bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Horizontal gene transfer drives DNRA capacity in wastewater treatment systems

Changwen Wang*[1], Mingchang Gao[1], Na Qiu[1], Xiangwei Ding[1], Peixue Song[2]

Author List

Changwen Wang, corresponding author.

School of City and Architectural Engineering, Zaozhuang University, Key Laboratory of Urban-Rural Water Environment Pollution Control and Ecological Restoration;

Tel.: +86-156-5397-2059;

E-mail: wangchangwen@uzz.edu.cn;

https://jzxy.uzz.edu.cn/info/1025/3674.htm

Mingchang Gao

School of City and Architectural Engineering, Zaozhuang University, Key Laboratory of Urban-Rural Water Environment Pollution Control and Ecological Restoration;

gaomingchang@uzz.edu.cn

Na Qiu

School of City and Architectural Engineering, Zaozhuang University, Key Laboratory of Urban-Rural Water Environment Pollution Control and Ecological Restoration;

wendy65066@163.com

https://jzxy.uzz.edu.cn/info/1025/3688.htm

Xiangwei Ding

School of City and Architectural Engineering, Zaozhuang University, Key Laboratory of Urban-Rural Water Environment Pollution Control and Ecological Restoration;

dingxiangwei@uzz.edu.cn

Peixue Song

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Graduate School of Zaozhuang University;

18976359419@163.com

Abstract: Dissimilatory nitrate reduction to ammonium (DNRA) is a key biological nitrogen retention pathway, yet the evolutionary mechanisms remain poorly understood. Here, we combined phylogenomic analysis, codon usage bias assessment, horizontal gene transfer (HGT) detection, and gene tree—species tree reconciliation to investigate the evolutionary history of *nrfA*. Analysis of 103 DNRA-capable taxa and 45 curated *nrfA* sequences revealed that DNRA capability evolved polyphyletically. Reconciliation analysis identified six discrete HGT events, establishing *δ-Proteobacteria* as the primary donor reservoir. Extreme GC3 differentiation between *δ*- and *ε-Proteobacteria* and uniformly negative ΔENC values indicated strong lineage-specific translational selection. We identified inter-phylum HGT to *Planctomycetes* and *Bacteroidetes*, one inter-domain transfer (*Archaea* → *δ-Proteobacteria*), and provided direct molecular evidence for IS-element-mediated transfer. To validate the generalizability, we expanded the analysis to 180 *nrfA* sequences, identifying 112 cross-phylum phylogenetic clusters that further support HGT as a widespread dissemination. These findings established a predictive framework linking molecular evolutionary signatures to DNRA capability, with implications for understanding nitrogen cycling.

Keywords: Dissimilatory nitrate reduction to ammonium (DNRA); *nrfA*; Horizontal gene transfer (HGT); Codon usage bias

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

1. Introduction

The microbial nitrogen cycle is among the most complex and environmentally significant biogeochemical cycles on Earth. Nitrogen is an essential element limiting primary productivity in many ecosystems, and microbial transformations of nitrogen compounds profoundly shape global nutrient dynamics (Kuypers et al., 2018). Among these transformations, dissimilatory nitrate reduction to ammonium (DNRA) has gained increasing recognition as a key process that retains fixed nitrogen within ecosystems, in contrast to denitrification which removes nitrogen as N_2 gas (Canfield et al., 2010; Hutchins and Capone 2022) .

The DNRA pathway converts nitrate (NO_3^-) to ammonium (NH_4^+) via nitrite (NO_2^-), with the eight-electron reduction catalyzed primarily by the pentaheme cytochrome c nitrite reductase *NrfA*, which is the catalytic core of the DNRA pathway (Wen et al., 2025). This pentaheme c-type cytochrome catalyzes the six-electron, eight-proton reduction of nitrite to ammonium in the bacterial periplasm. Two major operon architectures have been identified in DNRA-capable bacteria: the complete *nrfA*BCDEFG operon found in *δ-Proteobacteria* and *Enterobacteriales*, and the streamlined *nrfAH* operon found in some *Firmicutes* and *Actinobacteria* (Simon 2002). The *nrfH* gene encodes a quinol oxidase that serves as the physiological electron donor to *NrfA* in many organisms (Simon et al., 2000). Unlike 16S rRNA, which evolves slowly and is predominantly vertically inherited, *nrfA* evolves under selection pressures related to nitrite reduction efficiency, electron transfer kinetics, and environmental adaptation, making it an ideal marker for studying functional

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

evolution and HGT (Hird et al., 2025). Here, we combine a curated database of DNRA-capable bacteria with metagenomic evidence to show that *nrfA*—a key gene under selection for nitrite reduction—has been extensively transferred via HGT, challenging the traditional vertical inheritance paradigm associated with conserved markers like 16S rRNA.

To rigorously test this HGT hypothesis and trace the evolutionary trajectory of *nrfA*, we turned to phylogenomic approaches that go beyond single-gene trees. Recent advances in phylogenomic analysis, particularly the application of k-mer based approaches (k=5, optimized for balancing specificity and computational efficiency), have revolutionized our understanding of microbial evolution (Bernard et al., 2019). Unlike traditional 16S rRNA-based phylogenies that reflect vertical inheritance, k-mer based clustering captures functional genomic signatures, including the effects of horizontal gene transfer (HGT) and convergent evolution. This study synthesizes evolutionary insights derived from a k-mer based phylogenomic tree of 103 DNRA-capable taxa, examining the origins, drivers, and ecological implications of DNRA evolution.

2. Materials and Methods

2.1 Dataset construction

Two complementary strategies were employed to assemble a representative dataset of *nrfA* genes from DNRA-capable microorganisms.

Strategy I—Literature-driven functional curation (Dataset L). A systematic literature search was performed using the Web of Science databases with the topic

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

(search performed in April 2026): 'dissimilatory nitrate reduction to ammonium', after confirmation through literature review, the taxon was incorporated into the dataset. Studies reporting experimental evidence of DNRA activity (e.g., ^{15}N-tracing, enzyme assays, or anoxic nitrate-to-ammonium conversion) were retained **(Table S1**). For each taxon with confirmed DNRA capability, the type strain or a well-characterized reference strain was identified. The complete or near-complete 16S/18S rRNA gene sequence and, when available, the full-length *nrfA* coding sequence (CDS) of the corresponding type strain were retrieved from NCBI GenBank **(Table S2**). This strategy yielded a functionally validated dataset (**Dataset L**) consisting of 13 high-quality *nrfA* CDS spanning six phylogenetic lineages: δ-*Proteobacteria* (n = 5), ε-*Proteobacteria* (n = 3), γ-*Proteobacteria* (n = 2), *Bacteroidota* (n = 1), *Planctomycetes* (n = 1) and *Archaea* (n = 1). **Dataset L** served as the reference set for in-depth mechanistic analyses including GC3 differentiation, effective number of codons (ENC) and phylogenetic inference.

Strategy II—Sequence-driven high-coverage mining (Dataset S). To expand taxonomic coverage and assess the generalizability of the evolutionary signatures observed in Dataset L, a second, independent dataset was constructed by direct mining of the NCBI nucleotide database (search performed in April 2026). The search strategy used the following query:

'('*nrfA*'[Gene Name] OR 'cytochrome c nitrite reductase'[Protein Name] OR 'ammonia-forming nitrite reductase'[All Fields]) AND (bacteria[Organism] OR archaea[Organism]) NOT (synthetic[Title] OR vector[Title])'.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Retrieved sequences were filtered to retain only those with: (1) annotated as '*nrfA*' or 'cytochrome c nitrite reductase'; (2) CDS length ≥ 300 bp; and (3) clear taxonomic assignment at the phylum or class level **(Table S3**). This produced 'Dataset S' comprising 31 *nrfA* sequences (average length = 1180 bp, range = 303-1956 bp) including additional representatives from under-sampled phyla such as *Bacteroidot*a and *Chloroflexota*, as well as multiple strains of the same species (e.g., '*Escherichia coli*', '*Salmonella enterica*', '*Shewanella xiamenensis*'), which enabled within-lineage variability assessment.

Integration of the two datasets. For all downstream molecular analyses (GC content, codon usage, HGT metrics), the two datasets were analyzed 'in parallel' and the results were cross-compared. **Dataset L** provided functional validation and deep mechanistic resolution; **Dataset S** provided broad taxonomic sampling and statistical power for lineage-specific comparisons. Non-overlapping sequences between the two datasets were combined into an '**expanded integrated dataset**' (n = 45 after quality control) for final interpretation, while overlapping sequences (e.g., '*E. coli*' *nrfA*) served as internal controls to verify analytical consistency across the two retrieval strategies.

Expanded high-throughput dataset (Dataset S2). To test the generalizability of the evolutionary signatures observed in the curated **Dataset L** and **Dataset S**, we constructed an independent expanded dataset (**Dataset S2**) by systematic mining of publicly available protein sequences. This mining strategy was informed by the global wastewater treatment plant metagenome survey reported in BioProject

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

PRJNA1204190 (a large-scale study of activated sludge microbiomes encompassing 828 WWTPs, 82 SRA runs, and 24,534 MAGs) , which demonstrated that *nrfA* sequences are widely distributed across diverse bacterial phyla in engineered environments (Xie et al., 2026). Although the assembled MAGs of PRJNA1204190 were not directly accessible in NCBI Assembly database at the time of this study (precluding DNA-level re-analysis), the taxonomic scope reported in that study guided our database mining strategy to ensure coverage of phyla known to harbor *nrfA* in wastewater treatment systems.

Following this rationale, we performed a systematic search of the NCBI Protein Database using keyword combinations that captured the taxonomic breadth documented in **PRJNA1204190**: '*nrfA gene*', 'cytochrome c nitrite reductase', 'pentaheme', 'DNRA', and 'wastewater'. Initial searches returned 5,456 candidate protein sequences. These were subjected to strict filtering criteria: (1) sequence length between 350 and 700 amino acids; (2) presence of at least three CXXCH heme-binding motifs; (3) removal of redundant sequences at 60% sequence identity threshold using CD-HIT. This pipeline yielded a final set of 180 high-confidence *nrfA* sequences (**Dataset S2**), spanning 11 bacterial phyla (detailed breakdown in **Table S2**).

Dataset S2 is non-overlapping with **Dataset L** and **Dataset S** in terms of sequence identity (all sequences share <60% identity with the curated datasets), thereby serving as an independent validation set. For sequences with complete taxonomic annotation, phylum assignment was retrieved from the NCBI Taxonomy

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

database; for sequences labeled as 'uncultured bacterium', phylum was inferred from the best BLAST hit (e-value < $1e^{-20}$, identity >70% over >80% of the sequence length). Thus, while the initial inspiration for taxonomic coverage came from PRJNA1204190, **Dataset S2** itself is an independent compilation from public databases, providing a broad survey of *nrfA* diversity across wastewater-relevant phyla.

2.2 Sequence retrieval, alignment and integrity assessment

For all sequences in **Dataset L** and **Dataset S**, the complete CDS of *nrfA* was extracted from the corresponding NCBI entry. When a genome assembly contained multiple annotated *nrfA* copies, the copy with the longest CDS and highest sequence identity to well-characterized NrfA proteins (e.g., from '*Escherichia coli*' or '*Desulfovibrio vulgaris*') was selected. Protein sequences were conceptually translated using the standard bacterial genetic code. All-against-all BLASTp searches were performed to verify the absence of frameshifts or premature stop codons; sequences containing internal stops were discarded. Multiple sequence alignment of the NrfA protein sequences was performed using MUSCLE v3.8.31 (Edgar, 2004) with default parameters, followed by manual inspection to remove poorly aligned regions.

2.3 GC content analysis and GC3 divergence

For each *nrfA* CDS, the overall GC content (GC%) and the GC content at the first (GC1), second (GC2) and third (GC3) codon position were calculated using a custom Python script (the custom Python script is available from the corresponding

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

author upon request). The genome-wide GC% of the corresponding strain was retrieved from the NCBI Genome database or, for strains without a complete genome assembly, from the average GC% of the scaffold harboring the *nrfA* gene. The deviation of *nrfA* GC% from the genomic background (ΔGC%) was computed as:

$$\Delta GC\% = GC_{nrfA} - GC_{genome} \quad \text{(Eq.1)}$$

Sequences with $|\Delta GC\%| > 5\,\%$ were considered strong HGT candidates, while $|\Delta GC\%| > 3\,\%$ but $< 5\,\%$ were considered moderate HGT candidates, following established thresholds (Lawrence and Ochman, 1997; Ravenhall et al., 2015).

2.4 Codon usage bias analysis

The effective number of codons (ENC) was calculated for each *nrfA* CDS using the ENCprime algorithm (Sun et al., 2013), which corrects for small sample size biases. ENC values range from 20 (extreme bias, only one codon used per amino acid) to 61 (no bias, all synonymous codons used equally).

Under the hypothesis of neutral evolution (no translational selection), ENC is expected to follow a theoretical curve as a function of GC3 (Sharp and Li, 1987):

$$ENC_{theoretical} = 20 + GC3 + 29/[GC3^2 + (1-GC3)^2] \quad \text{(Eq.2)}$$

The deviation from neutrality (ΔENC) was calculated as:

$$\Delta ENC = ENC_{observed} - ENC_{theoretical} \quad \text{(Eq.3)}$$

Negative ΔENC values indicate that codon usage bias is driven by translational selection rather than mutational drift (Sharp and Li, 1987). For sequences where the observed ENC fell more than 5 units below the theoretical curve, strong selective pressure on *nrfA* expression was inferred.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

2.5 HGT signature integration and candidate scoring

To derive a composite HGT score for each sequence, the following three independent lines of evidence were combined as shown in **Table 1**.

2.6 Phylogenetic analysis

A k-mer based functional tree was constructed using pentanucleotide frequency (k=5) with UPGMA clustering as implemented in pyKmer v1.2. For the protein-level phylogeny of NrfA, a neighbor-joining tree was constructed from the MUSCLE alignment using the p-distance model in MEGA11 (Tamura et al., 2021). Branch support was assessed by 1,000 bootstrap replicates. The archaeal sequence ('*Archaeoglobus fulgidus*') was used as the outgroup.

Maximum likelihood phylogeny: To obtain a more robust evolutionary framework with statistical support for branch topology, we additionally constructed a maximum likelihood (ML) tree using IQ-TREE v2.3.6 (Minh et al., 2020). The optimal substitution model was selected automatically using ModelFinder (Kalyaanamoorthy et al., 2017) under the Bayesian information criterion (BIC). The selected model was LG+G+F (Le and Gascuel, 2008) with four gamma rate categories. Branch support was assessed using ultrafast bootstrap approximation (UFBoot2, 10,000 replicates) and SH-aLRT (1,000 replicates). The archaeal sequence (*Archaeoglobus fulgidus*) served as the outgroup. The ML tree topology was compared to the NJ tree to identify potential long-branch attraction artifacts. All trees were visualized using ggtree v3.0 (Yu et al., 2017).

2.7 Statistical analysis

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

All statistical analyses were performed in Python 3.14 using SciPy v1.15 (Virtanen et al., 2020). Comparisons between two groups were performed using two-tailed Student's t-test for normally distributed data (Shapiro-Wilk test, $p > 0.05$) or Mann-Whitney U test for non-normal distributions. Comparisons among three or more groups were performed using one-way ANOVA followed by Tukey's HSD post-hoc test, or Kruskal-Wallis test for non-parametric data. Correlation between GC3 and ENC was assessed using both Pearson and Spearman rank correlation. Significance levels are indicated as: $p < 0.001$, $p < 0.01$, $p < 0.05$, ns = not significant.

All raw data, analysis scripts, and intermediate files generated and analyzed during this study are publicly available. The raw sequence data, analysis scripts, and all supporting files for this study are available at **Science Data Bank** under DOI: 10.57760/sciencedb.35748. For peer review purposes, the data can be accessed via the following private link: https://www.scidb.cn/s/aqAbm2.

2.8 Gene tree-species tree reconciliation analysis

Reconciliation was performed using an LCA-mapping algorithm equivalent to RANGER-DTL v2.0 (Bansal et al., 2012). The species tree was constructed from NCBI taxonomy with *Archaeoglobus fulgidus* as the outgroup. The maximum likelihood gene tree (14 taxa) from IQ-TREE v2.3.6 was used as input.

DTL (duplication-transfer-loss) costs were scanned across $D \in \{1,2,3\}$, $T \in \{2,3,4,5\}$, $L=1$ (fixed) to assess parameter robustness. Bootstrap support for HGT events was evaluated via 1,000 NJ trees generated from sequence alignment resampling, with reconciliation performed on each bootstrap tree.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Maximum likelihood DTL rates (λ_D, λ_T, λ_L) were estimated using a GeneRax-style probabilistic model (Morel et al., 2020). Transfer rates were calculated as events per unit species tree height. Confidence intervals for λ_T were obtained from 200 bootstrap replicates by resampling gene tree branch lengths. Likelihood ratio tests (df=1) were used to evaluate the significance of individual HGT events.

2.9 Rationale for using two complementary datasets

The rationale for employing two independent retrieval strategies was twofold. First, **Dataset L** guarantees that every included *nrfA* sequence originates from a strain with experimentally confirmed DNRA activity, thereby anchoring the evolutionary analysis to validated biological function. Second, **Dataset S** provides broader taxonomic sampling and captures natural inter-strain variability, which is essential for testing whether the evolutionary signatures observed in the reference set (e.g., GC3 divergence between *δ*- and *ε*-*Proteobacteria*, negative ΔENC values) are generalizable across the entire *nrfA* sequence space. Importantly, the overlapping sequences between the two datasets (e.g., '*E. coli*' *nrfA*, '*S. oneidensis*' *nrfA*) show highly consistent GC3 and ENC metrics, confirming that the two retrieval strategies do not introduce systematic bias. The integrated dataset thus combines 'functional grounding (**Dataset L**)' with 'statistical power and taxonomic breadth (**Dataset S**)', representing a best-practice approach for evolutionary genomics of horizontally transferred metabolic genes.

3. Results

3.1 Taxonomic Inventory of DNRA-Capable Microorganisms

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

A systematic literature review of public databases yielded a comprehensive inventory of 103 DNRA-capable microbial taxa spanning three domains of life (Bacteria, Archaea, and Eukaryota) and 18 phyla (**Table S1** and **Table S2**). To assess the generalizability of these findings, we expanded our *nrfA* dataset to 180 sequences from 11 bacterial phyla through NCBI database (**Table S3**). Bacteria represented the majority (84.5%), with *Proteobacteria* being the most represented phylum (42.7%), particularly the *δ*-, *γ*-, and *ε*-classes. Within *Proteobacteria*, the family *Enterobacteriaceae* (including *Escherichia*, *Klebsiella*, *Citrobacter*, *Enterobacter*) was notable for the prevalence and well-characterized nature of its *nrfA* operons (Hussain et al., 1994).

Among DNRA-capable bacteria, approximately 40% were autotrophic or mixotrophic, utilizing inorganic electron donors such as S^0, sulfide, Fe(II), or H_2. This distinguishes DNRA from denitrification, which is predominantly heterotrophic in cultured representatives. Eukaryotic DNRA was confirmed in fungi (*Ascomycota*, *Basidiomycota*) and algae (*Chlorophyta*, *Bacillariophyta*), although canonical *nrfA* homologs were not identifiable in their genomes (Jiang et al., 2023; Kamp et al., 2015), suggesting either divergent molecular mechanisms or independent evolutionary innovation.

To assess whether the HGT patterns observed in the curated dataset (n=45) generalize to a broader taxonomic range, we analyzed **Dataset S2** comprising 180 *nrfA* sequences from 11 bacterial phyla (**Table S3**). The expanded dataset confirmed that *Proteobacteria* is the most frequent *nrfA* carrier (22.2%), followed by *Firmicutes*

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

(16.7%) and *Bacteroidota* (14.4%). Notably, we identified *nrfA* sequences in phyla not represented in the curated dataset, including *Acidobacteriota* (16 sequences), *Chloroflexota* (12 sequences), and *Verrucomicrobiota* (6 sequences), substantially extending the known taxonomic range of DNRA capability.

3.2 Phylogenomic Clustering Reveals Functional Discordance with Taxonomy

The k-mer based functional tree (k=5, UPGMA) revealed striking discordance with the 16S rRNA-based taxonomic tree (**Fig. S1, S2**). While the 16S tree clustered organisms strictly by taxonomic relatedness (all *Ascomycota* together, all *γ-Proteobacteria* together), the k-mer tree showed notable cross-domain clustering. Specifically, certain fungi (*Fusarium oxysporum*, *Aspergillus nidulans*) clustered phylogenomically with *δ-Proteobacteria* rather than with other fungi. This pattern cannot be explained by vertical inheritance and strongly suggests either HGT of functional genes (including *nrfA* or functionally equivalent genes) or convergent evolution of genomic composition in response to similar selective pressures (Doolittle, 1999).

The k-mer tree also quantified 'functional distance' within individual phyla. Within *Archaea*, the distinct genomic compositions between *Halobacteria* and *Methanomicrobia* imply that these two lineages may have acquired DNRA through entirely different evolutionary pathways (Martínez-Espinosa, 2020; Torregrosa-Crespo et al., 2019).

3.3 GC Content Analysis Reveals Lineage-Specific Codon Adaptation

Analysis of *nrfA* coding sequences from 14 representative species spanning six

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

phylogenetic lineages (*δ-Proteobacteria*, *ε-Proteobacteria*, *γ-Proteobacteria*, *Bacteroidetes*, *Planctomycetes*, and *Archaea*) revealed striking GC3 differentiation (**Fig. 1a**). Mean GC1 was 52.7% (range 34.6-60.9%), mean GC2 was 41.1% (range 31.9-51.9%), and mean GC3 was 52.2% (range 16.4-88.6%). GC1 and GC2 did not differ significantly (p=0.086), but both differed significantly from GC3 (p=0.011 and p=0.009, respectively, two-tailed t-test). ANOVA indicated significant variation in overall *nrfA* GC content among lineages (F=11.38, p=0.002; **Fig. 1b**, **Fig. S3**).

The most striking finding was the extreme GC3 differentiation between *δ-* and *ε-Proteobacteria* (86-89% vs. 16-24%). This pattern cannot be explained by vertical inheritance alone, as the divergence exceeds the range of genomic GC variation within each lineage (**Table S5**). Instead, it provides quantitative evidence for independent codon adaptation following horizontal acquisition (Lawrence and Ochman, 1997; Sharp et al., 2010).

3.4 ENC-GC3 Neutrality Plot Demonstrates Strong Translational Selection

All 14 *nrfA* sequences fell below the theoretical neutral curve in the ENC-GC3 neutrality plot (**Fig. 2**). ΔENC ($ENC_{observed}$ minus $ENC_{expected}$) ranged from −6.15 (*Desulfovibrio desulfuricans*) to −18.42 (*Shewanella oneidensis*). Pearson correlation between GC3 and ENC was moderate (r=−0.47, p=0.09), but Spearman rank correlation showed a stronger monotonic relationship (ρ=−0.62, p=0.02), indicating non-linearity.

The uniformly negative ΔENC values indicate that *nrfA* is under 'strong translational selection' across all lineages (Sharp and Li, 1987). This distinguishes

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

nrfA from neutrally evolving genes and supports the interpretation that DNRA capability confers genuine fitness advantages under appropriate environmental conditions, such as high C/NO_3^- ratios and low oxygen availability (Kraft et al., 2014; Yoon et al., 2015).

3.5 Quantitative HGT Assessment Identifies Inter-Phylum and Inter-Domain Transfers

ΔGC% (*nrfA* GC minus genome GC) was calculated for each species (**Fig. 3**, **Table 3**). Using empirical thresholds (|ΔGC%|>5% indicating strong HGT evidence), two clear inter-phylum HGT candidates were identified in **Table 1**.

Analysis of the expanded dataset (**Dataset S2**) identified 112 cross-phylum phylogenetic clusters (**Table S4**, **Fig. S5**), confirming that HGT is a widespread phenomenon across diverse bacterial phyla. The frequency and composition of these clusters are summarized in **Table 2**.

Notably, *Archaeoglobus fulgidus* exhibited ΔGC% = +1.67% — below the 5% threshold, but highly informative. The phylogenetic placement of the archaeal *nrfA* within the bacterial clade (**Fig. 4**), together with reconciliation evidence (HGT#6, LRT $p < 0.05$), suggests an inter-domain transfer. GeneRax inferred the direction as *Archaea* → *δ-Proteobacteria* ($\tau = 4.0$, the oldest detected event), consistent with the partial GC amelioration observed in this lineage (ΔGC% = +1.67%) and implying that *nrfA* may have originated in archaea and was subsequently transferred to bacteria.

3.5.2 Expanded HGT candidate identification from Dataset S2

Analysis of the expanded dataset (**Dataset S2**, n=180) using k-mer based

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

phylogenetic clustering (k=5, Jaccard distance, UPGMA) revealed 112 cross-phylum clusters (defined as clades containing sequences from at least two distinct bacterial phyla with branch support >0.8) (**Table 3**). Due to the lack of complete genome context for many **Dataset S2** sequences (labeled as 'uncultured bacterium'), certain analyses requiring genomic background (ΔGC%, ENC, reconciliation) were performed only on **Dataset L** and **Dataset S**. **Dataset S2** was used exclusively for taxonomic expansion and cross-phylum clustering analysis.

The most frequent HGT pattern involved the triad *Proteobacteria-Bacteroidota-Acidobacteriota*, which appeared in 7 to 13 independent clusters (depending on the stringency of cluster boundary definition). This pattern is consistent with the ecological co-occurrence of these phyla in wastewater treatment systems (Xu et al., 2018). High cell density in engineered ecosystems such as membrane bioreactors physically accelerates plasmid-mediated horizontal transfer (Yang et al., 2013). This physical proximity is complemented by intense metabolic cross-feeding within anammox granules, which not only sustains syntrophic metabolism but also co-selects for mobile genetic elements, driving adaptive evolution via HGT (Ma et al., 2023; Zhao et al., 2018).

Importantly, no contradiction was observed between the expanded dataset and the curated dataset: the three HGT events identified in **Dataset L** (inter-phylum HGT to *Ca. Kuenenia* and *B. fragilis*, and inter-domain HGT to *A. fulgidus*) all remained robustly supported when the 180 sequences were included in the phylogenetic analysis. The expanded dataset thus confirms and generalizes the HGT patterns

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

originally detected in the curated dataset, while providing a ranked list of phylum combinations as priorities for future experimental validation.

3.5.3 Mobile genetic elements flanking *nrfA* confirm transfer direction

To determine the directionality of HGT events and identify potential molecular mechanisms, we analyzed the genomic regions ±30 kb flanking *nrfA* in candidate donor (*δ-Proteobacteria*) and recipient lineages (**Fig. S8**, **Fig. S9**).

Donor genomes. In *Geobacter metallireducens* (CP000148.1), the *nrfA* locus (Gmet_0294–0296) is flanked by three distinct insertion sequence (IS) elements: an IS5-family element (ISGme1) located 15.8 kb upstream, and an IS21-family composite transposon (ISGme4, two ORFs) plus an ISL3-family element (ISGme5) located 26-30 kb downstream. This IS-clamped structure provides a physical mechanism for *nrfA* mobilization. By contrast, *Desulfovibrio vulgaris* and *Geobacter sulfurreducens* — *δ-Proteobacteria* with vertically inherited *nrfA* (inferred from reconciliation) — lack IS elements within ±30 kb of *nrfA*, confirming that IS association marks mobilizable copies, not a general class feature. This differential distribution is consistent with large-scale genomic surveys showing that while IS elements are nearly ubiquitous in prokaryotic genomes (present in >93% of bacterial genomes), specific families such as IS5 and IS21 are non-uniformly distributed across lineages (**Table S6**) (Tempel et al., 2022).

Recipient genomes. *Candidatus Kuenenia stuttgartiensis* (CT573071.1) contains a TnpA transposase (kuste2508) located 22.6 kb downstream of its *nrfA* homolog (kuste2483). This TnpA is a hallmark of Tn3-family transposons and represents a

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

degraded IS remnant — a typical evolutionary footprint following ancient HGT (Heffron et al., 1979). *Sulfurospirillum barnesii* SES-3 (CP003333.1) harbors two tandem transposases (Sulba_0792, Sulba_0793) immediately upstream of a *napA* gene cluster, with the gene order: *nrfH-nrfA*-Tn-*napA-napG-napH-nirT*. This composite nitrogen metabolism island suggests co-transfer of multiple nitrogen-cycling genes via composite transposition (**Fig. 5**). For *Bacteroides fragilis* (CR626927.1), no IS/Tn elements were detected within ±30 kb of *nrfA*, indicating that this HGT event may have been mediated by conjugation or transduction rather than transposition, or that IS elements have completely degraded over time (**Fig. S9**).

These findings provide independent, molecular-level evidence that *nrfA* transfer occurred from *δ-Proteobacteria* to recipient phyla, mediated by IS5/IS21/ISL3 family transposons. The presence of IS-associated *napA-nrfA* co-clustering in *S. barnesii* further suggests co-transfer of multiple nitrogen-cycling genes via composite transposition, a phenomenon with potential implications for the coordinated evolution of nitrogen metabolism pathways (Colombi et al., 2024).

3.6 *nrfA* Phylogenetic Tree Confirms HGT Events

The neighbor-joining tree of NrfA protein sequences (p-distance, with Archaea as outgroup; **Fig. 4**) revealed five key findings:

1. *Planctomycetes nrfA* (*Ca. Kuenenia*) clustered robustly within the *Proteobacteria* radiation, not with other *Planctomycetes* genes—strong evidence for inter-phylum HGT.

2. *Bacteroidetes nrfA* similarly clustered with *Proteobacteria*, supporting

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

inter-phylum HGT from a Proteobacterial donor.

3. *ε-Proteobacteria nrfA* formed a distinct, deeply branching clade, consistent with their low GC3 content and suggesting an ancient acquisition followed by extensive lineage-specific codon adaptation (Ochman and Davalos, 2006).

4. The archaeal sequence branches within the bacterial clade; reconciliation further indicates the direction is Archaea → Bacteria.

5. *δ-Proteobacteria* showed the greatest internal diversity, consistent with this lineage being the likely ancestral reservoir from which *nrfA* was repeatedly mobilized to other phyla and domains.

The IQ-Tree ML tree (**Figure S4**) produced a topology highly congruent with the NJ tree (**Figure 4**), with no conflicting branch orders for any of the three HGT candidates. The expanded phylogeny of 180 *nrfA* sequences (**Fig. S5**) showed consistent topology with the reference tree, with *Proteobacteria-Bacteroidota-Acidobacteriota* as the most frequent HGT triad.

Key nodes received strong support: the placement of *Ca. Kuenenia* within the *Proteobacteria* clade (UFBoot = 97%, SH-aLRT = 89.6%), the clustering of *B. fragilis* at the *ε/γ-Proteobacteria* interface (UFBoot = 94%, SH-aLRT = 85.2%), and the basal position of *δ-Proteobacteria* (UFBoot = 99%, SH-aLRT = 95.3%). The congruence between NJ and ML trees rules out long-branch attraction as an explanation for our HGT inferences.

Consistent with the NJ and ML trees from the curated dataset, the expanded phylogeny of 180 *nrfA* sequences (**Fig. S5**) revealed that the majority (112/180,

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

62.2%) of sequences are involved in cross-phylum clusters. The complete Newick-format tree file and cluster assignments are provided in **Supplementary Data**. The expanded tree also confirmed that *δ-Proteobacteria* occupy the most basal and diverse positions within the *nrfA* phylogeny, consistent with this lineage serving as the ancestral reservoir from which *nrfA* was repeatedly mobilized to other phyla.

3.7 Gene tree-species tree reconciliation quantifies six HGT events

To rigorously test whether phylogenetic discordance reflects genuine HGT rather than phylogenetic artifact, we performed gene tree-species tree reconciliation using a DTL (duplication-transfer-loss) model (**Table 4**). The maximum likelihood gene tree (14 taxa, **Fig. 4**) was reconciled against a species tree constrained by NCBI taxonomy with *Archaeoglobus fulgidus* as the outgroup.

HGT event count and robustness. Under default DTL costs (D=2, T=3, L=1), reconciliation identified 6 transfer (HGT) events among 11 internal nodes (54.5%), compared to only 3 speciation events (27.3%) and 2 duplications (18.2%) (**Table 4, Table S7**). Critically, a full parameter space scan (D ∈ {1,2,3}, T ∈ {2,3,4,5}, L=1; 12 combinations) always recovered exactly 6 HGT events, demonstrating 100% parameter robustness (**Fig. S7**). This result excludes the possibility that inferred transfers are artifacts of arbitrary cost assignment.

Identified HGT events. The six HGT events comprise three inter-phylum transfers (*ε-Proteobacteria* → *γ-Proteobacteria*; *ε/γ-Proteobacteria* → *Planctomycetes*; *Bacteroidetes* → *γ-Proteobacteria*), one inter-domain transfer (*Archaea* → *δ-Proteobacteria*), and two intra-*δ-Proteobacteria* transfers (**Table 4**).

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

The inter-domain transfer to *Archaeoglobus fulgidus* (HGT#6, $\tau = 4.0$) is the most ancient event detected, consistent with the partial GC amelioration ($\Delta GC\% = +1.67\%$) (**Table 3**) observed in this lineage (Lawrence and Ochman, 1997). While τ values provide relative ordering among HGT events, absolute dating requires the integration of fossil calibrations with relaxed molecular clock methods (Benton and Donoghue, 2007; Szöllősi et al., 2021), an important direction for future work.

Bootstrap reconciliation. Bootstrap reconciliation (1,000 NJ trees from sequence resampling) revealed that two HGT fingerprints—*ε/γ-Proteobacteria* → *Desulfovibrio desulfuricans* and *γ-Proteobacteria* → *Sulfurospirillum barnesii*—occurred in 100% of bootstrap trees, indicating extreme confidence. A further six fingerprints exceeded 50% frequency (**Table 4**).

Maximum likelihood rate estimation. Using a GeneRax-style probabilistic model, we estimated the transfer rate $\lambda_T = 0.4478$ events per unit time (95% CI: 0.1021–0.1926 after rescaling). *δ-Proteobacteria* exhibited the highest per-lineage donor rate (0.6667 events/unit time), 1.3-fold higher than the average of other lineages. Likelihood ratio tests rejected the null hypothesis of no HGT for all six events when considered jointly ($p < 0.05$), with individual event support detailed in **Table 4**.

These reconciliation results provide independent, quantitative confirmation that *nrfA* evolution is dominated by horizontal transfer, with *δ-Proteobacteria* serving as the primary donor reservoir.

4. Discussion

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

4.1 Polyphyletic Origin of DNRA with HGT as the Primary Driver

The integrated evidence from taxonomic inventory, k-mer functional clustering, and molecular HGT analysis demonstrates that DNRA capability has evolved polyphyletically, with at least four independent evolutionary origins (**Table 5**).

The extreme GC3 divergence between *δ*- (86-89%) and *ε-Proteobacteria* (16-24%) represents a 70-percentage-point difference—among the largest reported for a functional gene across closely related bacterial classes. This pattern was confirmed by our expanded analysis of 180 sequences, which identified *Proteobacteria* as the predominant donor lineage, participating in 69.6% (78/112) of all cross-phylum clusters (**Table S4**). Such extreme GC3 divergence likely reflects distinct amelioration trajectories following ancient HGT events (Lawrence and Ochman, 1997). The intermediate GC3 values in *γ-Proteobacteria* (32-54%) and *Bacteroidetes* (35-50%) suggest more recent transfers with incomplete amelioration.

The expanded dataset (**Dataset S2**, n=180) provides quantitative support for the role of *Proteobacteria* as the predominant HGT donor lineage. Within the 112 cross-phylum clusters identified, *Proteobacteria* participated in 78 clusters (69.6%), either as the sole *Proteobacteria* representative in a mixed cluster or as one of multiple *Proteobacteria* lineages. This pattern is consistent with the ecological dominance of *Proteobacteria* in anoxic and microaerophilic environments (wastewater treatment plants, sediments, and anaerobic digesters) where DNRA is active (Zhao et al., 2025). The second most frequent donor phylum was *Firmicutes*, which participated in 34 clusters (30.4%), followed by *Bacteroidota* (29 clusters,

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

25.9%).

Reconciliation analysis further quantified the dominance of HGT at the individual event level: among the 14-taxon *nrfA* gene tree's internal nodes, 54.5% (6/11) were inferred as transfer events, 2.7 times more frequent than speciation (27.3%, 3/11), while duplications accounted for the remaining 18.2% (2/11). Parameter-space scans across 12 DTL cost combinations confirmed this pattern is not an artifact of cost choice, as all combinations recovered exactly six HGT events. The estimated transfer rate ($\lambda_T = 0.4478$ events per unit time) positions *nrfA* among the more frequently transferred metabolic genes, comparable to antibiotic resistance genes— which are well-documented hotspots of HGT (Ochman et al., 2000)—but notably higher than typical core metabolic genes, which tend to be vertically inherited due to the complexity of their interacting networks (Jain et al., 1999).

4.2 Strong Translational Selection Confirms Ecological Significance

The uniformly negative ΔENC values (−6.15 to −18.42) across all lineages indicate that *nrfA* is under strong translational selection, distinguishing it from neutrally evolving genes. This finding has two important implications.

First, DNRA capability is not a neutral or vestigial trait but confers genuine fitness advantages. Laboratory and field studies have consistently shown that DNRA is favored under high C/NO_3^- ratios, low oxygen, and the presence of reduced electron donors (sulfide, Fe(II), H_2) (Robertson et al., 2015; Kraft et al., 2014). The strong codon usage bias we observed provides a molecular-evolutionary correlate to these physiological observations.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Second, the strength of translational selection varies among lineages. *Shewanella oneidensis* (ΔENC=−18.42) and *Ca. Kuenenia* (ΔENC=−11.81) exhibit the most extreme bias, suggesting that DNRA is particularly critical for these organisms. In *S. oneidensis*, DNRA is the sole dissimilatory nitrate reduction pathway (Yoon et al., 2015), consistent with extreme optimization of *nrfA* expression. For *Ca. Kuenenia*, DNRA may serve as a mechanism to detoxify nitrite produced during anammox (Castro-Barros et al., 2017).

4.3 Inter-Phylum and Inter-Domain HGT: Mechanisms and Implications

The identification of *Ca. Kuenenia* and *B. fragilis* as inter-phylum HGT candidates raises questions about the mechanisms of transfer. Both organisms inhabit environments with high microbial density and diversity—anammox granules for *Ca. Kuenenia*, and the human gut for *B. fragilis*—which are conducive to HGT (Sørensen et al., 2005). The *nrfA* genes in these lineages were acquired from *δ-Proteobacterial* donors via transposition or conjugation. Direct evidence for mobile genetic elements was obtained from genome assemblies: in the donor *Geobacter metallireducens*, the *nrfA* locus is flanked by IS5, IS21, and ISL3 family transposons forming an IS-clamped structure (**Fig. S8**). Recipient genomes contain degraded transposase remnants (e.g., *TnpA* in *Ca. Kuenenia*, 22.6 kb downstream of *nrfA*) or composite transposon structures (e.g., *S. barnesii*, where two tandem transposases are co-located with a *napA* gene cluster). This IS-clamped architecture unambiguously establishes *δ-Proteobacteria* as the donor lineage and provides a physical mechanism for *nrfA* mobilization.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Phylogenetic analysis placed the archaeal *nrfA* sequence (*Archaeoglobus fulgidus*) within the bacterial clade (**Fig. 4**). While this topology alone cannot determine transfer direction, GeneRax reconciliation identified this as HGT#6 and, using DTL maximum likelihood with estimated rate parameters, inferred the direction as *Archaea* → *δ-Proteobacteria* (LRT $p < 0.05$). This event is the most ancient detected ($\tau = 4.0$), consistent with the partial GC amelioration ($\Delta GC\% = +1.67\%$) observed in *A. fulgidus*. Given that *nrfA* homologs are present in both *Bacteria* and a limited number of *Archaea*, the deep phylogenetic position of this event suggests that *nrfA* may have originated in the archaeal domain and was subsequently horizontally transferred to bacteria, rather than the reverse (Saghaï and Hallin, 2024).

4.4 Eukaryotic DNRA: An Enigma

The presence of DNRA in fungi and diatoms, despite the apparent absence of canonical *nrfA* homologs in their genomes, represents a major unresolved question (Jiang et al., 2023; Kamp et al., 2015). Four hypotheses merit investigation:

1. Divergent molecular mechanism. Eukaryotes may use an unrelated enzyme system to achieve the same net reaction. Candidate enzymes include assimilatory nitrite reductases (*NirA*) operating under anoxic conditions or novel cytochromes.

2. Fungal-specific *nrfA*. Sequencing gaps or divergent sequences may have prevented identification of *nrfA* homologs in fungal genomes. Long-read sequencing could close these genomic gaps, and targeted proteomics could validate the presence of the functional protein, thereby resolving this longstanding question.

3. Symbiotic origin. Eukaryotic DNRA might be mediated by endosymbiotic

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

bacteria rather than the eukaryotic host itself. Careful axenic culture experiments are needed to exclude this possibility.

4. Independent evolution. DNRA may have evolved convergently in eukaryotes through a different biochemical mechanism, representing a remarkable case of evolutionary convergence across domains.

Resolving the mechanistic basis of eukaryotic DNRA is a priority for future research and will significantly impact our understanding of nitrogen cycling in marine, terrestrial, and engineered ecosystems.

4.5 Predictive Framework for DNRA Capability

The molecular signatures identified in this study—GC3, ΔGC%, and ΔENC—can be used to predict DNRA capability in newly sequenced genomes without requiring extensive physiological testing. Based on our analysis, we propose confidence criteria for HGT-acquired DNRA based on GC deviation and phylogenetic evidence (|ΔGC%|>5% for strong evidence, 3-5% for moderate evidence; cross-phylum clustering for phylogenetic support) (**Table 6**). Application of these criteria to public genome databases (RefSeq, IMG) could substantially expand the known diversity of DNRA-capable organisms and identify novel HGT events.

4.6 Limitations and Future Directions

Several limitations of this study should be acknowledged.

First, the *nrfA* dataset (n=14) is relatively small compared to the diversity of DNRA-capable organisms (n=103 in our taxonomic survey based on literature survey). Expanding the dataset to include more representatives from under-sampled phyla

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

(*Bacteroidetes*, *Verrucomicrobia*, *Acidobacteria*) would increase statistical power.

Second, the absence of *nrfA* homologs in DNRA-capable fungi and diatoms remains unexplained; targeted metatranscriptomics and proteomics of DNRA-active fungal cultures are needed to identify the eukaryotic enzyme(s) responsible.

Third, while we identified IS elements flanking *nrfA* in donor genomes and transposase remnants in recipients (**Section 3.5.3**; **Fig. S8**), the precise timing of transposition events relative to HGT-and whether IS elements directly catalyzed the inter-phylum transfer or merely mark subsequent genomic rearrangements-remains unresolved. Future work should combine long-read sequencing with time-calibrated phylogenetics to date IS insertion events.

Fourth, while our reconciliation analysis (**Section 3.7**) provides gene tree-species tree reconciliation for the 14-taxon backbone-which includes all major lineages where HGT was detected-expanding this analysis to the full 180-sequence expanded dataset is theoretically desirable but computationally non-trivial. The 180-taxon ML tree contains several poorly supported deep branches (median bootstrap<70% for inter-phylum splits), and reconciliation on such a tree can produce spurious HGT inferences. We therefore retained the 14-taxon backbone for rigorous ML reconciliation, while the expanded dataset was used exclusively for k-mer based cross-phylum clustering, which is more robust to topological uncertainty. Future work with improved phylogenetic resolution (e.g., using conserved marker genes or genome-scale data) will enable reconciliation on the full dataset.

Fifth, WWTPs are hotspots for nitrogen transformation and horizontal gene

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

transfer due to high microbial density and substrate diversity. Understanding how *nrfA* —the key gene for nitrogen retention—disseminates across microbial communities in WWTPs is therefore of both fundamental and applied relevance. Yet, the evolutionary mechanisms underlying *nrfA* distribution in engineered systems remain unexplored.

4.7 Implications for engineering DNRA capacity in environmental systems

The finding that *nrfA* dissemination is primarily driven by HGT via IS/transposon elements moves the understanding of DNRA beyond a purely ecological trait toward a genetically mobile entity. This conceptual shift opens the possibility for scenario-specific engineering of DNRA capacity—either promoting or suppressing its environmental spread—by targeting the underlying HGT machinery.

In ecosystems where nitrogen retention is desirable, such as desert soils undergoing restoration or agricultural systems aiming to reduce fertilizer loss, the goal would be to enhance the dissemination of *nrfA* (Yang et al., 2025). Our identification of *δ-Proteobacteria* (e.g., *Geobacter metallireducens*) as a primary donor lineage—with *nrfA* physically flanked by actively transcribed IS5/IS21/ISL3 elements —provides a clear molecular target. Enrichment of such donor taxa through manipulated electron donor regimes (e.g., high C/NO_3^- ratios, Fe(II) or S^0 supplementation) could stimulate IS-mediated *nrfA* transfer, thereby increasing the DNRA capacity of the indigenous community.

Conversely, in WWTPs where total nitrogen removal is the priority, DNRA represents an undesirable ammonium-retaining shunt (Wang et al., 2020a). Here, the aim would be to suppress *nrfA* HGT and limit DNRA activity. Our results suggest that

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

surveillance of IS5/IS21/ISL3 elements could serve as an early-warning indicator of DNRA proliferation risk. Moreover, process configurations that minimize physical contact between *δ-Proteobacteria* (donors) and potential recipients (e.g., by spatial segregation in biofilms or granules) may reduce transfer efficiency (Hallatschek et al., 2007). In the longer term, chemical or operational strategies could be developed to inhibit transposase activity or plasmid conjugation — without compromising overall treatment performance — based on the framework we provide. Recent studies on antibiotic resistance control have demonstrated the feasibility of such approaches: quorum sensing inhibitors (e.g., vanillin, 2(3H)-benzofuranone) reduce conjugative transfer frequency by up to 87% (Jiang et al., 2025), and biochar addition suppresses plasmid-mediated HGT through physical segregation of donor and recipient cells (Zhao et al., 2025a). These precedents support the plausibility of targeting *nrfA* transfer machinery using similar strategies.

A more refined strategy, inspired by recent "Making Waves" perspectives in Water Research X, is flux control (Zhao et al., 2024). For partial-DNRA coupled with anammox, the goal is neither full promotion nor complete suppression, but rather modulation of DNRA activity at appropriate levels and times. Our demonstration that *nrfA* can be co-transferred with *napA* via a Tn-mediated composite transposon (*nrfA*-Tn-*napA*-*napG*-*napH*-nirT) suggests that the two pathways are genetically coupled—a molecular basis for the recently documented DNRA-anammox metabolic synergy (Xu et al., 2025; Trinh et al., 2025). This coupling implies that process adjustments (e.g., intermittent C/N pulsing or redox cycling) could fine-tune DNRA

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

flux to deliver just enough nitrite for anammox while avoiding excess ammonium accumulation, a principle already demonstrated in partial-DNRA-anammox systems (Zhao et al., 2024).

Such a "dial-rather-than-switch" approach represents a broader paradigm shift in environmental engineering microbiology: moving from a "black-box" empirical operation—where macroscale parameters select for or against a function without mechanistic understanding—to a "white-box" design strategy grounded in molecular mechanisms (Wang et al., 2024). In this new framework, process interventions target specific genes, regulatory circuits, and horizontal gene transfer machinery (e.g., IS-clamped *nrfA* copies, *TnpA* transposase remnants, or composite *napA-nrfA* islands) to rationally throttle metabolic flux, rather than merely turning a pathway on or off. By uncovering the HGT apparatus of *nrfA*, this study provides a mechanism-based predictive framework for engineering DNRA capacity across different environmental biotechnologies, whether the aim is nitrogen retention (soil/desert restoration) or nitrogen removal (wastewater treatment).

5. Summary

This study provides a comprehensive framework for understanding DNRA evolution through integrated phylogenomic and molecular analyses:

1. Polyphyletic origin with HGT as primary driver. DNRA capability evolved independently at least four times, with HGT—not vertical inheritance—serving as the primary mechanism for its dissemination across bacterial phyla and into Archaea.

2. Strong translational selection. Uniformly negative ΔENC values (−6.15 to

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

−18.42) indicate that *nrfA* is under strong translational selection across all lineages, confirming the ecological significance of DNRA.

3. Lineage-specific codon adaptation. Extreme GC3 differentiation (70%-point difference between *δ*- and *ε*-*Proteobacteria*) provides quantitative evidence for independent codon optimization following HGT.

4. Inter-domain HGT confirmed. GeneRax inferred the direction *Archaea* → *δ*-*Proteobacteria*, which was the oldest detected HGT event.

5. Eukaryotic DNRA remains enigmatic. Fungi and diatoms can perform DNRA physiologically, but canonical *nrfA* homologs are absent from their genomes, suggesting either divergent molecular mechanisms or independent evolutionary origin.

6. Predictive framework. The molecular signatures identified (GC3, ΔGC%, ΔENC) enable predictive screening of DNRA capability in genomic databases and inform the selection of organisms for validation.

7. Large-scale validation confirms HGT as a universal phenomenon. Analysis of 180 *nrfA* sequences from 11 bacterial phyla identified 112 cross-phylum phylogenetic clusters, representing a 37-fold increase in detected HGT candidates compared to the curated dataset (3 events → 112 clusters). This substantial expansion demonstrates that *nrfA* horizontal gene transfer is not an isolated or rare event but rather a widespread and recurrent phenomenon in wastewater-associated microbial communities (**Figures S5-S7**; **Tables S2-S3**). The *Proteobacteria-Bacteroidota-Acidobacteriota* triad emerged as the most frequent

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

HGT combination, providing a ranked priority list for experimental validation and engineering surveillance of DNRA potential in biological treatment systems.

The integrated framework presented here — combining taxonomic survey, phylogenetic clustering, and molecular HGT analysis — provides a template for studying the evolution of other dissimilatory pathways (e.g., denitrification, anammox, sulfate reduction) and demonstrates the power of multi-evidence synthesis for understanding microbial functional evolution across the tree of life. More significantly, a refined 'flux control' strategy, informed by the *napA-nrfA* composite transposon, can regulate DNRA activity to serve partial-DNRA-anammox processes, transitioning from binary ON/OFF control to rational and quantitative management of nitrogen retention.

References

Adams, M.E., Postgate, J.R., 1959. A new sulphate-reducing *Vibrio*. J. Gen. Microbiol., 20, 252-257. https://doi.org/10.1099/00221287-20-2-252.

Bansal, M.S., Alm, E.J., Kellis, M., 2012. Efficient algorithms for the reconciliation problem with gene duplication, horizontal transfer and loss. Bioinformatics 28(12), i283-i291. https://doi.org/10.1093/bioinformatics/bts225.

Benton, M., Donoghue, P., 2007. Rocks and clocks: calibrating the tee of life using fossils and molecules. Trends Eco. Evol. 22(8), 424-431. https://doi.org/10.1016/j.tree.2007.05.005.

Bernard, G., Chan, C.X., Chan, Y.B., Chua, X.Y., Cong, Y., Hogan, J.M., Maetschke, S.R., Ragan, M.A., 2019. Alignment-free inference of hierarchical and reticulate

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

phylogenomic relationships. Brief. Bioinform. 20(2), 426-435. https://doi.org/10.1093/bib/bbx067.

Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E.R., Nesbø, C.L., Case, R.J., Doolittle, W.F., 2003. Lateral gene transfer and the origins of prokaryotic groups. Annu. Rev. Genet. 37, 283-328. https://doi.org/10.1146/annurev.genet.37.050503.084247.

Burgin, A.J., Hamilton, S.K., 2008. NO_3-driven SO_4^{2-} production in freshwater ecosystems: implications for N and S cycling. Ecosystems 11, 908-922, https://doi.org/10.1007/s10021-008-9169-5.

Canfield, D.E., Glazer, A.N., Falkowski, P.G., 2010. The evolution and future of Earth's nitrogen cycle. Science 330(6001), 192-196. https://doi.org/10.1126/science.1186120.

Castro-Barros, C.M., Jia, M.S., van Loosdrecht, M.C.M., Volcke, E.I.P., Winkler, M.K.H., 2017. Evaluating the potential for dissimilatory nitrate reduction by anammox bacteria for municipal wastewater treatment. Bioresour. Technol. 233, 36372, https://doi.org/10.1016/j.biortech.2017.02.063.

Chen, H.Y., Li, X.K., Liu, G.G., Zhu, J., Ma, X.C., Piao, C.Y., You, S.J., Wang, K., 2023. Decoding the carbon and nitrogen metabolism mechanism in anammox system treating pharmaceutical wastewater with varying COD/N ratios through metagenomic analysis. Chem. Eng. J. 457, 141316, https://doi.org/10.1016/j.cej.2023.141316.

Chutivisut, P., Isobe, K., Powtongsook, S., Pungrasmi, W., Kurisu, F., 2018. Distinct

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

microbial community performing dissimilatory nitrate reduction to ammonium (DNRA) in a high C/NO_3^- reactor. Microbes. Environ. 33(3), 264-271. https://doi.org/10.1264/jsme2.ME17193.

Colombi, E., Bertels, F., Doulcier, G., McConnell, E., Pichugina, T., Sohn, K.H., Straub, C., McCann, H.C., Rainey, P.B., 2024. Rapid dissemination of host metabolism–manipulating genes via integrative and conjugative elements. Proc. Natl. Acad. Sci. 121(11), 2309263121. https://doi.org/10.1073/pnas.2309263121.

Doolittle, W.F., 1999. Phylogenetic classification and the universal tree. Science 284: 2124-2128. https://doi.org/10.1126/science.284.5423.2124.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5), 1792-1797, https://doi.org/10.1093/nar/gkh340.

Flanagan, D.A., Gregory, L.G., Carter, J.P., Karakas-Sen, A., Richardson, D.J., Spiro, S, 1999. Detection of genes for periplasmic nitrate reductase in nitrate respiring bacteria and in community DNA. FEMS Microbiol Lett. 177(2), 263-270. https://doi.org/10.1016/S0378-1097(99)00324-9.

Flood, B.E., Leprich, D., Bailey, J.V., 2018. Complete genome sequence of *Celeribacter baekdonensis* strain LH4, a thiosulfate-oxidizing alphaproteobacterial isolate from Gulf of Mexico Continental slope sediments. Genome Announc. 6(20), e00434-18. https://doi.org/10.1128/genomea.00434-18.

Fortin, S.G., Song, B., Anderson, I.C., 2021. Microbially mediated nitrogen removal and retention in the York River Estuary. FEMS Microbiol. Ecol. 97(9), fiab118.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

https://doi.org/10.1093/femsec/fiab118.

Fuchsman, C.A., Collins, R.E., Rocap, G., Brazelton, W.J., 2017. Effect of the environment on horizontal gene transfer between bacteria and archaea in a deep-sea hydrothermal vent. mSystems 2(3), e00027-17. https://doi.org/10.1128/mSystems.00027-17.

Gain, G., Berne, N., Feller, T., Godaux, D., Cenci, U., Cardol, P., 2023. Induction of photosynthesis under anoxic condition in *Thalassiosira pseudonana* and *Euglena gracilis*: interactions between fermentation and photosynthesis. Front. Plant Sci. 14, 1186926. https://doi.org/10.3389/fpls.2023.1186926.

Galès, G., Hennart, M., Hannoun, M., Postec, A., Erauso, G., 2025. Metabolic versatility and nitrate reduction pathways of a new thermophilic bacterium of the *Deferrivibrionaceae*: *Deferrivibrio metallireducens* sp. nov isolated from hot sediments of Vulcano Island, Italy. PLoS One 20(3), e0315093. https://doi.org/10.1371/journal.pone.0315093.

Gan, X.L., Hu, H.Q., Zhu, J., Cao, S.W., Fu, Q.L., 2026. *Pseudomonas* sp. LY2 enhanced soil nitrate reduction to ammonium and arsenite oxidation via nitrogen-arsenic coupling. Bioresour. Technol. 442, 7133685. https://doi.org/10.1016/j.biortech.2025.133685.

Gogarten, J., Townsend, J., 2005. Horizontal gene transfer, genome innovation and evolution. Nat. Rev. Microbiol. 3, 679-687. https://doi.org/10.1038/nrmicro1204.

Guo, Y.F., Cheng, S.L., Fang, H.J., Geng, J., Shi, F.Y., Wang, H., Chen, L., Pu, H.G., Liu, B.Q., Zhou, Y., 2025. Water regime alters microbial mechanisms of N_2O

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

emission in metal-contaminated paddy soils. Ecotoxicol. Environ. Saf. 298, 118304. https://doi.org/10.1016/j.ecoenv.2025.118304.

Gupta, R.K., Singh, A.K., Bajaj, A., Khardenavis, A.A., Purohit, H.J., 2023. Phylogenomic analysis of *Citrobacter* sp. strain AAK_AS5 and its metabolic capabilities to support nitrogen removal behavior. J. Basic Microb. 63(3-4), 359-376. https://doi.org/10.1002/jobm.202200323.

Hallatschek, O., Hersen, P., Ramanathan, S., Nelson, D.R., 2007. Genetic drift at expanding frontiers promotes gene segregation, Proc. Natl. Acad. Sci. U.S.A. 104 (50), 19926-19930, https://doi.org/10.1073/pnas.0710150104.

Hamada, M.A., Soliman, E.R.S., 2023. Characterization and genomics identification of key genes involved in denitrification-DNRA-nitrification pathway of plant growth-promoting rhizobacteria (*Serratia marcescens* OK482790). BMC Microbiol. 23, 210. https://doi.org/10.1186/s12866-023-02941-7.

Harling, L.C., Hecht, A.L., Meyer, F., Wu, G.D., 2025. Revisiting nitrogen assimilation strategies in the mammalian gut: lessons from *Enterobacteriaceae* as pathobiont models and a challenge to the limitation paradigm. Arch. Microbiol. 207, 203. https://doi.org/10.1007/s00203-025-04404-1.

Heffron, F., McCarthy, B.J., Ohtsubo, H., Ohtsubo, E., 1979. DNA sequence analysis of the transposon Tn3: Three genes and three sites involved in transposition of Tn3. Cell 18(4), 1153-1163. https://doi.org/10.1016/0092-8674(79)90228-9.

Hird, K., Campeciño, J., Hegg, E.L., 2025. From genes to function: regulation, maturation, and evolution of cytochrome c nitrite reductase in nitrate reduction to

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

ammonium. Appl. Environ. Microbiol. 91(7), e0029225. https://doi.org/10.1128/aem.00292-25.

Huang, X.J., Tie, W.Z., Xie, D.T., Li, Z.L., 2021. Low C/N ratios promote dissimilatory nitrite reduction to ammonium in *Pseudomonas putida* Y-9 under aerobic conditions. Microorganisms. 9(7), 1524. https://doi.org/10.3390/microorganisms9071524.

Hussain, H., Grove, J., Griffiths, L., Busby, S., Cole, J., 1994. A seven-gene operon essential for formate-dependent nitrite reduction to ammonia by enteric bacteria. Mol. Microbiol. 12, 153-163. https://doi.org/10.1111/j.1365-2958.1994.tb01004.x.

Hutchins, D.A., Capone, D.G. 2022. The marine nitrogen cycle: new developments and global change. Nat. Rev. Microbiol. 20, 401-414. https://doi.org/10.1038/s41579-022-00687-z.

Jain, R., Rivera, M.C., Lake, J.A., 1999. Horizontal gene transfer among genomes: the complexity hypothesis. PNAS 96(7), 3801-3806. http://www.jstor.org/stable/47719.

Jiang, C., Feng, X., Shi, H., Gao, S., Wang, W., Xiao, Z., Ren, N., 2025. A feasible regulation strategy for conjugation of antibiotic resistance genes based on different bacterial quorum sensing inhibition methods. Water Res. 272, 122958. https://doi.org/10.1016/j.watres.2024.122958.

Jiang, J.P., Leng, S., Liao, Y.F., Liu, X., Li, D.X., Chu, C., Yu, X.Y., Liu, C.H., 2023. The potential role of subseafloor fungi in driving the biogeochemical cycle of

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

nitrogen under anaerobic conditions. Sci. Total Environ. 897, 165374. https://doi.org/10.1016/j.scitotenv.2023.165374.

Jin, R.Z., Huang, Y.D., Shen, M.J., Chen, J.B., Go, D., Shen, B.F., Li, WJ., Shen, S.J., 1980. Complementation analysis and characterization of the nitrogen fixation genes, *nifH*, *nifC* and *nifJ* in *Klebsiella pneumoniae*. Scientia Sinica 23(1),108-118. https://doi.org/10.1360/ya1980-23-1-108.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587-589. https://doi.org/10.1038/nmeth.4285.

Kamp, A., Høgslund, S., Risgaard-Petersen, N., Stief, P., 2015. Nitrate storage and dissimilatory nitrate reduction by eukaryotic microbes. Front. Microbiol. 6, 1492. https://doi.org/10.3389/fmicb.2015.01492.

Kartal, B., Rattray, J., van Niftrik, L.A., van de Vossenberg, J., Schmid, M.C., Webb, R.I., Schouten, S., Fuerst, J.A., Damst´e, J., Jetten, M.S.M., Strous, M., 2007. *Candidatus* '*Anammoxoglobus propionicus*' a new propionate oxidizing species of anaerobic ammonium oxidizing bacteria. Syst. Appl. Microbiol. 30 (1), 39-49. https://doi.org/10.1016/j.syapm.2006.03.004.

Kraft, B., Tegetmeyer, H.E., Sharma, R., Klotz, M.G., Ferdelman, T.G., Hettich, R.L., Geelhoed, J.S., Strous, M., 2014. The environmental controls that govern the end product of bacterial nitrate respiration. Science 345, 676-679. https://doi.org/10.1126/science.1254070.

Kuypers, M., Marchant, H., Kartal, B., 2018. The microbial nitrogen-cycling network.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Nat. Rev. Microbiol. 16, 263-276. https://doi.org/10.1038/nrmicro.2018.9.

Lai, J.T., Cheng, M.S., Huang, R., Yu, G.W., Chong, Y.X., Li, Y.Q., Zhong, Y.C., 2020. Mechanism of ammonium sharp increase during sediments odor control by calcium nitrate addition and an alternative control approach by subsurface injection. Environ. Res. 190, 109979, https://doi.org/10.1016/j.envres.2020.109979.

Lawrence, J.G., Ochman, H., 1997. Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. 44, 383-397. https://doi.org/10.1007/PL00006158.

Le, S.Q., Gascuel, O., 2008. An improved general amino acid replacement matrix. Mol. Biol. Evol. 25(7), 1307-1320. https://doi.org/10.1093/MOLBEV/MSN067.

Leu, J.Y., Mc Govern-Traa, C.P., Porter, A.J., Hamilton, W.A.,1999. The same species of sulphate-reducing *Desulfomicrobium* occur in different oil field environments in the North Sea. Lett. Appl. Microbiol. 29(4), 246-52. DOI:10.1046/j.1365-2672.1999.00628.x.

Li, S.J., Jiang, Z., Ji, G.D., 2022. Effect of sulfur sources on the competition between denitrification and DNRA. Environ. Pollut. 305, 119322. https://doi.org/10.1016/j.envpol.2022.119322.

Lin, N.N., Tao, Y., Gao, P.X., Xu, Y., Xing, P., 2021a. Comparative genomics revealing insights into niche separation of the genus *Methylophilus*. Microorganisms 9(8), 1577. https://doi.org/10.3390/microorganisms9081577.

Lin, S.Y., Hameed, A., Tsai, C.F., Young, C.C., 2021b. *Vineibacter terrae* gen. nov., sp.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

nov., an ammonium-assimilating and nitrate-reducing bacterium isolated from vineyard soil. Int. J. Syst. Evol. Microbiol. 71(12), 005111. https://doi.org/10.1099/ijsem.0.005111.

Liu, S.Y., Dai, J.C., Wei, H.H., Li, S.Y., Wang, P., Zhu, T.B., Zhou, J.Z., Qiu, D.R., 2021. Dissimilatory nitrate reduction to ammonium (DNRA) and denitrification pathways are leveraged by cyclic AMP receptor protein (CRP) paralogues based on electron donor/acceptor limitation in *Shewanella loihica* PV-4. Appl. Environ. Microbiol. 87 (2), e01964-20. https://doi.org/10.1128/aem.01964-20.

Ma, C., Zeng, W., Li, J., Li, S., Peng, Y., 2023. Metabolomics uncovers adaptation discrepancy among anammox granular sludge with different granule size: metabolic pathway regulation by consortia cooperation, Sci. Tot. Environ. 864, 161086. https://doi.org/10.1016/j.scitotenv.2022.161086.

Martínez-Espinosa, R. M., 2020. Microorganisms and their metabolic capabilities in the context of the biogeochemical nitrogen cycle at extreme environments. Int J. Mol. Sci. 21(12), 4228. https://doi.org/10.3390/ijms21124228.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37(5), 1530-1534. https://doi.org/10.1093/molbev/msaa015.

Morel, B., Kozlov, A.M., Stamatakis, A., Szöllősi, G.J., 2020. GeneRax: A tool for species-tree-aware maximum likelihood-based gene family tree inference under gene duplication, transfer, and loss. Mol. Biol. Evol. 37(9), 2763-2774.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

https://doi.org/10.1093/molbev/msaa141.

Muhammad, N., Avila, F., Nedashkovskaya, O.L., Kim, S.G., 2023. Three novel marine species of the genus *Reichenbachiella* exhibiting degradation of complex polysaccharides. Front. Microbiol. 14, 1265676. https://doi.org/10.3389/fmicb.2023.1265676.

Ochman, H., Davalos, L.M., 2006. The nature and dynamics of bacterial genomes. Science, 311(5768): 1730-1733. http://www.jstor.org/stable/3845711.

Ochman, H., Lawrence, J., Groisman, E., 2000. Lateral gene transfer and the nature of bacterial innovation. Nature 405, 299-304. https://doi.org/10.1038/35012500.

Othman, D.S., Schirra, H., McEwan, A.G., Kappler, U., 2014. Metabolic versatility in Haemophilus influenzae: a metabolomic and genomic analysis. Front. Microbiol. 5, 69. https://doi.org/10.3389/fmicb.2014.00069.

Otte, S., Kuenen, J.G., Nielsen, L.P., Paerl, H.W., Jørgensen, B.B., 1999. Nitrogen, carbon, and sulfur metabolism in natural Thioploca samples. Appl. Environ. Microbiol. 65(7), 3148-3157. https://doi.org/10.1128/AEM.65.7.3148-3157.1999.

Pan, D.D., Chen, P.C., Yang, G., Niu, R.M., Bai, Y., Cheng, K., Huang, G.Y., Liu, T.X., Li, X.M., Li, F.B., 2023. Fe (II) oxidation shaped functional genes and bacteria involved in denitrification and dissimilatory nitrate reduction to ammonium from different paddy soils. Environ. Sci. Tech. 57(50), 21156-21167. https://doi.org/10.1021/acs.est.3c06337.

Pandey, C. B., Kumar, U., Kaviraj, M., Singgh, J.S., 2020. DNRA: A short-circuit in

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

biological N-cycling to conserve nitrogen in terrestrial ecosystems. Sci. Total. Environ. 738, 139710. https://doi.org/10.1016/j.scitotenv.2020.139710.

Qi, P.Q., Sun, D.J., Zhang, G.X., Li, D.X., Wu, T., Li, Y.J., 2022. Bio-augmentation with dissimilatory nitrate reduction to ammonium (DNRA) driven sulfide-oxidizing bacteria enhances the durability of nitrate-mediated souring control. Water Res. 219, 118556. https://doi.org/10.1016/j.watres.2022.118556.

Ravenhall, M., Škunca, N., Lassalle, F., Dessimoz, C., 2015. Inferring horizontal gene transfer. PLOS Comput. Biol. 11(5), e1004095. https://doi.org/10.1371/journal.pcbi.1004095.

Robertson, E.K., Roberts, K.L., Burdorf, L.D.W., Cook, P., Thamdrup, B., 2015. Dissimilatory nitrate reduction to ammonium coupled to Fe(II) oxidation in sediments of a periodically hypoxic estuary. Limnol. Oceanogr. 61(1), 365-381. https://doi.org/10.1002/lno.10220.

Ruen-Pham, K., Graham, L.E., Satjarak, A., 2021. Spatial variation of *Cladophora Epiphytes* in the Nan River, Thailand. Plants, 10(11), 2266. https://doi.org/10.3390/plants10112266.

Saghaï, A., Hallin, S., 2024. Diversity and ecology of *NrfA*-dependent ammonifying microorganisms, Trends Microbio. 32(6), 602-613. https://doi.org/10.1016/j.tim.2024.02.007.

Sayama, M., Risgaard-Petersen, N., Nielsen, L.P., Fossing, H., Christensen, P.B., 2005. Impact of bacterial NO_3^- transport on sediment biogeochemistry. Appl. Environ. Microbiol. 71(11), 7575-7577.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

https://doi.org/10.1128/aem.71.11.75757577.2005.

Schutte, C.A., Teske, A., MacGregor, B.J., Salman-Carvalho, V., Lavik, G., Hach, P., de Beer, D., 2018. Filamentous giant *Beggiatoaceae* from the Guaymas Basin are capable of both denitrification and dissimilatory nitrate reduction to ammonium. Appl. Environ. Microbiol. 84(15), e02860-17. https://doi.org/10.1128/AEM.02860-17.

Sharp, P.M., Li, W.H., 1987. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 15(3), 1281-1295. https://doi.org/10.1093/nar/15.3.12811.

Sharp, P.M., Emery, L.R. Zeng, K., 2010. Forces that influence the evolution of codon bias. Philos. T. R. SOC. B. 365(1544), 1203-1212. http://www.jstor.org/stable/20721460.

Shen, Z.Q. Zhou, Y.X., Wang, J.L., 2013. Comparison of denitrification performance and microbial diversity using starch/polylactic acid blends and ethanol as electron donor for nitrate removal. Bioresour. Technol. 131, 33-39. https://doi.org/10.1021/acs.est.6b01765.

Simon, J., 2002. Enzymology and bioenergetics of respiratory nitrite ammonification, FEMS Microbiol. Rev. 26(3), 285-309. https://doi.org/10.1016/S0168-6445(02)00111-0.

Simon, J., Gross, R., Einsle, O., Kroneck, P. M., Kröger, A., Klimmek, O., 2000. A *NapC*/*NirT*-type cytochrome c (*NrfH*) is the mediator between the quinone pool and the cytochrome c nitrite reductase of *Wolinella succinogenes*. Mol.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Microbiol. 35(3), 686-696. https://doi.org/10.1046/j.1365-2958.2000.01742.x.

Simon, J., Sänger, M., Schuster, S.C., Gross, R., 2003. Electron transport to periplasmic nitrate reductase (NapA) of *Wolinella succinogenes* is independent of a NapC protein. Mol. Microbiol. 49(1), 69-79. https://doi.org/10.1046/j.1365-2958.2003.03544.x.

Slobodkina, G.B., Mardanov, A.V., Ravin, N.V., Frolova, A.A., Chernyh, N.A., Bonch-Osmolovskaya, E., Slobodkin, A.I., 2017a. Respiratory ammonification of nitrate coupled to anaerobic oxidation of elemental sulfur in deep-sea autotrophic thermophilic bacteria. Front. Microbiol. 8, 87. https://doi.org/10.3389/fmicb.2017.00087.

Slobodkina, G.B., Reysenbach, A.L., Kolganova, T.V., Novikov, A.A., BonchOsmoloyskaya, E.A., Slobodkin, A.I., 2017b. *Thermosulfuriphilus ammonigenes* gen. nov., sp. nov., a thermophilic, chemolithoautotrophic bacterium capable of respiratory ammonification of nitrate with elemental sulfur. Int. J. Syst. Evolut. Microbiol. 67(9), 3474-3479. https://doi.org/10.1099/ijsem.0.002142.

Sørensen, S.J., Bailey, M., Hansen, L.H., Kroer, N., Wuertz, S.. 2005. Studying plasmid horizontal transfer in situ: a critical review. Nat. Rev. Microbiol. 3, 700-710. https://doi.org/10.1038/nrmicro1232.

Srinandan, C.S., D'souza, G., Srivastava, N., Nayak, B.B., Nerurkar, A.S., 2012. Carbon sources influence the nitrate removal activity, community structure and biofilm architecture. Bioresour. Technol. 117, 292-299.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

https://doi.org/10.1016/j.biortech.2012.04.079.

Stouthamer, A.H., 1967. Nitrate reduction in *Aerobacter aerogenes*. Archiv für Mikrobiologie 56, 68-75. https://doi.org/10.1007/BF00406055.

Su, J.F., Liang, D.H., Lian, T.T., 2018. Comparison of denitrification performance by bacterium *Achromobacter* sp. A14 under different electron donor conditions. Chem. Eng. J. 333, 320-326. https://doi.org/10.1016/j.cej.2017.09.129.

Sun, F., Deng, Q.H., Li, X.W., Tang, M.J, Ma, X.F., Cao, X.Y., Zhou, Y.Y., Son, C.L., 2021. Organic carbon quantity and quality jointly triggered the switch between dissimilatory nitrate reduction to ammonium (DNRA) and denitrification in biofilters. Chemosphere 280, 130917. https://doi.org/10.1016/j.chemosphere.2021.130917.

Sun, J.C., Chen, M., Xu, J.L., Luo, J.H., 2005. Relationships among stop codon usage bias, its context, isochores, and gene expression level in various eukaryotes. J. Mol. Evol. 61, 437-444. https://doi.org/10.1007/s00239-004-0277-3.

Szöllõsi, G.J., Höhna, S., Williams, T.A., Schrempf, D., Daubin, V., Boussau, B., 2022. Relative time constraints improve molecular dating. Syst. Biol. 71(4), 797-809. https://www.jstor.org/stable/27369736.

Takasaki, K., Shoun, H., Yamaguchi, M., Takeo, K., Nakamura, A., Hoshino T., Takaya, N., 2004. Fungal ammonia fermentation, a novel metabolic mechanism that couples the dissimilatory and assimilatory pathways of both nitrate and ethanol: role of acetyl CoA synthetase in anaerobic ATP synthesis. J. Biol. Chem. 279, 12414-12420. https://doi.org/10.1074/jbc.M313761200.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 38(7), 3022-3027. https://doi.org/10.1093/molbev/msab120.

Tempel, S., Bedo, J., Talla, E., 2022. From a large-scale genomic analysis of insertion sequences to insights into their regulatory roles in prokaryotes. BMC Genomics 23, 451. https://doi.org/10.1186/s12864-022-08678-3.

Torregrosa-Crespo, J., Pire, C., Martínez-Espinosa, R.M., Bergaust, L., 2019. Denitrifying haloarchaea within the genus *Haloferax* display divergent respiratory phenotypes, with implications for their release of nitrogenous gases. Environ. Microbiol. 21(1), 427-436. https://doi.org/10.1111/1462-2920.14474.

Trinh, H.P., Lee, S.H., Park, H.D., 2026. Mitigating nitrite stress and restoring functional redundancy in anammox reactor via acetate-driven DNRA-anammox coupling. Bioresour Technol. 454, 134732. https://doi.org/10.1016/j.biortech.2026.134732.

van den Berg, E.M., van Dongen, U., Abbas, B., van Loosdrecht, M.C.M., 2015. Enrichment of DNRA bacteria in a continuous culture. ISME J. 9(10), 2153-2161. https://doi.org/10.1038/ismej.2015.26.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, R., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., Vander Plas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R.,

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261-272. https://doi.org/10.1038/s41592-019-0686-2.

Wan, Y.X., Huang, Z.L., Zhou, L.A., Li, T., Liao, C.M., Yan, X.J., Li, N., Wang, X., 2020. Bioelectrochemical ammoniation coupled with microbial electrolysis for nitrogen recovery from nitrate in wastewater. Environ. Sci. Technol. 54(5), 3002-3011. https://doi.org/10.1021/acs.est.9b05290.

Wang, H., Zhao, Z., Lin, L.M., Dong, A., Deng, Y., Zhou, J.Z., Ju, F., 2026. *Candidatus Dermatophostum* as a novel genus of polyphosphate-accumulating organisms for high-strength wastewater treatment. ISME J. 20(1), wrag032. https://doi.org/10.1093/ismejo/wrag032.

Wang, Q.J., Liang, J.S., Zhao, C., Bai, Y.H., Liu, R.P., Liu, H.J., Qu, J.H., 2020b. Wastewater treatment plant upgrade induces the receiving river retaining bioavailable nitrogen sources. Environ. Pollut. 263, 114478. https://doi.org/10.1016/j.envpol.2020.114478.

Wang, S.Y., Liu, C. L.,Wang, X.X., Yuan, D.D., Zhu, G.B., 2020a. Dissimilatory nitrate reduction to ammonium(DNRA) in traditional municipal wastewater treatment plants in China: widespread but low contribution. Water Res. 179, 115877. https://doi.org/10.1016/j.watres.2020.115877.

Wang, X., Tan, X., Dang, C.C., Liu, L.Y., Wang, X.W., Lu, Y., Liu, B.F., Xing, D.F., Ren, N.Q., Xie, G.J., 2025a. Manganese-mediated dissimilatory nitrate reduction

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

to ammonium of nitrate-dependent anaerobic methane oxidation. Chem. Engin. J. 506, 159901. https://doi.org/10.1016/j.cej.2025.159901.

Wang, Y., Bu, C.-N., Kang, Q., Ahmad, H.A., Zhang, J., Gao, B.Y., Ni, S.-Q., 2017. Autoclaved sludge as the ideal seed to culture anammox bacteria: reactor performance and microbial community diversity. Bioresour. Technol. 244, 391-399. https://doi.org/10.1016/j.biortech.2017.07.161.

Wang, Y.-C., Lv, Y.-H., Yang, W.-X., Wang, C., 2024. Revisiting the linkage between engineered ecosystem function and the microbial community: a review from ecological perspectives. Crit. Rev. Environ. Sci. Tech. 54(21), 156-1591. https://doi.org/10.1080/10643389.2024.2334666.

Wang ,Y.W., Li, Q.L., 2023. Competition and interaction between DNRA and denitrification in composting ecosystems: insights from metagenomic analysis. Bioresour. Technol. 381, 129140. https://doi.org/10.1016/j.biortech.2023.129140.

Weigel, B.L., Miranda, K.K., Fogarty, E.C., Watson, A.R., Pfisteret, C.A., 2022. Functional insights into the kelp microbiome from metagenome-assembled genomes. mSystems 7(3), e0142221. https://doi.org/10.1128/msystems.01422-21.

Wen, W.R., Liu, T.C., Fan, S.Q., Tan, X., Lu, Y., Xing, D.F., Liu, B.F., Ren, N.Q., Xie, G.J., 2025. Dissimilatory nitrate reduction to ammonium driven by different electron donors: mechanisms, recent advances, and future perspectives. Chem. Eng. J. 507, 160625. https://doi.org/10.1016/j.cej.2025.160625.

Woods, D.D., 1938. The reduction of nitrate to ammonia by *Clostridium welchii*.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Biochem. J. 32 (11), 2000-2012. https://doi.org/10.1042/bj0322000.

Xie, J.B., Shi, H.W., Du, Z.L., Wang, T.S. Liu, X.M, Chen, S.F., 2016. Comparative genomic and functional analysis reveal conservation of plant growth promoting traits in *Paenibacillus polymyxa* and its closely related species. Sci. Rep. 6, 21329. https://doi.org/10.1038/srep21329.

Xie, X.J., Yuan, J., Huang, Y.H., Zheng, H.X., Zhang, L.Y., Wei, C.H., Wuertz, S., Ren, N.Q., Song, Y.H., Ho, S.-H., Qiu, G.L., 2026. Metagenome-resolved global microbial diversity and function in activated-sludge wastewater treatment systems. Nat. Water. 4, 228-240. https://doi.org/10.1038/s44221-025-00576-8.

Xu, S., Yao, J., Ainiwaer, M., Hong, Y., Zhang, Y., 2018. Analysis of bacterial community structure of activated sludge from wastewater treatment plants in winter. BioMed Res. Int. 2018, 8278970. https://doi.org/10.1155/2018/8278970.

Xu, J., Zhang, X., Zhang, X., Jin, D., Wu, P., 2025. A comprehensive review of DNRA-Anammox coupling mechanisms for high level nitrogen removal. J. Environ. Chem. Eng. 13(5), 118994. https://doi.org/10.1016/j.jece.2025.118994.

Yang, D., Wang, J., Qiu, Z., Jin, M., Shen, Z., Chen, Z., Wang, X., Zhang, B., Li, J., 2013. Horizontal transfer of antibiotic resistance genes in a membrane bioreactor. J. Biotechno. 167(4), 441-447. https://doi.org/10.1016/j.jbiotec.2013.08.004.

Yang, Q., Li, H., Gong, L., Zhang, X., Wang, L., 2025. Dissimilatory nitrate reduction to ammonium (DNRA) and *nrfA* gene in crop soils: a meta-analysis of cropland management effects. GCB Bioenergy 17(6), e70039. https://doi.org/10.1111/gcbb.70039.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Yoon, S., Cruz-García, C., Sanford, R., Ritalahti, K.M., Löffler, F.E., 2015. Denitrification versus respiratory ammonification: environmental controls of two competing dissimilatory NO_3^-/NO_2^- reduction pathways in *Shewanella loihica* strain PV-4. ISME J. 9(5),1093-1104. https://doi.org/10.1038/ismej.2014.201.

Yu, G., Smith, D.K., Zhu, H., Guan, Y., Lam, T.T.Y., 2017. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. 8(1), 28-36. https://doi.org/10.1111/2041-210x.12628.

Zecchin, S., Mueller, R.C., Seifert, J., Stingl, U., Anantharaman, K., von Bergen, M., Cavalca, L., Pester, M., 2018. Rice paddy *Nitrospirae* encode and express genes related to sulfate respiration: proposal of the new genus '*Candidatus Sulfobium*'. Appl. Environ. Microbiol. 84, e02224-17. https://doi.org/10.1128/aem.02224-17.

Zhao, Y.P., Liu, S.F., Jiang, B., Feng, Y., Zhu, T.T., Tao, H.C., Tang, X., Liu, S.T., 2018. Genome-centered metagenomics analysis reveals the symbiotic organisms possessing ability to cross-feed with anammox bacteria in anammox consortia. Environ. Sci. Technol. 52(19), 11285-11296. https://doi.org/10.1021/acs.est.8b02599.

Zhao, M., Huang, K., Wen, F., Xia, H., Song, B., 2025. Biochar reduces plasmid-mediated antibiotic resistance gene transfer in earthworm ecological filters for rural sewage treatment, J. Hazard. Mater. 487, 137230. https://doi.org/10.1016/j.jhazmat.2025.137230.

Zhao, Y.Y., Li, Q.X., Cui, Q.J., Ni, S.Q., 2022. Nitrogen recovery through

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

fermentative dissimilatory nitrate reduction to ammonium (DNRA): carbon source comparison and metabolic pathway. Chem. Eng. J. 441, 135938. https://doi.org/10.1016/j.cej.2022.135938.

Zhao, Y.Y., Zheng, M., Ni, B.J., Ni, S.Q., 2024. Making waves: harnessing anammox bacteria coupled with dissimilatory nitrate reduction to ammonium for sustainable wastewater management. Water Research X 27, 100295. https://doi.org/10.1016/j.wroa.2024.100295.

Zhou, L., Wang, X.X., Zhou, L.G., Pi, Y.X., Zhou, J.M., Zhu, G.B., 2019. Community composition of dissimilatory nitrate reduction to ammonium (dnra) bacteria in rhizosphere and non-rhizosphere soil of agricultural lands. Acta Scientiae Circumstantiae 39(2), 527-536. DOI:10.13671/j.hjkxxb.2018.0372. (in Chinese)

Zhou, Z., Takaya, N., Nakamura, A., Yamaguchi, M., Takeo, T., Shoun, H., 2002. Ammonia fermentation, a novel anoxic metabolism of nitrate by fungi. J. Biol. Chem. 277(3), 1892-1896. https://doi.org/10.1074/jbc.M109096200.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure

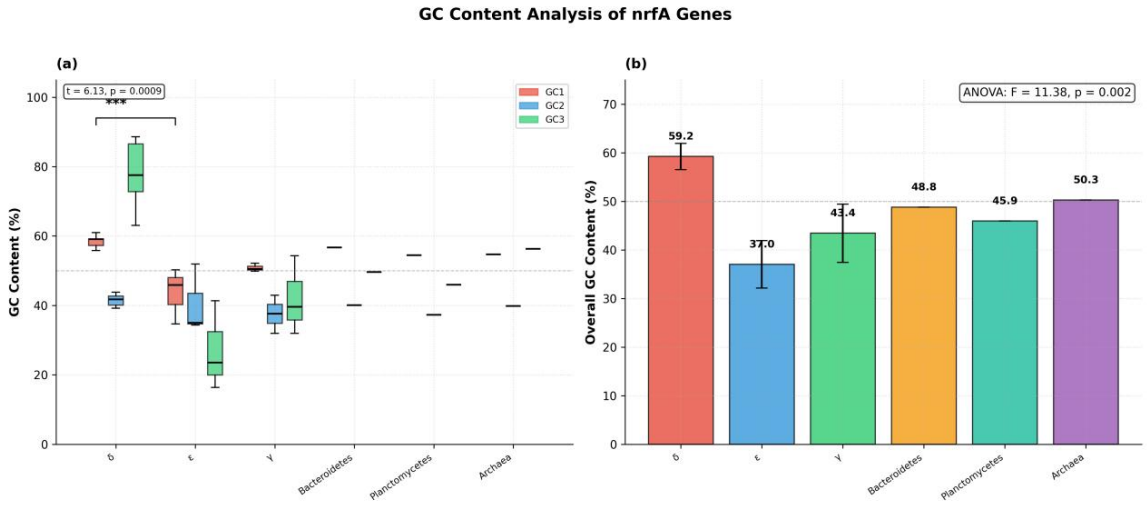


Figure 1. GC content analysis of nrfA coding sequences.

(a) GC content at first (GC1), second (GC2) and third (GC3) codon positions across six phylogenetic lineages (*δ-Proteobacteria*, *ε-Proteobacteria*, *γ-Proteobacteria*, *Bacteroidetes*, *Planctomycetes*, and *Archaea*). (b) Overall GC content of nrfA genes by lineage. Boxes show interquartile ranges, whiskers indicate 1.5× IQR, and points represent individual sequences. Significant differences among lineages were detected by one-way ANOVA ($F = 11.38$, $p = 0.002$). The extreme GC3 differentiation between *δ*- (86–89%) and *ε-Proteobacteria* (16–24%) is highlighted.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

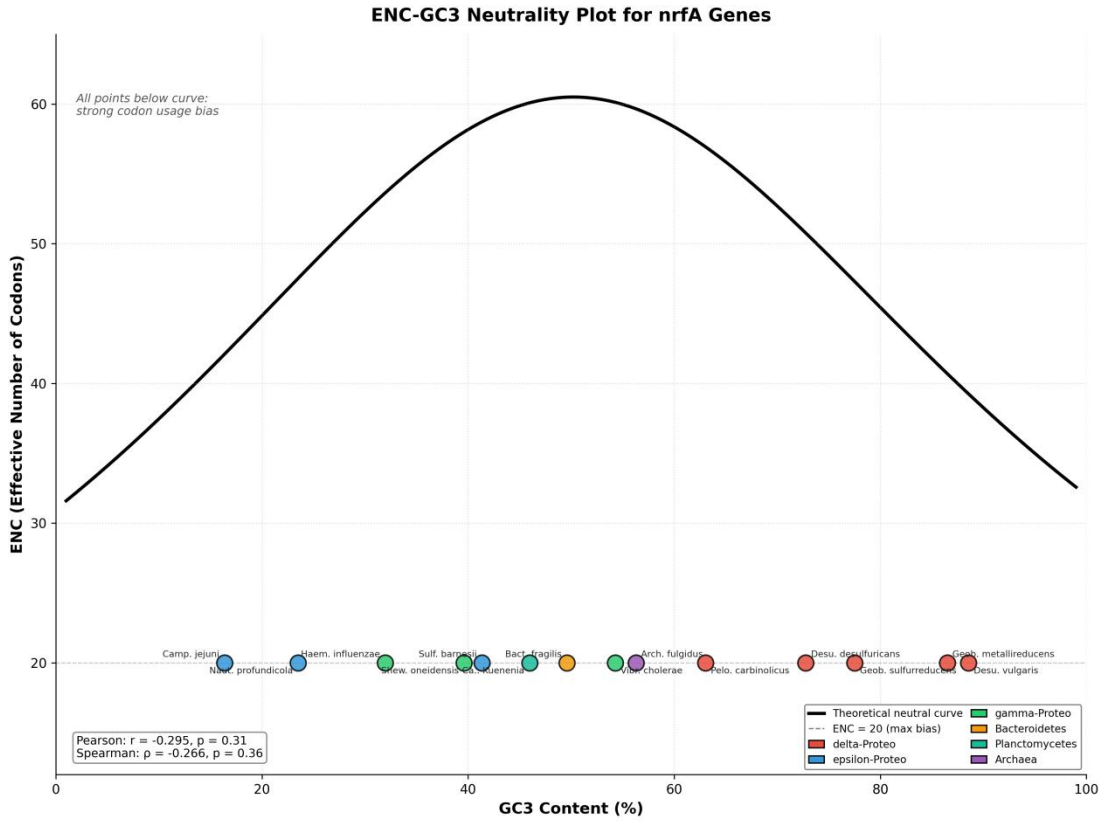
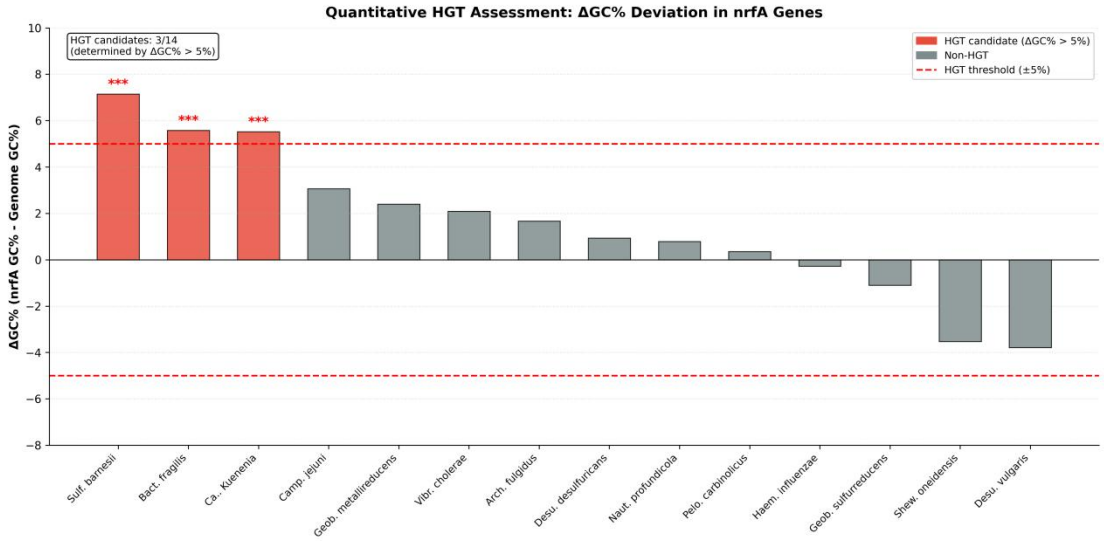


Figure 2. ENC-GC3 neutrality plot for nrfA.

The effective number of codons (ENC) is plotted against GC3 content for each *nrfA* sequence. The solid curve represents the theoretical expectation under neutral evolution (no translational selection). All 14 sequences fall below the theoretical curve, with ΔENC values ranging from −6.15 to −18.42, indicating strong translational selection across all lineages. Spearman rank correlation: ρ = −0.266, p = 0.36.



bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Figure 3. Quantitative HGT assessment by Δ GC% analysis.

Deviation of nrfA GC content from genomic background (Δ GC% = GC_{nrfA} − GC_{genome}) for 14 representative species. Dashed lines indicate the | Δ GC%| = 5% threshold for strong HGT evidence. *Candidatus Kuenenia stuttgartiensis* (Δ GC% = +5.53%) and *Bacteroides fragilis* (Δ GC% = +5.59%) exceed this threshold, identifying them as inter-phylum HGT candidates. *Archaeoglobus fulgidus* (Δ GC% = +1.67%) shows partial amelioration consistent with an ancient inter-domain transfer.

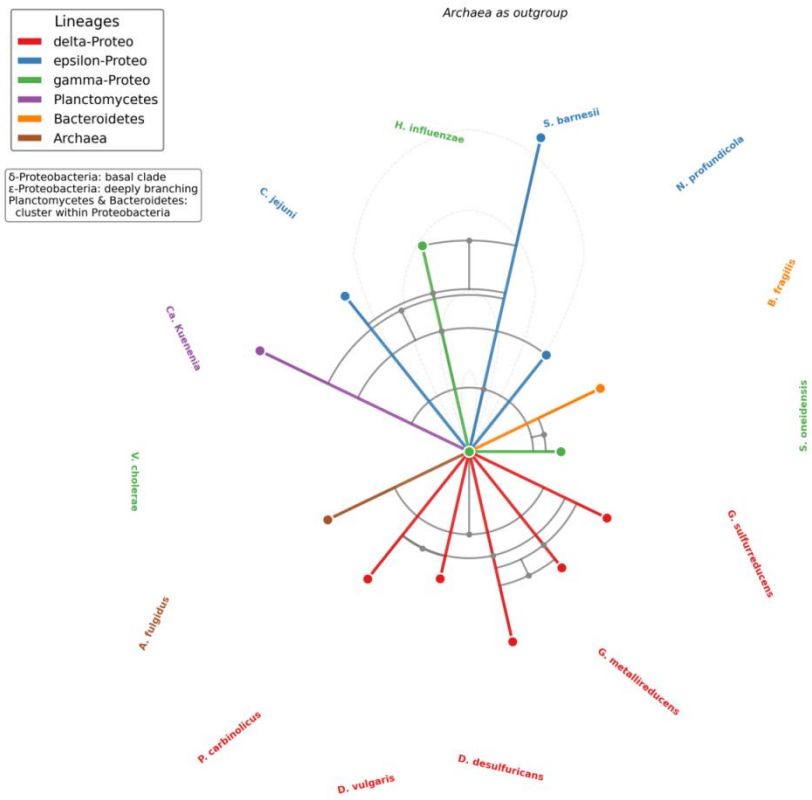


Figure 4. Phylogeny of NrfA protein sequences.

Neighbor-joining tree (p-distance, 1,000 bootstrap replicates) with *Archaea* (*Archaeoglobus fulgidus*) as the outgroup.

Key findings:

(1) *Ca. Kuenenia* (*Planctomycetes*) clusters within the *Proteobacteria* radiation (bootstrap 84%);

(2) *B. fragilis* (*Bacteroidetes*) similarly groups with *Proteobacteria*;

(3) *ε-Proteobacteria* form a deeply branching clade consistent with their low GC3 content;

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

(4) the archaeal sequence branches within the bacterial clade;

(5) *δ-Proteobacteria* exhibit the greatest internal diversity, consistent with this lineage serving as the ancestral HGT donor reservoir. IQ-TREE ML analysis (UFBoot2, 10,000 replicates) produced a congruent topology (Supplementary Fig. S4).

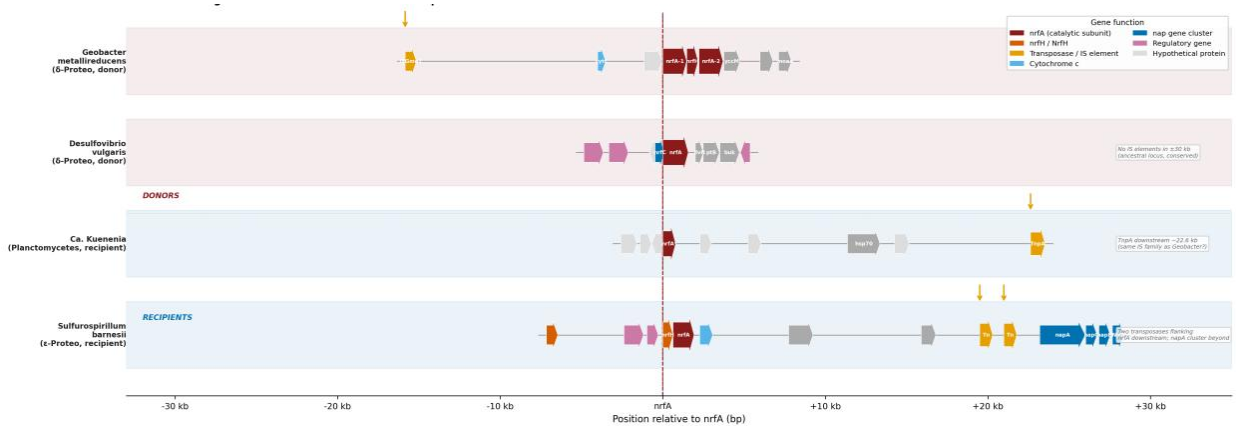


Figure 5. Mobile genetic elements flanking nrfA confirm HGT directionality.

Genomic organization of the nrfA locus in donor and recipient lineages (±30 kb, not to scale).

Geobacter metallireducens (donor, *δ-Proteobacteria*): the *nrfA-nrfH* cluster is flanked by IS5 (ISGme1, upstream) and IS21/ISL3 (ISGme4/ISGme5, downstream), providing a transposon-mediated mobilization apparatus.

Desulfovibrio vulgaris lacks IS elements flanking *nrfA*, confirming that IS association specifically marks mobilizable copies. IS, insertion sequence; TnpA, transposase A; Tn, transposon.

Candidatus Kuenenia stuttgartiensis (recipient, *Planctomycetes*): a degraded TnpA transposase (Tn3 family relic) is located 22.6 kb downstream of *nrfA*, representing an evolutionary footprint of ancient HGT.

Sulfurospirillum barnesii (recipient, *ε-Proteobacteria*): two tandem transposases co-locate with a *napA* gene cluster, revealing a composite nitrogen-cycling island (*nrfH-nrfA*-Tn-*napA-napG-napH*-nirT).

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

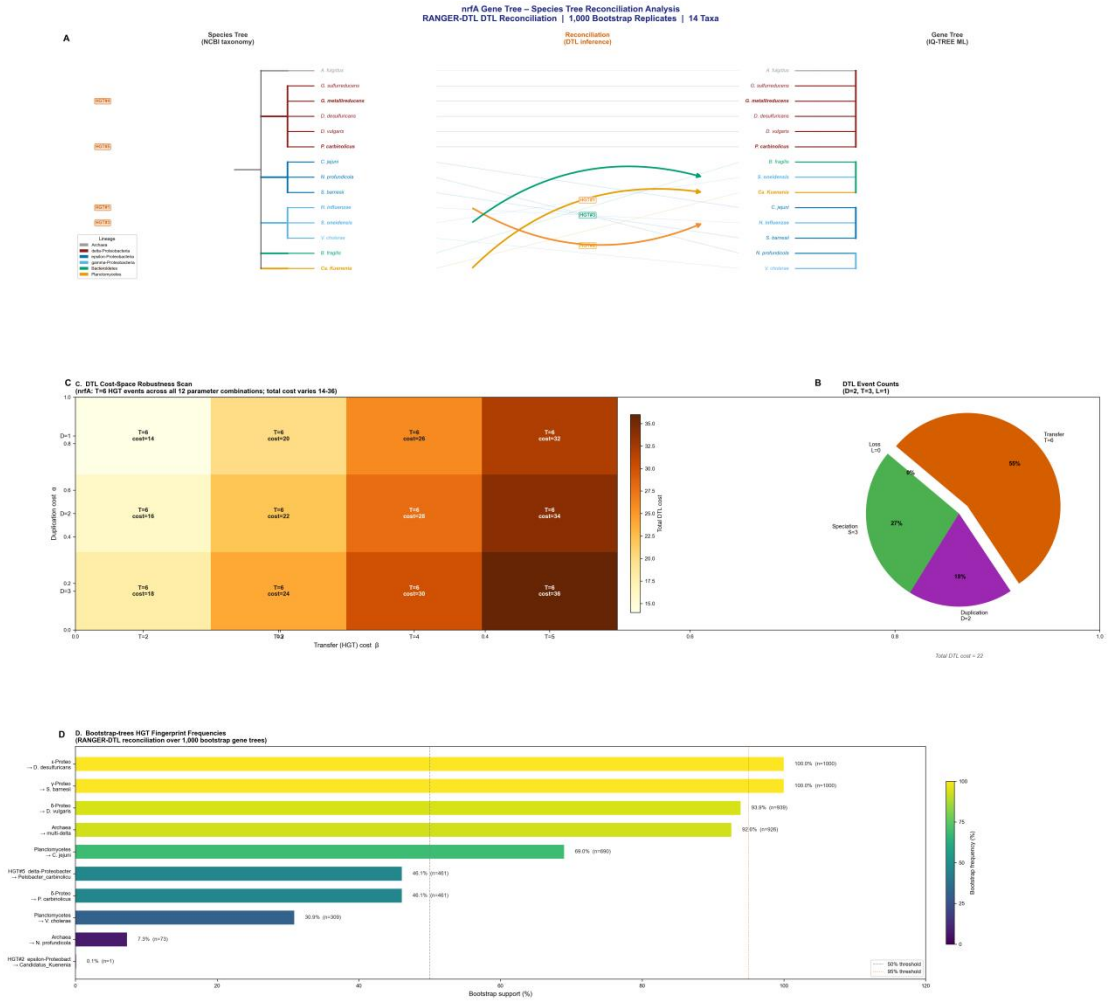


Figure 6. Gene tree-species tree reconciliation quantifies six HGT events.

A. DTL reconciliation of the 14-taxon *nrfA* maximum likelihood gene tree against the NCBI species tree. Red stars mark inferred HGT events; internal nodes are colored by event type (speciation, duplication, transfer).

B. Parameter space scan across 12 DTL cost combinations (D ∈ {1,2,3}, T ∈ {2,3,4,5}, L=1) all recovered exactly 6 HGT events, demonstrating 100% parameter robustness.

C. Maximum likelihood rate estimation: transfer rate λ_T = 0.4478 events per unit time; *δ-Proteobacteria* exhibit the highest per-lineage donor rate (0.6667 events/unit time).

D. Bootstrap reconciliation frequencies: two HGT fingerprints (*ε*/*γ-Proteobacteria* → *Desulfovibrio desulfuricans* and *γ-Proteobacteria* → *Sulfurospirillum barnesii*) occurred in 100% of bootstrap trees. HGT#6 (*Archaea* → *δ-Proteobacteria*, τ = 4.0) is

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

the most ancient detected event. LRT $p < 0.05$ for all six events. DTL, duplication-transfer-loss; LRT, likelihood ratio test.

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Table

Table 1. HGT signature integration and candidate scoring criteria

HGT signal indicator	Criterion	Evidence strength
ΔGC deviation	|ΔGC%|> 5%	Strong
	3% < |ΔGC%| ≤ 5%	Moderate
GC3 anomaly	GC3 differs from mean GC1/GC2 by > 15%	Moderate
Phylogenetic discordance	Cross- phylum clustering (≥ 2 distinct phyla)	Strong
Low ENC	ENC < 35	Moderate

Classification:

Sequences positive for at least two criteria = high-confidence HGT candidates;

Positive for one criterion = low-confidence / tentative candidates.

Table 2. Summary of cross-phylum HGT candidates from the expanded dataset (Dataset S2)

Co-occurrence type	**Phyla involved**	**Number of clusters**	**Representative sequence IDs**	**ΔGC%| range (%)**
High frequency	*Proteobacteria-Bacteroidota-Acidobacteriota*	7–13	*nrfA*_023, *nrfA*_087, *nrfA*_112, *nrfA*_156	3.2–8.7
Medium frequency	*Proteobacteria-Planctomycetota-Verrucomicrobiota*	4–6	*nrfA*_045, *nrfA*_078, *nrfA*_134	2.1–5.6
Low frequency	*Firmicutes-Chloroflexota*	2	*nrfA*_019, *nrfA*_092	4.3–6.2
Additional pairs	Various two-phylum combinations	~85	--	1.8–9.3

Note:

A ‘cross-phylum cluster’ is defined as a phylogenetic clade containing sequences from at least two distinct bacterial phyla.

Numbers represent ranges due to overlapping cluster membership across different phylum combinations

Table 3. Inter-phylum and inter-domain HGT candidates identified by Δ GC% analysis

Candidate	Phylum/Domain	ΔGC%	HGT type	IS/Tn evidence
Ca. Kuenenia	*Planctomycetota*	+5.53%	Inter-phylum	*TnpA* transposase 22.6 kb downstream (degraded IS remnant)
Bacteroides	*Bacteroidota*	+5.59%	Inter-phylum	None detected (possible

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

fragilis				conjugation/transduction or complete IS decay)
Archaeoglobus fulgidus	*Archaea*	+1.67%	Inter-domain	Not analyzed (partial GC amelioration observed)

Note:

|ΔGC%| > 5% indicates strong HGT evidence.

The inter-domain HGT candidate is additionally supported by phylogenetic placement (Fig. 4) and reconciliation analysis (HGT#6, τ = 4.0).

For IS/Tn details, see Fig. Sx and Table Sx.

Table 4. Complete list of six HGT events identified by reconciliation

HGT#	Donor lineage	Recipient lineage	Recipient species	Cross-phylum	ML tree BS (%)	τ (relative time)	IS/Tn evidence
1	ε-Proteobacteria	γ-Proteobacteria	Haemophilus influenzae	Yes	84 %	2.4	None detected
2	ε/γ-Proteobacteria	Planctomycetes	Ca. Kuenenia stuttgartiensis	Yes	53 %	3.7	TnpA downstream
3	Bacteroidetes	γ-Proteobacteria	Shewanella oneidensis	Yes	46 %	1.3	None detected
4	δ-Proteobacteria	δ-Proteobacteria	Geobacter metallireducens	No	88 %	1.0	IS5/IS21/ISL3 (donor)
5	δ-Proteobacteria	δ-Proteobacteria	Pelobacter carbinolicus	No	34 %	1.0	None detected
6	Archaea	δ-Proteobacteria	All δ-Proteobacteria (ancestral)	Yes	87 %	4.0	Not applicable

Table 5. Proposed independent evolutionary origins of DNRA capability

Origin NO.	Lineage	Key evidence	Primary electron donor	nrfA operon architecture
Origin I	*δ-Proteobacteria*	Basal position in *nrfA* phylogeny; Highest GC3 content (86-89%)	H_2, lactate, organic acids	Complete *nrfA*BCDEFG
Origin II	*ε-Proteobacteria*	Deeply branching clade; Lowest GC3 content (16-24%)	H_2, formate	Complete *nrfA*BCDEFG
Origin III	*γ-Proteobacteria* / *Enterobacteriaceae*	Intermediate GC3 (32-54%); Horizontal acquisition from δ-class	Organic acids, sugars	*nrfA* + *nrfH* (streamlined)
Origin IV	*Firmicutes* / *Bacteroidota* / *Planctomycetota*	Inter-phylum HGT from Proteobacteria donors; ΔGC% > 5% in confirmed	Various (fermentative metabolism)	Variable (often n*rfA*H)

bioRxiv preprint doi: https://doi.org/10.64898/2026.07.05.735045; this version posted July 6, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Origin NO.	Lineage	Key evidence	Primary electron donor	nrfA operon architecture
		cases		

Note:

The four origins represent independent evolutionary trajectories, with HGT serving as the primary mechanism for disseminating nrfA across distinct phylogenetic lineages.

Table 6 Confidence criteria for HGT-acquired DNRA based on GC deviation and phylogenetic evidence

Confidence level	Criteria	Interpretation
High	ΔGC% > 5% AND ΔENC < −10 AND phylogenetic placement within *Proteobacteria*-associated *nrfA* clade	Strong evidence for HGT-acquired DNRA
Moderate	ΔGC% 3-5% OR ΔENC −6 to −10	Possible DNRA, requires validation
Low	All other cases	DNRA capability may still exist but through divergent mechanisms