Environmental and
Biogeochemical Processes

https://doi.org/10.48130/ebp-0026-0014

Original Research

Open Access

Revealing the genetic diversity of DNA polymerase genes in the Napahai plateau wetland

Xiaolian Shen#, Lingling Xiong#, Jingting Sun, Ting Wang and Xiuling Ji*

Received: 28 May 2026

Revised: 29 July 2026

Accepted: 3 September 2026

Published online: 24 September 2026

Abstract

Wetlands, as the "kidneys of the earth," play distinctive roles in environmental processes. Located in the Three Parallel Rivers region of Yunnan Province, China, the Napahai plateau wetland stands out as a unique ecosystem. In this study, we investigated the diversity and phylogenetic distribution of DNA polymerase (Pol) genes, key enzymes in maintaining genome stability, across a range of environments and host sources. In total, 1,222 DNA polymerase gene sequences were selected for subsequent analysis, of which 104 were derived from the Napahai plateau wetland, 578 from other habitats, and 540 from other sources. Phylogenetic trees were constructed to dissect and elucidate the diversity of genetic inheritance. Phylogenetic analysis showed that PolA, PolB, and PolC family genes each formed distinct phylogenetic clusters. Notably, the Napahai sequences consistently formed cohesive, unique subclusters across all families. Principal coordinate analysis (PCoA) and nonmetric multidimensional scaling (NMDS) were applied to explore the distribution of DNA polymerase genes from different habitats. The PCoA ordination showed that sequences from the Napahai plateau wetland formed a separate cluster from those of other environments. This pattern was also observed in the NMDS ordination (stress = 0.0546, R^2 = 0.9739). The distinct ecological conditions of Napahai distinguish these genes from other wetlands and marine environments. The unique DNA polymerase gene clusters identified in the Napahai wetland contributed to an expanded comprehension of DNA polymerase functionality. It underscored the significant genetic diversity of DNA polymerases in the Napahai plateau wetland, showcasing both diversity and individuality within the gene pool. These findings offer novel insights into the spatial distribution and molecular ecology of DNA polymerases in the Napahai plateau wetland. Additionally, they provide valuable supplementary information for future investigations into wetland ecology.

Keywords: Metagenome, DNA polymerase, Plateau wetland, Genetic diversity

Highlights

- DNA polymerase genes in the Napahai plateau wetland exhibit distinct spatial and geographical differences compared with other ecological environments. Their genetic clustering is unique, showing differential distribution across various habitats and sources.
- Metagenomic data analysis reveals that DNA polymerase genes in the Napahai plateau wetland possess abundant species resources, high diversity, and a certain degree of uniqueness, thereby expanding our understanding of the molecular ecology of these genes in wetland ecosystems.
- Future investigations should aim to explore the applicability of these marker genes in diverse wetlands or terrestrial environments.

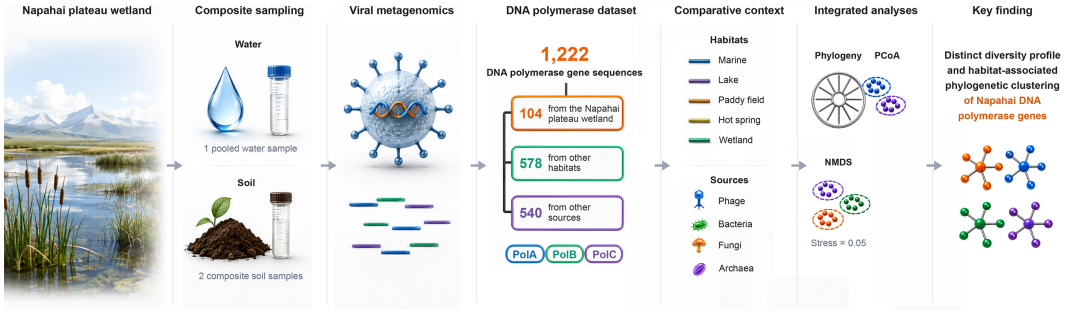
# Authors contributed equally: Xiaolian Shen and Lingling Xiong

* Correspondence: Xiuling Ji (jixiuling1023@126.com)

Full list of author information is available at the end of the article.

https://doi.org/10.48130/ebp-0026-0014

Graphical abstract



Introduction

DNA polymerases (Pol) play a vital role in maintaining genome stability within living organisms. DNA polymerase genes serve as essential molecular markers for examining the genetic diversity of viruses and microorganisms, alongside other widely used markers such as the viral capsid protein genes *g20*[1] and *g23*[2], as well as the cyanophage-carried photosynthesis genes *psbA* and *psbD*[3]. It is noteworthy that DNA polymerase genes exhibit significant variation among viruses, bacteria, and eukaryotes[4]. Through phylogenetic analysis and the examination of primary protein sequences, these genes are classified into seven families: A, B, C, D, X, Y, and RT. The structural conservation of DNA polymerases is critical for essential cellular functions such as DNA replication and repair, and their structural and kinetic properties have been extensively researched[5].

PolA family genes (*polA*) are gene markers for cyanophage diversity[6]. PolA is one of the DNA polymerase families utilized by certain phages for genome replication, although many phages use PolB or phage-specific polymerases. A study explored the flexibility of the PolA active site in accommodating bulky nucleotide analogs[7]. The prediction of PolA peptides from the Chesapeake Bay viral metagenome suggested that PolA was a molecular marker for observing planktonic virus diversity and composition[8]. PolA genes from phages have been identified in various environments, including freshwater, marine environments, hot springs, and soils, spanning a diverse range of viral taxonomies[8,9]. PolB family genes play a crucial role in DNA replication and repair. Baptiste et al.[10] found that DNA polymerase β exhibits distinct advantages in filling single nucleotide gaps and performing helicase-independent functions. PolB demonstrates remarkable diversity and widespread distribution in both cells and viruses. The diversity of PolB in Archaea is particularly noteworthy[11]. Kazlauskas et al. identified a new, widespread group of bacterial PolB and characterized six new groups of PolB in Archaea[11]. PolB can be utilized to study the population-level diversity of Megaviridae in natural environments[12]. PolC is the main chromosomal replication enzyme in Gram-positive bacteria, where it collaborates with DnaE to complete bidirectional replication[13,14]. This evolutionarily conserved enzyme exists as a monomer with intrinsic proofreading activity[15]. Restarting replication after replication–transcription conflicts requires recombination proteins and primosome assembly[16]. Following a successful restart, PolC re-engages at the fork to resume chromosomal DNA synthesis[17].

Most previous studies have focused on DNA polymerase diversity in marine and freshwater viral communities[18,19], whereas information regarding DNA polymerase diversity in plateau wetland ecosystems remains scarce. Kuznetsova et al.[5] discussed the highlights of the kinetic and conformational dynamics of DNA polymerases from all known polymerase families during DNA synthesis. Wetlands, as the "kidneys of the earth," play a crucial role in biogeochemical processes, including pollutant degradation and methanogenesis. They are considered to be hotspots for studying microbial ecology. However, assessments of DNA polymerase diversity in wetland viral metagenomes remain limited, despite wetlands being recognized as hotspots of microbial diversity and ecological processes.

The Napahai plateau wetland, located in the core of the Transverse Mountains and the Three Parallel Rivers area, represents a seasonal swamp wetland with low latitude and high altitude in the Jinsha River Basin. The absence of waterways between wetlands contributes to its unique biological characteristics. Rich in species and featuring diverse habitats, it is recognized as one of the global hotspots for biodiversity conservation[20], possessing high scientific research value. Analysis of the diversity and distribution of DNA polymerase genes in the Napahai wetland using viral metagenomics is crucial for revealing the role of DNA polymerase in the biological evolution of wetland ecosystems.

Materials and methods

Sample collection

On May 23, 2019, water and soil samples were gathered within the coordinates of 27°48'34"–27°53'36" N, 99°38'24"–99°40'38" E in the Napahai plateau wetland (Supplementary Fig. S1). Six sampling points were chosen at a soil depth ranging from 5 to 8 cm, with each yielding 5 kg of soil. This depth was selected because the surface layer (0–5 cm) was characterized by dense meadow root mats and abundant gravel, which made it difficult to obtain intact, homogeneous soil cores. The 5–8-cm depth provided a more uniform soil matrix while remaining within the biologically active layer, consistent with sampling practices for rocky or root-dense wetland soils. The six soil samples were designated S1 to S6, and S1–S3 were homogenized to form Soil 1 (upstream), whereas S4–S6 were homogenized to form Soil 2 (downstream), with each composite sample weighing 400 g. Water samples (W1–W3) were collected from three sampling points along the lake of the Napahai plateau wetland. Each sample, totaling 25 kg, was collected and subsequently mixed to form a single representative water sample. Before homogenization, all individual soil and water samples were analyzed for elemental content. After homogenization, each composite sample was divided into three aliquots for independent DNA extraction. Various parameters, including longitudes, latitudes, altitudes, temperatures, atmospheric pressure, and pH values, were recorded. The elemental content of the samples was analyzed by the Agricultural Product Quality Supervision, Inspection, and Testing Center (Kunming) of the Ministry of Agriculture (Supplementary Table S1).

DNA extraction and viral metagenomic sequencing

First, 0.5-g soil samples were mixed with a buffer. The supernatant was concentrated to 10 mL using a tangential flow filter system. The water samples were collected and filtered through 4.0-μm and 0.22-μm filter membranes and then concentrated using $FeCl_3$ precipitation. After being shaken and settled, the samples were filtered through 0.8-μm filter membranes and then resuspended in fresh 0.1 M EDTA–0.2 M $MgCl_2$–0.2 M buffer (pH 6.5). The concentrated viruses were purified using DNase, which was then inactivated with 5 mM ethylenediaminetetraacetic acid (EDTA)[21]. Next, viral genomes were extracted from the purified samples using the QIAamp MinElute Virus Spin Kit (USA) and further purified with a MicroElute™ DNA Clean-up Kit (China). The viral genomes were then amplified and quantitatively analyzed using the GenomiPhi HY DNA Amplification Kit (China). Finally, the HiSeq 2500 high-throughput sequencing platform was used to conduct 125-bp paired-end sequencing on the sample DNA, with an insert length of 300 bp. The sequencing process was performed by Guangdong MAGIGENE Biotechnology Co., Ltd.

The software SOAPnuke was used for quality control[22]. Clean-read alignment against a designated host genome and subsequent removal of the host sequences were performed using BWA[23] (v0.7.17, default parameters). To extract viral component information, cleaned reads after decontamination were aligned to virus reference data (separated from the NCBI Nucleotide database) using BWA, with a filter length less than 80% of the total length of reads. Viruses were classified according to annotations from the NCBI Taxonomy Database. MEGAHIT[24] v1.0 (default parameters) was utilized to assemble quality-controlled reads from metagenomics, with the exclusion of short contigs (< 1,000 bp). The high-quality reads were aligned to the assembled sequences, and their utilization was calculated using the BWA software[23]. These statistics were combined to evaluate the assembly's impact. For predicting the genes' open reading frames (ORFs), we used the MetaGeneMark software tool[25], and the nonredundant gene set sequences were aligned with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for gene homology alignment using Diamond software (e-value ≤ 0.001). We also extracted all DNA polymerase (EC: 2.7.7.7) nucleotide sequences from metagenomic datasets.

Phylogenetic analysis of DNA polymerase genes

We conducted a phylogenetic analysis to explore the evolutionary origins of the genes. First, we obtained the amino acid sequences of DNA polymerase from viral metagenomic data in the Napahai plateau wetland. We performed searches in the NCBI database to retrieve relevant reference amino acid sequences from diverse environments, including marine, lake, hot spring, wetland, and paddy fields, as well as from different sources such as bacteria, phages, fungi, and Archaea. Subsequently, we constructed phylogenetic trees based on the amino acid sequences from the Napahai plateau wetland and other habitats. We performed multiple sequence alignment using MAFFT v7.505[26]. We constructed phylogenetic evolutionary trees using the Linux IQ-TREE[27], with the evolutionary models selected by IQ-TREE (-m MF -T AUTO -B 1000). Finally, we visualized the trees using iTOL[28].

Statistical analysis

The reference sequences of the DNA polymerase genes were obtained from a variety of environments and sources, including those from the Napahai plateau wetland. Amino acid sequences were analyzed using Mothur[29] software to calculate UniFrac distances, perform operational taxonomic unit (OTU) typing, and assess the alpha genetic diversity indices within the community, such as the Chao, Shannon, and Simpson values, where higher Chao and Shannon values and lower Simpson values indicate greater diversity. Following this, the R package ggplot2 was utilized for principal coordinate analysis (PCoA) and visualization.

However, because of the composite nature of the soil samples (n = 2) and the pooled water sample (n = 1), the ordination analyses (PCoA, nonmetric multidimensional scaling [NMDS]) presented in this study were used for exploratory data visualization and hypothesis generation only. Formal statistical inference was not performed on these data, as the sample size is insufficient for robust multivariate hypothesis testing. The physicochemical properties of the individual field samples (S1–S6 and W1–W3) were measured independently to characterize the environmental heterogeneity within each composite, but these measurements do not constitute biological replicates for community-level analyses.

Results

Phylogenetic analysis of DNA polymerase genes in different habitats

To enhance the comprehension of the genetic diversity of DNA polymerase genes within the Napahai plateau wetland, we constructed circular phylogenetic trees using amino acid sequences sourced from a spectrum of diverse habitats (Fig. 1). PolA family genes were segregated into a single prominent cluster, encompassing all sequences originating from the Napahai plateau wetland (Fig. 1a). PolB family genes were clustered separately into two clusters (Fig. 1b). PolC family genes exhibited clustering into two distinct groups: One associated with hot springs and the other with paddy fields. Remarkably, regardless of whether they were from hot springs or paddy fields, the sequences from the Napahai plateau wetland consistently formed cohesive clusters of their own (Fig. 1c). In conclusion, compared with the relevant sequences from different habitats, the individual clustering of DNA polymerase genes was more obvious in the Napahai plateau wetland.

Phylogenetic analysis of DNA polymerase genes from different sources

To gain insights into the diverse origins of DNA polymerase genes within the Napahai plateau wetland, comprehensive phylogenetic trees were constructed using sequences from different sources (Fig. 2). The results showed that most of the PolA genes clustered, and the sequences from the Napahai plateau wetland were relatively clustered and showed certain uniqueness (Fig. 2a). As can be seen in Fig. 2b, PolB sequences from the Napahai plateau wetland were divided into two distinct clusters. One clustered with archaeal sequences; the other clustered with phage sequences. PolC genes primarily segregated into two main clusters: One cluster associated with phage sequences, whereas the other cluster associated with the sequences from other phages and Archaea (Fig. 2c). DNA polymerase genes from the Napahai plateau wetland formed notably differentiated clusters, co-clustering with sequences from other sources while retaining habitat-specific subclusters.

Principal coordinate analysis of DNA polymerase genes from the Napahai plateau wetland

The amino acid sequences of DNA polymerase genes from the Napahai plateau wetland and other ecological environments, as well as different sources, were analyzed using PCoA. The Bray–Curtis calculation was performed using Mothur software[29], and the distribution

https://doi.org/10.48130/ebp-0026-0014

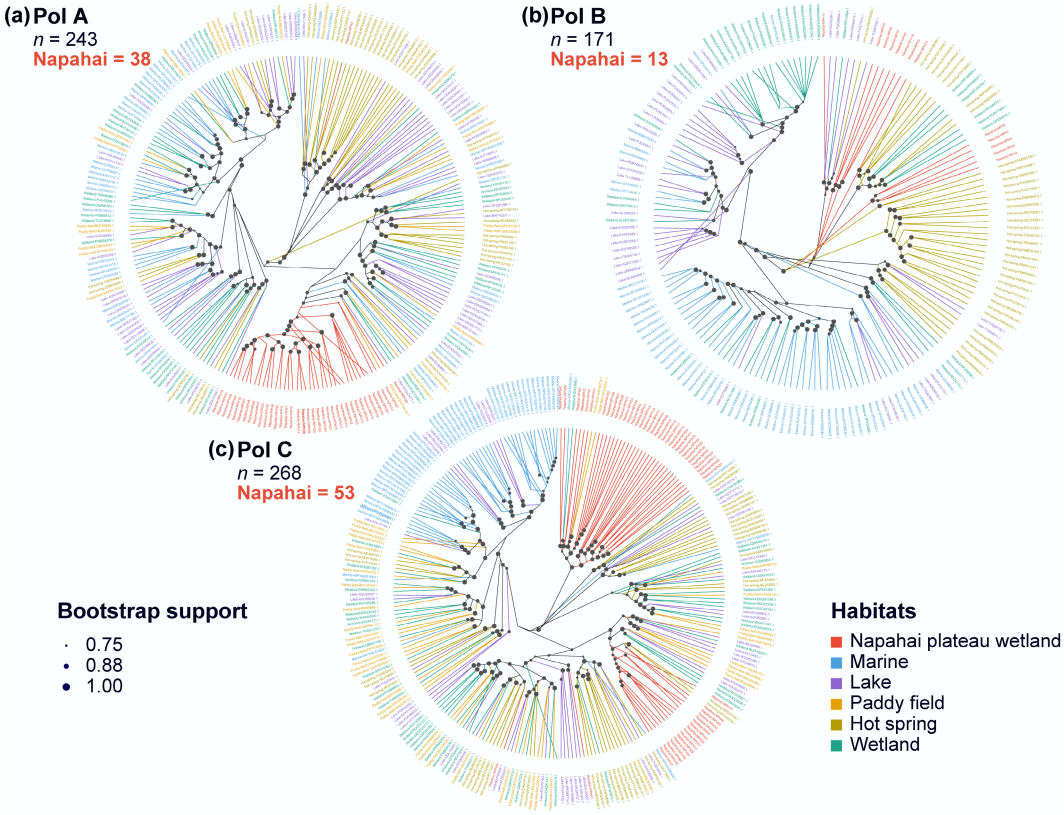


Fig. 1 Phylogenetic analysis of DNA polymerase (Pol) across diverse habitats. The phylogenetic trees of (a) PolA, (b) PolB, and (c) PolC family genes, with different colors representing different habitats.

patterns of DNA polymerase genes were explored and visualized using the R language. As shown in Fig. 3, the ordination plots illustrated the compositional distribution of microbial communities from different environments. The PCoA analysis showed that PolA sequences from the Napahai plateau wetland formed a separate cluster from those of other environments (Fig. 3a). The PolB gene sequences were predominantly distributed in the fourth quadrant, forming a relatively aggregated cluster (Fig. 3b). The PolC gene sequences were mainly distributed in the first and fourth quadrants; a few formed independent clusters, although most clustered together with sequences from other habitats (Fig. 3c). PCoA analysis of PolA genes from different sources showed that Napahai PolA sequences clustered together with those from phages and bacteria, while also forming distinct clusters separate from other sources (Fig. 3d). The PolB sequences appeared to be relatively dispersed across the ordination space (Fig. 3e). The PolC sequences were primarily divided into two major clusters, with one group clustering with phage sequences in the fourth quadrant (Fig. 3f). Overall, the ordination plots revealed that the distribution patterns of DNA polymerase genes from the Napahai plateau wetland exhibited a degree of uniqueness compared with other environments, which was consistent with the phylogenetic analysis results. These observations suggest that the DNA polymerase genes of the Napahai plateau wetland harbor a distinct diversity profile.

The results of the alpha diversity analysis of DNA polymerase genes in different habitats and host origins are shown in Supplementary Fig. S2. Common mathematical measures of species alpha diversity, such as Chao, Shannon, and Simpson diversity indices, were based on the OTU dataset to characterize genetic diversity[30]. Overall, the alpha diversity indices (Chao, Shannon, and Simpson) revealed distinct patterns among the three DNA polymerase gene families (PolA, PolB, and PolC) across different habitats and host origins. These observed differences in alpha diversity among the three gene families suggest that their distribution and prevalence may vary across habitats and host sources, with the Napahai wetland showing a notably distinct diversity pattern.

Nonmetric multidimensional scaling analysis in the Napahai plateau wetland

The nucleotide sequences of DNA polymerase genes from the Napahai plateau wetland and other eco-environments (see Supplementary Table S2) were analyzed using NMDS. The Bray–Curtis distance was calculated using Mothur software[29], and the distribution patterns of DNA polymerase genes were explored using the R language. As depicted in Fig. 4, the NMDS ordination yielded a stress value of 0.0545949 and an R^2 value of 0.973878. The ordination plot illustrated that samples from similar environments tended to cluster together in the NMDS space. The Napahai plateau wetland's DNA polymerase genes formed a separate cluster, distinct from those of other environments. Although positioned relatively close to the Northeast China wetland, the two groups still showed discernible differences in community composition in the ordination space.

Discussion

In this study, DNA polymerase genes were selected for an in-depth exploration of diversity and phylogenetic relationships in the Napahai plateau wetland. Using metagenomic data, we examined the evolution of DNA polymerases across diverse locations and observed distinct clustering patterns specific to this wetland. PolA sequences from the

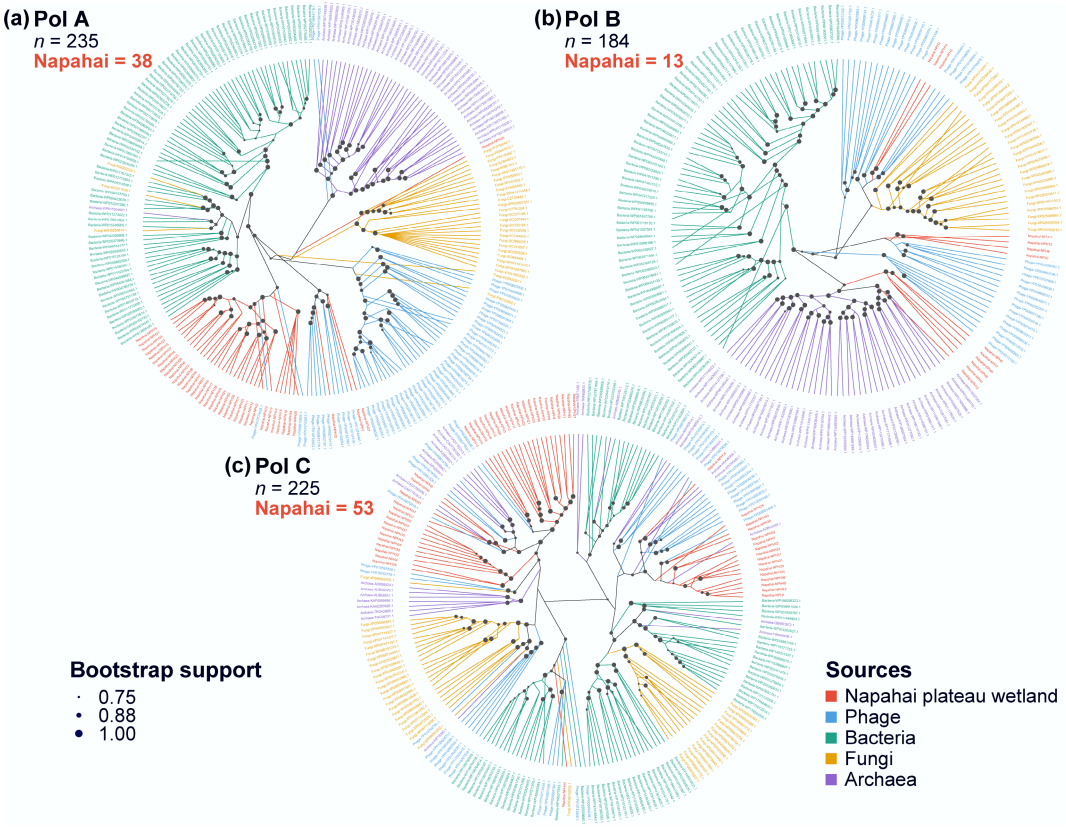


Fig. 2 Phylogenetic analysis of DNA polymerases from different sources. Phylogenetic trees of (a) PolA, (b) PolB, and (c) PolC family genes, with different colors indicating different sources.

Napahai plateau wetland clustered with those from diverse habitats while retaining distinct subclusters, reflecting the local and regional biogeography and showcasing the evident indigenous biodiversity and uniqueness of DNA polymerases in this ecosystem. These characteristics may be intricately linked to the distinct geographical and environmental factors shaping the originality of the Napahai plateau wetland.

The Napahai plateau wetland, situated in the Jinsha River Basin of the Three Parallel Rivers area amid the extended mountains of the Qinghai–Tibet Plateau, offers a typical environment distinguishable from other wetlands. Chen et al.[31] reported the diversity, dynamic population, and seasonal variations of cyanophages using the DNA polymerase gene in Chesapeake Bay. The DNA polymerase sequences of cyanopodoviruses from global oceans and the South China Sea were phylogenetically categorized into Subclusters II, III, VIII, IX, and XI[32], whereas sequences from the Chesapeake Bay formed groups within Subclusters I–X[31], indicating that the distribution of DNA polymerase genes in the open oceans differs from that in Chesapeake Bay[32]. Thus, the phylogenetic distribution of DNA polymerase genes in the Napahai plateau wetland differed from that in paddy fields and marine environments (Fig. 1).

Phylogenetic tree analysis showed that DNA polymerase genes from the Napahai plateau wetland formed distinct clades, suggesting a unique evolutionary lineage associated with this geographical environment. Consistently, NMDS analysis illustrated that DNA polymerase genes from wetland environments, including both the Northeast China wetland and the Napahai plateau wetland, exhibited both distinctions and connections in the ordination space. These two wetland habitats formed separate clusters yet were positioned relatively close to each other compared with non-wetland habitats, suggesting a degree of compositional similarity that may reflect shared selective pressures associated with wetland ecosystems. The clustering patterns observed in both phylogenetic and ordination analyses imply that DNA polymerase genes may distinguish microbial communities in different locations and potentially serve as a biogeographical marker for specific habitats, such as wetlands and rice fields in Northeast China[19,33,34]. The distinct phylogenetic distribution of DNA polymerase genes in the Napahai wetland may be related to its unique environmental conditions. As previously described, the Napahai plateau wetland is characterized by distinct geographical and climatic conditions that set it apart from other wetland ecosystems. This unique climatic regime, shaped by the uplift of the Qinghai–Tibet Plateau and influenced by warm and humid airflows from the Indian Ocean, combined with prolonged daylight duration and high solar ultraviolet intensity, may contribute to the distinctive composition of DNA polymerase genes observed in the Napahai plateau wetland samples. The observed differences in DNA polymerase gene composition between the Napahai plateau wetland and other habitats suggest that geographical and climatic factors, along with associated microbial community dynamics, may play a role in shaping the genetic diversity of microbial communities.

Schmidt et al.[35] discovered that a new type of A-family DNA polymerase was prevalent in marine planktonic viruses, and similar sequences were found in various remote geographical locations, indicating widespread distribution across different environments and a potential role in the global biogeochemical cycle. This genetic exchange in aquatic environments is significant in the geochemical cycle[36]. Nasko et al.[8] discovered more than 3,000 *polA* sequences in the global metagenome of aquatic viruses, suggesting that

https://doi.org/10.48130/ebp-0026-0014

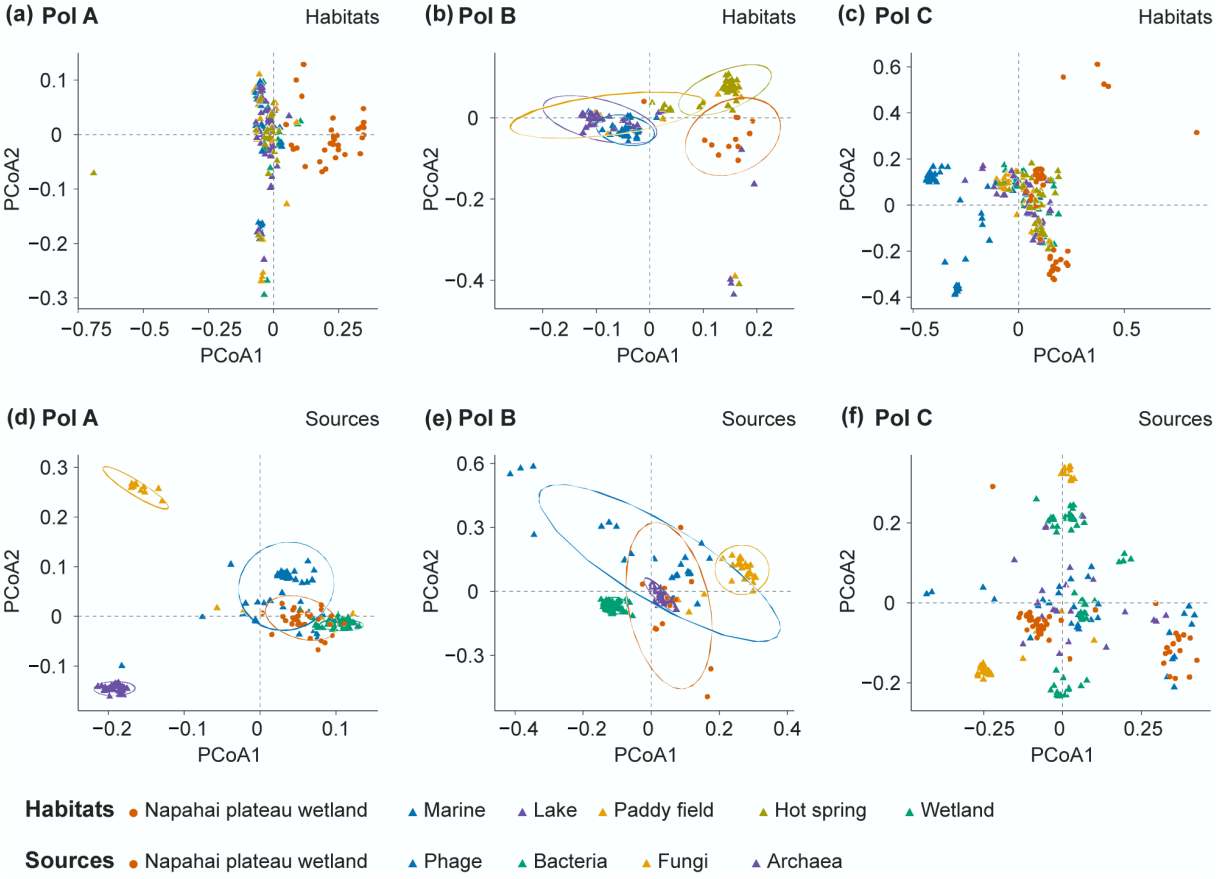


Fig. 3 Principal coordinate analysis of DNA polymerase genes in the Napahai plateau wetland. Panels (**a**) and (**d**) present the PCoA for PolA family genes from different habitats and different origins. Panels (**b**) and (**e**) present the PCoA for PolB family genes from different habitats and different origins. Panels (**c**) and (**f**) present the PCoA for PolC family genes from different habitats and different origins.

differences in biochemical, geographical, and seasonal factors between aquatic sites were closely linked to microbial biodiversity. In this study, most PolA sequences in the Napahai plateau wetland formed distinct clusters, although a few clustered with phages (Fig. 2a).

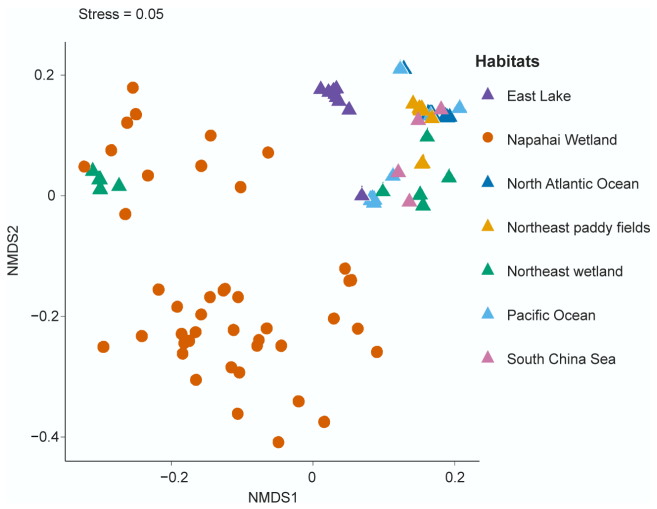


Fig. 4 NMDS analysis of DNA polymerase genes from different habitats. The stress value (0.0545949) and R^2 (0.973878) indicate a reliable ordination configuration.

PolB, a widespread replicative DNA polymerase in Archaea, eukaryotes, and many DNA viruses, has been extensively studied, but its origin and evolution remain unclear[11]. Feng et al.[37] found that PolB could not only replicate intact DNA with high fidelity but also efficiently extend mismatched and damaged DNA substrates. Cyanophages can serve as donors for gene transfer, permuting the DNA polymerase gene of the host cells[38]. In this paper, the phylogenetic analysis showed that the PolB sequences from the Napahai plateau wetland were significantly different from those of the marine environment, and the distribution pattern was likely to be different (Fig. 1b). Across different sources, PolB sequences originating from the Napahai plateau wetland were divided into two clusters: One cluster was close to Archaea, and the other was clustered with the phages (Fig. 2b).

Huang et al.[39] considered PolC to be an information macromolecule that could be used to infer phylogenetic relationships among bacteria and could provide an additional phylogenetic marker for bacterial phylogeny. The analysis of PolC revealed that in the Napahai plateau wetland environment, the genetic diversity and geographical distribution of PolC were distinct (Fig. 1c). In the Napahai plateau wetland, PolC sequences formed distinct clusters, with some co-clustering with phage and archaeal homologs (Fig. 2c). It is noteworthy that PolC rarely clustered with fungi and bacteria, possibly because of geographical barriers. The flow of PolC genes may be restricted by physical and chemical factors such as the salinity gradient, oxygen content, and geographical conditions, as well as ecological barriers such as antagonistic biological invasion. These barriers

may hinder the inter-ecosystem exchange of microbes, leading to a unique succession of PolC genes in the Napahai plateau wetland and significant differences between different habitats and sources. The widespread distribution of PolC genes and the variability in their community composition reflect habitat-specific separation in the local environments. Although the PolC genes in different geographical environments possesses the same source, some become distinct from each other through different living conditions and evolutionary trends. PolC homologs from diverse environments exhibit significant sequence divergence and functional differentiation, reflecting adaptive evolution in response to habitat-specific selective pressures. Different DNA polymerase genes are maintained in habitats such as oceans, wetlands, and lakes (Fig. 1), which may result from various environmental types where they undergo several selection pressures and evolve to adapt to their respective environments. Genetic differentiation of DNA polymerase genes between wetland and marine habitats may reflect community differences among these habitats, as well as variation in the hydrodynamic conditions. This variance affects the parameters, including turbulence, water, light, and nutrient availability.

In this study, we examined and compared the biogeographic distribution of DNA polymerase genes between the Napahai plateau wetland and other environments. Our findings showed that DNA polymerase genes were clustered differently in various habitats and sources, possibly caused by differences in gene abundance, biological characteristics, and the biogeographic environment. The Napahai plateau wetland, located at the intersection of East Asia, South Asia, and the Qinghai–Tibet Plateau, is renowned as the "World's biological gene bank". The unique DNA polymerase gene cluster is influenced by the geographical isolation and alpine landforms of glaciers. However, we are uncertain whether these gene differences occur within individual genes or involve larger DNA fragment exchanges. Metagenomic studies have limitations in revealing the ecological, physiological, and evolutionary roles of wetland DNA polymerase genes, and further experimental research on polymerase gene systems is necessary.

Leveraging metagenomic data and incorporating sequences from various habitats and sources, this study broadens our understanding of DNA polymerase genes, revealing their widespread distribution in wetlands. The NMDS ordination suggested a relatively high genetic abundance of DNA polymerase genes in the Napahai plateau wetland, indicating a degree of diversity and distinctiveness among the DNA polymerase genes in this habitat. This observation is somewhat expected, given that the Napahai plateau wetland is an enclosed inland wetland. It is crucial to note that the DNA polymerase sequences were derived from only nine wetland samples in this study. Thus, we cannot definitively assert the extensive presence of DNA polymerase genes in other wetlands. Future investigations should aim to explore the applicability of these marker genes in diverse wetlands or terrestrial environments. The genetic diversity of DNA polymerase in the Napahai plateau wetland is examined in detail, providing new insights into the spatial distribution and molecular ecology of DNA polymerase in this unique wetland ecosystem.

Conclusions

In summary, this study investigated the diversity and phylogenetic distribution of DNA polymerase genes (PolA, PolB, and PolC) in the Napahai plateau wetland using a metagenomic approach. Phylogenetic and ordination analyses consistently revealed that the Napahai sequences formed distinct, cohesive clusters separate from those of other environments, with PolA, PolB, and PolC each exhibiting unique clustering patterns. These observations suggest that the distinct geographical and climatic conditions of the Napahai wetland, shaped by uplift of the Qinghai–Tibet Plateau and influenced by the Indian Ocean's monsoons, may have driven the unique evolutionary trajectory of these genes. Our findings broaden the understanding of DNA polymerase diversity in wetland ecosystems and highlight the potential of these genes as biogeographical markers. However, given the limited sample size (n = 2 for soil, n = 1 for water), these results should be considered preliminary and hypothesis-generating, warranting further validation with increased biological replication. This study provides valuable baseline data and novel insights into the spatial distribution and molecular ecology of DNA polymerases in the Napahai plateau wetland, contributing to future research on wetlands' microbial ecology and functional gene evolution.

Supplementary information

It accompanies this paper at: https://doi.org/10.48130/ebp-0026-0014.

Ethical statements

Not applicable.

Author contributions

The authors confirm their contributions to the paper as follows: Xiaolian Shen: design, writing – original draft; Xiaolian Shen, Lingling Xiong: investigation; Xiaolian Shen, Lingling Xiong, Jingting Sun, Ting Wang: data curation; Xiuling Ji: supervision, writing – review and editing. All authors reviewed the results and approved the final version of the manuscript.

Data availability

The metagenomic sequencing data from the Napahai plateau wetland weredeposited with links to BioProject accession number PRJNA846364 in the NCBIBioProject database (www.ncbi.nlm.nih.gov/bioproject/term=PRJNA846364).

Acknowledgments

No acknowledgement are applicable for this work.

Funding

This research was supported by the National Natural Science Foundation of China (32560297) and the Young Talents Support Program of Yunnan Province (YNWR-QNBJ-2019-202).

Declarations

Competing interests

The authors declare no competing interests.

Author details

Life Science and Technology & Medical School, Kunming University of Science and Technology, Kunming 650500, China

https://doi.org/10.48130/ebp-0026-0014

References

[1] Almutairi A, Suttle CA, Gustavsen JA. 2025. The hypersaline northwestern Arabian Gulf contains a phylogenetically diverse and highly uneven community of viruses related to cyanophages and pelagiphages. *Aquatic Microbial Ecology* 91:1–14

[2] Potapov S, Belykh O, Krasnopeev A, Gladkikh A, Kabilov M, et al. 2018. Assessing the diversity of the *g23* gene of T4-like bacteriophages from Lake Baikal with high-throughput sequencing. *FEMS Microbiology Letters* 365:fnx264

[3] Fuchsman CA, Carlson MCG, Prieto DG, Hays MD, Rocap G. 2021. Cyanophage host-derived genes reflect contrasting selective pressures with depth in the oxic and anoxic water column of the Eastern Tropical North Pacific. *Environmental Microbiology* 23:2782–2800

[4] Laatri S, El Khayari S, Qriouet Z. 2024. Exploring the molecular aspect and updating evolutionary approaches to the DNA polymerase enzymes for biotechnological needs: a comprehensive review. *International Journal of Biological Macromolecules* 276:133924

[5] Kuznetsova AA, Fedorova OS, Kuznetsov NA. 2022. Structural and molecular kinetic features of activities of DNA polymerases. *International Journal of Molecular Sciences* 23:6373

[6] Maidanik I, Kirzner S, Pekarski I, Arsenieff L, Tahan R, et al. 2022. Cyanophages from a less virulent clade dominate over their sister clade in global oceans. *The ISME Journal* 16:2169–2180

[7] Purohit V, Grindley NDF, Joyce CM. 2003. Use of 2-aminopurine fluorescence to examine conformational changes during nucleotide incorporation by DNA polymerase I (Klenow fragment). *Biochemistry* 42:10200–10211

[8] Nasko DJ, Chopyk J, Sakowski EG, Ferrell BD, Polson SW, et al. 2018. Family A DNA polymerase phylogeny uncovers diversity and replication gene organization in the virioplankton. *Frontiers in Microbiology* 9:3053

[9] Palmer M, Hedlund BP, Roux S, Tsourkas PK, Doss RK, et al. 2020. Diversity and distribution of a novel genus of hyperthermophilic *Aquificae* viruses encoding a proof-reading family-A DNA polymerase. *Frontiers in Microbiology* 11:583361

[10] Baptiste BA, Baringer SL, Kulikowicz T, Sommers JA, Croteau DL, et al. 2021. DNA polymerase β outperforms DNA polymerase γ in key mitochondrial base excision repair activities. *DNA Repair* 99:103050

[11] Kazlauskas D, Krupovic M, Guglielmini J, Forterre P, Venclovas Č. 2020. Diversity and evolution of B-family DNA polymerases. *Nucleic Acids Research* 48:10142–10156

[12] Li Y, Hingamp P, Watai H, Endo H, Yoshida T, et al. 2018. Degenerate PCR primers to reveal the diversity of giant viruses in coastal waters. *Viruses* 10:496

[13] Dervyn E, Suski C, Daniel R, Bruand C, Chapuis J, et al. 2001. Two essential DNA polymerases at the bacterial replication fork. *Science* 294:1716–1719

[14] McHenry CS. 2011. Bacterial replicases and related polymerases. *Current Opinion in Chemical Biology* 15:587–594

[15] Evans RJ, Davies DR, Bullard JM, Christensen J, Green LS, et al. 2008. Structure of PolC reveals unique DNA binding and fidelity determinants. *Proceedings of the National Academy of Sciences of the United States of America* 105:20695–20700

[16] Million-Weaver S, Samadpour AN, Merrikh H. 2015. Replication restart after replication-transcription conflicts requires RecA in *Bacillus subtilis*. *Journal of Bacteriology* 197:2374–2382

[17] Carrasco B, Torres R, Moreno-Del Álamo M, Ramos C, Ayora S, et al. 2024. Processing of stalled replication forks in *Bacillus subtilis*. *FEMS Microbiology Reviews* 48:fuad065

[18] Finke JF, Suttle CA. 2019. The environment and cyanophage diversity: insights from environmental sequencing of DNA polymerase. *Frontiers in Microbiology* 10:167

[19] Li X, Sun Y, Liu J, Yao Q, Wang G. 2019. Molecular diversity of cyanopodoviruses in two coastal wetlands in Northeast China. *Current Microbiology* 76:863–871

[20] Lu M, Ren Y, Wang S, Tian K, Sun X, et al. 2019. Contribution of soil variables to bacterial community composition following land use change in Napahai plateau wetlands. *Journal of Environmental Management* 246:77–84

[21] Hurwitz BL, Deng L, Poulos BT, Sullivan MB. 2013. Evaluation of methods to concentrate and purify ocean virus communities through comparative, replicated metagenomics. *Environmental Microbiology* 15:1428–1440

[22] Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, et al. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience* 7:gix120

[23] Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25:1754–1760

[24] Li D, Luo R, Liu CM, Leung CM, Ting HF, et al. 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102:3–11

[25] Zhu W, Lomsadze A, Borodovsky M. 2010. *Ab initio* gene identification in metagenomic sequences. *Nucleic Acids Research* 38:e132

[26] Nakamura T, Yamada KD, Tomii K, Katoh K. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics* 34:2490–2492

[27] Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32:268–274

[28] Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 47:W256–W259

[29] Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, et al. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75:7537–7541

[30] Morris EK, Caruso T, Buscot F, Fischer M, Hancock C, et al. 2014. Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories. *Ecology and Evolution* 4:3514–3524

[31] Chen F, Wang K, Huang S, Cai H, Zhao M, et al. 2009. Diverse and dynamic populations of cyanobacterial podoviruses in the Chesapeake Bay unveiled through DNA polymerase gene sequences. *Environmental Microbiology* 11:2884–2892

[32] Huang S, Wilhelm SW, Jiao N, Chen F. 2010. Ubiquitous cyanobacterial podoviruses in the global oceans unveiled through viral DNA polymerase gene sequences. *The ISME Journal* 4:1243–1251

[33] Cai L, Chen Y, Xiao S, Liu R, He M, et al. 2023. Abundant and cosmopolitan lineage of cyanopodoviruses lacking a DNA polymerase gene. *The ISME Journal* 17:252–262

[34] Wang X, Liu J, Yu Z, Jin J, Liu X, et al. 2016. Novel groups of cyanobacterial podovirus DNA polymerase (*pol*) genes exist in paddy waters in Northeast China. *FEMS Microbiology Ecology* 92:fiw192

[35] Schmidt HF, Sakowski EG, Williamson SJ, Polson SW, Wommack KE. 2014. Shotgun metagenomics indicates novel family A DNA polymerases predominate within marine virioplankton. *The ISME Journal* 8:103–114

[36] Xia H, Li T, Deng F, Hu Z. 2013. Freshwater cyanophages. *Virologica Sinica* 28:253–259

[37] Feng X, Zhang B, Gao Z, Xu R, Liu X, et al. 2022. A well-conserved archaeal B-family polymerase functions as an extender in translesion synthesis. *mBio* 13:e0265921

[38] Chénard C, Wirth JF, Suttle CA. 2016. Viruses infecting a freshwater filamentous cyanobacterium (*Nostoc* sp.) encode a functional CRISPR array and a proteobacterial DNA polymerase B. *mBio* 7:e00667-16

[39] Huang YP, Ito J. 1999. DNA polymerase C of the thermophilic bacterium *Thermus aquaticus*: classification and phylogenetic analysis of the family C DNA polymerases. *Journal of Molecular Evolution* 48:756–769

Copyright: © 2026 by the author(s). Published by Maximum Academic Press, Fayetteville, GA. This article is an open access article distributed under Creative Commons Attribution License (CC BY 4.0), visit https://creativecommons.org/licenses/by/4.0/.