nature communications

Article

https://doi.org/10.1038/s41467-026-70744-1

The hidden role of rhizospheric viruses in promoting nitrogen fixation in soils

Received: 6 April 2024

Accepted: 3 March 2026

Published online: 18 March 2026

Check for updates

Dong Zhu[1,2,3,9] ✉, Wen Zhang[1,9], Jose Luis Balcazar[4], Danrui Wang[1], Mingming Sun[1] ✉, Feng Hu[1], Josep Penuelas[5,6,7] & Yong-Guan Zhu[8]

Biological nitrogen fixation is a cornerstone of terrestrial nitrogen cycling, traditionally attributed to bacterial nitrogenase activity. However, the potential contribution of rhizospheric viruses remains largely unexplored. Here, we reveal the global distribution of nitrogen-fixing genes, with widespread detection of *nifA*, *nifL*, *nifU*, and *nifH* in both bacteria and viruses, and identify *nifU* as a viral auxiliary metabolic gene (AMG). Analysis of viral communities in rhizosphere and bulk soils cultivated with cowpea showed that viral *nifU* expression was significantly upregulated in rhizosphere soils. Using $^{15}N_2$ stable-isotope tracing and virus transplantation experiments, we demonstrate that virus-encoded nitrogen-fixing AMGs, horizontally transferred from bacteria such as *Azospirillum thermophilum* (70–99% homology), increased nitrogenase activity from 1.79 to 3.14 nmol C_2H_4 g^{-1} dry soil h^{-1}. This enhancement was accompanied by shifts in bacterial community composition, with the relative abundance of the nitrogen-fixing genus *Azotobacter* reaching 90.8%. These results uncover a previously hidden role of rhizospheric viruses in promoting bacterial nitrogen fixation, suggesting that viral-mediated gene transfer could be leveraged to enhance nitrogen cycling in soils and inform sustainable soil management strategies.

Nitrogen (N) is an essential nutrient that underpins plant growth, crop productivity, and global food security[1,2]. Despite its abundance in the atmosphere as dinitrogen (N_2), plants cannot directly access it, relying instead on biologically available forms such as ammonium and nitrate[3]. The natural provision of these forms depends largely on biological nitrogen fixation (BNF), a microbial-mediated process that converts atmospheric N_2 into plant-assimilable nitrogen. In terrestrial ecosystems, a substantial portion of fixed nitrogen is supplied through symbiosis between leguminous plants and rhizobia, while free-living diazotrophs contribute additional nitrogen independently[4]. The efficiency of nitrogen fixation in soils is fundamentally determined by the genetic potential and activity of nitrogen-fixing microbial communities, particularly the nitrogenase complex and its associated genes (e.g., *nifH*, *nifD*, *nifK*, and *nifU*)[4,5]. However, factors regulating microbial nitrogen fixation in natural soils, beyond classical nutrient and environmental constraints, remain poorly understood. This highlights a key knowledge gap in understanding the full spectrum of biotic regulators of soil nitrogen cycling.

Viruses are the most abundant biological entities on Earth and are increasingly recognized as key regulators of microbial evolution and

[1]Soil Ecology Lab, Jiangsu Provincial Key Laboratory of Coastal Saline Soil Resources Utilization and Ecological Conservation, Jiangsu Collaborative Innovation Center for Solid Organic Waste Resource Utilization and Jiangsu Key Laboratory for Solid Organic Waste Utilization, Nanjing Agricultural University, Nanjing, China. [2]State Key Laboratory of Regional and Urban Ecology, Ningbo Observation and Research Station, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, China. [3]Zhejiang Key Laboratory of Pollution Control for Port-Petrochemical Industry, CAS Haixi Industrial Technology Innovation Center in Beilun, Ningbo, China. [4]Catalan Institute for Water Research (ICRA-CERCA), Girona, Spain. [5]CSIC, Global Ecology Unit CREAF-CSIC-UAB, Barcelona, Catalonia, Spain. [6]CREAF, Bellaterra, Barcelona, Catalonia, Spain. [7]Global Change Research Institute of the Czech Academy of Sciences, Brno, Czech Republic. [8]State Key Laboratory of Regional and Urban Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing, China. [9]These authors contributed equally: Dong Zhu, Wen Zhang. ✉e-mail: dzhu@iue.ac.cn; sunmingming@njau.edu.cn

soil biogeochemical processes[6]. By mediating horizontal gene transfer (HGT) and expressing auxiliary metabolic genes (AMGs), viruses can reprogram host metabolism and modulate major nutrient-cycling pathways, including carbon, phosphorus, and nitrogen cycling[7,8]. Soil viromic studies have revealed a diverse repertoire of viral AMGs involved in carbon fixation (e.g., *rbcL*, *ppdK*, *TKT*), phosphorus metabolism and transport (*phoA*, *phoH*, *phoB*), and the degradation of organochloride pesticides (e.g., L-DEX), highlighting the capacity of viruses to directly influence host metabolic functions[9–11]. Within the nitrogen cycle, viral genes associated with nitrogen assimilation and denitrification, such as *gdh*, *norQ*, and *norD*, have been detected in soil viromes[10,11]. More recently, viruses have been shown to horizontally transfer denitrification-related AMGs (*norB*, *norR*, *nirK*) to bacterial hosts, with potential consequences for denitrification rates and N_2O emissions[12]. Despite these advances, viral AMGs linked to nitrification and denitrification represent only a subset of nitrogen-cycling processes, and no AMGs directly involved in biological nitrogen fixation have yet been reported in soil viromes. Consequently, whether rhizosphere viruses can facilitate bacterial nitrogen fixation through the encoding or transfer of nitrogen-fixation genes (NFGs) remains unknown, limiting our understanding of viral contributions to soil nitrogen availability and plant nutrition.

Here, we hypothesize that viral communities in soils, particularly in the rhizosphere of nitrogen-fixing plants, encode AMGs involved in nitrogen fixation, and that these genes are actively expressed in situ. We further postulate that viruses could enhance nitrogen fixation activity through horizontal transfer of nitrogen-fixing AMGs. To test these hypotheses, we combine global database mining, field virome and metatranscriptome analyses, and experimental virus-transplant microcosms coupled with $^{15}N_2$ stable-isotope probing (Fig. S1). This multi-scale approach allows us to examine: (i) the global distribution of NFGs in viruses and bacteria, (ii) in situ expression of viral AMGs in rhizosphere soils, and (iii) the functional impact of viruses on microbial nitrogen fixation. Our integrated analyses reveal that viral *nifU* is globally widespread, actively expressed in the rhizosphere, and functionally contributes to enhanced nitrogen fixation, highlighting an overlooked role of viruses in shaping soil nitrogen cycling.

Results

Global distribution patterns of NFG-carrying bacteria and viruses

To establish the global prevalence of nitrogen-fixing genes (NFGs) in viral communities, we first screened 8,606,551 vOTUs from the IMG/VR v4 database. Only 2996 vOTUs encoded at least one NFG, indicating that NFG-carrying viruses constitute a rare component of the global virome (~0.03%). Despite this rarity, their distribution was strongly non-random at the global scale. NFG-encoding viruses exhibited pronounced geographic clustering (χ^2 = 13,156, $p < 0.001$; Supplementary Data 1), with the vast majority detected in North America, Europe, and Asia, together accounting for more than 93% of all observations (Fig. 1a). A similar large-scale biogeographic pattern was observed for nitrogen-fixing bacteria (χ^2 = 18,896, $p < 0.001$), whose highest abundances also occurred in these regions. This geographic concordance suggests that viral and bacterial nitrogen-fixation potentials are structured by shared environmental and ecological drivers at continental scales.

Across habitats, both NFG-carrying viruses and bacteria were preferentially associated with terrestrial environments (Fig. 1b). Viral NFGs were detected across a broad range of ecosystems, with freshwater and host-associated environments contributing the largest fractions, but soil-related habitats consistently emerged as a major reservoir. In contrast, nitrogen-fixing bacteria showed a much stronger habitat specialization, being overwhelmingly enriched in agricultural soils and plant-associated niches, particularly root nodules and bulk soil. These contrasting habitat distributions indicate that, while viral NFGs occur across multiple environments, soils represent a key ecological intersection where viral and bacterial nitrogen-fixation potentials co-occur. At the gene level, viral NFGs were dominated by a small subset of the nitrogen-fixation machinery. Four genes, including *nifA*, *nifL*, *nifU*, and *nifH*, accounted for most viral detections and were consistently enriched across habitats, particularly in soils.

Phylogenetic analyses revealed that these genes were not randomly distributed among viral lineages: members of the families *Kyanoviridae*, *Nudiviridae*, and *Bronfenbrennervirinae* emerged as the principal carriers of NFGs, with *nifU* being strongly associated with *Kyanoviridae* (Fig. S2). In contrast, bacterial NFGs displayed much greater taxonomic breadth, with no single gene dominating the repertoire and with major contributions from well-known diazotrophic genera, including *Pseudomonas*, *Klebsiella*, and *Bradyrhizobium* (Fig. S3). Together, these results demonstrate that viruses carrying NFGs form a rare but non-random component of the global virome. Their distribution is geographically concentrated, functionally biased toward a small subset of NFGs, and taxonomically restricted compared with bacteria. The recurrent co-occurrence of viral and bacterial NFGs in soils highlights terrestrial ecosystems as key hotspots where viral AMGs may intersect with microbial nitrogen fixation.

Virus-host relationships and viral auxiliary metabolic genes

To investigate interactions between nitrogen-fixing bacteria and viruses encoding NFGs, we predicted putative hosts for vOTUs carrying NFGs. Using the iPHop approach, we linked 480 vOTUs to 1,145 bacterial hosts (Supplementary Data 2), representing 16.0% of all NFG-carrying vOTUs. This substantially improved host coverage compared with existing annotations in the IMG/VR v4 database (Supplementary Data 3). The top 20 most abundant virus–host pairs are shown in Fig. 2. Among predicted hosts, *Klebsiella* (n = 1,201) and *Pseudomonas* (n = 403) were predominant, indicating that Proteobacteria were the dominant bacterial hosts of NFG-carrying viruses. In several cases, distinct viruses encoding the same NFG were linked to the same host species, suggesting that these hosts may represent hotspots for viral gene exchange (see Discussion). For example, two viruses carrying *nifA* (IMGVR_UViG_3300024336_000758 and IMGVR_UViG_3300028274_000577) were both linked to the putative host *Stutzerimonas*.

Among the 480 vOTUs with predicted hosts, viruses from the families *Mesoniviridae* and *Kyanoviridae* were most abundant. While host assignments are based on computational inference, we observed cases in which individual viruses were linked to phylogenetically distinct hosts, for example, the *nifU*-carrying virus IMGVR_UViG_3300008111_001506 was associated with both *Nitrogeniibacter aestuarii* and *Pseudomonas mucidolens*. Although such patterns could reflect limitations of current host prediction approaches, they raise the possibility that some viruses might facilitate horizontal transfer of NFGs across divergent bacteria. This potential mechanism should be treated as a testable hypothesis and warrants further experimental investigation. Environmental distribution of host-linked viruses was dominated by freshwater (1,914 host linkages), followed by marine (310) and soil (217) environments. Notably, prophages, temperate viruses integrated into host genomes, accounted for 49.7% of the linked viruses (Fig. 2), indicating widespread genomic integration. The most common bacterial genera hosting prophages were *Pseudomonas*, *Aeromonas*, *Methylobacterium*, *Bradyrhizobium*, and *Mesorhizobium*.

To determine whether NFGs were true viral AMGs, we used VIBRANT and DRAM-v to screen a total of 2,996 vOTUs for viral-encoded NFGs (Fig. 3a). We identified one confirmed viral AMG, *nifU*, in 37 vOTUs, representing 1.23 % of NFG-carrying vOTUs and 0.00043 % of the total global virome. Viral identity was supported by the presence of viral-like promoter and terminator signals flanking *nifU*, and domain analysis indicated the potential to encode a functional protein (LDF = 7.55, GMQE = 0.86). While expression data were unavailable, these genomic signatures strongly support *nifU* as a bona fide viral AMG. Adjacent to the *nifU* gene, we detected iron-sulfur cluster

https://doi.org/10.1038/s41467-026-70744-1

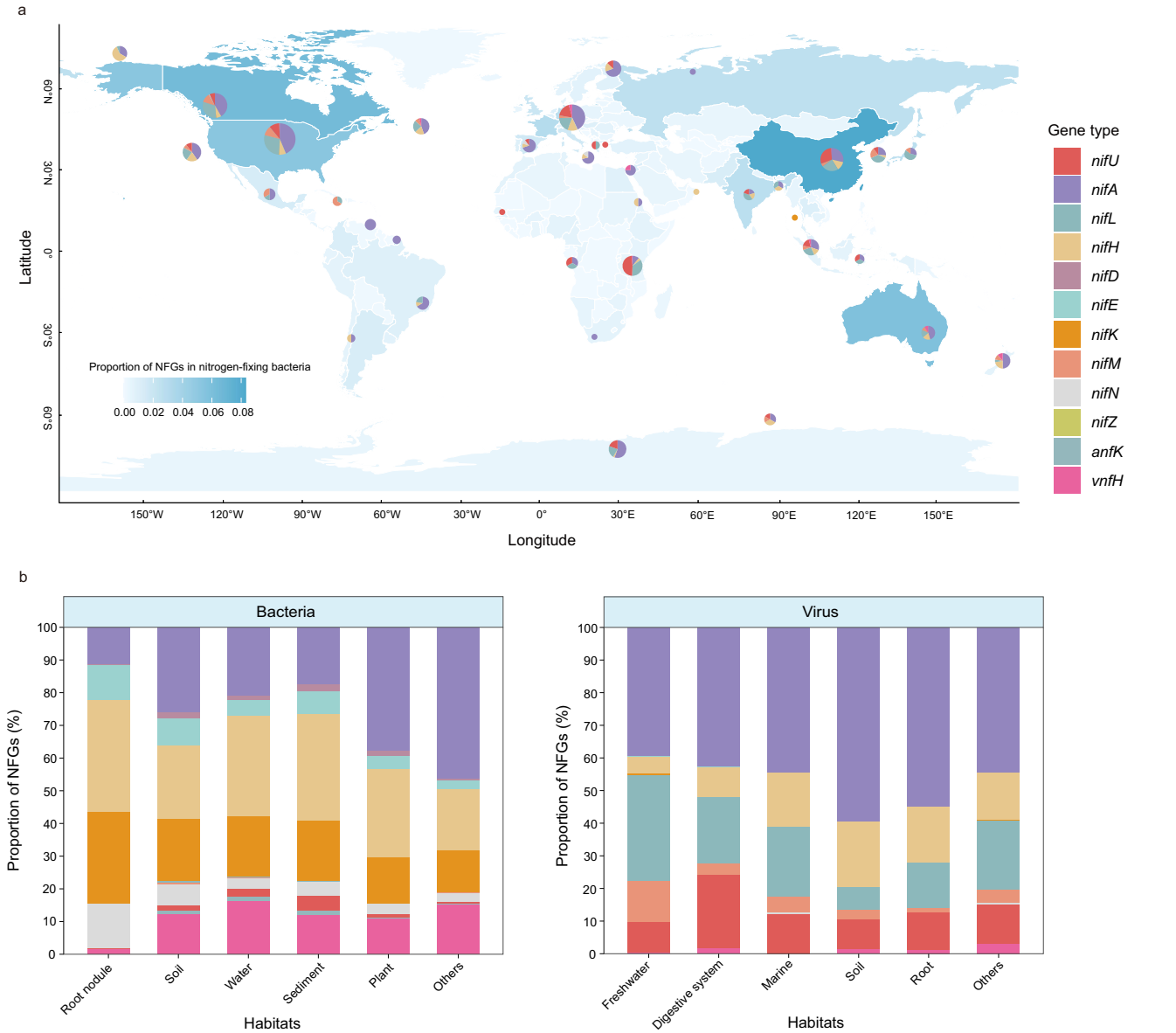

Fig. 1 | Global distribution and habitat characteristics of bacteria and viruses encoding nitrogen-fixing genes. a Global distribution of nitrogen-fixing genes (NFGs) in bacteria and viruses. The world map shows the relative abundance of bacteria carrying distinct NFGs across different geographic regions. Overlaid pie charts indicate the relative proportions of viruses encoding NFGs in the corresponding regions. Map outlines generated using Natural Earth data. **b** Habitat enrichment of NFG-carrying bacteria and viruses. Bar charts showing the proportions of NFGs detected in the top five habitats with the highest overall abundance of bacteria and viruses. Source data are provided as a Source Data file.

binding domains, consistent with motifs necessary for assembling iron-sulfur clusters, critical components of nitrogen-fixation enzymes[13]. Phylogenetic analysis revealed that viral NifU proteins clustered with bacterial homologues, with the highest similarity (89% protein identity) to *Zoogloea* sp. LCSB751 (Betaproteobacteria) (Fig. 3b). This high homology suggests that the viral *nifU* gene was likely acquired from bacterial hosts via HGT and retained functional capacity during virus-host co-evolution. Together, these results highlight the complexity of virus–bacteria interactions in nitrogen cycling and underscore the potential role of viral AMGs such as *nifU* in modulating host nitrogen metabolism.

Viral AMGs associated with nitrogen fixation in the cowpea rhizosphere

To explore the role of virus-encoded nitrogen-fixing AMGs in supporting bacterial nitrogen fixation, we analyzed rhizosphere and bulk soils from a long-term cowpea cultivation site in Hengxi, Nanjing. Viral metagenomic analysis using vConTACT 2.0 clustered 2917 vOTUs into 44.9% clustered and 39.1% singleton viral clusters (VCs) (Fig. S4; Supplementary Data 4). Comparison with NCBI Viral RefSeq and previously published agricultural soil viromes revealed 1016 unique VCs in our soil samples, with minimal overlap (12 and 30 VCs, respectively)[14,15], indicating a large proportion of previously uncharacterized viruses in these farmland soils (Fig. S5). The virus taxonomic assignment further revealed that the cowpea rhizosphere (S1, S3) contained a higher diversity of viral families than bulk soils (S2 and S4) (Fig. S6).

Analysis of 9308 soil vOTUs using VIBRANT and DRAM-v identified 1365 vOTUs carrying AMGs. Forty-four distinct AMG types were annotated, spanning glycan biosynthesis, cofactor and vitamin metabolism, carbohydrate metabolism, and energy metabolism (Supplementary Data 5). Notably, the nitrogen-fixation-associated gene *nifU* was detected as a virus-encoded AMG, with higher abundance in

https://doi.org/10.1038/s41467-026-70744-1

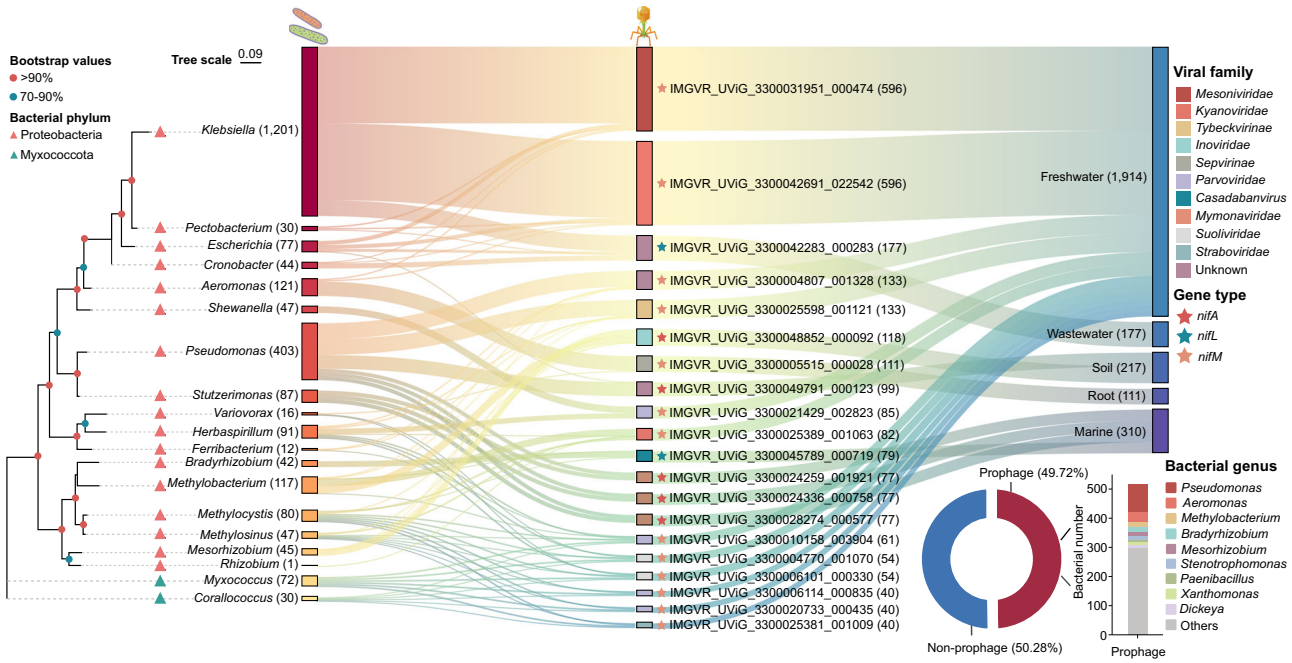

Fig. 2 | Predicted virus-host associations of nitrogen-fixing bacteria. Left: Phylogenetic tree of nitrogen-fixing bacteria inferred from 16S rRNA gene sequences. Coloured triangles indicate different bacterial phyla (red, Proteobacteria; blue, Myxococcota). Coloured circles at nodes represent bootstrap value intervals (red, >90%; blue, 70%–90%). Middle: Number of viruses carrying NFGs identified across distinct habitats. Stars indicate the presence of NFGs. Right: Habitat distribution of these viruses. The pie chart shows the relative proportions of prophages versus non-prophages among all NFG-containing viruses. The bar graph indicates the number of predicted bacterial hosts associated with the prophages. Source data are provided as a Source Data file.

cowpea rhizosphere under organic fertilization (S1) compared to the other soils ($p < 0.05$, Fig. 4a). Co-occurrence analysis revealed significant associations of *nifU* with other viral AMGs, including *DPM1* (glycan biosynthesis), *cobS* and *acpT* (cofactor metabolism), *ppdK* (carbohydrate metabolism), and *coxA* (energy metabolism) (Fig. S7). Protein structure modelling via SWISS-MODEL predicted a structurally intact NifU protein with 99.8% confidence (Fig. 4b). PCR amplification confirmed *nifU* presence in both viral and bacterial DNA from rhizosphere samples (S1), but not in bulk soil (S2) (Fig. 4c).

Metatranscriptomic profiling showed that *nifU* expression was enriched 13.6–18.9-fold in rhizosphere soil with organic fertilization (S1) relative to bulk soils (S2 and S4) (Fig. 4d) ($p < 0.05$; Supplementary Data 6 and 7). The upregulation of viral *nifU* in rhizosphere versus bulk soil was more pronounced under organic fertilization ($\log_2FC = 4.24$, $p = 0.03$) than without ($\log_2FC = 3.77$, $p = 0.0025$). Transcript-specific mapping indicated that *nifU* expression was predominantly bacterial (95.7%), with a minor but detectable viral contribution (4.3%) (Fig. S8). These results suggest that viruses carrying *nifU* may act as compensatory contributors, facilitating host bacterial nitrogen fixation in the rhizosphere.

$^{15}N_2$ stable-isotope probing combined with viral transplantation suggests a potential role of viral AMGs in soil nitrogen fixation

To examine whether viral AMGs are associated with enhanced soil nitrogen fixation activity, we established soil microcosms with or without cowpea rhizosphere virus transplantation under $^{15}N_2$ or $^{14}N_2$ atmospheres (see the Methods section, Soil microcosms). This experimental design enabled us to assess whether viral transplantation coincides with changes in nitrogen fixation–related processes, rather than to resolve a direct mechanistic role. After 35 days of incubation, soils receiving a 10-fold virus suspension (NV10) exhibited significantly higher total soil nitrogen concentrations (36.43 ± 4.40 mg kg^{-1}) compared with the corresponding non-amended controls (16.86 ± 3.11 mg kg^{-1}). Nitrogenase activity, measured by the acetylene reduction assay, was also significantly elevated in the NV10 treatment (3.14 ± 0.16 nmol C_2H_4 g^{-1} dry soil h^{-1}) relative to controls (1.79 ± 0.07 nmol C_2H_4 g^{-1} dry soil h^{-1}) ($p < 0.05$; Fig. 5a). These increases indicate that viral transplantation was associated with enhanced nitrogen fixation activity in the soil microcosms.

To assess active nitrogen fixation at the microbial population level, we conducted DNA-SIP targeting the conserved nitrogenase gene *nifH*. In $^{15}N_2$-labelled treatments, the peak relative abundance of *nifH* shifted toward higher buoyant density fractions compared with $^{14}N_2$ controls, consistent with assimilation of ^{15}N into nitrogen-fixing biomass (Fig. 5b). The maximum *nifH* signal occurred at buoyant densities of 1.712 and 1.720 g mL^{-1} in the NV and NV10 treatments, respectively, compared with 1.709 and 1.705 g mL^{-1} in the corresponding $^{14}N_2$ controls. Sequencing of 16S rRNA genes from the heavy DNA fractions with the highest *nifH* abundance in the $^{15}N_2$-labelled soils revealed a strong dominance of *Azotobacter* in the NV10 treatment (90.8% relative abundance; Fig. 5c), indicating that actively nitrogen-fixing bacterial taxa were enriched in virus-amended microcosms. Metagenomic sequencing of light and heavy DNA fractions from the $^{15}N_2$-labelled soils yielded 841 contigs. Annotation using DRAM-v identified a single viral contig carrying the nitrogen-fixation–related AMG *nifU* in the heavy DNA fraction of the NV10 treatment. Promoter prediction and protein structure modelling suggested that this viral *nifU* retains coding potential (Fig. 5d; LDF = 3.01, GMQE = 0.92), although functional activity cannot be directly inferred from these analyses alone.

Phylogenetic comparison of *nifU* sequences from public viral databases, cowpea rhizosphere viromes, the ^{15}N-labelled virus transplant soil, and bacterial genomes recovered in this study revealed high sequence similarity among viral *nifU* variants (up to 99%) and moderate similarity (70–73%) to bacterial homologues (Fig. S9). This pattern is consistent with historical HGT between nitrogen-fixing bacteria and viruses, a process through which viruses may acquire AMGs, but

https://doi.org/10.1038/s41467-026-70744-1

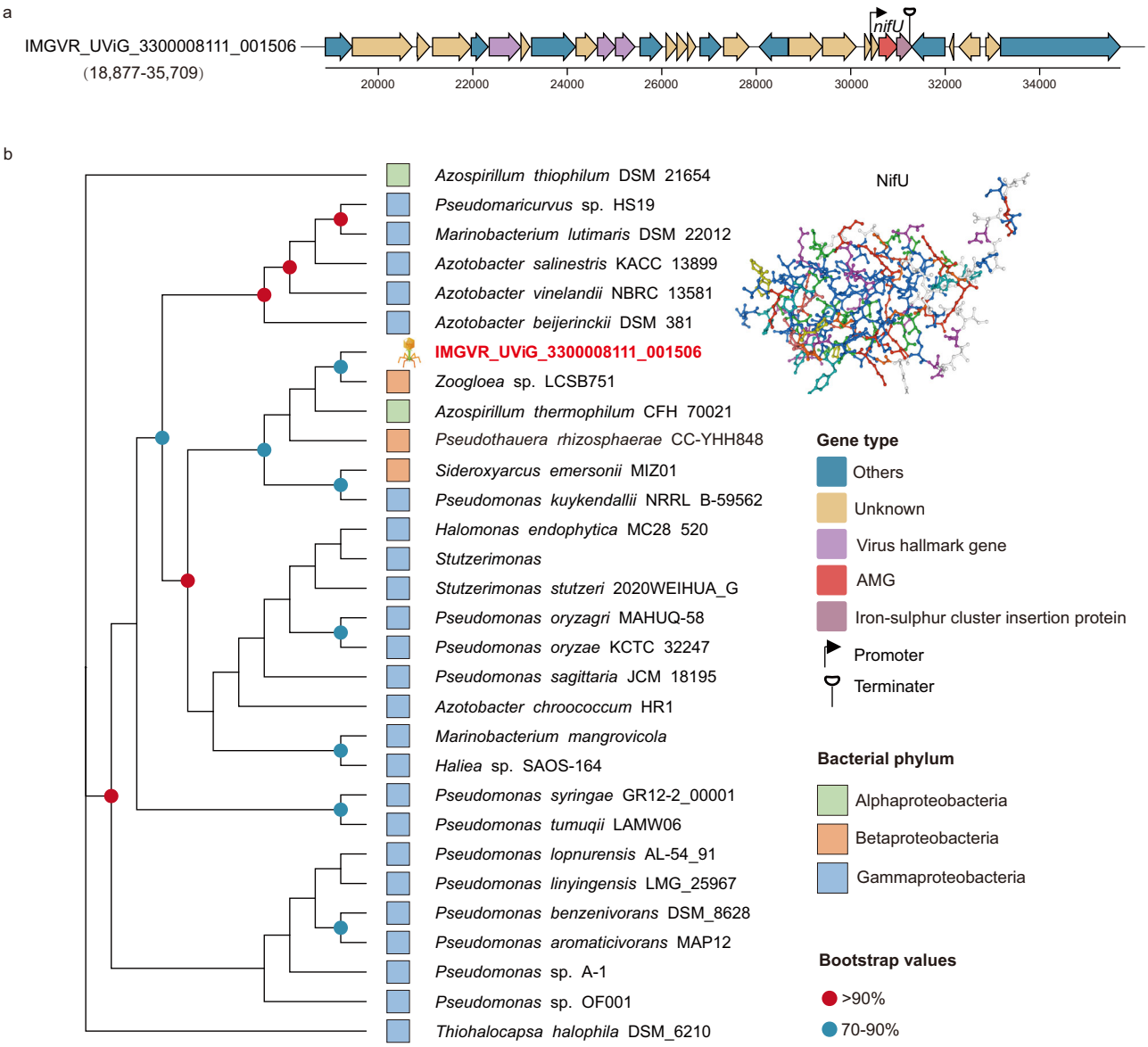

Fig. 3 | Global distribution and characteristics of nitrogen-fixing viral auxiliary metabolic genes (AMGs). a Composition of nitrogen-fixing AMG clusters encoded by viruses detected across global metagenomes, highlighting the diversity of viral NFGs. **b** Phylogenetic relationships of viral sequences carrying nitrogen-fixing AMGs, alongside the predicted structure of the NifU protein encoded by these viruses. Coloured circles at nodes represent bootstrap value intervals (red, >90%; blue, 70–90%). Source data are provided as a Source Data file.

does not establish the direction or timing of transfer. Collectively, the combined SIP, metagenomic, and viral transplantation results provide correlative evidence linking viral addition to enhanced nitrogen fixation activity and the presence of a viral *nifU* AMG within isotopically labelled, actively metabolizing microbial communities. While these findings are consistent with a potential contribution of viral AMGs to soil nitrogen cycling, this approach has inherent limitations. Specifically, it cannot demonstrate active viral infection or confirm that the identified AMGs were functionally expressed in situ during the experiment. Therefore, our findings, while suggestive, highlight the need for future studies that directly track viral infection dynamics and AMG transcription to establish mechanistic causality.

Discussion

This study combined global public database mining, soil multi-omic analyses, $^{15}N_2$ DNA-SIP, and a virus transplantation experiment to investigate whether viruses may be involved in soil nitrogen fixation. Across public virome datasets, viruses encoding NFGs represented only a very small fraction of total viral populations. However, after normalization, these viruses were consistently detected across geographically diverse regions, particularly in North America, Europe, and Asia (Fig. 1), suggesting that although rare, NFG-carrying viruses are globally distributed and persist across ecosystems. Such rarity, coupled with broad distribution, is characteristic of genes with potential long-term evolutionary relevance rather than numerical dominance at the ecosystem scale[16–18]. Viruses carrying NFGs were detected across freshwater, marine, soil, and rhizosphere habitats, indicating that virus-associated nitrogen metabolism is not restricted to a single ecosystem type. Similar co-occurrence patterns of bacterial and viral NFGs within shared habitats suggest ecological coupling between host bacteria and viruses, consistent with previous evidence for virus-mediated HGT. Public database analyses identified viral genomes encoding regulatory genes (*nifA* and *nifL*), a biosynthetic gene (*nifU*), and the structural gene (*nifH*)[19,20]. The presence of both *nifA* and *nifL* in

https://doi.org/10.1038/s41467-026-70744-1

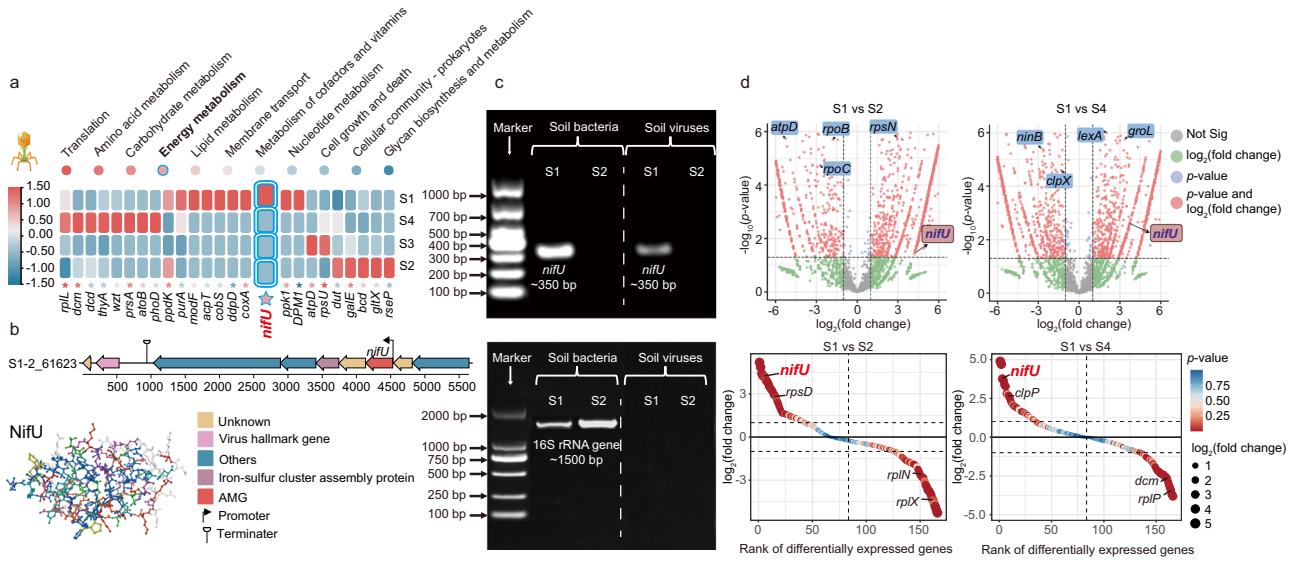

Fig. 4 | Transcriptomic validation of viral nitrogen-fixing auxiliary metabolic genes (AMGs) in rhizosphere and bulk soils. a Distribution and relative abundance of virus-encoded AMGs classified by KEGG Level 2 pathways. Stars represent AMGs, circles represent KEGG Level 2 pathways, and the colours of both symbols indicate the specific pathway. Among these, *nifU* is the only annotated nitrogen-fixing AMG detected in viral genomes. **b** Schematic representation of the viral genomic context surrounding *nifU* in rhizosphere soil viromes, along with the predicted NifU protein it encodes. **c** PCR amplification of *nifU* from environmental DNA extracted from rhizosphere soils, showing the expected band size of ~350 bp for both bacterial and viral templates. Experiments were repeated two times with similar results. **d** Transcriptomic fold changes (log_2FC) of significantly differentially expressed genes in cowpea rhizosphere soil (S1, organic-fertilizer treatment) relative to bulk soil with organic fertilizer (S2) and bulk soil without organic fertilizer (S4). Positive log_2FC values indicate strong upregulation compared to the referenced control, whereas negative values indicate significant downregulation. Differential expression analysis was performed using DESeq2 with a two-sided Wald test. For visualization in this figure, genes with a nominal p-value < 0.05 were considered significant. Note that multiple testing correction (Benjamini–Hochberg FDR) was applied, and results using adjusted p-value (p_{adj} < 0.05) are provided in Supplementary Data 6. Source data are provided as a Source Data file.

viral genomes raises the possibility that viruses may influence host nitrogen metabolism by modulating regulatory balance and energy allocation, particularly under fluctuating environmental conditions.

Phylogenetic analyses further revealed that NFG-carrying viruses span multiple viral families (Fig. S2), indicating that nitrogen-related functions are distributed across diverse viral lineages rather than being confined to a single clade. Such diversity supports the hypothesis that viral NFGs likely arise through repeated HGT events from bacterial hosts rather than representing a conserved viral trait[21]. Host prediction analyses identified *Klebsiella* and *Pseudomonas* as dominant putative hosts for NFG-carrying viruses (Fig. 2), consistent with their well-documented nitrogen-fixing capacity in soils[22–24]. The detection of both proviruses and viruses with broad host ranges suggests that viral-mediated transfer of nitrogen-related genes among host taxa is possible[25,26]. For instance, a *nifU*-carrying virus (IMGVR_UViG_3300008111_001506) was predicted to infect both *Nitrogeniibacter aestuarii* and *Pseudomonas mucidolens*, illustrating a potential route for interhost gene dissemination. Moreover, multiple viruses carrying the same NFG were linked to the same host taxa, suggesting that certain bacterial lineages may function as hubs for genetic exchange[27]. Collectively, these patterns support the hypothesis that viruses may contribute to the diversification and redistribution of NFGs through HGT[28], thereby influencing the evolutionary trajectory of nitrogen-fixing microbial communities[29,30].

Among the viral NFGs identified, *nifU* was consistently annotated as a viral AMG by both VIBRANT and DRAM-v (Fig. 3a). Nitrogen-related viral AMGs have been previously reported in marine, wetland, and agricultural systems, including genes involved in nitrification (e.g., *amoC*), denitrification (e.g., *nirK*), and assimilatory nitrite reduction (e.g., *nirA*)[10,31,32]. In contrast, *nifU* has not been previously reported as a viral AMG in soil viromes, to our knowledge, making this study the identification of *nifU* as a virus-encoded AMG in soil and rhizosphere environments. The higher abundance of viral *nifU* in rhizosphere compared with bulk soils (p < 0.05; Fig. 4a) is consistent with prior observations that viral reprogramming of host metabolism is stronger in rhizosphere environments, where microbial activity and host availability are elevated[33]. Correspondingly, cowpea rhizosphere soils harboured more diverse viral families than bulk soils (Fig. S6), reflecting the resource-rich and microbially dynamic nature of the rhizosphere[34]. The recurrent detection of viral *nifU* across public databases, native rhizosphere viromes, and virus-transplanted soils suggests that this gene is conserved and maintained within viral populations. Phylogenetic analyses showed high similarity between viral *nifU* sequences identified here and those from nitrogen-fixing bacteria such as *Azospirillum thermophilum* (Fig. S9), supporting a host-derived origin via HGT. Given that NifU functions as a scaffold for iron–sulphur cluster assembly[35], viral *nifU* could potentially enhance host nitrogen fixation efficiency if expressed during infection or integrated during lysogeny[36]. Such integration and metabolic modulation have been documented for other viral AMGs in lysogenic systems[37,38]. However, it is important to emphasize that our data indicate association rather than direct causation.

Consistent with this interpretation, metatranscriptomic analyses showed elevated *nifU* expression in the rhizosphere, accompanied by increased MBN and nitrogenase activity following virus transplantation (Fig. 4d and 5a). DNA-SIP further revealed a shift of *nifH* toward heavier fractions under ^{15}N labelling, with *Azotobacter* dominating the labelled DNA in virus-transplanted treatments (Fig. 5b, c). These results suggest that viral communities may indirectly enhance the ecological competitiveness of nitrogen-fixing bacteria, potentially by restructuring host community composition or modulating host metabolism[39,40], rather than directly driving nitrogen fixation through viral gene expression alone. Several limitations should be acknowledged. First, the rarity of NFG-carrying viruses constrains their likely impact on ecosystem-scale nitrogen budgets, suggesting their primary role may be evolutionary or regulatory rather than quantitative. Second, AMG prediction relies on

https://doi.org/10.1038/s41467-026-70744-1

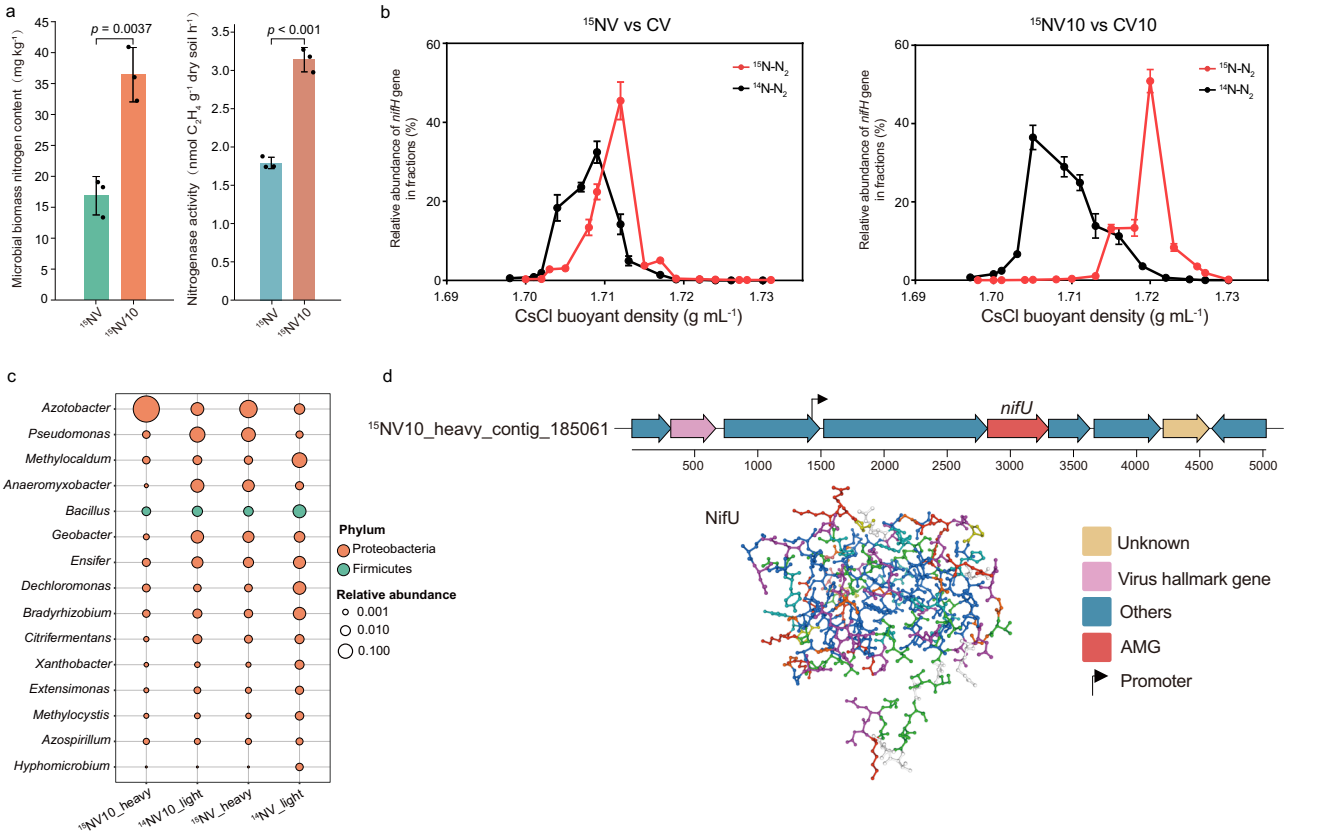

Fig. 5 | ^{15}N-DNA stable isotope probing (DNA-SIP) validates viral nitrogen-fixing auxiliary metabolic genes (AMGs) in agricultural soils. a Microbial biomass nitrogen (MBN) and nitrogenase activity were measured after soil microcosm incubation under different treatments. Data are presented as mean values ± standard deviation (SD). Statistical significance for data analysis was assessed using a two-sided *t*-test ($n = 3$). **b** Relative abundance of the *nifH* gene across CsCl gradient fractions following incubation with $^{15}N_2$ (red line) and $^{14}N_2$ (black line). Left: soil-only and bacteria-only treatments. Right: soil amended with bacteria plus a 10-fold virus addition. Twenty-four fractions were analyzed per gradient. *nifH* abundance in each fraction is expressed as a proportion of the total *nifH* gene abundance across all fractions. Data are presented as mean values ± SD. **c** Bacterial community composition of the most abundant genera in heavy-DNA fractions of the $^{15}N_2$ treatment and light-DNA fractions of the $^{14}N_2$ treatment. **d** Viral contigs carrying nitrogen-fixing AMGs and the predicted structure of the viral NifU protein detected in $^{15}N_2$-treated heavy fractions. Source data are provided as a Source Data file.

bioinformatic tools and curated databases, which may misclassify genes or overlook context-dependent functionality. Third, DNA-SIP and virus transplantation experiments establish correlation and enrichment, but cannot disentangle direct viral gene expression from indirect community-level effects. Finally, controlled infection experiments and targeted mutagenesis are required to directly test the functional contribution of viral *nifU* to host nitrogen fixation.

Taken together, our findings expand current views of viral AMGs beyond carbon and photosynthetic metabolism and suggest that viruses may participate in nitrogen cycling by redistributing key functional genes and reshaping microbial communities. Rather than acting as dominant drivers of nitrogen fixation, viruses likely influence nitrogen metabolism through rare but persistent gene-transfer events and modulation of host ecology. Incorporating viral processes into soil nitrogen-cycling models may therefore improve understanding of long-term ecosystem resilience and evolutionary dynamics. While speculative at present, targeted manipulation of viral–microbial interactions could eventually complement biological nitrogen-fixation strategies, although such applications require substantial further validation.

Methods

Collection of bacterial genomes and screening for nitrogen-fixing genes

We first downloaded 251,800 bacterial genomes from the RefSeq database (National Center for Biotechnology Information) (https://www.ncbi.nlm.nih.gov/refseq/). These genomes constituted our initial dataset for bacterial analysis. For each genome, we searched for the presence of NFGs. NFGs were defined as genes encoding core structural components or regulatory factors of the nitrogenase complex responsible for biological nitrogen fixation, including *nifA*, *nifD*, *nifE*, *nifH*, *nifK*, *nifL*, *nifM*, *nifN*, *nifU*, *nifZ*, *anfK*, and *vnfH*. Bacterial genomes were then screened for completeness and contamination using CheckM (v1.2.3) (https://ecogenomics.github.io/CheckM/) with thresholds of >80% completeness and <10% contamination[41]. Of the 26,175 genomes that carried at least one of the 12 NFGs, 16,245 genomes met these quality criteria and were retained as candidate nitrogen-fixing bacterial genomes for subsequent analyses (Supplementary Data 8). This two-stage process, broad gene detection followed by genome quality filtering, ensured that downstream analyses focused on high-confidence nitrogen-fixing genomes.

Retrieval and quality control of viral genomes carrying nitrogen-fixing genes

Viral genomes were retrieved from the IMG/VR v4 database (https://img.jgi.doe.gov/vr/) to identify uncultivated viral genomes (UViGs) potentially carrying NFGs[42]. Candidate UViGs were initially selected based on sequence similarity to the 12 defined NFGs using the following thresholds: bit score > 40, E-value < 1×10^{-5}, and sequence identity ≥ 20%. To ensure high-quality data, NFG-carrying UViGs were then subjected to quality assessment using CheckV (v0.8.1) (https://github.com/CheckV/checkv) under default parameters[43], retaining

only medium- and high-quality genomes in line with established viromics standards. Next, NFG-carrying UViGs were clustered into viral operational taxonomic units (vOTUs) using Minimum Information about an Uncultivated Virus Genome (MIUViG) criteria: ≥95% average nucleotide identity (ANI) and ≥85% aligned fraction (AF)[44]. Only vOTUs ≥5 kilobases in length were retained to minimize assembly artifacts and ensure analytical robustness[43,45]. Following this workflow, a total of 2996 quality-assured NFG-carrying vOTUs were obtained for downstream analyses (Supplementary Data 9).

Compilation of geographic and habitat metadata

For both datasets (16,245 nitrogen-fixing bacterial genomes and 2996 viral vOTUs), we compiled associated geographic location and habitat origin metadata where available. For bacterial genomes, geographic distribution was aggregated by continent and ocean; for viral vOTUs, habitat categories were similarly annotated. To enable comparative analyses across regions and gene categories, we normalized counts by calculating the percentage of each bacterial population carrying NFGs relative to the total nitrogen-fixing bacterial dataset per region. An analogous normalization approach was applied to viral NFGs across habitats.

Viral phylogenetic tree and species annotation

To investigate evolutionary relationships among the 2996 NFG-carrying vOTUs, we constructed a phylogenetic tree based on 19 conserved protein markers[46]. Marker sequences were concatenated and aligned; columns with <50% representation across vOTUs were removed using trimAl (v1.4) prior to alignment with MUSCLE (v5.1)[47,48]. The final concatenated alignment was used to infer a phylogenetic tree with FastTree (v2.1.11) under the default model[49]. Taxonomic assignment of vOTUs at the family level was performed using PhaGCN2, following the most recent virus taxonomy standards approved by the International Committee on Taxonomy of Viruses[50].

Viral Host prediction and prophage identification

Host prediction for the 2996 vOTUs was conducted using iPHoP (v1.3.3) under default parameters, which reports candidate hosts and associated reliability scores (Supplementary Data 2)[51]. Prophage regions within the 16,245 nitrogen-fixing bacterial genomes were identified using Prophage Hunter[52]. Only regions classified as active were retained. To remove redundancy, prophage sequences on the same genomic scaffold were deduplicated, followed by clustering with CD-HIT (v4.7)[53].

Viral AMG identification and structural analysis

To identify AMGs among the 2996 viral vOTUs, we applied both DRAM-v (v1.2.0) and VIBRANT (v1.2.0) with default settings (including the –virome flag for VIBRANT). DRAM-v annotations were queried against KEGG, RefSeq, Pfam, dbCAN, VOGdb, and CAZymes databases; candidate AMGs were retained if they had an auxiliary score <4 and were not flagged as problematic[54]. VIBRANT AMG hits required annotation as an AMG with a v-score ≥1[55]. Results from both tools were merged, identifying 4972 potential AMGs (Supplementary Data 5). Of these, only *nifU* was confirmed as a viral AMG related to nitrogen fixation and retained for further characterization. To evaluate the functional potential of the viral *nifU* protein, we modelled its quaternary structure using SWISS-MODEL, retaining models with GMQE > 0.5[56]. Putative upstream promoters and downstream transcriptional terminators were predicted using BPROM and FindTerm, respectively, with BPROM results filtered using an LDF threshold >2.75[31].

Soil sample collection

Soil samples were collected in July 2023 from a cowpea field with a long history of cultivation at Hengxi Agroecological Park, Nanjing, Jiangsu Province, China (31.726485°N 118.734189°E). Four sample types were collected: cowpea rhizosphere (S1) and bulk soil (S2) from plots fertilized with organic fertilizer, and cowpea rhizosphere (S3) and bulk soil (S4) from unfertilized plots. For each type, three replicates of 3 kg were collected. Samples were stored at 4 °C in 1-L polypropylene tubes during transport and subsequently stored at −80 °C until further analysis.

Soil viromics analysis

Viral communities were characterized to assess composition and diversity across the four soil types. Viral DNA was extracted in triplicate from 0.25 g of each soil sample using the E.Z.N.A.® Viral DNA Kit (Omega Bio-tek, Norcross, USA), which efficiently recovers viral DNA while minimizing co-extraction of inhibitors[57]. DNA quality was evaluated using a NanoDrop 2000 spectrophotometer and 1% agarose gel electrophoresis. The presence of the viral *nifU* gene was confirmed by PCR in both viral and total microbial fractions. Total microbial DNA was separately extracted from the same soils using the FastDNA Spin kit for soil (MP Biomedicals, OH, USA). Primer sequences and PCR conditions are provided in Supplementary Data 10[58]. All viral DNA samples (S1–S4) were used to construct sequencing libraries (Shanghai Biozeron Biotechnology Co., Ltd.), followed by 150-bp paired-end sequencing on an Illumina NovaSeq 6000 platform. Raw reads were quality-filtered using Trimmomatic v0.36[59], and eukaryotic sequences were removed by aligning reads to the NCBI RefSeq Eukaryotic Genomes database using BWA-MEM v0.7.17[60].

Viral sequences were assembled with Megahit v1.1.2 (--presets meta-large, –min-contig-len 300)[61], and contigs >5 kb were clustered with CD-HIT (–c 0.95, –aS 0.8) to generate a nonredundant dataset[62]. High-confidence viral contigs were identified by retaining those present in the final-viral-combined.fa output of VirSorter2[63] and having a DeepVirFinder p-value ≤ 0.05[64]. Taxonomic annotation was assigned using HMMER against the ViPhOG database[65], and genome quality was assessed with CheckV according to MIUViG standards (95% ANI, 85% aligned fraction). A total of 9308 vOTUs were retained for downstream analyses. Viral taxonomy and ecological similarity were further assessed with vConTACT2[66], including comparison against RefSeq and previously reported soil vOTUs[14,15,67]. Viral abundance was calculated by mapping clean reads to vOTUs using BWA-MEM and computing RPKM values. Viral AMGs and predicted protein structures were analyzed as described in the Methods section, Viral AMG identification and structural analysis.

Soil metatranscriptome analysis

To investigate microbial gene expression and viral AMG activity, total RNA was extracted from 0.1 g of each soil replicate using TRIzol®, with technical triplicates for each sample to minimize processing bias. RNA integrity and quantity were assessed by 1% agarose gel electrophoresis, NanoDrop 2000, and Qubit 4.0. Libraries were constructed using a whole-genome shotgun approach and sequenced on the Illumina NovaSeq 6000 platform. Low-quality and short reads were filtered with Trimmomatic. Noncoding rRNA sequences were removed by mapping to Rfam v12.2 using Bowtie2 (v2.33)[68]. Remaining reads were de novo assembled with Trinity (v2.4.0; parameters: --pre_correction --mink 20 --maxk 60 --step 10)[69].

Open reading frames (>200 bp) were predicted with Prodigal (v2.6.3)[70], clustered into unigenes, and annotated using BLASTx against eggNOG, GO, CAZymes, and KEGG databases. Gene expression levels were quantified by mapping clean reads to unigenes with Bowtie2, and viral AMG transcripts were mapped using BBmap (v38.70) at 100% identity and coverage[71–73]. Expression was normalized as RPKM, accounting for gene length and sequencing depth[74]. To distinguish viral from bacterial *nifU* transcripts, a combined reference database of viral vOTUs and bacterial genomes was constructed. Reads uniquely mapping to viral *nifU* were classified as viral transcripts, while remaining reads were assigned as bacterial transcripts. Relative

abundances of viral and bacterial *nifU* genes were calculated separately using RPKM.

Soil microcosms

Soil microcosm experiments were conducted to simulate the effects of viruses on nitrogen fixation by soil bacterial communities and to evaluate their ecological functions. Rhizosphere soil (S1) from *Vigna unguiculata* (100 g) was mixed with 100 mL of virus buffer (68 mM NaCl, 10 mM $MgSO_4$, 10 mM Tris-HCl, pH 7.5) and transferred to a 500-mL conical flask. After thorough mixing, the suspension was shaken at 4 °C for 20 min and centrifuged at 4500 × g for 20 min. The supernatant was subsequently filtered using a tangential-flow filtration (TFF) system with pore sizes of 0.2 µm (Sartorius VF20P7) and 30 kDa (Sartorius VF20P2). The 0.2-µm filtrate separated bacterial (~27 mL) and viral (~66 mL) communities, while the 30-kDa filtration concentrated extracellular, free-living viruses (~1.5 mL). Concentration of bacterial and viral communities ensured sufficient microbial abundance and virus titre in small-volume microcosms, improving sensitivity for downstream SIP and sequencing analyses and maximizing microbe-virus interactions[75].

To collect inducible temperate viruses, a subset of pure bacterial communities obtained by TFF was treated with mitomycin C (1 µg mL^{-1} final concentration) and incubated in the dark at room temperature for 24 h with shaking at 200 rpm. Mitomycin C-induced viruses were subsequently concentrated (~0.8 mL) and combined with free-living viruses to create a representative viral inoculum (~2.4 mL) for transplantation experiments[76]. Autoclaved soils (50 mL per sample, 121 °C, 90 min, two consecutive days) were used as sterile microcosm matrices[77]. Comparison of available nitrogen and phosphorus concentrations before and after autoclaving confirmed that sterilization did not significantly alter soil properties ($w = 6–9$, $p > 0.05$). DNA-SIP microcosms were constructed in 50-mL serum vials containing 5 g of sterilized S1 soil. For the treatment group, 1.2 mL of the extracted bacterial community and 1 mL of a tenfold concentrated viral suspension were added. For the control group, 1.2 mL of the bacterial community and 1 mL of sterile water were added to maintain equal total volume. Serum bottles were sealed with butyl plugs, flushed three times with helium, and filled with O_2 and N_2 in a 1:4 ratio, using either $^{15}N_2$ for labelled groups or $^{14}N_2$ for unlabelled controls[76].

Four microcosm types were established to differentiate viral effects and isotope labelling: sterilized soil with bacteria plus viral suspension (^{15}NV10), sterilized soil with bacteria alone under $^{15}N_2$ (^{15}NV), sterilized soil with bacteria plus viral suspension under $^{14}N_2$ (CV10), and sterilized soil with bacteria alone under $^{14}N_2$ (CV). Three independent replicates were prepared for each type and incubated in the dark at 25 °C for 35 days. This incubation period, based on prior DNA-SIP studies and pilot experiments, was sufficient to allow detectable ^{15}N incorporation, active nitrogen fixation, and bacterial turnover while minimizing DNA degradation[78]. Soil microcosms were destructively sampled after 35 days of incubation, and samples were immediately stored at −80 °C for subsequent analyses.

Nitrogenase activity and microbial biomass nitrogen

To assess the functional impact of viruses on nitrogen fixation and microbial nitrogen transformation, nitrogenase activity and microbial biomass nitrogen (MBN) were measured. Nitrogenase activity was determined using the acetylene-reduction method[8]. Fresh soil (1 g dry weight) was incubated in a 20-mL serum vial with 0.1 mL of 0.1 M glucose and a reducing agent (sulphite). After injection of 1 mL of acetylene gas (>99.99%), vials were incubated at 25 °C for 24 h, and C_2H_4 production was quantified using a Thermo Trace 1300 gas chromatograph. Nitrogenase activity was expressed as nmol C_2H_4 g^{-1} dry soil h^{-1}. MBN was quantified using the chloroform fumigation-extraction method[79]. Briefly, 2 g of soil were split into fumigated and control groups. The fumigated samples were exposed to chloroform under vacuum for 24 h, lysing microbial cells. Total extractable nitrogen in the fumigated and control groups was measured using a 0.5 M K_2SO_4 extraction, and MBN was calculated as the difference multiplied by a calibration factor (1.85), expressed in mg MBN kg^{-1} dry soil.

DNA extraction and SIP fractionation

Total DNA was extracted from 0.5 g of 35-day soil samples from both $^{15}N_2$ and $^{14}N_2$ treatments using the FastDNA Spin Soil Kit (MP Biomedicals) according to the manufacturer's instructions. DNA quality and concentration were assessed with a NanoDrop 2000 spectrophotometer. Genomic DNA (~5 µg) was separated by CsCl gradient ultracentrifugation into "heavy" (^{15}N-DNA) and "light" (^{14}N-DNA) fractions using OptiSeal polyallomer tubes (Beckman Coulter) at 182,500 × *g* and 20 °C for 48 hours[80]. Fractions were collected, and the refractive index was measured to identify buoyancy, followed by CsCl removal using Amicon® Ultra ultrafiltration. *nifH* gene abundance in each fraction was quantified using qPCR with PolF (5′-TGCGAYCCSAARGCBGATCC-3′) and PolR (5′-ATSGCCATYTCRCCGGA-3′) primers[81]. Purified DNA from SIP experiments was then used for amplicon and metagenomic sequencing.

16S rRNA amplicon sequencing and analysis

Bacterial community composition and diversity were determined using the V4–V5 hypervariable regions of the 16S rRNA gene, amplified from both heavy and light DNA fractions using 515F (5′-GTGCCAGCMGCCGCGGTAA-3′) and 907 R (5′-CCGTCAATTCMTTTRAGTTT-3′) primers[82]. Amplicons were sequenced on a NovaSeq 6000 platform (PE250). Sequence quality control, including filtering, denoising, deduplication, merging, and chimera removal, was performed in QIIME 2 (v2024.10)[83]. Amplicon sequence variants (ASVs) were generated using DADA2[84], with truncating parameters set to 220 bp for forward reads and 205 bp for reverse reads. ASVs with frequencies <2 were removed, and the remaining ASVs were taxonomically classified with a naive Bayesian classifier trained on the SILVA 138 database[85].

Microcosmic soil metagenomic analysis

Metagenomic sequencing was performed on pooled heavy DNA fractions with the highest *nifH* gene abundance per replicate. Libraries were prepared using the Nextera XT DNA Library Prep Kit (Illumina) and sequenced on a NovaSeq 6000 platform (PE250). Raw reads were trimmed with Trimmomatic to remove adapters and retain high-quality reads (>36 bp, Q > 30)[59]. Reads were assembled using MEGAHIT with k-mer sizes of 21–141 (--meta)[61] and scaffolds >1.5 kb were binned into metagenome-assembled genomes (MAGs) using MetaWRAP v1.2.2, employing MetaBAT2, MaxBin2, and CONCOCT[86–88]. Binning results were refined, and MAG integrity and contamination were assessed using CheckM; MAGs with >60% completeness and <10% contamination were retained. Dereplication of genomes was performed with dRep v2.3.2, and taxonomic classification of genomes was performed using GTDB-Tk (release 03-RS86)[89,90]. MAGs were annotated with Prodigal (-p meta) and KEGG using DIAMOND[91], while nitrogen-fixing MAGs were identified via NCycDB using HMM models ($e < 10^{-5}$)[92]. Viral sequences were predicted using VirSorter2 and DeepVirFinder, with AMG identification and protein structure predictions described in the Methods section Viral AMG identification and structural analysis.

Data analysis and visualization

Differential gene expression between treatments was assessed using DESeq in R (v4.0.3), with significance defined as $p \le 0.05$ and $|\log_2(\text{fold change})| \ge 1$. KEGG enrichment analyses were conducted using Fisher's Exact Test. The distribution patterns of bacteria and viruses containing NFGs across geographical regions and habitat types were evaluated using Pearson's Chi-squared test (χ^2) on the raw observed counts. For

significant associations ($p < 0.05$), the strength of the association was quantified using Cramér's V as the effect size. Full contingency tables, including observed and expected counts alongside standardized residuals, are provided in Supplementary Data 1. with data visualized using R and Chiplot (https://www.chiplot.online).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All raw sequence data of the viromes and metatranscriptomes generated in this research have been deposited in the NCBI's Sequence Read Archive (SRA) database and can be accessed under project accession no. PRJNA1089531. All data are publicly accessible and can be downloaded from https://www.ncbi.nlm.nih.gov/sra/PRJNA1089531. All raw sequences from 16S rRNA amplicon sequencing and metagenomic sequencing in the ^{15}N-N_2 stable-isotope tracing and virus transplantation experiments have also been deposited in the SRA and can be accessed under project accession no. PRJNA1328377. The global genomic sequences of bacteria and viruses containing NFGs used for the global-scale analysis in this study were obtained from public databases. Specifically, bacterial NFG-containing genomes were retrieved from the NCBI RefSeq database, and viral NFG-containing genomes were retrieved from the IMG/VR database. All accession numbers for the bacterial genomes (RefSeq assembly accession numbers) and viral genomes (UViG IDs) analyzed in this study are provided in Supplementary Data 11. To further facilitate reproducibility, the complete compiled sequence datasets used for our analysis have been made available on figshare. The bacterial sequences can be accessed via https://doi.org/10.6084/m9.figshare.25523014 and https://doi.org/10.6084/m9.figshare.25546690, and the viral sequences via https://doi.org/10.6084/m9.figshare.25538410. The supplementary and source data generated in this study are provided in the Supplementary Information/Source Data file. Supplementary Data is provided with this paper. Source data are provided with this paper.

Code availability

The code used for our analyses has been deposited in the Zenodo database with https://doi.org/10.5281/zenodo.18638369.

References

1. Ladha, J. K. et al. Achieving the sustainable development goals in agriculture: The crucial role of nitrogen in cereal-based systems. *Adv. Agron.* **163**, 39–116 (2020).
2. Yuan, Z. Y. & Chen, H. Y. H. A global analysis of fine root production as affected by soil nitrogen and phosphorus. *Proc. Biol. Sci.* **279**, 3796–3802 (2012).
3. Ladha, J. K. et al. Biological nitrogen fixation and prospects for ecological intensification in cereal-based cropping systems. *Field Crops Res.* **283**, 108541 (2022).
4. Markmann, K. & Parniske, M. Evolution of root endosymbiosis with bacteria: how novel are nodules? *Trends Plant Sci.* **14**, 77–86 (2009).
5. Rubio, L. M. & Ludden, P. W. Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Annu. Rev. Microbiol.* **62**, 93–111 (2008).
6. Li, W. B. et al. Host-virus coevolution drives soil microbial function succession along a millennium land reclamation chronosequence. *J. Adv. Res.* **71**, 297–306 (2025).
7. Coutinho, F. H. et al. New viral biogeochemical roles revealed through metagenomic analysis of Lake Baikal. *Microbiome* **8**, 163 (2020).
8. Frazão, N., Sousa, A., Lässig, M. & Gordo, I. Horizontal gene transfer overrides mutation in *Escherichia coli* colonizing the mammalian gut. *Proc. Natl. Acad. Sci. USA.* **116**, 17906–17915 (2019).
9. Lu, J. -n et al. DNA viral community enhances microbial carbon fixation capacity via auxiliary metabolic genes in contaminated soils. *Nat. Commun.* **16**, 9984 (2025).
10. Huang, X. et al. Soil nutrient conditions alter viral lifestyle strategy and potential function in phosphorous and nitrogen metabolisms. *Soil Biol. Biochem.* **189**, 109279 (2024).
11. Zheng, X. X. et al. Organochlorine contamination enriches virus-encoded metabolism and pesticide degradation associated auxiliary genes in soil microbiomes. *ISME J* **16**, 1397–1408 (2022).
12. Fang, L. F. et al. Ciprofloxacin- driven purifying selection on viral genomes accelerates soil N_2O production. *Proc. Natl. Acad. Sci. USA.* **122**, e2503199122 (2025).
13. Lund, T., Kulkova, M. Y., Jersie-Christensen, R. & Atlung, T. Essentiality of the *Escherichia coli* YgfZ protein for the In Vivo Thiomethylation of Ribosomal Protein S12 by the RimO Enzyme. *Int. J. Mol. Sci.* **24**, 4728 (2023).
14. Liao, H. et al. Response of soil viral communities to land use changes. *Nat. Commun.* **13**, 6027 (2022).
15. Santos-Medellin, C. et al. Viromes outperform total metagenomes in revealing the spatiotemporal patterns of agricultural soil viral communities. *ISME J* **15**, 1956–1970 (2021).
16. Pedrós-Alió, C. The rare bacterial biosphere. *Ann. Rev. Mar. Sci.* **4**, 449–466 (2012).
17. Lynch, M. D. J. & Neufeld, J. D. Ecology and exploration of the rare biosphere. *Nat. Rev. Microbiol.* **13**, 217–229 (2015).
18. Jousset, A. et al. Where less may be more: how the rare biosphere pulls ecosystems strings. *ISME J* **11**, 853–862 (2017).
19. Yuvaniyama, P., Agar, J. N., Cash, V. L., Johnson, M. K. & Dean, D. R. NifS-directed assembly of a transient [2Fe-2S] cluster within the NifU protein. *Proc. Natl. Acad. Sci. USA.* **97**, 599–604 (2000).
20. Boyer, N. R. et al. Structural insights into redox signal transduction mechanisms in the control of nitrogen fixation by the NifLA system. *Proc. Natl. Acad. Sci. USA.* **120**, e2302732120 (2023).
21. Pleška, M., Lang, M., Refardt, D., Levin, B. R. & Guet, C. C. Phage-host population dynamics promotes prophage acquisition in bacteria with innate immunity. *Nat. Ecol. Evol.* **2**, 359–366 (2018).
22. Zhang, Y., Xu, Q., Wang, G. & Shi, K. Mixed Enterobacter and Klebsiella bacteria enhance soybean biological nitrogen fixation ability when combined with rhizobia inoculation. *Soil Biol. Biochem.* **184**, 109100 (2023).
23. Ogunseitan, O. A., Sayler, G. S. & Miller, R. V. Application of DNA probes to analysis of bacteriophage distribution patterns in the environment. *Appl. Environ. Microbiol.* **58**, 2046–2052 (1992).
24. Silby, M. W., Winstanley, C., Godfrey, S. A. C., Levy, S. B. & Jackson, R. W. Pseudomonas genomes: diverse and adaptable. *FEMS Microbiol. Rev.* **35**, 652–680 (2011).
25. Zhang, J. et al. Expansion of the plaquing host range and improvement of the absorption rate of a T5-like salmonella phage by altering the long tail fibers. *Appl. Environ. Microbiol.* **88**, e0089522 (2022).
26. Huang, D. et al. Enhanced mutualistic symbiosis between soil phages and bacteria with elevated chromium-induced environmental stress. *Microbiome* **9**, 150 (2021).
27. Touchon, M., Bernheim, A. & Rocha, E. P. C. Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J.* **10**, 2744–2754 (2016).
28. Howard-Varona, C. et al. Phage-specific metabolic reprogramming of virocells. *ISME J.* **14**, 881–895 (2020).
29. Emerson, J. B. et al. Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* **3**, 870–880 (2018).
30. Breitbart, M. & Rohwer, F. Here a virus, there a virus, everywhere the same virus? *Trends Microbiol* **13**, 278–284 (2005).
31. Gazitúa, M. C. et al. Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. *ISME J.* **15**, 981–998 (2021).

32. Zhu, Y., Li, R. L., Yan, S., Li, Y. Y. & Xie, S. G. Copper contamination determined the impact of phages on microbially-driven nitrogen cycling in coastal wetland sediments. *J. Hazard. Mater.* **490**, 137870 (2025).
33. Muscatt, G. et al. Crop management shapes the diversity and activity of DNA and RNA viruses in the rhizosphere. *Microbiome* **10**, 181 (2022).
34. Cheng, Z. et al. Interactive dynamics between rhizosphere bacterial and viral communities facilitate soybean fitness to cadmium stress revealed by time-series metagenomics. *Soil Biol. Biochem.* **190**, 109313 (2024).
35. Tao, J. J., Wang, S. S., Liao, T. H. & Luo, H. W. Evolutionary origin and ecological implication of a unique *nif* island in free-living *Bradyrhizobium* lineages. *ISME J.* **15**, 3195–3206 (2021).
36. López-Torrejón, G. et al. Expression of a functional oxygen-labile nitrogenase component in the mitochondrial matrix of aerobically grown yeast. *Nat. Commun.* **7**, 11426 (2016).
37. Howard-Varona, C., Hargreaves, K. R., Abedon, S. T. & Sullivan, M. B. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J* **11**, 1511–1520 (2017).
38. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).
39. Suttle, C. A. Marine viruses – major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
40. Jover, L. F., Effler, T. C., Buchan, A., Wilhelm, S. W. & Weitz, J. S. The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nat. Rev. Microbiol.* **12**, 519–528 (2014).
41. Chklovski, A., Parks, D. H., Woodcroft, B. J. & Tyson, G. W. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods* **20**, 1203–1212 (2023).
42. Camargo, A. P. et al. IMG/VR v4: an expanded database of uncultivated virus genomes within a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.* **51**, D733–D743 (2023).
43. Nayfach, S. et al. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2020).
44. Roux, S. et al. Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat. Biotechnol.* **37**, 29–37 (2018).
45. Liao, H. et al. Mesophilic and thermophilic viruses are associated with nutrient cycling during hyperthermophilic composting. *ISME J* **17**, 916–930 (2023).
46. Low, S. J., Džunková, M., Chaumeil, P.-A., Parks, D. H. & Hugenholtz, P. Evaluation of a concatenated protein phylogeny for classification of tailed double-stranded DNA viruses belonging to the order Caudovirales. *Nat. Microbiol.* **4**, 1306–1315 (2019).
47. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
48. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
49. Piñeiro, C., Abuín, J. M., Pichel, J. C. & Ponty, Y. Very Fast Tree: speeding up the estimation of phylogenies for large alignments through parallelization and vectorization strategies. *Bioinformatics* **36**, 4658–4659 (2020).
50. Jiang, J.-Z. et al. Virus classification for viral genomic fragments using PhaGCN2. *Brief. Bioinform.* **24**, bbac505 (2023).
51. Arumugam, M. et al. iPHoP: An integrated machine learning framework to maximize host prediction for metagenome-derived viruses of archaea and bacteria. *PLoS Biol.* **21**, e3002083 (2023).
52. Song, W. et al. Prophage Hunter: an integrative hunting tool for active prophages. *Nucleic Acids Res.* **47**, W74–W80 (2019).
53. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).
54. Shaffer, M. et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res.* **48**, 8883–8900 (2020).
55. Kieft, K., Zhou, Z. & Anantharaman, K. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* **8**, 90 (2020).
56. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
57. Yang, K. et al. Rhizosphere phage communities drive soil suppressiveness to bacterial wilt disease. *Microbiome* **11**, 16 (2023).
58. Rösch, C., Mergel, A. & Bothe, H. Biodiversity of denitrifying and dinitrogen-fixing bacteria in an acid forest soil. *Appl. Environ. Microbiol.* **68**, 3818–3829 (2002).
59. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
60. Peng, Y., Leung, H. C. M., Yiu, S. M. & Chin, F. Y. L. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).
61. Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015).
62. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
63. Guo, J. et al. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome* **9**, 37 (2021).
64. Ren, J. et al. Identifying viruses from metagenomic data using deep learning. *Quant. Biol.* **8**, 64–77 (2020).
65. Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. Massive expansion of human gut bacteriophage diversity. *Cell* **184**, 1098–1109.e9 (2021).
66. Jang, H. B. et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **37**, 632–639 (2019).
67. Haft, D. H. et al. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res.* **46**, D851–D860 (2018).
68. Wegner, M. & Kaulich, M. ReCo: automated NGS read-counting of single and combinatorial CRISPR gRNAs. *Bioinformatics* **39**, btad448 (2023).
69. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
70. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
71. Wu, R. et al. Moisture modulates soil reservoirs of active DNA and RNA viruses. *Commun. Biol.* **4**, 992 (2021).
72. Luo, X. Q. et al. Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. *Microbiome* **10**, 190 (2022).
73. Yoshida, T. et al. Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan. *ISME J.* **12**, 1287–1295 (2018).
74. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
75. Xie, Z. X. et al. Metaproteomics of marine viral concentrates reveals key viral populations and abundant periplasmic proteins in the oligotrophic deep chlorophyll maximum of the South China Sea. *Environ. Microbiol.* **20**, 477–491 (2017).

76. Knowles, B. et al. Variability and host density independence in inductions-based estimates of environmental lysogeny. *Nat. Microbiol.* **2**, 17064 (2017).
77. Acuña-Rodríguez, I. S. et al. The role of the soil microbiome in the colonisation of glacier forefields by Antarctic pearlwort (*Colobanthus quitensis*) under current and future climate change scenarios. *Soil Biol. Biochem.* **188**, 109249 (2024).
78. Li, Y. et al. Characterization of diazotrophic root endophytes in Chinese silvergrass (*Miscanthus sinensis*). *Microbiome* **10**, 186 (2022).
79. Zhou, Z. et al. The biogeography of soil microbiome potential growth rates. *Nat. Commun.* **15**, 9472 (2024).
80. Kapili, B. J., Barnett, S. E., Buckley, D. H. & Dekas, A. E. Evidence for phylogenetically and catabolically diverse active diazotrophs in deep-sea sediment. *ISME J.* **14**, 971–983 (2020).
81. Poly, F., Monrozier, L. J. & Bally, R. Improvement in the RFLP procedure for studying the diversity of *nifH* genes in communities of nitrogen fixers in soil. *Res. Microbiol.* **152**, 95–103 (2001).
82. Liao, H. et al. Herbicide selection promotes antibiotic resistance in soil microbiomes. *Mol. Biol. Evol.* **38**, 2337–2350 (2021).
83. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
84. Callahan, B. J. et al. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
85. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
86. Uritskiy, G. V., DiRuggiero, J. & Taylor, J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**, 158 (2018).
87. Kang, D. D. et al. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
88. Wu, Y. W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
89. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de–replication. *ISME J.* **11**, 2864–2868 (2017).
90. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-TK: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2019).
91. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
92. Tu, Q., Lin, L., Cheng, L., Deng, Y. & He, Z. NCycDB: a curated integrative database for fast and accurate metagenomic profiling of nitrogen cycling genes. *Bioinformatics* **35**,1040–1048 (2018).

Acknowledgements

We thank Yunling Wu and Shujian Yuan for the valuable advice on the viral genome analyses and data interpretation. This work was supported by the National Natural Science Foundation of China (42277115 to M.M.S.); the Key R&D Project of Jiangsu Province (Modern Agriculture; BE2022322 to M.M.S.); the Fundamental Research Funds for the Central Universities (KJJQ2025028 to M.M.S.); Jiangsu Provincial Agricultural Independent Innovation Project (CX(24)3102 to M.M.S.); and the China Postdoctoral Science Foundation (2024M761435 to D.R.W.). This work was also supported by the Bioinformatics Center of Nanjing Agricultural University.

Author contributions

D.Z. and W.Z. constructed viral genome data and made the figures. W.Z., M.M.S., and J.P. performed bacteria metagenome analyses and virus-host linkage analysis. D.Z., W.Z., and J.L.B. assisted bacterial metatranscriptome analyses. D.Z., D.R.W., and J.P. analyzed the viral AMG annotation and carried out PCR amplification experiments. M.M.S., F.H. and Y.G.Z. supervised the study. W.Z., M.M.S., and Y.G.Z. wrote the manuscript. All authors discussed the design, the methods, the results, and the manuscript, and finally read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-70744-1.

Correspondence and requests for materials should be addressed to Dong Zhu or Mingming Sun.

Peer review information *Nature Communications* thanks Ruonan Wu, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026