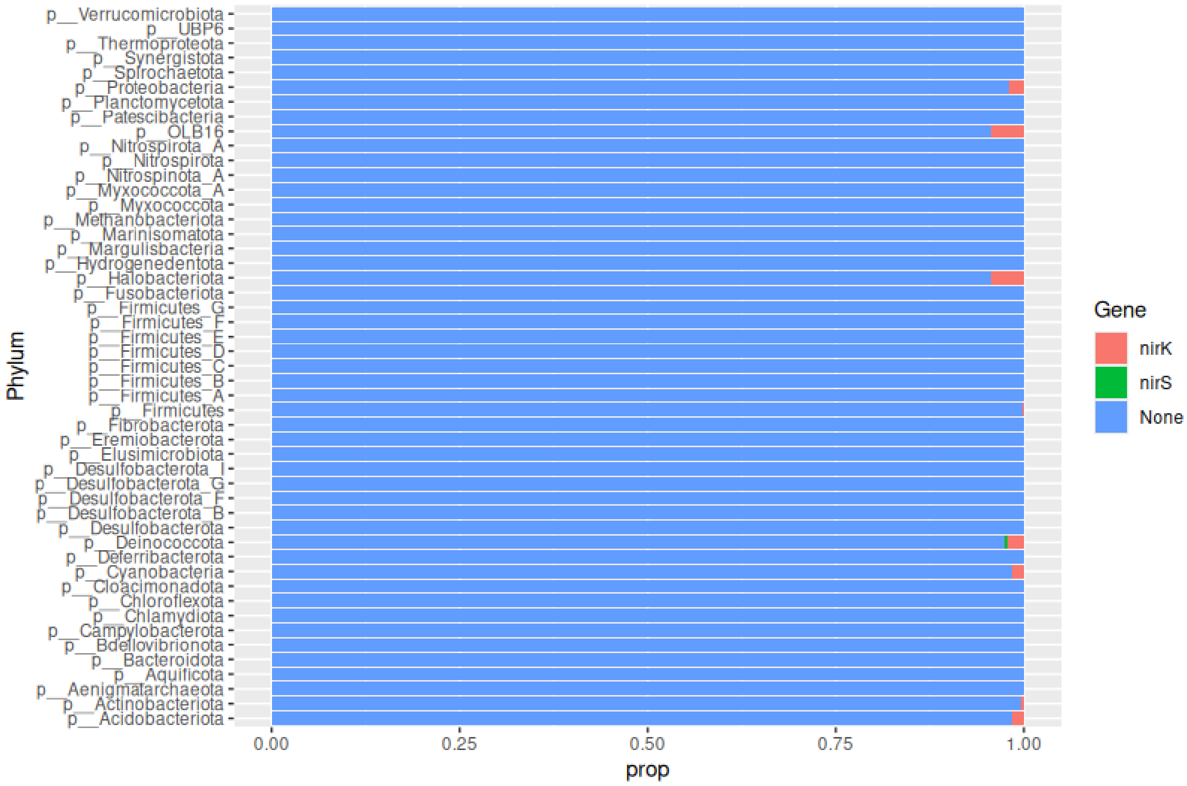

Figure 4. Plasmids with no phylum assignment were overwhelmingly negative for nitrite reductase genes, with only 0.01% carrying *nirK* and 0.01% carrying *nirS*. Overall, *nirK* was the most widespread across phyla and the most common in absolute numbers within Proteobacteria, whereas *nirS* was restricted to Proteobacteria and Deinococcota at measurable levels.

Absolute counts per phylum

When absolute counts are examined (Figure 5), Proteobacteria are found to dominate, contributing approximately 43% of all nir-positive plasmids (≈2,200), with both genes present but exhibiting a strong skew toward *nirK*.

An even larger share—about ~56% (≈2,850)—is unassigned at the phylum level, and *nirK* is again the predominant form in this group. The remaining phyla (e.g., Actinobacteriota, Firmicutes, Bacteroidota, Cyanobacteria) each account for <2–3% of nir-positive plasmids. Overall, nir-encoding plasmids are concentrated in Proteobacteria and taxonomically unresolved entries, with only minor contributions from other phyla.

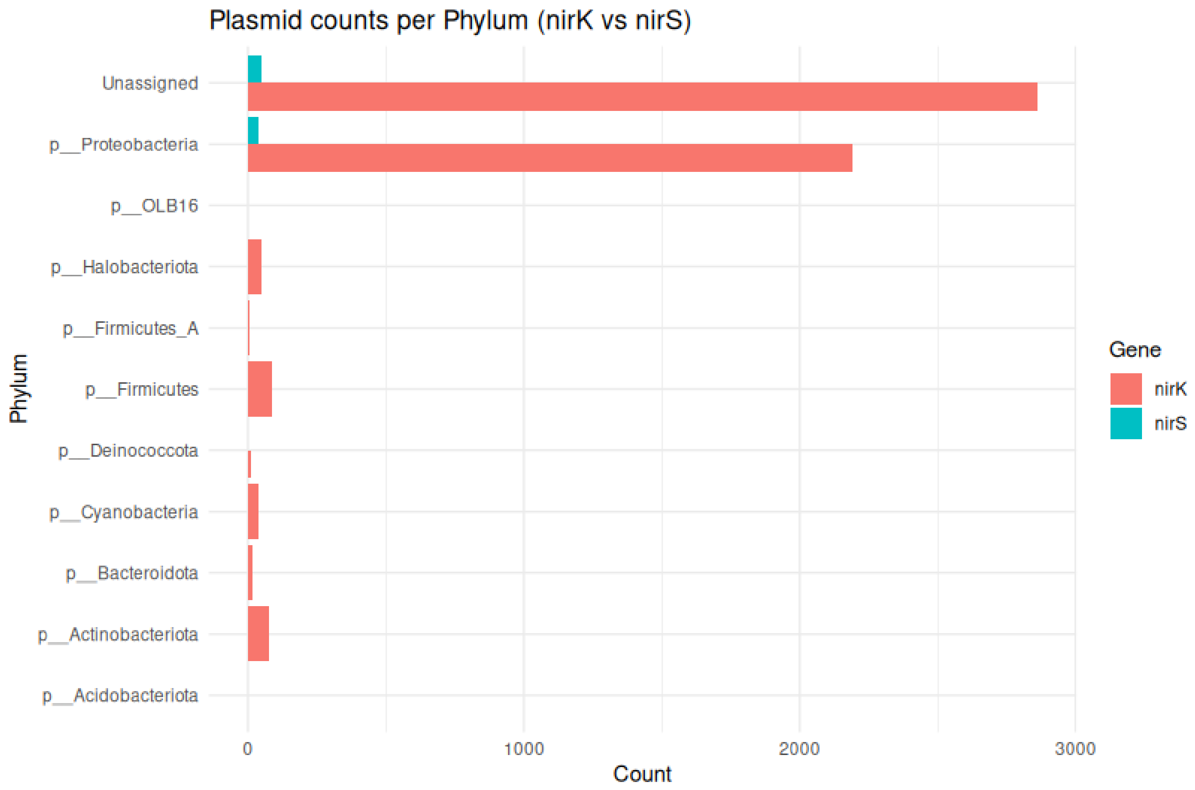

Figure 5. Distribution of plasmids carrying *nirK* and *nirS* across bacterial phyla. Bars show the absolute counts of plasmids per phylum, with "Unassigned" indicating plasmids lacking confident taxonomic placement.

Biome–phylum interaction

The combined biome–phylum analysis (Figure 6 and 7) demonstrated that *nirK*-positive plasmids in Proteobacteria were found across diverse biomes, including engineered wastewater systems, host-associated samples, and multiple environmental contexts. *nirS*-positive plasmids were concentrated in a small number of biomes, mostly engineered and host-associated. The heatmap highlights the much wider ecological breadth of *nirK* compared to *nirS*.

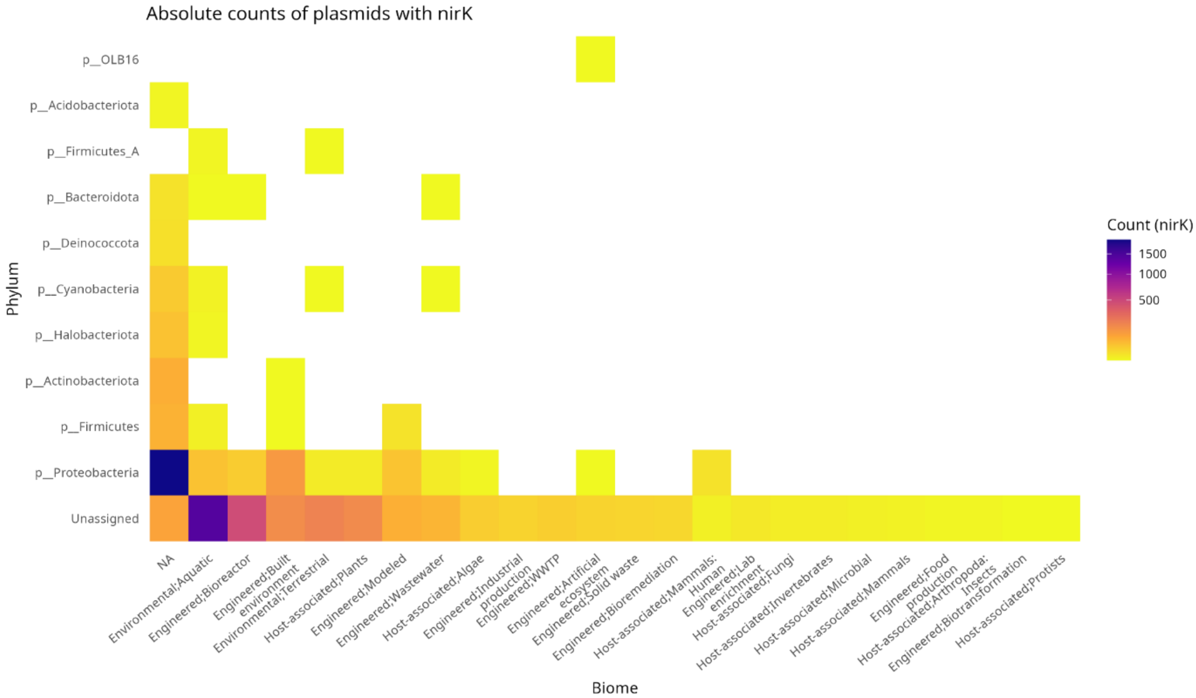

Figure 6: Heatmap showing the distribution of plasmid-borne *nirK* across phyla (rows) and biomes (columns). Darker colors indicate higher absolute counts of *nirK*-positive plasmids within each phylum–biome combination. Across all plasmids (699,973), *nirK* is rare but clearly more frequent than *nirS* (0.762% vs 0.013%). Within complete, biome-annotated plasmids (n = 129,282), *nirK* is elevated in specific sub-biomes rather than uniformly across "engineered": Environmental;Aquatic ~1.33% (OR 4.97, 95% CI 4.6–5.3, FDR ≈ 0) and Engineered;Industrial production ~2.77% (OR 6.50, 95% CI 4.2–9.7, FDR ~3.7×10^{-12}). By contrast, *nirK* is depleted in Host-associated;Mammals ~0.015% (OR 0.03, FDR < 1×10^{-30}). At the macro-biome level, "engineered" as a whole is not higher than others (0.70% vs 0.78%, OR 0.90, p = 0.001), indicating that the signal comes from specific engineered sub-biomes. Taxonomically, detections span multiple phyla but are most frequent in Proteobacteria, with many nirK-positive plasmids unassigned at the phylum level.

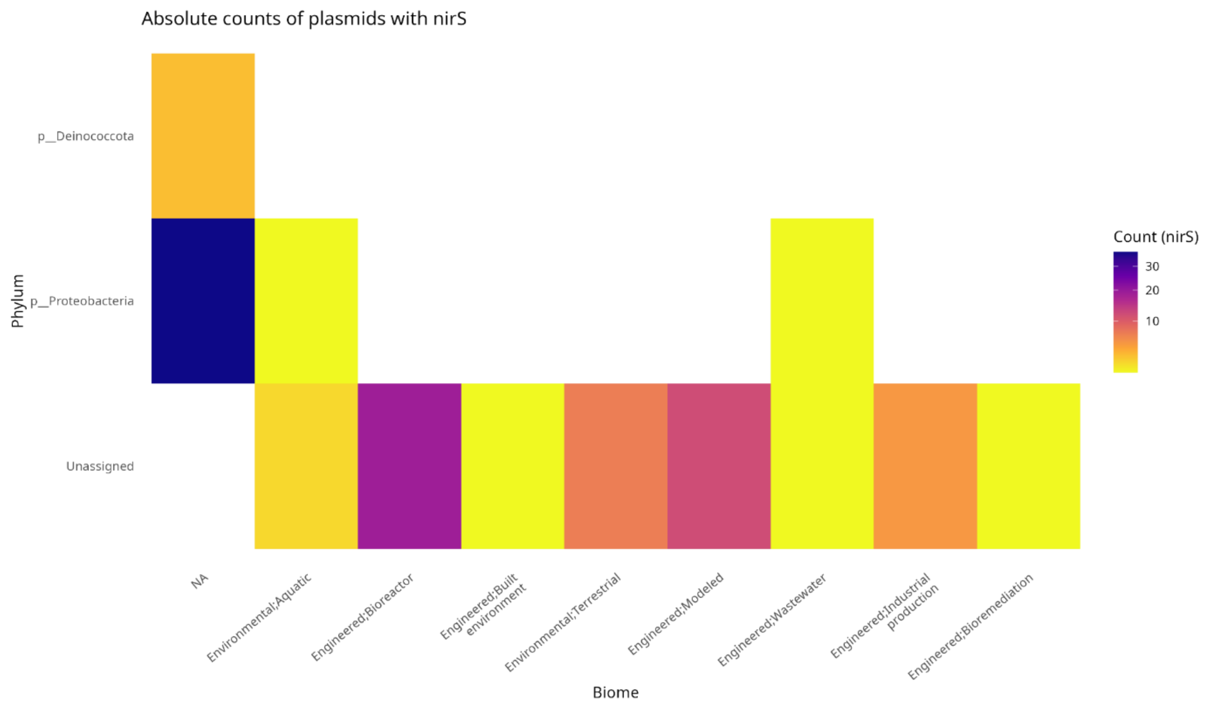

Figure 7. Heatmap showing the distribution of plasmid-borne nirS across phyla (rows) and biomes (columns). Darker colors indicate higher absolute counts of *nirS*-positive plasmids within each phylum–biome combination. *nirS* is very rare overall (0.013% of 699,973 plasmids) and largely confined to engineered niches. The strongest signals occur in Engineered;Industrial production ~0.55% (OR 86.7, 95% CI 27–217, FDR 1.4×10^{-7}) and Engineered;Bioreactor ~0.052% (OR 11.1, 95% CI 5.9–20.3, FDR 1×10^{-10}); most other biomes show no increase after correction. At the macro-biome level, *nirS* is modestly enriched in engineered contexts overall (0.025% vs 0.009%, OR 2.65, 95% CI 1.7–4.1, $p = 8.3\times10^{-6}$). Where taxonomy is resolved, detections are primarily Proteobacteria; elsewhere they are unassigned. Overall, *nirS* remains restricted in both ecology and taxonomy, in sharp contrast to the broader footprint of *nirK*.

Discussion

This study provides the first systematic comparison of nitrite reductase genes (*nirK* and *nirS*) across multiple classes of mobile genetic elements (MGEs). A clear asymmetry was observed: *nirK* was consistently more prevalent than *nirS* across all datasets, both in absolute counts and proportional representation. Among plasmids, ~0.7% encoded *nirK* compared to only ~0.02% encoding *nirS*, representing more than a 30-fold difference. Genome surveys show that nirK is more common than *nirS* across bacterial genomes, whereas *nirS* occurs at lower overall genomic prevalence and in fewer taxa (e.g., Jones et al., 2008; Graf et al., 2014). This genome-level baseline helps explain why plasmid screens recover more *nirK* than *nirS*: a larger chromosomal donor pool increases opportunities for capture onto plasmids, even before considering differences in transferability. Functionally, NirK may be more transferable because its copper cofactor requirements are simpler than the complex cytochrome cd_1 structure of NirS, which requires specialized heme d_1 biosynthesis enzymes such as NirN and NirF (Bothe et al., 2000; Cantera & Stein, 2007). The reliance of *nirS* on auxiliary maturation proteins likely constrains its horizontal mobility. Taken together, these observations suggest that *nirK* is not only more widespread at the genomic level but also better suited to movement through HGT pathways, making it a more versatile player in shaping nitrogen cycling capacity across microbial communities.

The prevalence of nir genes was strikingly uneven across mobile genetic elements. In the plasmid dataset (IMG-PR), hundreds of *nirK* and dozens of *nirS*-positive plasmids were identified. By contrast, viral and ICE datasets showed almost no hits: only two *nirK*-positive phages were detected among more than 180,000 viral genomes (IMG-VR, PhageScope), and ICEberg 3.0 yielded a single *nirK*-positive integrative element with no *nirS*. This imbalance emphasizes that plasmids are, under current sampling, the primary vehicles for mobilization of nitrite reductase genes, even though the possibility of rare but ecologically relevant viral or ICE associations cannot be excluded.

Across datasets, detections of nitrite reductase genes were concentrated on plasmids, whereas the viral and ICE sets yielded only isolated *nirK* and no *nirS*. This is interpreted cautiously. Part of the pattern is likely attributable to database scope and curation: the ICEberg 3.0 catalogue has been recognized as relatively small and biased toward model organisms (Liu et al., 2019), IMG-VR has been enriched for marine viruses (Roux et al., 2020), and PhageScope has remained under-sampled for many environments (R. H. Wang et al., 2023). Biological and mechanistic factors could also favor plasmid carriage: plasmids typically have greater cargo capacity than most phages, can capture multi-gene cassettes (including *nirS*-associated maturation genes such as *nirN*/*nirF*), and spread via conjugation, which efficiently transfers long DNA segments across diverse hosts; maintenance systems and broad-host-range backbones may further support persistence (Norman et al., 2009). ICEs can, in principle, mobilize large regions as well, but their detection depends on accurate boundary calls and they are less comprehensively catalogued than plasmids. For viruses, genome size and packaging constraints likely limit retention of denitrification genes, even though auxiliary metabolic genes have been reported in oxygen-depleted waters (e.g., viral *nirK*/*norB*) and may play focused ecological roles (Gazitúa et al., 2020b). Taken together, our data indicate a plasmid-centered signal under current sampling and plausible mechanistic advantages, while leaving open the possibility that expanded ICE and viral coverage could revise the relative contributions of different MGEs. Future efforts integrating long-read sequencing, improved viral genome curation, and direct experimental assays of phage-host interactions could help resolve whether viruses contribute more substantially to nitrite reductase mobility than currently apparent.

Ecological stratification revealed strong enrichment of plasmid-borne *nirK* in engineered and aquatic biomes. In engineered sub-biomes such as wastewater, bioreactors, and industrial production systems, *nirK* frequencies ranged from ~1–3%, compared to <0.5% in most other contexts. Statistical enrichment tests confirmed that these niches exhibited 3–6-fold higher odds of harboring *nirK*-positive plasmids relative to the database background. By contrast, mammalian-associated plasmids showed extreme depletion, with only 0.015% encoding *nirK*. This pattern is consistent with ecological theory: environments with high microbial density and strong selective pressures (engineered systems) favor HGT, while host-associated niches dominated by pathogens or commensals provide fewer opportunities or incentives for nitrite reductase acquisition.

These results resonate with the idea that engineered ecosystems are HGT hotspots due to high microbial density, frequent cell–cell contact, and selective pressures imposed by fluctuating nutrient and oxygen conditions (Frost et al., 2005; Smillie et al., 2011). Similar hotspots for denitrification genes have been described in wastewater treatment plants, where nir genes are strongly selected under anoxic, nitrate-rich phases (Hallin et al., 2006). Interestingly, the enrichment of *nirK* in natural aquatic biomes suggests that plasmid-mediated nitrite reduction may also be ecologically important in oxygen minimum zones (OMZs) and aquatic sediments, where denitrifiers play a central role in nitrogen loss (Ward et al., 2009; Thamdrup & Dalsgaard, 2002). This convergence of engineered and aquatic enrichment signals highlights a common theme: plasmid-borne *nirK* is favored in environments where oxygen limitation coincides with nitrate/nitrite availability, regardless of whether the setting is human-made or natural.

In contrast, *nirS* appeared largely confined to engineered settings. Signals were most pronounced in industrial production and bioreactor systems, with only sporadic detections elsewhere. This pattern suggests that plasmid-borne *nirS* is not broadly disseminated but instead favored under the strong, recurrent selection regimes characteristic of human-modified environments—conditions that likely help overcome the structural and regulatory hurdles associated with cytochrome cd_1 assembly and function. Its near-absence from expansive host-associated collections further underscores a narrow ecological range, implying that the fitness benefits of maintaining *nirS* on plasmids are context-dependent and rarely realized outside engineered niches. The ecological specificity of plasmid-borne *nirS* raises interesting questions about whether its presence is linked to particular microbial consortia or industrial substrates that create unique selection windows.

Phylum-level patterns. Most detections mapped to Proteobacteria, consistent with their known role in denitrification (Zhang et al., 2023). Outside Proteobacteria, *nirK* occurred sporadically in Firmicutes, Actinobacteriota, Cyanobacteria, and Deinococcota, whereas *nirS* detections were rare. In Deinococcota, nir-positive plasmids were few; among these positives, ~81% carried *nirK* and ~19% carried *nirS*. Because both the number of nir-positive plasmids and the total Deinococcota plasmid count are small, these ratios should be interpreted cautiously and do not imply high within-phylum prevalence. Scattered detections in non-Proteobacteria may reflect rare cross-phyla transfers via broad-host-range plasmids (Norman et al., 2009; Smalla et al., 2015). These isolated events nevertheless demonstrate the potential for occasional long-distance HGT events, which could have outsized evolutionary importance if they introduce nitrite reductase capacity into previously incapable lineages.

In contrast, *nirS*-positive plasmids were overwhelmingly restricted to Proteobacteria, underscoring its narrow taxonomic distribution. Chromosomal surveys have found *nirS* across multiple bacterial phyla, though less widespread than *nirK* (Graf et al., 2014; Pold et al., 2024 preprint). Thus, the plasmid data suggest that nirS mobility is far more constrained than chromosomal diversity would predict, reinforcing the hypothesis that structural complexity and dependence on auxiliary genes hinder its transferability. This divergence between chromosomal diversity and plasmid mobility is a key theme: while *nirS* may exist in multiple lineages at the genome level, its capacity to move horizontally is curtailed, limiting its role in dynamic ecological redistribution.

An important consideration is the denominator: both *nirK* and *nirS* frequencies were calculated against the same total number of plasmids (~11,281 in IMG-PR), ensuring comparability. However, database composition is uneven. IMG-PR is biased toward pathogens, Proteobacteria, and host-associated systems (Chen et al., 2021). Consequently, the low prevalence of nir genes in mammals, insects, and human-associated plasmids may partly reflect over-representation of taxa not typically engaged in denitrification. Conversely, the enrichment in wastewater and industrial samples appears robust, as these biomes represent smaller fractions of the database yet yield higher relative frequencies. These findings highlight the need to interpret enrichment results in light of database sampling bias. Future database expansions that more evenly capture soil, freshwater, and engineered biomes will be critical for confirming whether the trends reported here are representative of true ecological distributions or artifacts of current sampling bias.

The dominance of *nirK* on plasmids suggests that copper nitrite reductases may act as modular, transferable elements of nitrite reduction metabolism, enabling rapid microbial adaptation in fluctuating redox environments. By contrast, the rarity and ecological confinement of *nirS* support the idea that cytochrome cd_1 nitrite reductases remain largely chromosomal, with limited plasmid dissemination. This asymmetry has significant implications for nitrogen cycling. In engineered environments such as wastewater treatment plants, plasmid-mediated *nirK* transfer could stabilize nitrite reduction capacity across diverse microbial hosts, potentially affecting N_2O emissions (Domeignoz-Horta et al., 2016b; Roux et al., 2018, Kim & Price, 2011). In natural aquatic systems, plasmid-borne *nirK* may contribute to microbial resilience under hypoxia and nitrate enrichment, expanding the ecological range of nitrite reduction. In both cases, the ecological consequences are profound: gene mobility alters not only the adaptability of individual microbes but also the collective function of entire communities, influencing greenhouse gas fluxes and nutrient turnover (Butterbach-Bahl et al., 2013).

At the evolutionary scale, the presence of *nirK* across multiple phyla and biomes suggests a history of recurrent horizontal transfers, while *nirS* remains evolutionarily constrained. This divergence in mobility may partly explain why *nirK* has a broader taxonomic footprint in microbial genomes worldwide (Graf et al., 2014). Understanding the structural and ecological barriers limiting nirS dissemination will be key to predicting the stability of nitrite reduction pathways under global change. By integrating phylogenetic, ecological, and mechanistic perspectives, future work can determine whether *nirK*'s high mobility and *nirS*'s restricted dissemination represent fixed evolutionary strategies or dynamic traits shaped by environmental pressures.

Beyond the specific findings, the methodology applied in this thesis provides insight into how gene mobility studies can be conducted systematically at the interface of bioinformatics and microbial ecology. The combination of stringent HMM-based detection, motif filtering, and phylogenetic validation minimized false positives and provided confidence that the sequences classified as *nirK* or *nirS* represent bona fide nitrite reductases. Nevertheless, this conservative approach may under-detect divergent or novel variants, particularly in viral and archaeal lineages that remain poorly sampled. This raises the possibility that our results represent a lower-bound estimate of nitrite reductase mobility, with true diversity and prevalence still to be uncovered as sequencing expands across underrepresented environments. In this sense, database-driven surveys must be seen as complementary to targeted cultivation and metagenomic assembly approaches, which can recover highly divergent genes and contextualize them in their genomic neighborhoods (Alneberg et al., 2018).

The ecological implications of plasmid-dominated *nirK* mobility are significant when considered in the context of global change. Denitrification is a major sink for reactive nitrogen and a key source of nitrous oxide (N_2O), a greenhouse gas with nearly 300-fold greater warming potential than CO_2. If *nirK* continues to spread across microbial taxa and biomes through plasmid exchange, this could stabilize the capacity for nitrite reduction in engineered and aquatic systems, buffering ecosystem function against perturbations such as fluctuating oxygen or nitrate supply. At the same time, the mobility of *nirK* could contribute to unpredictable shifts in N_2O emissions, since the balance between nitrite reduction, nitric oxide reduction, and N_2O reduction depends on whether the full denitrification pathway is present. Thus, understanding how plasmid-mediated nitrite reduction integrates into broader community metabolism is essential for predicting climate feedback.

From an evolutionary perspective, the asymmetry between *nirK* and *nirS* mobility also illustrates broader principles of gene transfer. Genes with simpler cofactor requirements, fewer accessory proteins, and more modular operon structures appear more likely to succeed as "mobile" units. In contrast, genes with complex structural or maturation dependencies—like *nirS* with its reliance on heme d_1 biosynthesis—are less easily mobilized, confining their distribution despite ecological demand. This echoes patterns seen in other metabolic pathways, where auxiliary metabolic genes carried by MGEs tend to be those with relatively simple biochemical integration (Anantharaman et al., 2016). The *nirK*/*nirS* case therefore provides a model for understanding which metabolic functions are most amenable to horizontal dissemination.

Finally, future research should move beyond cataloguing presence/absence toward testing function. Experimental validation of plasmid-encoded *nirK* activity, for example through heterologous expression or functional assays in model hosts, would establish whether these mobile copies confer measurable fitness advantages. Comparative genomics could identify conserved gene neighborhoods—such as the co-occurrence of *nirK* with conjugation machinery or other denitrification genes—that might act as transferable cassettes. Coupling these approaches with metagenomic time-series data from wastewater plants, agricultural soils, or oxygen minimum zones would allow direct observation of *nirK* dynamics under selective pressure. Together, these directions highlight the need for interdisciplinary integration: bioinformatics to detect mobility, microbial physiology to test function, and ecosystem science to assess impacts on nitrogen fluxes.

Ethical aspects, gender perspectives, and impact on the society

This study was conducted entirely using publicly available genomic databases (IMG-PR, IMG-VR, PhageScope, ICEberg 3.0) and computational analyses. As such, no experiments involving animals, human subjects, or pathogenic exposure were performed, and no ethical approval was required. The absence of wet-lab experimentation means that issues of biosafety (e.g., handling of pathogens, antibiotics, or genetically modified organisms) were not relevant in this context. Nonetheless, the research touches indirectly on genes involved in greenhouse gas production (notably N_2O), which raises dual-use considerations: while understanding the mobility of denitrification genes can help mitigate emissions and inform environmental management, the same knowledge could hypothetically be misused in synthetic biology. Careful data stewardship and adherence to open scientific norms reduce this risk.

With respect to sex and gender perspectives, this project did not involve human or animal subjects, and therefore does not directly address biological sex or social gender dimensions. However, reflecting on gender perspectives in science, it is worth noting that environmental challenges such as nitrogen pollution and climate change have differential societal impacts across communities, including vulnerable populations. The integration of diverse perspectives in environmental microbiology research is therefore important to ensure equitable outcomes and inclusive problem-solving, even if the present study itself is computational and theoretical.

The broader societal impact of this work lies in its contribution to understanding how mobile genetic elements disseminate key metabolic genes involved in the nitrogen cycle. By identifying patterns of horizontal transfer of nitrite reductase genes (*nirK* and *nirS*), this research improves our understanding of microbial adaptation under fluctuating redox conditions. Such insights are relevant for environmental management, for example in wastewater treatment plants, where denitrification is harnessed to reduce nitrate levels but can also generate N_2O, a potent greenhouse gas. Clarifying the mobility of these genes may help design microbial communities or engineering strategies that enhance nitrogen removal efficiency while minimizing emissions. At a broader level, these findings add to the knowledge base needed to predict microbial responses to global change, supporting both climate mitigation efforts and sustainable ecosystem management.

Future perspectives

This thesis provides new insights into the prevalence and ecological distribution of plasmid-borne *nirK* and *nirS* genes, but it also raises important questions that merit further investigation. A promising direction is to examine not only single-gene presence but also the full operon context. Many *nirS* genes rely on accessory proteins (e.g., NirN, NirF), and co-occurrence patterns could clarify whether plasmid-mediated transfer involves complete functional cassettes or fragmented modules. Similarly, gene neighborhood analyses and synteny comparisons could help identify conserved mobile gene clusters and assess whether plasmids act as vehicles for integrated denitrification pathways.

From an evolutionary perspective, further work should apply phylogenetic congruence analyses (e.g., Robinson–Foulds distances, tanglegrams, reconciliation models) to compare gene trees with host species trees. This would provide direct evidence of horizontal transfer events and their frequency across lineages. Expanding to other denitrification genes beyond *nirK* and *nirS* (e.g., *norB*, *nosZ*) would give a more holistic view of how plasmids and other MGEs shape nitrogen cycling capacities.

Finally, the applied implications of plasmid-borne denitrification genes should be explored. In wastewater treatment and bioremediation, understanding the transferability of *nir* genes could inform microbial community design strategies aimed at improving nitrogen removal while reducing N_2O emissions. In natural ecosystems such as oxygen minimum zones, detecting plasmid-mediated denitrification may help predict microbial resilience under climate-driven redox fluctuations. By integrating genomic data mining, evolutionary analysis, and environmental microbiology, future research can bridge the gap between molecular mechanisms of gene transfer and their global biogeochemical consequences.

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Dr. Grace Pold (SLU), for her invaluable guidance, encouragement, and constructive feedback throughout the course of this thesis. Her expertise and continuous support have been instrumental in shaping this work. I am also sincerely thankful to Dr. Maryam Rezaei Somee for her insightful advice, helpful discussions, and kind support during my research.

I would also like to acknowledge the support of the Molecular Biotechnology program at the University of Skövde and the Department of Ecology and Genetics at SLU, which provided an excellent academic environment and resources for my work. My sincere thanks also go to the PDC Center for High Performance Computing (KTH) for providing computational resources that made the large-scale analyses in this thesis possible.

Finally, I am deeply grateful to my family and friends for their patience, encouragement, and unwavering support throughout my studies.

References

Alneberg, J., Bjarnason, B. S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., … & Quince, C. (2018). Binning metagenomic contigs by coverage and composition. Nature Methods, 11(11), 1144–1146. https://doi.org/10.1038/nmeth.3103.

Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., … & Banfield, J. F. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nature Communications, 7, 13219. https://doi.org/10.1038/ncomms13219.

Averill, B. A. (1996). Dissimilatory nitrite and nitric oxide reductases. Chemical Reviews, 96(7), 2951–2964. https://doi.org/10.1021/cr950056p

Bellanger, X., Payot, S., Leblond-Bourget, N., & Guédon, G. (2014). Conjugative and mobilizable genomic islands in bacteria: Evolution and diversity. FEMS Microbiology Reviews, 38(4), 720–760. https://doi.org/10.1111/1574-6976.12058.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bothe, H., Ferguson, S. J., Newton, W. E., & Eady, R. R. (2000). Biology of the Nitrogen Cycle. Elsevier.

Brüssow, H., Canchaya, C., & Hardt, W. D. (2004). Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. Microbiology and Molecular Biology Reviews, 68(3), 560–602. https://doi.org/10.1128/MMBR.68.3.560-602.2004.

Butterbach-Bahl, K., Baggs, E. M., Dannenmann, M., Kiese, R., & Zechmeister-Boltenstern, S. (2013). Nitrous oxide emissions from soils: How well do we understand the processes and their controls? Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1621), 20130122. https://doi.org/10.1098/rstb.2013.0122.

Cantera, J. J. L., & Stein, L. Y. (2007). Role of nitrite reductase in the ammonia-oxidizing pathway of Nitrosomonas europaea. Archives of Microbiology, 188(4), 349–354. https://doi.org/10.1007/s00203-007-0255-4.

Change, N. I. P. O. C. (2023). Climate Change 2021 – The Physical science basis. https://doi.org/10.1017/9781009157896

Chen, I. M. A., et al. (2021). IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. Nucleic Acids Research, 49(D1), D751–D763. https://doi.org/10.1093/nar/gkaa939.

Domeignoz-Horta, L., Putz, M., Spor, A., Bru, D., Breuil, M., Hallin, S., & Philippot, L. (2016). Non-denitrifying nitrous oxide-reducing bacteria - An effective N2O sink in soil. Soil Biology and Biochemistry, 103, 376–379. https://doi.org/10.1016/j.soilbio.2016.09.010.

Eddy, S. R. (2011). Accelerated profile HMM searches. PLoS Computational Biology, 7(10), e1002195. https://doi.org/10.1371/journal.pcbi.1002195.

Einsle, O., & Kroneck, P. M. (2004). Structural basis of denitrification. Biological chemistry, 385(10), 875–883. https://doi.org/10.1515/BC.2004.115

Frost, L. S., Leplae, R., Summers, A. O., & Toussaint, A. (2005). Mobile genetic elements: The agents of open source evolution. Nature Reviews Microbiology, 3(9), 722–732. https://doi.org/10.1038/nrmicro1235.

Galloway, J. N., Dentener, F. J., Capone, D. G., Boyer, E. W., Howarth, R. W., Seitzinger, S. P., Asner, G. P., Cleveland, C. C., Green, P. A., Holland, E. A., Karl, D. M., Michaels, A. F., Porter, J. H., Townsend, A. R., & V Osmarty, C. J. (2004). Nitrogen cycles: past, present, and future. Biogeochemistry, 70(2), 153–226. https://doi.org/10.1007/s10533-004-0370-0

Galloway, J. N., Townsend, A. R., Erisman, J. W., Bekunda, M., Cai, Z., Freney, J. R., … & Sutton, M. A. (2008). Transformation of the nitrogen cycle: Recent trends, questions, and potential solutions. Science, 320(5878), 889–892. https://doi.org/10.1126/science.1136674.

Gazitúa, M. C., Vik, D. R., Roux, S., Gregory, A. C., Bolduc, B., Widner, B., Mulholland, M. R., Hallam, S. J., Ulloa, O., & Sullivan, M. B. (2020b). Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. The ISME Journal, 15(4), 981–998. https://doi.org/10.1038/s41396-020-00825-6.

Graf, D. R. H., Jones, C. M., & Hallin, S. (2014). Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N_2O emissions. PLoS ONE, 9(12), e114118. https://doi.org/10.1371/journal.pone.0114118.

Hallin, S., ThrobäCk, I. N., Dicksved, J., & Pell, M. (2006). Metabolic Profiles and Genetic Diversity of Denitrifying Communities in Activated Sludge after Addition of Methanol or Ethanol. Applied and Environmental Microbiology, 72(8), 5445–5452. https://doi.org/10.1128/aem.00809-06

Johnson, C. M., & Grossman, A. D. (2015). Integrative and Conjugative Elements (ICEs): what they do and how they work. Annual Review of Genetics, 49(1), 577–601. https://doi.org/10.1146/annurev-genet-112414-055018.

Jones, C. M., Stres, B., Rosenquist, M., & Hallin, S. (2008). Phylogenetic analysis of nitrite, nitric oxide, and nitrous oxide respiratory enzymes reveal a complex evolutionary history for denitrification. Molecular Biology and Evolution, 25(9), 1955–1966. https://doi.org/10.1093/molbev/msn146

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., & Crook, D. W. (2009). Genomic islands: Tools of bacterial horizontal gene transfer and evolution. FEMS Microbiology Reviews, 33(2), 376–393. https://doi.org/10.1111/j.1574-6976.2008.00136.x.

Kim, P., & Price, N. D. (2011). Genetic Co-Occurrence Network across Sequenced Microbes. PLoS Computational Biology, 7(12), e1002340. https://doi.org/10.1371/journal.pcbi.1002340

Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., … & Ou, H. Y. (2019). ICEberg 2.0: An updated database of bacterial integrative and conjugative elements. Nucleic Acids Research, 47(D1), D660–D665. https://doi.org/10.1093/nar/gky1123.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, … & Schleifer, K. H. (2004). ARB: A software environment for sequence data. Nucleic Acids Research, 32(4), 1363–1371. https://doi.org/10.1093/nar/gkh293.

Norman, A., Hansen, L. H., & Sørensen, S. J. (2009). Conjugative plasmids: Vessels of the communal gene pool. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1527), 2275–2289. https://doi.org/10.1098/rstb.2009.0037.

Parks, D. H., Chuvochina, M., Waite, D. W., Rinke, C., Skarshewski, A., Chaumeil, P. A., & Hugenholtz, P. (2020). A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nature Biotechnology, 38(9), 1079–1086. https://doi.org/10.1038/s41587-020-0501-8.

Philippot, L. (2002). Denitrifying genes in bacterial and archaeal genomes. Biochimica et Biophysica Acta, 1577(3), 355–376. https://doi.org/10.1016/S0167-4781(02)00420-7.

Pold, G., Bonilla-Rosso, G., Saghaï, A., Strous, M., Jones, C. M., & Hallin, S. (2024). Phylogenetics and environmental distribution of nitric oxide forming nitrite reductases reveals their distinct functional and ecological roles. ISME Communications. https://doi.org/10.1093/ismeco/ycae020

Roux, S., Adriaenssens, E. M., Dutilh, B. E., Koonin, E. V., Kropinski, A. M., Krupovic, M., Kuhn, J. H., Lavigne, R., Brister, J. R., Varsani, A., Amid, C., Aziz, R. K., Bordenstein, S. R., Bork, P., Breitbart, M., Cochrane, G. R., Daly, R. A., Desnues, C., Duhaime, M. B., . . . Eloe-Fadrosh, E. A. (2018). Minimum Information about an Uncultivated Virus Genome (MIUViG). Nature Biotechnology, 37(1), 29–37. https://doi.org/10.1038/nbt.4306

Roux, S., Páez-Espino, D., Chen, I. A., Palaniappan, K., Ratner, A., Chu, K., Reddy, T. B. K., Nayfach, S., Schulz, F., Call, L., Neches, R. Y., Woyke, T., Ivanova, N. N., Eloe-Fadrosh, E. A., & Kyrpides, N. C. (2020). IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. Nucleic Acids Research, 49(D1), D764–D775. https://doi.org/10.1093/nar/gkaa946

Shintani, M., Sanchez, Z. K., & Kimbara, K. (2015). Genomics of microbial plasmids: Classification and identification based on replication and transfer systems and host range. Frontiers in Microbiology, 6, 242. https://doi.org/10.3389/fmicb.2015.00242.

Smalla, K., Jechalke, S., & Top, E. M. (2015). Plasmid detection, characterization, and ecology. Microbiology Spectrum, 3(1). https://doi.org/10.1128/microbiolspec.PLAS-0038-2014.

Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., & Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. Nature, 480(7376), 241–244. https://doi.org/10.1038/nature10571

Soucy, S. M., Huang, J., & Gogarten, J. P. (2015). Horizontal gene transfer: Building the web of life. Nature Reviews Genetics, 16(8), 472–482. https://doi.org/10.1038/nrg3962.

Thamdrup, B., & Dalsgaard, T. (2002). Production of N2through anaerobic ammonium oxidation coupled to nitrate reduction in marine sediments. Applied and Environmental Microbiology, 68(3), 1312–1318. https://doi.org/10.1128/aem.68.3.1312-1318.2002

Touchon, M., De Sousa, J. a. M., & Rocha, E. P. (2017). Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. Current Opinion in Microbiology, 38, 66–73. https://doi.org/10.1016/j.mib.2017.04.010

Varin, T., Lovejoy, C., Jungblut, A. D., Vincent, W. F., & Corbeil, J. (2011). Metagenomic Analysis of Stress Genes in Microbial Mat Communities from Antarctica and the High Arctic. Applied and Environmental Microbiology, 78(2), 549–559. https://doi.org/10.1128/aem.06354-11

Wang, R. H., Yang, S., Liu, Z., Zhang, Y., Wang, X., Xu, Z., Wang, J., & Li, S. C. (2023). PhageScope: a well-annotated bacteriophage database with automatic analyses and visualizations. Nucleic Acids Research, 52(D1), D756–D761. https://doi.org/10.1093/nar/gkad979

Wang, Y., Qi, L., Huang, R., Wang, F., Wang, Z., & Gao, M. (2020). Characterization of denitrifying community for application in reducing nitrogen: A comparison of nirK and nirS gene diversity and abundance. Applied Biochemistry and Biotechnology, 192(1), 22–41. https://doi.org/10.1007/s12010-020-03250-9.

Ward, B. B., Devol, A. H., Rich, J. J., Chang, B. X., Bulow, S. E., Naik, H., … & Jayakumar, A. (2009). Denitrification as the dominant nitrogen loss process in the Arabian Sea. Nature, 461(7260), 78–81. https://doi.org/10.1038/nature08276.

Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. Springer. https://doi.org/10.1007/978-3-319-24277-4.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … & Yutani, H. (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686. https://doi.org/10.21105/joss.01686.

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023a). dplyr: A grammar of data manipulation. R package version 1.1.2. https://CRAN.R-project.org/package=dplyr.

Wickham, H., & Girlich, M. (2023). tidyr: Tidy Messy Data. R package version 1.3.0. https://CRAN.R-project.org/package=tidyr.

Zhang, I. H., Sun, X., Jayakumar, A., Fortin, S. G., Ward, B. B., & Babbin, A. R. (2023). Partitioning of the denitrification pathway and other nitrite metabolisms within global oxygen deficient zones. ISME Communications, 3(1). https://doi.org/10.1038/s43705-023-00284-y

Zumft, W. G. (1997). Cell biology and molecular basis of denitrification. Microbiology and Molecular Biology Reviews, 61(4), 533–616. https://doi.org/10.1128/MMBR.61.4.533-616.1997.

Appendix

Percentages of genomes in various phyla encoding *nirK* and *nirS*

phylum	*nirK*	*nirS*
p__Acidobacteriota	5.3	0.9
p__Actinomycetota	15.8	0.1
p__Armatimonadota	0.7	0
p__Bacillota	4.9	0.1
p__Bacillota_A	0	0
p__Bacillota_B	0.3	0
p__Bacillota_C	0.3	0
p__Bacillota_D	0	0
p__Bacillota_E	15.5	0
p__Bacillota_F	0	0
p__Bacillota_G	4.8	0
p__Bacillota_H	0	0
p__Bacteroidota	10.5	0.4
p__Bdellovibrionota	23.7	5
p__Campylobacterota	0.8	21.5
p__Chloroflexota	24.6	4.9
p__Cyanobacteriota	0.4	0.2
p__Deinococcota	14	16.7
p__Delongbacteria	9.1	0
p__Desulfobacterota	0.5	0.2
p__Desulfobacterota_B	17.2	1.1
p__Desulfobacterota_C	0	0
p__Desulfobacterota_D	4	0
p__Desulfobacterota_E	0	0
p__Desulfobacterota_F	2.2	7.6
p__Desulfobacterota_G	0	0

p__Dictyoglomota	0	0
p__Dormibacterota	2.2	0
p__Gemmatimonadota	31.2	0.8
p__Methylomirabilota	16.7	19
p__Muirbacteria	0	0
p__Myxococcota	19.9	4.1
p__Myxococcota_A	40	28.6
p__Nitrospinota	75.5	12.2
p__Nitrospinota_B	0	0
p__Nitrospirota	44.7	3.2
p__Nitrospirota_A	41.7	0
p__Nitrospirota_B	0	0
p__Omnitrophota	5.1	0
p__Patescibacteria	6.8	0
p__Planctomycetota	2.3	1.9
p__Pseudomonadota	15	12.2
p__SAR324	14.3	10.7
p__Spirochaetota	7.4	0.8
p__Sumerlaeota	5.3	0
p__ Verruche microbiota	12.8	0.1