**Reviewer 1**

We are extremely thankful for these constructive comments, we have carefully addressed all comments and suggestions, please see detailed revisions below.

**Comment 1**

Line 35, add related references on "…… a pool of carbon and nitrogen". How about the carbon and nitrogen storage in the glaciers?

*Response*

The organic carbon stored in the global glacier was estimated on an order of 6 Pg (Hood et al., 2015), but the estimate for nitrogen stored in the global glacier is not yet available. We have amended the sentence for clarification

*Amended manuscript*

It has been estimated that six Pg of carbon are stored in global glaciers. These carbon may be released into downstream ecosystems with glacier runoff (Hood et al., 2015), influencing key elemental cycling in downstream ecosystems.

**Comment 2**

Line 35, "The six Pg of carbon…with glacier runoff." Do you really mean that every year six PgC can be released due to glacial runoff? This sentence is not so clear?

*Response*

This is an estimated number of the total amount of carbon stored in global glaciers. Certainly, not all glaciers will melt and release the carbon stored within. We have rewritten the sentence to clarify this.

*Amended manuscript*

It has been estimated that six Pg of carbon are stored in global glaciers. These carbon may be released into downstream ecosystems with glacier runoff (Hood et al., 2015), influencing key elemental cycling in downstream ecosystems.

**Comment 3**

Line 44-54, how about the "englacial and subglacial ecosystems" as you mentioned in the first sentence of this paragraph?

*Response*

We have added data from the ice core and subglacial ice. Therefore, englacial and subglacial ecosystems are covered in the dataset. We have also added additional sentences to elaborate on the activity differences between microorganisms in supraglacial ecosystem and englacial ecosystems.

*Amended manuscript*

Compared with other glacier-related habitats, the microorganisms in supraglacial ecosystems are the most active, due to their exposure to external environment and ambient temperature.

**Comment 4**

Line 55, Is the glacier surface meltwater belonged to "supraglacial ecosystem"?

*Response*

Glacier surface melting water is a part of the supraglacial ecosystem.

**Comment 5**

Line 61, Add a "." at the end of this sentence.

*Response*

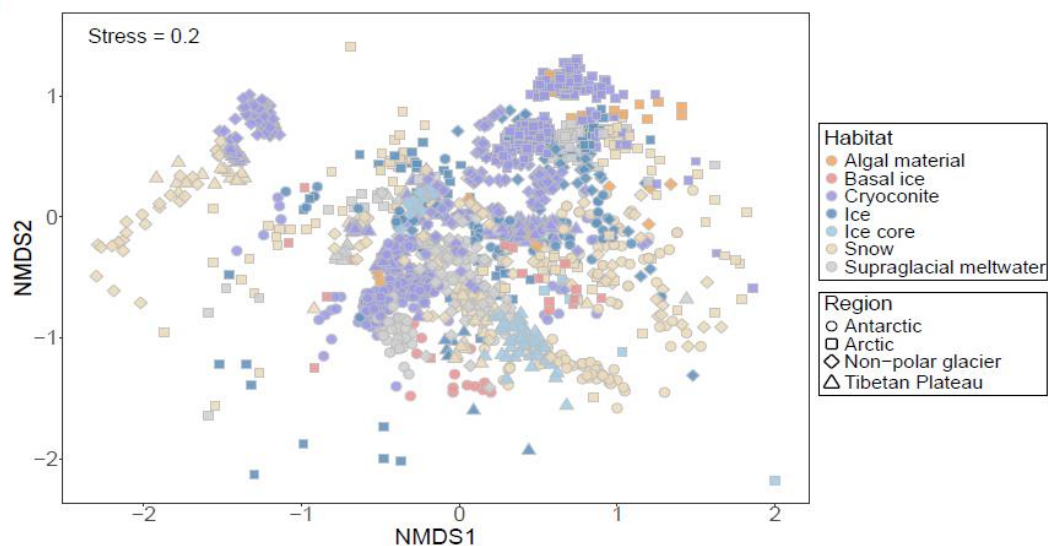We have added the missing "." at the end of the sentence.

**Comment 6**

Figure 2(a), the legends for Ice (Inverted triangle) and Snow (Regular triangle) should be separated.

*Response*

We have increased the space between the figure legends so that these symbols are separated.



**Comment 7**

Figure 3, What is the captions for (a) and (b) respectively? Do you mean they have the same captions, and the meaning of a,b,c on the figure are also the same?

*Response*

The results on potential pathogens have been removed by the comments from reviewer #3.

**Reviewer 2**

The manuscript offers a dataset on microbial communities from glaciers in Antarctica, the Arctic, and Tibet, comprising 815 amplicon sequence data, 952 cultured bacterial genome data, and 208 metagenomic data. This dataset, covering diverse habitats like ice, snow, and cryoconite, is instrumental in understanding microbial diversity, taxonomy, community structure, and genetic function in glacial environments. However, there are critical issues concerning the dataset's completeness need to be addressed.

*Response*

We appreciate the reviewer's suggestions, and we have provided detailed additional information on the sediment sampling process, please see comments below.

**Comment1**

One major concern is the apparent lack of comprehensiveness in the metagenomic data presented in Table S2. Notably, the authors' previous study (Zhang et al., 2023, Microbiome, https://doi.org/10.1186/s40168-023-01621-y), which analyzed 88 metagenomes from 26 glacier cryoconites, is not included. This omission is surprising and detracts from the dataset's perceived completeness.

*Response:*

The 88 metagenomes in Zhang et al., (2023, Microbiome) are all from published papers, all these dataset have been included in the present work.

They include

Franzetti A，Tagliaferri I，Gandolfi I, et al. Light-dependent microbial metabolisms drive carbon fluxes on glacier surfaces[J]. ISME J，2016，10: 2984-2988.

Bellas C M，Schroeder D C，Edwards A, et al. Flexible genes establish widespread bacteriophage pan-genomes in cryoconite hole ecosystems[J]. Nat Commun，2020，11: 4403

Zhang B，Chen T，Guo J, et al. Microbial mercury methylation profile in terminus of a high-elevation glacier on the northern boundary of the Tibetan Plateau[J]. Sci Total Environ，2020，708: 135226

Hauptmann A L，Sicheritz-Pontén T，Cameron K A, et al. Contamination of the Arctic reflected in microbial metagenomes from the Greenland ice sheet[J]. Environmental Research Letters，2017，12: 074019

Liu YQ, Ji MK, Yu T, et al., A genome and gene catalog of glacier microbiomes. Nature Biotechnology, 2022. 40(9): 1341.

Murakami T, Takeuchi N, Mori H,et al., Metagenomics reveals global-scale contrasts in nitrogen cycling and cyanobacterial light-harvesting mechanisms in glacier cryoconite. MICROBIOME, 2022. 10(1).

**Comment 2**

Furthermore, the absence of citations for other significant metagenomics studies (below) is a critical oversight, especially given the reliance of the current paper's results on metagenomic analysis. Incorporating a broader range of relevant literature is essential to bolster the study's credibility and thoroughness.

Varliero et al. (2021) Frontiers in Microbiology, doi.org/[10.3389/fmicb.2021.627437](10.3389/fmicb.2021.627437)
Melanie C. Hay et al., (2023) Microbial genomics, doi.org/[10.1099/mgen.0.001131](10.1099/mgen.0.001131)
Bellas et al. (2020) Nature Communications. doi.org/[10.1038/s41467-020-18236-8](10.1038/s41467-020-18236-8)
Edwards et al. (2013) Environ. Res. Lett. DOI [10.1088/1748-9326/8/3/035003](10.1088/1748-9326/8/3/035003)

*Response*

We thank you for reminding us of the oversight of these important references.

- For Varliero et al. (2021) Frontiers in Microbiology, doi.org/10.3389/fmicb.2021.627437. We have included the six metagenomes from glacier habitats from this study (PRJEB41174, ERR4837082-ERR4837084, ERR4837105, ERR4837106, and ERR4837128).

- For Melanie C. Hay et al., (2023) Microbial genomics, doi.org/10.1099/mgen.0.001131. We have included six samples from this study is from the glacier habitats, we have included these in our dataset (PRJEB59067, ERR10878199-ERR10878204)

- For Bellas et al. (2020) Nature Communications. doi.org/10.1038/s41467-020-18236-8, we have included six metagenomes that are available from NCBI SRA archive in our dataset (SRR8842250, SRR8842249, SRR8842248, SRR12327455, SRR12327363, and SRR12350504). But for mgm4491734.3, it is deposited in MG-RAST and the server is no longer active. We could not retrieve this data anymore.

- For Edwards et al. (2013) Environ. Res. Lett. DOI 10.1088/1748-9326/8/3/035003. The data is deposited in MG-RAST, for the same reason, this is no longer available.

These additional metagenomes have been added to the current dataset.

**Comment 3**

Regarding the amplicon sequencing data in Table S1, the quantity appears insufficient for a study claiming to be a comprehensive database of DNA data from glaciers. To enhance the robustness of the dataset, it is advisable to include data from at least 10 or more amplicon sequencing studies focusing on glaciers. The current dataset's overrepresentation of data from the Tibetan Plateau might introduce a geographical bias in the study's findings.

*Response:*

We thank you for raising this concern. We performed an additional search in the NCBI Biosample database, using the keyword <Glacier OR ice OR snow OR cryoconite>, with sample type being DNA and instrument of Illumina, we retrieved 225,378 SRA entries initially. Then the results were carefully filtered manually to remove non glacier habitats (such as glacier forefield and ice cave), metagenome data, and primers that do not amplify the V4 region of the 16S rRNA gene (i.e., only those amply V3V4, V4, and V4V5 were retained) (Table S1).

We then only kept samples with more than 5000 reads for analysis. This ended with 2039 samples from 66 bioprojects.

Please see the Table 1 in the attached file.

This ended with similar number samples for both region and major habitats (snow, ice, cryoconite, and supraglacial meltwater).

**Comment 4**

Additionally, there is a need for accurate and comprehensive citation of all data sources in Table S1. Currently, it seems that only three references from the authors of this paper are cited. Ensuring that all data sources are correctly and comprehensively cited is crucial for maintaining the integrity of the research and providing clear, traceable scientific evidence.

*Response*

We have generated a comprehensive data source for all samples, which includes longitude, latitude, geographic location, habitats, project ID, and sample ID as a supplementary table.

Please see Table S1 and S2 in the attached files.

**Reviewer 3**

We would like to express our sincere gratitude to you for the immense dedication and time invested in this article, as well as for providing many valuable suggestions, which have greatly contributed to the quality of the manuscript.

**Comment 1**

Formatting:

Some headers contain letters that aren't bolded.

*Response:*

We have checked the format of the entire manuscript, and ensured that all text has been formatted consistently

**Comment 2**

Table S2 has cited data from Trivedi et al. in the future.

*Response*

We are sorry for the mistake due to the misused auto-filling function of Excel. It has been corrected as Trivedi et al., 2020.

**Comment 3**

Figure 1; The Alps are not considered the Arctic. Is the microbial richness calculated with all of the taxonomic identifications from metagenomes, cultures and amplicons?

*Response*

We have regrouped the samples, now the samples were grouped as Antarctic, Arctic, Tibetan Plateau, and other non-Polar glaciers.

**Comment 4**

In sentence, "Metagenomic assemblies were binned using MetaBAT 2 (v2.12.1) (Kang et al., 2019), MaxBin 2 (v2.2.7) (Wu et al., 2016), and VAMB (v2.0.1) (Nissen et al., 2021) respectively" respectively should be replaced with separately.

*Response*

We have changed the sentence as the reviewer suggested.

*Amended manuscript*

Metagenomic assemblies were binned using MetaBAT 2 (v2.12.1) (Kang et al., 2019), MaxBin 2 (v2.2.7) (Wu et al., 2016), and VAMB (v2.0.1) (Nissen et al., 2021) separately.

**Comment 5**

In "These MAG together with the downloaded isolate genomes were dereplicated using the thresholds of 30% aligned fraction and a genome-wide average nucleotide identity (ANI) threshold of 95% …" makes it sound like the MAGs and the cultured genomes

were dereplicated against each other instead of the method being applied to both sets of data, which I think is what is supposed to be conveyed.

*Response*

We have rephrased the sentence for clarity.

*Amended manuscript*

The obtained MAGs were combined with the isolate genomes, and these genomes were dereplicated using the thresholds of 10% aligned fraction and a genome-wide average nucleotide identity (ANI) threshold of 95%.

**Comment 6**

In "Spatially, 69.7% of all samples (n = 568) were from Tibetan glaciers, 24% (n = 196) were from Antarctic glaciers, while those from Arctic glaciers were slightly under-represented (6.3%, n = 51)." The arctic being slightly under-represented in an under statement.

*Response*

We have re-run the SRA entry filtering process and greatly improved the number of amplicon sequencing retrieved from the SRA dataset. Additionally, we have updated our data analysis pipeline, to allow a more samples can be kept after quality filtering. Please see attached Table 1 for details.

Thus, the Arctic data are no longer underrepresented

**Comment 7**

A main table displaying the number of samples in the cryoconite (sediment and water), ice and snow of the Arctic, Antarctica and Tibetan Plateau would help understanding.

*Response*

We have added a table in the main text to clarify the number of samples in the dataset. Please see attached Table 1 for details.

**Comment 8**

These sentences could be combined and refined for clarity. "We identified ubiquitous phylotypes for each region-habitat pair (i.e., identified in more than 55% of samples). There were five phylotypes identified as ubiquitous in all region-habitat pairs (Table S8), affiliated with Gammaproteobacteria (Comamonadaceae) or Actinobacteria (Microbacteriaceae).

*Response*

We rephrased the sentence according to the new results.

*Amended manuscript*

We then attempted to identify ubiquitous phylotypes for each region-habitat pair, defined as those present in more than 55% of samples with a relative abundance greater than 0.1%. However, we found no phylotypes that were common across any

regions or habitats (**Table S11**). The number of ubiquitous phylotypes varied, ranging from nine in Arctic ice to 129 in Tibetan-supraglacial meltwater. Notably, we did not identify any ubiquitous phylotypes in Arctic-snow or other non-polar-snow. Most of the ubiquitous phylotypes were associated with Gammaproteobacteria (29% of the total), followed by Bacteroidetes (19%), Alphaproteobacteria (16%), Cyanobacteria (13%), and Actinobacteria (11%). This distribution may indicate their ability to disperse and adapt to different environmental conditions.

**Comment 9**

Section 3.1.4 on Potential pathogens. Details are needed on what constitutes a pathogenic organism that made up this curated database.

*Response*

We have removed the potential pathogen sections due to the read length of amplicon sequencing does not meet the requirement for species-level classification

**Comment 10**

Line 195 – 196, "We propose that this could be explained by the similar selection mechanisms for long-distance dispersal survival and host-immune evasion" Is a bold claim without reporting how many of these phylotypes have the genes for the teichoic acid or reporting how many of these snow and ice pathogens were staphylococcus.

*Response*

The pathogen prediction part has been removed.

**Comment 11**

Line 203 "Of these dereplicated ORFs, 47.8% (29,947,128) were functional annotations using eggnog" should be 'functionally annotated'.

*Response*

We have corrected the sentence

*Amended manuscript*

Of these dereplicated ORFs, 47.8% (29,947,128) were functionally annotated using eggNOG.

**Comment 12**

Line 212, "…likely representing novel species" this cannot be claimed from gOTUs.

*Response*

We have rephrased the sentence as "Notably, 87% of the gOTUs were unable to be classified at the species level (**Fig. S5**), reflecting the genomic novelty of glacial microbiome."

**Comment 13**

Line 213-214. This is the first time South American glaciers are being considered alone.

*Response*

We have now classified as South American glaciers as other non-polar glaciers.

**Comment 14**

Figure 4b is difficult to read and discern and determine what habitat they come from. 4d would be improved by adding the locations were each of these gOTUs are present and if they relate to the organisms (phyla level) to the ones in figure 2b. 4e is written as 4f in the caption, and is difficult to apply these virulence factors to the

*Response*

We have modified the Fig. 4B, so that the number of genes from each habitats are displayed. For Fig. 4d, it is difficult to separate the genes by habitat as most of these genes are recovered from cultured isolated, thus the habitat information is not available. For Fig. 4e, we have added the taxonomy information for each gene.

Please see attached Fig. 4 for details.

**Comment 15**

Line 230 "relatively rare" add the percentage that these nitrogen fixation and nitrification genes make up.

*Response*

We have added the percentage of these genes relative to all nitrogen cycling genes identified

*Amended manuscript*

In comparison, genes involved in nitrogen fixation (*nifH*) and nitrification (*hao*) were relatively rare, with only 678 (0.49% of the nitrogen cycling-related genes) and 240 (0.17%) unique genes identified, respectively.

**Comment 16**

The Antimicrobial resistance genes section is fascinating and a great addition to the study.

*Response*

We thank you for acknowledging the values of this result

**Comment 17**

Line 268; The 4GDB link in inaccessible.

*Response*

We have changed the provider of the website, with a new address of https://nmdc.cn/4gdb/, the website is fully functional.

**Reviewer 4**

This manuscript proposes a dataset for supraglacial prokaryotic communities from glaciers in the Arctic, Antarctica, and from the Tibetan Plateau. It comprises amplicon sequence data, cultured bacterial genome data, and metagenomic data from the three main supraglacial habitats: snow, ice and cryoconite holes. The authors' research reveals a higher prokaryotic diversity on glaciers from Tibet in comparison to Arctic and Antarctic glaciers. Furthermore, the study of potential pathogens could help identify potential biohazards in glacial communities. This dataset is the first step in offering a study setting for a worldwide view of the prokaryotic community composition for supraglacial environments and covers microbial diversity, taxonomy, community structure, and genetic functions in glacial environments.

*Response*

Thank you for your thoughtful and constructive feedback on our manuscript, we appreciate the time and effort you dedicated to reviewing our work. We are grateful for your positive comments highlighting the significance of our dataset and its potential to provide a worldwide perspective on the prokaryotic community composition in supraglacial environments. We have responded your comments in a point-by-point style, please see our response below.

**Comment 1**

However, major concerns can be raised regarding the completeness, comparability and exploitation of the dataset. Among them, the lack of proper sourcing and credit attribution for data used but not produced by the authors. Table S1 should be completed replacing "NCBI SRA database" by proper credit when possible, including Millar et al., 2021, Webster-Brown et al., 2015, and others. Some samples, such as s2016ZPGSN, cannot be found on NCBI.

*Response*

We have added the manuscript DOI and author names wherever possible, otherwise they are cited as "unpublished data".

Please see table S1 and S2 in the attached file for details.

**Comment 2**

Driving conclusions on the bacterial composition and diversity from glacier samples all over the world, comparing studies performed over several years, seems risky without addressing and assessing first key points such as the use of different DNA extraction pathways, the evolution of sequencing techniques and the depth of sequencing they offer, and the different sequencing primer pairs used. The lack of proper credit to the authors and producers of the data retrieved from NCBI complicates further the access of the reader to such discrepancies in the assembled dataset.

We agree that different primers, sequencing platform, depth, and strategy will influence the sequencing results. We have provided this information in the supplementary table. Please see attached Table S2 for details.

**Comment 3**

Furthermore, ~70% of the samples presented in this study come from the Tibetan plateau, inducing a considerable bias in the estimation of the bacterial diversity worldwide. PERMANOVA tests should be paired with beta dispersion tests to start addressing this issue. For a report on the three Poles, the under-representation in Arctic samples is concerning for the interpretation of this study's results.

*Response*

We have revised the samples included in the study, the sample number bias is now minimized. We also performed the permdisp analysis, significant differences in the distance to the centroid were found among most regions (except between Non-polar and Tibetan glaciers). Therefore, within group variations significantly influenced the between group community similarity.

*Amended manuscript*

PERMDISP analysis showed that snow samples exhibited a significantly (at the threshold of $P = 0.05$) higher dispersal from centroid (67.4%) compared to cryoconite (65.1%). Comparatively, those of ice core and algae were lower (57.9% and 58.9%, respectively), possibly due to the low sample numbers.

**Comment 4**

L98: Could you provide rarefaction curves justifying the use of this threshold?

*Response*

We have removed the pathogen prediction part, as this method could be unreliable.

**Comment 5**

L148-150: a higher diversity found in Tibetan cryoconite holes compared to snow and ice could be due to a sampling bias as well (which proportion of the Tibetan samples are from cryoconite holes?), which should be addressed.

*Response*

We agree that sampling bias heavily impact our results. In the revised manuscript, we refined the dataset included in the dataset and revised the analysis methods. The results dataset is much more balanced in sample number.

Please see table S1 in the attached files for details.

Minor revisions:

**Comment 6**

L15-16: as the authors study the prokaryotic community, it would be good to stick with the terms "bacterial and archaeal" or "prokaryotic" instead of "microbial" over the course of this manuscript.

*Response*

We have replaced microbial to "prokaryotic" where appropriate.

**Comment 7**

L22: "which could be attributed to the similar adaptation mechanisms for microbial survival in aerosol and immune evasion" this seems like a far-fetched conclusion.

*Response*

We have removed the pathogen prediction results, due to its inaccuracy.

**Comment 8**

L35-36: to reformulate. 6 petagrams (Pg) is the total organic carbon contained in glacier ice, and not all is released during the yearly seasonal melting.

*Response*

We have reformatted the sentence as "It has been estimated that six Pg of carbon are stored in global glaciers. These carbon may be released into downstream ecosystems with glacier runoff (Hood et al., 2015), influencing key elemental cycling in downstream ecosystems."

**Comment 9**

L41-43: grammatical changes needed

*Response*

We have rephrased the sentence as

"As microorganisms are the key driver of carbon and nitrogen transformation in glacier ecosystems, knowledge of their biogeography and functions can greatly enhance our understanding of the biogeochemical cycling in glacial ecosystems and aid in predicting the impact of climate change."

**Comment 10**

L49-50: citation needed

*Response*

We added appropriate citation for this statement

*Amended manuscript*

Algae and Cyanobacteria are the primary producers in supraglacial ecosystems, with other heterotrophic microorganisms participating in the transformation and degradation of endogenous and exogenous nutrients (Hotaling et al., 2017, Anesio et al., 2017).

**Comment 11**

L55-56: what is the justification for the mean microbial abundance in surface meltwater to increase with enhanced glacier retreat?

*Response*

We have added additional citation for this statement. In Segawa et al., 2005, the authors investigated the season variations in microbial biomass and found that the cell number of bacteria significantly increased during the melting season.

*Amended manuscript*

It was estimated that the mean microbial abundance in glacier surface meltwater is $10^4$ cells mL$^{-1}$ (Stevens et al., 2022), this quantity may further increase with the enhanced climate warming (Segawa et al., 2005).

**Comment 12**

L57-58: "which are not commonly monitored in the environment but have the potential to enter the environment" this needs enhanced clarity

*Response*

We have removed this sentence, as this is clearly redundant and only causes ambiguity.

**Comment 13**

L55-61: this part would benefit from a re-writing linking clearly the different findings to each other, to knowledge gaps and to the authors' hypotheses.

*Response*

We thank you for the kind suggestion. Nevertheless, this work presents a collection of data, and it is not aimed to answer any specific question(s). Therefore, we did not provide any hypotheses for this work. The section you mention is to briefly introduce what this dataset could be used for.

**Comment 14**

L69-70: archiving biological data from endangered environments such as glacier ecosystems is invaluable. However (and unfortunately), qualifying it of method allowing for the preservation of biodiversity is highly debatable.

*Response*

We have revised the sentences without stressing that our work could preserve biodiversity.

*Amended manuscript*

The dataset archives glacial-specific microorganisms and unique genes in digital form, thus representing an invaluable resource for bioprospecting.

**Comment 15**

L267: the link provided is not accessible

*Response*

We have changed the provider of the website, with a new address of https://nmdc.cn/4gdb/, the website is fully functional.