nature portfolio

Peer Review File

Diversity and biogeography of the glacier-fed stream bacterial microbiome

Corresponding Author: Professor Tom Battin

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
Towards a global biogeography of the glacier-fed stream benthic microbiome
Ezzat et al.

The paper presents a unique data base on world-wide microbial characteristics and diversity in glacier-fed streams and gives very interesting and valuable insights. My expertise lies more in glaciers and glacier runoff, and I thus comment mainly on these aspects.

In sum, I find the glacier-related aspects of the contribution a bit surficial, acknowledging though that the main messages and focus of the paper are for the most part not depending on these aspects. Still, I think the paper would benefit from a more careful and clear treatment and discussion of glacier-related aspects, as detailed in the following. I number my comments in random sequence, not according to priority, just to ease response.

(1) The first sentence (Line 26) might be a bit misleading. The 60% referred to depend much on the climate scenario applied. Also, these '60%' refer to a sample (i.e. global distribution of glaciers) that is likely not similar to the representativeness of the GFS sampled in the current study. Many glaciers are located in regions that are not sampled, e.g. in the Arctic. Also, in particular Arctic glaciers often have no GFS, they are calving directly into the ocean. In reality, the numbers of glaciers projected to disappear that better represent your GFS sample might actually be significantly higher. I recommend to connect the present GFS study and the current (see below reference Hugonnet et al 2021) and projected global glacier change more specific, not just on a gross global level.

(2) I was surprised to not read about the concept of 'peak water', at least in the discussion of the paper (e.g. DOI:10.1038/s41558-017-0049-x ;Huss and Hock, 2018). I.e. the stream-flow below glaciers does not at all linearly respond to glacier size changes. That means glacier area or glaciation % as used here are not necessary precise indicators of the glacial-hydrological conditions of a catchment. This leads to my next comment:

(3) Glacier climatology and thus run-off characteristics can vary a lot in space and time. Could part of the microbial diversity also reflect climatological effects, less than ecological/microbial? Or, how would such glacier-climatological and glacier-hydrological effects impact on your statistics? For instance, there are large differences in run-off and potential ecological importance of glacier melt between glaciers in very arid regions (critical dry-season run-off) or monsoon-type glaciers (relatively unimportant glacier runoff compared to rain). Such differences can actually happen over comparable small distances, e.g. southern and northern side of the Himalayan ridge. Similarly, the ratio between glacier-melt and rain contribution and its seasonality could be an important constraint for GFS microbial communities. Further, about your elevation parameter: Is it normalized, e.g. with latitude? Else, it is rather a climatological parameter. Or is there some other effect expected for micobes from absolute elevation? (BTW, extented data figure 6: what is the elevation anomaly? Against what?). A good overview of measured (not modelled as in your reference 1) global variation of glacier mass responses is given by DOI: 10.1038/s41586-021-03436-z (Hugonnet et al. 2021)

(4) Line 85. Through glacier changes, e.g. increased run-off until 'peak water' and deglaciation of unconsolidated sediments, many GFS are actually exposed to much more severe impacts as listed here, for instance debris flows, which I guess will have severe impacts on the microbial communities.

(5) Paragraph following L 289. See above (3-4) on the environmental effects beyond glacier size and %.

(6) I could imagine that the glacier type, even within a region, could play an important role for microbes, beyond glacier and runoff climatology (3). E.g. debris-covered glacier, glacier with frontal lake etc.

(7) L 371-373. As commented above under (1) and (2), glacier change is quite diverse and not just a linear decline with warming atmosphere. So, the statement about ASV diversity and evolutionary lines being at risk should be formulated more careful and precise.

(8) L 375: glacier shrinkage is actually in principle not irreversible. Colder/wetter climate would within a few years to decades be able to restore a glacier. For this reason there are actually glaciers growing in a few regions. (How realistic it is that mankind will manage to reach a climate trajectory where glaciers can be stable or even recover, is a different topic, though).

(9) In sum, the final paragraph is quite surficial and speculative (see e.g. (7) and (8)).
(10) Extended data Fig 2a: What is the glacier area parameter? Average glacier size of the GFS investigated? Else? Same for the glazierisation %. Average? Else?

End

Referee #2

(Remarks to the Author)
The manuscript of Ezzat et al reports 16S rRNA gene sequencing from a relatively high number of glacier fed stream ecosystems, and attempts to correlate these data with geographic and geochemical data to draw conclusions regarding global biogeographic patterns of diversity. The authors targeted specifically the benthic component of the streams. This is an important topic because, as the authors point out, glaciers are melting at an anomalously fast rate and their disappearance will have important implications for the ecology of the cryosphere. The conclusions regarding microdiversity are quite similar to a recent study published by many of the same authors, and therefore not necessarily a novel result (https://doi.org/10.1038/s41396-021-01106-6). What is new here is the high number of samples processed from a global dataset, and it is impressive and I think with the correct controls it could be a very nice study. Unfortunately however, there are fundamental controls that were missing (see below) that in my opinion preclude publication.

The authors main conclusions are that (1) a large microbiome fraction is endemic to single mountain ranges, and (2) a very small portion is disproportionately abundant and forms a cosmopolitan core. Because of the extreme dispersal potential of microbes, such questions regarding biogeography are very difficult questions to answer in microbial ecology, and require careful controls. Unfortunately, there were many controls for contamination and checking ASV false positive rates (using mock communities) that were not done and the absence of these controls prevents the conclusions regarding core communities. Moreover, several taxa of their "core community" are well known as ubiquitous contaminants in molecular biology reagent kits (Salter et al 2014). This is concerning and the authors need to sequence extraction blanks to make sure that those taxa (e.g., Polaromonas) are indeed coming from the samples and not coming from the kits. The authors use a PCR primer combination that is fairly unconventional (why did they not use the Earth Microbiome Project primers?) and do not provide any information on whether their protocol was calibrated using mock communities and extraction blanks for contaminations. Some samples had a very low disproportionate sequencing depth compared to others, and this precludes conclusions regarding endemic taxa (they could be in the other samples just not detected due to a low sequencing depth.

Here are some more specific comments regarding my concerns on this matter:

line 123-126: The "rare biosphere" (doi: 10.1073/pnas.0605127103) is a controversial idea in microbial ecology that has been criticized (doi: 10.1111/j.1462-2920.2009.02051.x) to be biased by ASV clustering algorithms and sequencing errors that lead to high rates of false positive ASVs present at low abundance (https://www.nature.com/articles/nmeth.2604). It was bad with 454 sequencing, and has improved with more accurate illumina sequencing but the problem remains depending on how the ASVs are clustered and the degree of cross sample contamination, non-sample contamination, sample bleeding, and systemic errors in particular ASV picking algorithms. That is why all 16S based methods need to be ground truthed with empirical data from mock communities to figure out which parameters give an ASV richness that is correct and not overestimated. I cannot find any indication in the methods or manuscript that this was done for the authors 16S protocol reported here.

lines 168-169: If you define "endemic" as being present in a specific location, and not present in any others, then in my opinion, it is not possible to test this using 16S amplicon data unless a very high sequencing depth was applied to all samples. But, looking at the methods some samples had as low as 10,800 reads and therefore many ASVs could easily been missed.

lines 190-195: Some of the "core microbiome" is possibly the result of cross-sample contamination if all the samples were extracted and PCR amplified in the same lab. Particularly if you are including ASVs that are as lows 0.1% of total reads. Did the authors perform checks for cross sample contamination? Or laboratory, or kit contamination? I don't see any indication in the methods.

lines 199-204: Have you checked your data against contamination from the "kitome"? e.g., Salter et al 2014. Many of the group you write here are well represented as ubiquitous contaminants in commonly used molecular biology kits and DNA extraction kits.

line 253: Polaromonas is one of the most commonly found contaminants of molecular biology kits (Salter et al., 2014 BMC Biology). The fact that this is found to be an abundant cosmopolitan taxa of the "core community" is concerning, the authors need to sequence extraction blanks and make sure that polaromonas ASVs are not a contaminant of their PCR and DNA extraction reagents.

lines 286-287: Or perhaps it is an indication that there is some noise in the data (contamination, or ASV false positives?).

General comment on the 16S amplicon methods: Have the authors used control samples to determine the extent of kit or laboratory contamination (see Salter et al 2014), sample bleeding, contamination, ASV false positive rate, or cross sample contamination during library preparation? What is the false positive rate of ASVs using this approach? Many of the ASV clustering algorithms over estimate microbial richness, and the ASV false positive rate needs to be determined empirically using a mock community (https://www.nature.com/articles/nmeth.2604) for any 16S sequencing and analysis protocol. The authors do not report on this, or describe any of these controls, therefore I have to assume that they were not done.

lines 473-475: Why did you use these primers, as opposed to the more conventional 515F/806R primers that are used by the Earth Microbiome Project (https://earthmicrobiome.org/protocols-and-standards/16s/) ? Have these 341f/785r primers been tested on a mock community by the authors and the ASV cutoff empirically determined?

lines 496-498: What was the identity cutoff used to classify ASVs? 97% , 98%, or 99% ?

line 507: It seems like a stretch to make conclusions about endemicity for samples that had as low as 10812 reads. With such low sequencing depth many ASVs at lower relative abundance could easily have been missed in the sequencing libraries. The authors allude to this , indicating they themselves are aware of the problem, on lines 157-159. Yet, they still draw very strong conclusions about endemicity in the manuscript. In my opinion such strong conclusions about endemicity are not support by the authors data for reasons described above.

line 508: By "at least 50%" do you mean that the ASV had to be detected in at least 2 out of 3 replicates?

Referee #3

(Remarks to the Author)
Climate warming has caused significant loss of the world's glaciers, with great impacts on glacier-fed streams (GFS). However, our understanding of the biodiversity and biogeography on the biofilm microbiome of GFS is limited. Thus, the authors collected samples from about 150 GFS from world's major mountain ranges and determined bacterial distribution and biogeography using 16S amplicon-sequencing. Their results suggested that the global GFS benthic microbiome is quite different from other cryospheric microbiomes with very high diversity. About 60% of the total amplicon sequence variants are endemic, with very small core (0.4%). Further analysis indicated that the community diversity and changes are mainly shaped by dispersal limitation and environmental selection. Although the ecosystem examined is interesting, I have several major concerns on the novelty, significance, and conclusions of this study. First, it appears that the experimental results on high microbial diversity in this system just represent incremental improvements compared to previous studies by Wilhelm et al (2013, ISME J) and Ezzat et al (2022, AEM), and are not novel enough to warrant a publication in Nature, which generally publishes groundbreaking through, transformative studies in various fields. Second, the significance of the experimental results is not clear. The importance of dispersal limitation and environmental selection in shaping the community diversity and distribution is kind of expected given the large differences of continental level geographic locations and environmental conditions. Third, there are serious flaws on methodologies and analyses (see below for details) and thus the results/conclusions are not reliable. In addition, the experimental analyses are preliminary, just based on 16S amplicon sequencing. No functional structure differences among these systems are presented. Other omics approaches such as shotgun sequencing, genechip, metabolomics beyond compositional profiling to assess the changes of functional traits could be considered. Moreover, it is not clear whether the community composition changes affect the functions of the glacier ecosystems, which is a critical question in ecology and climate change biology. There is no attempt to link community structure to the functions of these ecosystems.

I have various more detailed questions on the methodologies. First, the experimental design and sampling is problematic. The authors collected three samples from each GFS with distance of average 300m as ecological replicates. Because of spatial heterogeneity, many more biological replicates are needed to represent the microbial community diversity of the GFS examined. Although 148 GFS were selected, only 418 sequencing libraries were constructed, which ended up only 1 or 2 libraries for some GFS. Rigorously speaking, the authors should spend more efforts in constructing sequencing libraries for all GFS with equal amount of sampling efforts, which is critical to the subsequent endemic analysis. Second, the sequence preprocessing is also problematic. The authors used the average of three (two or one) ecological replicates to represent the diversity of each GFS. Since GFS is a big area, it will make more sense to use them as independent biological replicates (rather than as technical replicates) so that rigorous statistical tests can be performed to assess whether GFS X is significantly different from GFS Y. Third, Data2 was used to preprocess the sequencing data. Since Data2 tends to remove

many true ASVs, it is advisable to also use other sequencing preprocessing approaches such as dBlur, Unoise, and even UParse. Although the exact results obtained with different approaches will be different, the general trends should be consistent for the same datasets with different preprocessing methods. It is important to ensure that the experimental results are not dependent on the methods used for sequence preprocessing. Fourth, defining endemism is very tricky and challenging. The conclusion that ~60% ASVs are endemic is problematic. An ASV which is not detected in other samples could be due to many other reasons beyond their “extinction” in these samples. One could be because sequencing depth is far not enough. Some OTUs which are not detected could be primarily due to shallow sequencing and/or random sampling processes. Thus, to determine “true” endemism, much deeper sequencing (several magnitudes higher) is needed to ensure that an ASVs is present only in one GFS but not in other GFSs. Spatial heterogeneity will also have significant effects on the detection of endemic ASVs. If more samples are analyzed from each GFS, less endemism would be expected. Another reason could be due to the sequence preprocessing method (Data2) used. Data2 tends to remove many true ASVs, and rarefaction curve is easier to be saturated. I guess that the degree of the endemism will be greatly affected by the sequencing preprocessing methods. Fifth, the authors estimated the community assembly processes for each mountain range. However, the authors used an out of dated approach to determine community assembly processes, which tends to overestimate selection and underestimate “drift”. The recently developed more advanced approaches could be considered (Ning et al. 2020, Nature Communication, 11:4717) , Sixth, the estimated drift is very low for most of the mountain ranges (undetermined in Extended Fig. 11), which is contradictory to db-RDAs and VPA analyses that showed at least 52% community variations could not be explained. Theoretically, the unexplained variations are generally due to unmeasured environmental variables and ecological drift. In addition, the authors discussed the importance of horizontal gene transfer in shaping community structure. This is highly speculative because no experimental data are presented. To address this type of questions, as mentioned above, functional traits-based approaches are needed to determine the diversity and distributions of microbial functional genes.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
My expertise lies more in glaciers and glacier runoff, and I thus comment mainly on these aspects.

My concerns regarding the initial version of the manuscript have been addressed to a large extent and the glacier-related aspects of the paper have improved.

For the main text there remains one question in my view. The authors agree with me on the highly dynamic nature of the glacial environment (daily + seasonal runoff, runoff peaks/floods, debris flows, sediment reworking, glacial action, etc.). To what extent can the (cumulative) history of such actions have influenced the microbial communities found? For instance, the timing of glacier retreat from the sample location. How long is the sample location free of glacier cover? Is there are history of heavy stream reworking, not necessarily easily visible on site? Etc. I assume such (cumulative) history can be very different from site to site. I recommend a sentence or two discussing the potential influence of such environmental history on the study results, or why it is expected to not play a role, resp.

Supp. Inf.:

There is a parameter "permafrost" in several tables that I don't find explained. In the main text it is referred to as "permafrost soil". How is it defined? From which data? What would be its influence on microbial communities? You typically don't have permafrost under glaciers.

Tab2: what refers "glacier" to? Area? %? Similar for snow. I find Tab2 and parameters used too little explained.

Referee #2

(Remarks to the Author)
The authors have done a lot more work in attempts to improve their manuscript, which includes new mock community analysis, analysis of blanks, and resequencing of many samples. This extra effort is appreciated, and I recognize the hard work that they have put in to make the manuscript better. However, there unfortunately still remain points that need to be clarified from the analysis of the blanks and the mock community which I provide below. I know this may seem like a moving target, but I assure the authors it is not. Rather, I am encouraging the authors to take necessary steps to ensure that the data are of highest quality (presumably the authors also want this as well). As a global dataset, I think it meets the high mark required for Nature. But, I do not think that the authors claims for endemicity can be proven with their data - and therefore in

its current form I cannot support the manuscript as it stands for publication. If the authors are willing to change their stance on this, I would be more positive. I think it comes down to what the word "endemic" means, which I discuss in more detail below.

Here is why the authors data do not support conclusions about endemic taxa: the authors write on lines 150-151 that their analysis predicts that they only sampled 73.6% of the total ASVs present in the mountain ranges. That means that about 25% of the taxa were not detected. It is therefore entirely possible that some ASVs are not actually endemic to a specific geographic region as the authors claim, but rather simply not detected in other geographically separated mountain ranges (because they are part of the undetected 25%). Surely the authors know this, and would agree with this point. I suggest to remove all claims of endemicity from the manuscript, because this is not possible to prove unless all taxa have been found in all samples (not the case here). If the authors publish this dataset in Nature, and make claims about endemicity when only 75% of ASVs are detected, this will surely result in a critical response from the scientific community. The authors define "endemic ASV" on lines 231-232 as "being specific to a single mountain range". I suggest changing the word "endemic" to "specific" throughout the entire manuscript, that might be an easy way to soften these strong conclusions.

Related to the endemicity topic, on lines 253-260 (and Figure 3a) If you only sampled each GFS 3 times (and some even less), I don't think you can statistically show that ASVs detected in only one GFS is truly specific to that GFS (more sampling is needed within the GFSs to show this).

I disagree with Reviewer 3 that the results are incremental, to my knowledge this is the first global dataset on GFSs with data spanning all of the major glacier covered mountain ranges on all continents. As a global dataset, I think it meets the high mark required for Nature. I agree with Reviewer 3 that functional, or ecological information is the end goal of any microbial ecology study, and I think that the added metagenomic data in figure 1 strengthen the study even more. I agree with Reviewer 3 original comments, that the original methodologies were standing on weak ground. However, the authors have reanalyzed their data with multiple pipelines and apparently obtained similar results. I don't think the main results are a spurious result of clustering algorithms. There are still some things to check however with the mock community and inadequate sequencing depth of the blanks (see below for details).

Where I agree most strongly with Reviewer 3, is in our shared critical opinion on the authors use of the word "endemism". I agree with Reviewer 3 that the GFS sampling could have been better (3 samples per GFS is a bit low and some GFSs had less than 3 samples). This is another reason, that precludes the authors from drawing conclusions about endemicity (maybe if you took more samples, you would have found the 'endemic' taxa in other mountain ranges) - I think Reviewer 3 and I are in agreement on that. However, I get the impression that the authors have a different understanding of the word "endemic" than I do (and apparently Reviewer 3 as well). Their results show that many GFS ASVs are highly specific to certain geographic areas, and it might be a matter of them just changing the word "endemic" to "specific" (or something similar). The authors cannot prove with their data that ASVs are not present in different mountain ranges because they only detected on average 75% of the ASVs (lines 150-151). The authors must know this and I think should therefore agree.

Here I provide my comments on the new results of the mock community and the blanks:

Comments on the mock community results:
It is good that the authors finally included a mock community here, in Figure S13. However, the authors need to show the total number of ASVs detected using the different algorithms compared to the true number of species in the mock community. Does the number of species in the mock community equal the number of ASVs after sequencing ? This is a critical point the authors need to address. The stacked histogram they presented does not show this.

The authors say that in the mock community they "did not find any signature of contamination" what does this mean exactly? Are the authors claiming that no other ASVs other than the mock community were detected? This is very hard to believe, since contaminating ASVs in mock communities are always detected at low relative abundance in sequencing runs (see papers by Robert Edgar).

Comments on the blank results:
It is a good start, that the authors included blanks to check for contaminants. However, they state in the rebuttal that they only obtained an average of 486 reads per blanks. This is very, very low, (the authors must know this) and the blanks need to be resequenced to a depth at least similar to their samples before it can be compared. Its entirely possible if the blanks are sequenced to the same depth as the samples (>20,000 reads) then more ASVs are found. The current extremely low sequencing depth of the blanks makes it impossible to reliably compare to the samples.

In the blank sequencing results (SI figure 13), please update the stacked histogram plot currently with the "other" label by showing all the taxonomic groups that were detected. In their rebuttal the authors state they only found 43 different ASVs. These should be possible to group into the different genera as colors and plot. Readers deserve to know what was found in each blank sequencing replicate and what the "others" actually are (Polaromonas for example). I am very curious what is in the "others" particularly since some of the blank samples have almost 100% "others", and the authors write in the rebuttal that they found one Polaromonas ASV but don't label it in this figure but rather group it with "others". This is a bit conspicuous.

I understand, that it may be a possibility (but the authors didn't point it out) that the blanks are so clean, as to only produce a faint band on a gel resulting in the prepared libraries being extremely low concentration and did not sequence well. That would explain the extremely low sequencing depth for the blanks. Can the authors explain this, or at least show a picture of the gel with the blank compared to the samples?

More general comments on the role of (high) microbial biomass for the interpretation of the authors results. I think this is an important issue that the authors should discuss in more detail in regards to interpreting their data.

lines 113-114: I disagree with this statement. The biomass of the GFS (10^5 - 10^6) that is shown in Table S1 is actually quite high in my opinion, because it is comparable to seawater where you have about 1 million cells per mL. Compared to more extreme sedimentary environments like the deep subseafloor biosphere (where cell counts are as low as 10^2 per gram, (Kallmeyer et al., 2012 PNAS; Inagaki et al., 2015 Science), the GFSs cell counts the authors show in table S1 are very high. I would therefore not classify these GFSs as "low biomass". These high cell counts actually support the authenticity of the authors 16S data, because at these high cell concentrations contamination should not be a major issue. Indeed, the authors refer to stream "biofilms" throughout the manuscript. You cannot have a biofilm, and low biomass. Biofilms are by definition high biomass. The authors need to correct this in the manuscript.

lines 519-521: Is this the first time you used this sonication and flow cytometry approach to count bacterial cells from sediments? If you've done it before please cite the protocol here. I would expect that it underestimates cell counts by at least an order of magnitude since not all cells will become detached from the sediment grains. It would be good to at least discuss this important point. e.g., that the sampled ecosystems are actually not that low biomass, 10^6 per gram is quite high in my opinion (its similar to what you have in seawater). This is actually a good thing for the authors, and makes concerns regarding contamination much less of an issue. So, its a point that will strengthen the results quite substantially, and the authors should point this out in the manuscript.

Some additional comments on the new meta genomic data:

lines 128-129: I would keep the ordination in Figure 1 showing the meta genomic data differences between the samples, but get rid of the heat map that only has two columns. Its misleading and gives the impression only two meta genomes are being compared (but in reality its many many more in the ordination).

lines 129-136: This is all very speculative and a very general statement based on gene annotation data from meta genomes (not activity measurements were made). I suggest removing it, since its not needed to make the points. Rather, just keep the ordination showing that the GFS microbiome has a completely different functional profile compared to the other cryosphere environments and let the 16S tell the rest of the story about vanishing "specific" GFS microbiomes.

Referee #3

(Remarks to the Author)
I appreciated the author’s efforts with very long detailed responses to my previous comments. However, the authors failed to address most of my major concerns despite long rebuttal letter. In the following, I will describe them in details.

1. Novelty of this study

In the rebuttal letter, the authors listed five points for novelty of this study in terms of sampling, analyses, and conclusions. It seems that the authors misunderstood my previous comments. I have no doubt that some new information is presented. Actually, new knowledge or information has to present for any publications. Otherwise, there is no point to publish. But, the key question is whether this study is novel enough to warrant a publication in Nature, the most prestigious multidisciplinary scientific journal, which publishes top 1-2% groundbreaking, transformative discoveries in various fields. I searched Web of Science using the key words related to global microbial diversity and biogeography. There are hundreds, if not thousands, of publications on these topics. Thus, to my knowledge and standards, this study definitely does not meet the standards by Nature as a transformative groundbreaking study in microbial ecology. Also, a paper on similar ecosystems was published with sequencing approach in Nature Communications in 2022 by some of the same authors (https://www.nature.com/articles/s41467-022-30816-4). This further devalued the novelty and importance of this current study for Nature.

The authors also mentioned Tara Ocean as justification. I do not think it is comparable. The first study of Tara Ocean was published approximately 10 years ago. During last decade, microbiology has been developed so quickly. As time proceeds, the standards become much higher. For instance, 20 years ago, genome sequencing from a pure culture can be published in Nature, Science, Nature sister journals, and other decent journals. Now it can be only published typically as a genome announcement. Similarly, many of this type of community analyses can be published in broad impact multidisciplinary journals such as Nature and Science several years ago, but now only in typical professional journals.

In the rebuttal letter, it seems that the authors misunderstood my previous comment on 16S-based studies. My point is that given the rapid advance in microbial ecology, 16S sequencing is so easy as a routine tool, advanced microbial ecology studies have to be moved beyond 16S sequencing, either with additional omics and other experimental approaches (i.e., isotope), and/or theoretical analyses. Also, the authors cited similar papers published 10 years ago in PNAS and ISME J, and current paper in Ecology and Evolution to justify this study in Nature. Yes, as I mentioned before, I have no doubt on the certain merits of this study, but they do not meet the high standards set up by Nature. As I know, many top professional journals like ISME J now have editorial policy to directly reject 16S-based diversity studies without external review. However, this does not mean that 16S-based approaches are not important. They are parts of the important toolboxes, but the research questions must be beyond diversity per se (either experimental or theoretical) to be justified for publications in prestigious journals, particularly for Nature.

2. Significance of the results

The authors found that the GFS are different from other studies with many unknown diversity, and with biogeographic patterns, this is not surprising. It is kind of expected based on numerous microbial studies from soils, ocean, freshwater, groundwater, wastewater treatment plants, as well as human microbiomes with high throughout omics technologies. More important and interesting questions are whether the diversity and distributions are important to the functions of GFS ecosystems. It seems that no efforts were attempted to explore the importance of the microbial diversity to the functions of the GFS ecosystems.

As I mentioned previously, the importance of dispersal limitation and environmental selection in shaping the community diversity and distribution is also expected and not surprising at all from theoretical perspectives. Because the samples are from different continentals, dispersal has to be limited. Since environmental conditions are quite different across samples, environmental filtering has to play roles. Numerous global studies on various habitats have supported such theoretical predictions. From this angle, this study does not provide new transformative data/information/thoughts as the authors claimed in the rebuttal letter.

In addition, in the rebuttal letter, the authors highlighted the importance of the results related to endemism and functional redundancy. Honestly, I still have serious concerns on this type of analyses and robustness of the conclusions as before (see below). Even if we assume that the analyses on endemism are reliable, do the results represent groundbreaking, transformative discoveries in microbial ecology? The answer is not. Given extremely high microbial diversity, continental scale, and dramatic differences in environmental conditions, high degree of endemism is theoretically expected, which are also demonstrated by numerous previous studies in various habitats with soils, groundwater, freshwater, wastewater, and so on. (e.g., doi: 10.1002/mbo3.644, doi.org/10.1186/s40168-023-01572-4, doi.org/10.1073/pnas.1919139117, doi: 10.1038/ncomms12083, doi.org/10.1038/s41564-019-0426-5, doi: 10.1128/mSystems.00803-19, https://doi.org/10.1093/femsec/fiz128). Furthermore, given the extremely high microbial diversity, existence of functional redundancy in the GFS microbial communities and its importance to functional resilience is not surprising at all. It is kind of expected from both ecological theories and other empirical studies.

3. Experimental design problem of this study

The authors argued that 149 GFS were sampled, representing global most comprehensive study. This is an over-statement, and it is not a comprehensive study. Actually, as I mentioned before, this study has some serious flaws in sampling design. It seems that Reviewer 1 had similar concerns. It is not clear whether the 149 GFS samples represent the global diversity of GFS. Given the huge area of GFS, more sites need to be selected to cover the habitat diversity of a given GFS. For instance, the sites selected from Alaska are very close based on the map. Since Arctic is huge area and highly heterogeneous, it is hard to justify that the selected few adjacent sites can fully represent the Arctic GFS. How and what sites are selected are critical to addressing the research questions, particularly related to endemism. I do understand the logistic difficulties to sample more sites, but the data collected are hard to address the type of questions related to “true” endemism. More importantly, given the high spatial heterogeneity, three replicates from each site are far not enough to reflect the microbial community diversity of the GFS examined. In their rebuttal letter, the authors mentioned that they focused on the GFSs as close as possible to the glacier snout to avoid environmental gradient. Yes, this will help minimize the variations among replicates, but jeopardize the sample representation of a particular GFS. Thus, to consider spatial heterogeneity, many more samples (not three in this study) from each GFS should have been collected. In addition, this is just a snapshot study. To determine “true” endemism, temporal sampling of representative sites are needed because ecosystem dynamics are not linear and hence time is a critical factor in controlling microbial diversity and distribution.

4. Methodological problems for quantifying endemism

The author claimed that one novel aspect of this study is high degree of endemism. As I mentioned above, this is also expected because of dispersal limitation due to continental sampling scales, and environmental selection owing to huge differences of environmental conditions. More importantly, quantifying endemicity is very tricky, which depends on sample numbers, sample representation, sequencing approaches, sequencing depth and bioinformatic processing approaches, as well as taxonomic resolutions of the molecular markers used for analysis. Reviewer 2 also had similar concerns, particularly related to contaminations. Beyond contaminations, I would like to highlight some issues again. First, spatial heterogeneity will have significant effects on the detection of endemic ASVs. If more samples are analyzed from each GFS, less endemism would be expected. As I mentioned above, it seems that the 149 GFS examined may not cover the global habitat diversity of GFS, and three replicates from each GFS do not reflect the spatial heterogeneity. Also, microbial community abundances in the GFS ecosystems could be quite dynamic. If time is considered and more spatial samples are analyzed, the picture of

endemism could be quite different. In fact, the experimental designs and the current availability of samples do not allow the authors to address questions related to endemism although it is an interesting topic. Second, as I mentioned previously, the sequencing depth is shallow. Although the authors resequenced some samples and increased sequencing depth to 25,526 sequences per sample for analyses. It is still not clear whether the sequencing depth enough to represent the diversity of the communities examined, which is critical to address the questions related to endemism. Instead of using DATA2, the authors used Deblur to preprocess the data, which is also very conservative. Many sequences could be removed prior to analyses, which could greatly overestimate the estimation of endemism because such removed sequences are in the samples, but they are simply not counted for estimation. Third, sequencing approaches are problematic in terms of reproducibility as demonstrated by numerous studies since last decade (e.g., doi:10.1111/vop.12854, doi:10.1093/bioinformatics/btaa926, doi:10.2306/scienceasia1513-1874.2021.073, doi:10.1186/s40168-020-00998-4, doi:10.1016/j.ebiom.2021.103649, doi:10.1038/nbt.3780, doi:10.1371/journal.pone.0099414, doi: 10.1016/j.mimet.2013.07.015, doi: 10.1016/j.mimet.2013.07.015, doi:10.1128/mBio.02288-14. https://doi.org/10.1371/journal.pone.0043093). For instance less than 40% overlaps were detected among technical replicates (i.e., the same DNA samples are sequenced multiple times) even with very simple communities (doi.org/10.1371/journal.pone.0043093). The low reproducibility is an inherent problem of sequencing approaches due to random sampling processes (doi:10.1128/mBio.00324-13). Although reproducibility is less problematic in addressing questions related to community structure differences, it is critical to determining endemicity because information on presence/absence of OTUs (or ASVs) is needed for estimating endemicity. For instance, some missed ASVs in a sample could be due to low reproducibility problem of the sequencing approaches. They are in the sample, but are simply not sampled. In addition, taxonomic resolution of the molecular markers used is critical to understanding endemism. As an example, if one focuses on the taxonomic resolution at kingdom (e.g. bacteria, archaea, fungi) or phylum (e.g., acidobacteria, proteobacteria) levels, there could be no or very little endemism. It is well known that the sequenced region of 16S gene generally provides the resolution at genus or family levels. Ideally, to define “true” endemism, molecular markers with the taxonomic resolutions at species/strain levels are needed. In a word, given all sampling and methodological problems, I really doubt the reliability and robustness of the conclusions on endemism in GFS ecosystems.

5. Problems related to shotgun sequencing data and functional redundancy

I mentioned that examining functional genes-based traits could be more interesting. The authors provided a very limited shotgun sequencing information in this study. I appreciated such efforts, but I also have some concerns. It seems that the majority of such data will be presented in another manuscript. To be honest, I do not think this helps to strengthen this current study because based on my experiences, unlike other typical microbiological journals such as Applied and Environmental Microbiology, Environmental Microbiology, mSystems, and etc, Nature only publishes original, first hand data for original research papers (could be different from meta-analysis papers). Such data should be merged with this paper to really strengthen this type of studies.

Second, no information on shotgun sequencing methods is provided in M&M except KEGG functional analyses. I could not find the data mentioned in Supplementary Table 8 with this link, doi:10.5281/zenodo.10029840. Thus, I could not judge the appropriateness of the methodologies and the reliability and robustness of the conclusion with respect to functional redundancy.

In the rebuttal letter, the authors indicated that this study “shows signatures of functional redundancy as a response to the GFS environment”. Because functional redundancy is critical to ecological stability, it is a central but challenging topic in ecology. Thus, I am very interested in this part and searched all three files (main text, extended figs and supplementary). Unfortunately, I could find such signatures as the authors mentioned in the rebuttal letter. Based on my own experiences, defining and quantifying functional redundancy are very difficult, particularly for microbial communities. Identifying signatures for functional redundancy is even more difficult in microbial ecology.

In conclusion, the revised manuscript still has serious issues in terms of novelty, significance, experimental design, experimental analyses, data analyses, and data interpretation. More importantly, the experimental design and the data generated by the authors could not allow the authors to address the research questions in mind (e.g. endemism, functional redundancy). The results presented in this study does definitely not represent groundbreaking advance in microbial ecology to warrant a publication in Nature, one of the most two prestigious multidisciplinary scientific journals, along with Science. I have no doubt that this study can be published in other professional journals, but not suitable to Nature.

Version 2:

Reviewer comments:

Referee #1

(Remarks to the Author)
I am satisfied with the responses of the authors to my comments and the respective changes made in the manuscript. As final thought, the authors might want to consider to at least include one sentence that summarizes the discussion that the authors give in their response in the paragraph "This reviewer also mentions the potential for legacy effects from streambed ... rather speculative in nature.". In order to at least roughly indicate that there are potential influences that the study could - well understood - not consider. The fact that the referee wonders might indicate that other readers could well wonder the same.

Referee #2

(Remarks to the Author)
The revised manuscript of Ezzat et al has been substantially improved and I was pleased to read and see that the authors took many of my suggestions into consideration. Namely, there is now a very extensive analysis of mock communities and ASV calibrations and the authors did a good job using their mock communities to pick the correct algorithm for their data (Deblur). I think that their work in this area sets the bar for quality control for future 16S based microbial ecology studies, Im not aware of any study that did a more extensive checking on this than the current study. Moreover, I was furthermore pleased to see that the authors took my comments about the word "endemic" seriously, and have changed the word to "specific" or "unique". That being said, I think that the discussion and introduction should have a stronger connection to the historical debate on microbial biogeography, namely that the authors results may shed new light on the original hypothesis proposed by Bass-Becking in 1934 that "everything is everywhere". This hypothesis has been debated ever since, with some arguing that endemic species may not exist because of limitless dispersal potential of microbes. The citation provided in reference 16 is a good recent review on the topic of microbial biogeography, but I think the last part of the Introduction of this manuscript (and also in the discussion where the results are presented) should provide a little bit more on the historical perspective and long standing debate. For example, here are some papers to familiarize with past examples of the debate

(doi:10.1111/j.1462-2920.2006.01017.x)
(doi: 10.1126/science.1070710)
(doi: 10.1126/science.1070710)

Perhaps, a nice example of the debate is well documented in this article

(DOI: 10.1641/0006-3568(2004)054[0885:RFFAF]2.0.CO;2)

Where those authors stated "We therefore remain unconvinced by the limited evidence for microbial biogeographies"

Its seems this new manuscript by Ezzat et al on global GFS biogeography is a very strong dataset that adds alot of value to this historical debate and therefore I think the authors could improve the discussion a bit by adding it into the context. I know that the main focus is on the vanishing unique cryosphere biomes, but this is an additional aspect of the work that will be highly impactful. While the current manuscript does not prove endemicity, it provides very strong evidence for highly specific communities that have undergone strong selective pressures resulting in very clear patterns of global biogeography. This discovery in itself on a global scale in the cryosphere environment is very novel in my opinion.

In general, I think that this manuscript is vastly improved from the first submission and is well on its way to be published in Nature. It meets the high bar for novelty, impact, and quality in my opinion. Aside from the points above, I have a series of additional comments below where I identified a few incorrect statements in relation to new references that were added. I also point out a few regions of the text and figures that are unclear and provide suggestions for improvement.

Specific comments:

line 70: typo 'fulfil', change to "fulfill".

lines 89-95: This section could benefit from a couple of sentences introducing the nearly 100 year old debate on 'everything is everywhere, but the environment selects' that was originally proposed by Bass-Becking and was reignited decades ago when new molecular techniques began to be applied to microbial communities (see comments above).

line 125: What metric was used to conclude that these pathways "drive" the segregation? Was a statistic used in the ordination for this?

line 135: I think this statement could be made more precise, particularly what "The cryosphere was long thought to be devoid of microbial life...." means. I don't agree with the statement as written, because the existence of microbial life in the cryosphere goes back at least 70 years (https://doi.org/10.2307/2257289). Perhaps, instead of 'long thought to be devoid of life' the authors could rephrase to say the microbial cryosphere is gaining more attention in recent years.

line 145: Do you mean generalizable conclusions about global microbial biogeography?

lines 146-148: This acknowledgement is appreciated, as is the new tone of the manuscript in regards to topics of specificity and endemicity.

line 151: Im not sure if "harshness" is the right word here, to be more specific it is a matter of low primary productivity and organic matter that keeps the microbial biomass low.

lines 157-159: This statement is inaccurate and needs a revision. The concentrations of the GFS cells reported here are actually lower (not higher, as is currently written) than what Ref 25 reports. Ref 25 reports cell concentrations of surface deep sea sediments in the range of 10^7 - 10^9 cells g^{-1}. The authors manuscript reports orders of magnitude lower, around 10^6 - 10^7 cells g^{-1}. What the authors measured in GFSs is however higher than the extreme subseafloor biosphere that exists down to several kilometers below the seafloor where cell concentrations drop to levels 10^5 - 10^2 cells g^{-1}

(https://doi.org/10.1038/s41579-018-0046-8). Can the authors please revise this text for accuracy. I think its important to place the measured microbial biomass into context. Please explicitly state the statistics on concentrations of microbial cells is in the text (in addition to citing Supplemental Table 1), and provide a comparison to other "high" and "low" biomass biomes.

line 212: Could you provide a brief explanation of what Levin's niche breath is, for the reader. Has this ever been applied to microbial communities before, in order to classify specialists and generalists as is depicted in Extended Data 11?

line 223: I suggest using "range specific ASVs", to make it clear which ASVs are specific to the ranges as opposed to those specific to GFSs.

lines 226-231: I suggest to state here that these calibrations were all made using comparisons to mock communities and therefore the decision to use Deblur is very robust.

line 238-239: This is very interesting indeed!

lines 243-246: This statement is appreciated, and provides a much more balanced view and interpretation of the results in my opinion. The dataset provides a fresh insight into this age old debate, and may stimulate new research in this area.

lines 273-275: Were any of these organisms isolated from these locations, or is this only based on 16S gene fragments in sequencing libraries? If they have been isolated that is additional evidence supporting their presence (as opposed to laboratory contaminants or cross sample contaminants from PCR products).

lines 317-323: I suggest that the authors also point out, that just because a KO is shared between different locations, does not mean that the exact same function is occurring. Recent evolutionary events such as point mutations, single amino acid changes in key functional parts of the protein, can have a significant effect on protein or enzyme function in different strains of bacterial species that colonize different habitats (doi: 10.1126/science.1218198). Therefore, its possible that recent evolution in the genomes of ASVs has taken place and that even though they share KOs, the encoded proteins are not exactly the same and might have different functional profiles between ASVs inhabiting different ranges or GFSs. I suppose this is something for a future project, but nevertheless I think its important to point out. That the conclusion of "functional redundancy" might not apply if this was the case.

lines 368-370: Do these results also potentially reflect recent speciation events in range specific ASVs?

lines 380-381: What do you mean by "fine-scale phylogenetic structure"? Is this referring to "micro diversity"?

Figure 3 panel d: It looks like the core microbial community is actually the most abundant part of the community compared to the specific and indicator ASVs. This result did not come out strongly to me in the text. Can the authors please highlight this important point a bit more in the discussion (if I am interpreting this correctly). Why do the specific ASVs tend to be less abundant than the core ASVs?

Figure 5: This is a really nice figure. However, for the reader I think it might be a good idea in the text to briefly define what "phylogenetic agglomeration" means (combining all ASVs within a certain larger hierarchical group?). Also, is it possible that DL and HoS colors (blue and red) are mis labeled in panel B? It looks like HoS increases with distance but DL decreases (which is different from what is written in the legend).

Extended data 5: The top dendrogram can be removed since it is only connecting two sample groupings.

Extended data 7: These individual plots are so small they are barely readable. Does this figure really add something critical to the story? Aside from a brief mention on lines 164-165 it doesn't feature in the story. Its unclear to me what the authors mean when they say that extended data 7 suggests truncation of the rare ASVs. What do you mean 'truncation' in this part of the text? Do you mean removing? What does each plot show? Are they different samples? Please provide the units and scale for X and Y axes. I could not find this information in the legend either.

Extended data 10a (heat map): I think that this heat map tells alot more the story about uniqueness and potential endemism compared to some of the current main text figures. I would consider making the heat map (or a version thereof) a main text figure. Because, it is the only figure I have seen so far that is actually showing the distribution of the "specific" ASVs across the samples.

Referee #3

(Remarks to the Author)
I reviewed this MS twice and raised various critical concerns. The authors have addressed my comments to some degree. In the previous two rounds of reviews, I have been focused more on high level of questions, particularly novelty, significance, and potential impacts based on Nature standards. Although I still have concerns, I will leave these issues upon Editor's decisions. In this round of review, I am more focused on details about experimental analyses, data analyses, and data interpretation and presentation for how to improve the quality of this study. After more seriously thinking of this study, although habitats are unique, but the analyses seem very shallow. I have various additional comments and suggestions for improvements. Please notice that my comments are not presented in a logical order.

1. Although habitats studied are interesting, this study is quite descriptive. Right now, it is a fish-expedition, and no hypothesis is presented. Based on numerous such types of studies plus their previous experiences with this habitat, the authors should be able to formulate some interesting and attractive hypotheses to test.
2. Two of the main drawbacks of amplicon sequencing approaches are low reproducibility and poor quantification. Due to the randomness nature, quantitation problem should be less serious for shogun sequencing approaches. The authors could recover phylogenetic markers (e.g. 16S, 18S and etc) from shotgun sequencing data and then compare them with PCR-based amplicon sequence data to assess the consistency and reliability of the experimental datasets.
3. Because this is a global analysis, it will be interesting to see whether the GFS microbial communities exhibit some general macroecological patterns such as latitudinal diversity patterns (LDGs), and species abundance patterns (SAD). But, no such analyses are presented.
4. Another very attractive theoretical question to address is how high is the microbial diversity across global GFSs, i.e., the extent of microbial biodiversity. This has recently become an interesting hot topic in global biodiversity studies. The authors can look at some representative studies (Locey et al. 2016. Proc. Natl Acad. Sci. 113, 5970–5975; Wu et al. 2019. Nature Microbiology, 4:1183–1195; Lennon and Locey 2020, Biology Direct, 15:5), and other similar studies which cited the 2016 PNAS paper).
5. In global biodiversity pattern section, it will be interesting to perform real global biodiversity analysis by comparing the GFS microbial communities to representative microbial communities from other habitats such as lakes, ocean, sediments, and soils to experimentally demonstrate the uniqueness of the GFS microbial communities under the extreme conditions. This will also help to understand whether GFSs microbial communities are more closely related to lakes, oceans soils or they totally stand alone.
6. As I mentioned previously, the authors should address whether the GFS microbiome diversity is important to the functions of GFS ecosystem. The authors measured some basic geochemical variables in C, N and cycling, which can be used as proxy of community functions. Thus, the authors could assess whether the biodiversity in the GFSs is important to the ecosystem functions to some degree by performing various statistical analyses.
7. The authors used iCAMP to assess the relative importance of different community assembly processes in controlling the GFS microbial community diversity and distribution. Their results (Fig. 5b) revealed that HoS increases with spatial distance, but dispersal limitation decreases with spatial distances. These results are counteractive to intuition. I expect that DL would increase with geographic distance, whereas HoS would decrease. The authors should check whether they perform iCAMP correctly, and if so, they should give some appropriate explanations. In addition, the authors could test how different community assembly processes (e.g. selection, drift) change with environmental conditions of the GFS systems (Ning et al. 2024. Nature Microbiology, 9:490-501).
8. The authors also use db-RDA to partition community variations. Their results indicated that 45.1% variations could not be explained, which could be due to unmeasured environmental variables, biotic interactions and/or drift. The authors should compare the results with the estimations of drift from iCAMP to assess the consistency across different approaches. But I failed to locate the data for drift's estimation.
9. Line 258. The section title, Taxonomy of the GFS microbiome, is not very attractive. It could be more interesting and informative to change it to global core bacterial community or global core microbiomes if other microbial groups are included (see Point 12). Also, the authors could test whether the core microbial communities are important to the functions of GFS ecosystems by using various statistical analyses.
10. As I mentioned before as well as indicated by Reviewer 2, the experimental design and the data generated by the authors could not allow the authors to address the research questions in mind (e.g. endemism, functional redundancy). It is good to not use the words endemic or endemicity, which will help to avoid potential confusion and misunderstanding. Using specific or specificity also problematic because, to me, specific or specificity and endemic or endemicity means same within this context. The authors can easily get away from this problem without using them. For example, this sentence "ASV specificity was not uniformly distributed across mountain ranges" can be easily rewrote as ASV was not uniformly detected across mountain ranges.
11. The authors could perform structural equation modeling (SEM) analysis to assess how important various environmental factors/climate change factors are in shaping the diversity and distributions of the global GFS microbiomes.
12. The habitats are unique, but the analyses are not as comprehensive as it should have been. The authors examined just the diversity of bacteria. As a comprehensive study, it will be very interesting to determine the diversity and distributions of other groups of microbiota such as archaea, fungi, and protists. Archaea could be important in this habitat from biodiversity perspective. But, the currently used primers are not good to recover archaea sequences. To examine the diversity and distribution of archaea, archaea-specific primers must be used. The diversity of fungi and protists can be easily done by sequencing 18S genes. In fact, the shotgun sequences could include sequences from all these groups of microorganisms although the majority could be bacteria. Thus, it will be nice to use amplicon-sequencing approach to also examine the diversity of archaea, fungi, and protists in these important habitats to make the data more comparable and comprehensive.

13. It is not easy to understand the meanings of the titles of some subsections, e.g. “Taxonomy of the GFS microbiome”, “The phylogenetic structure of beta-diversity and specificity”, “Drivers of GFS microbiome beta-diversity”. They are not straightforward, and it is hard to figure out what they really mean from ecological standpoints. Based on the text, I think they should simply use more common terms, as I mentioned above, e.g., global core bacterial community, community assembly mechanisms, environmental drivers of the GFS microbiomes.

Version 3:

Reviewer comments:

Referee #2

(Remarks to the Author)
The authors have satisfactorily addressed all of my comments, and in light of the numerous changes they have made over several revisions - which have substantially improved the quality of the study - I hereby recommend publication. I would also like to congratulate the authors on a job well done! I think this study will have a major impact.

Referee #3

(Remarks to the Author)
See my confidential comments to the Editors below

Rebuttal: Manuscript# 2023-03-04207 (Nature)
Ezzat *et al.* - Towards a global biogeography of the glacier-fed stream benthic microbiome

Referees' comments:

Referee #1 (Remarks to the Author):
Towards a global biogeography of the glacier-fed stream benthic microbiome
Ezzat et al.

The paper presents a unique data base on world-wide microbial characteristics and diversity in glacier-fed streams and gives very interesting and valuable insights. My expertise lies more in glaciers and glacier runoff, and I thus comment mainly on these aspects.

In sum, I find the glacier-related aspects of the contribution a bit surficial, acknowledging though that the main messages and focus of the paper are for the most part not depending on these aspects. Still, I think the paper would benefit from a more careful and clear treatment and discussion of glacier-related aspects, as detailed in the following. I number my comments in random sequence, not according to priority, just to ease response.

AUTHORS: We are grateful to this referee for her/his overall positive impression on our work, despite her/his expertise (in glaciology) differing from ours (GFS microbial ecology and biogeochemistry). We agree that the glaciological-related aspects were somewhat superficially addressed, which (as the reviewer noted) was due to the main points of the paper not relying on these aspects, and thus we did not want to detract from the main narrative of the manuscript by dedicating more space to these more 'tangential' elements. However, in light of the reviewer's comments, we would like to now take the opportunity to expand on our research approach and furthermore improve the clarity of our manuscript.

First, we are aware of the non-linearity of seasonal peak flow, the spatial variation in mountain glacier melt patterns, and diurnal runoff dynamics. However, to accurately quantify this variability would be non-trivial, requiring time-series data from each site to simultaneously capture both glaciological/environmental aspects as well as biological - an effort that is far out of the scope of the present study. Indeed, the goal of our study was to provide a first comprehensive spatial survey of this disappearing ecosystem, with the primary aims of quantifying microbial diversity and biogeographical patterns, as well as potential drivers. With the long-standing experience (more than 20 years) on stream biofilms in streams in general, and glacier-fed streams in particular, we have a reasonable understanding of the environmental factors that may shape biofilm communities in GFSs (also acknowledging their

spatiotemporal heterogeneity). The environmental parameters measured for this study result from that knowledge and what is actually feasible given the scope and logistical constraints of our work.

Secondly, it is critical to understand that the GFS environmental template not only depends on the glacier *sensu stricto*, as the GFS environmental template also depends on catchment processes and properties beyond the edge of the ice. The water physicochemical properties that we have used across all our GFSs worldwide integrate the combined glacier and catchment properties. It is exactly because of the process complexity — of the glacier and its catchment — that GFS ecosystem scientists/biogeochemists have used these integrated indicators for decades. The following papers may serve as a representative sample of their use, both for system understanding *per se* and predictive capacity of GFS functioning as glaciers melt: doi.org/10.1016/j.scitotenv.2019.04.221, doi.org/10.1016/S0883-2927(01)00123-8, doi.org/10.1039/C3EM00243H, doi.org/10.1002/hyp.7197, doi.org/10.1073/pnas.1619807114.

As a final consideration, one major finding from our work is the relatively strong influence of spatial effects in structuring GFS communities, which indicates the importance of dispersal limitation for community assembly. Therefore, while some additional glaciological variables could potentially help to explain community structure variability within the variation partitioning analyses (remembering that these are all the same ecosystem types to start with, and assuming the new variables would be both important and not already accounted for; please see our comment below), they are effectively irrelevant to biogeographical patterns if these patterns are due to dispersal limitation and not environmental sorting. Thus, quantifying, and incorporating, additional glaciological metrics are likely to have limited influence on the narrative as a whole, despite the great effort it would take to quantify them.

(1) The first sentence (Line 26) might be a bit misleading. The 60% referred to depend much on the climate scenario applied. Also, these '60%' refer to a sample (i.e. global distribution of glaciers) that is likely not similar to the representativeness of the GFS sampled in the current study. Many glaciers are located in regions that are not sampled, e.g. in the Arctic. Also, in particular Arctic glaciers often have no GFS, they are calving directly into the ocean. In reality, the numbers of glaciers projected to disappear that better represent your GFS sample might actually be significantly higher. I recommend to connect the present GFS study and the current (see below reference Hugonnet et al 2021) and projected global glacier change more specific, not just on a gross global level.

AUTHORS: We are grateful for this comment. Given the focus of our manuscript on the GFS microbiome, not on glacier shrinkage *per se*, we were obviously not careful enough when

selecting the reference. We have changed that first sentence of the abstract now so that the information is more general, and we added the Hugonnet reference.

(2) I was surprised to not read about the concept of 'peak water', at least in the discussion of the paper (e.g. DOI:10.1038/s41558-017-0049-x ;Huss and Hock, 2018). I.e. the stream-flow below glaciers does not at all linearly respond to glacier size changes. That means glacier area or glaciation % as used here are not necessary precise indicators of the glacial-hydrological conditions of a catchment. This leads to my next comment:

AUTHORS: Thank you for this comment. As stream scientists, we are well aware of the peak water concept (as outlined in the Huss and Hock 2018 paper), as well as the imperfect relationships between % glaciation and glacier area with hydrology and streamwater characteristics (more below). While peak water is certainly important for conceptualizing how these ecosystems may change in the future, we find it too tangential to insert into the discussion, especially given the text space that would be necessary to adequately explain the concept, and the limited utility the added text would ultimately have in putting our results into context (especially true given the paper's greater microbiological focus, coupled with the higher importance of dispersal limitation over environmental filtering).

In our manuscript, we present a unique global survey of microbiome data and environmental metadata, which includes basic data on glacier extent. While it would of course be optimal to directly relate communities to measured hydrological indices (necessary to directly relate to peak water), simply quantifying runoff is a daunting task because of temporal and spatial variability in runoff. For example, the same glacier can have several outlets (that is, channels that collect meltwater), and therefore modeled runoff data from such a glacier cannot be properly attributed to its various GFSs. We are currently working on a different study (actually with one of the authors of the Huss and Hock paper) using runoff data. In doing so, we further realized that the uncertainty related to this approach is substantial, particularly for our specific question addressed here. Furthermore, even if it were possible to model runoff for a given GFS, it is a major question how much variability one runoff number would explain in the context of the dynamic and complex hydrographs these streams exhibit both daily and seasonal timescales. A second problem is (assuming we are able to generate reliable hydrological data for each GFS), without a timeseries of biological data to complement it, it is impossible to interpret how these communities would respond to changes in runoff both within and across seasons, let alone speculate how they might change on either side of the 'peak water' threshold (i.e. what would be necessary for an informed discussion).

Given our 'snapshot' nature of data collection (which was necessary in order to capture a wide range of spatially disparate sites to fulfill the goals of the study), it is much more ecologically and statistically meaningful for us to correlate communities with conditions as we

found them in the field, and use proxies to approximate hydrological conditions for a given GFS. While we acknowledge that glacier area, relative catchment coverage, etc, are imperfect indicators of catchment conditions, we nonetheless found them to be highly correlated with some streamwater variables (e.g. water temperature), and they are widely used in the GFS ecology literature as proxies for glacier influence on the downstream ecology and biogeochemistry (see e.g., doi.org/10.1038/NCLIMATE1435, doi.org/10.1038/s41559-017-0426-x, doi.org/10.1038/s41558-021-01004-x). Coupled with their 'ease' in calculation, they are therefore almost certainly the most straightforward, and arguably most useful, in making broad comparisons between catchments for our purposes.

(3) Glacier climatology and thus run-off characteristics can vary a lot in space and time. Could part of the microbial diversity also reflect climatological effects, less than ecological/microbial? Or, how would such glacier-climatological and glacier-hydrological effects impact on your statistics? For instance, there are large differences in run-off and potential ecological importance of glacier melt between glaciers in very arid regions (critical dry-season run-off) or monsoon-type glaciers (relatively unimportant glacier runoff compared to rain). Such differences can actually happen over comparable small distances, e.g. southern and northern side of the Himalayan ridge. Similarly, the ratio between glacier-melt and rain contribution and its seasonality could be an important constraint for GFS microbial communities. Further, about your elevation parameter: Is it normalized, e.g. with latitude? Else, it is rather a climatological parameter. Or is there some other effect expected for microbes from absolute elevation? (BTW, extented data figure 6: what is the elevation anomaly? Against what?). A good overview of measured (not modeled as in your reference 1) global variation of glacier mass responses is given by DOI: 10.1038/s41586-021-03436-z (Hugonnet et al. 2021)

AUTHORS: Clearly, several of these points relate to the previous comment, which we have addressed above (i.e., specifically regarding the temporally-dynamic nature of GFS hydrology and possible community responses). We do agree with the reviewer that microbial diversity (and other community dimensions) could be influenced by climate-related parameters. Mountain climates change with latitude and altitude, which we have included as a spatial factor in our analyses — in the variance partitioning, for instance (i.e., linear trend). Another example is the elevational change in diversity, which for communities of plants, animals and microorganisms is typically attributed to temperature gradients. Here, we used elevation normalized by latitude to explore patterns of alpha-diversity. Specifically, we fit a global spline to account for latitudinal differences in the position of the glacier snouts, and related diversity to residual elevation. Therefore, we have accounted for climate, though without going into the mechanistic underpinnings, as these would be beyond the scope of this work. Moreover, while the pivotal role of microorganisms for climate change biology is undisputed (doi.org/10.1038/s41579-019-0222-5), the physiological versatility and eco-evolutionary adaptability of microbes may uncouple microbial communities from climatic processes — at

least climatic processes on longer time scales, such as climate change doi.org/10.1007/s10533-023-01015-0).

Finally, it would not be statistically possible to isolate the potential effects of climate (as suggested here by the reviewer) on our microbial parameters. That said, encouraged by this comment (and in agreement that precipitation patterns specifically could be a relevant variable for the GFS microbiome), we have now added snow cover days and mean annual precipitation as (bio-)climatic variables to the variance partitioning. We found that mean annual precipitation was among the significant variables retained by the models to explain changes in GFS community structure across mountain ranges. However, the addition of this variable had only a limited effect on the total explained variance.

(4) Line 85. Through glacier changes, e.g. increased run-off until 'peak water' and deglaciation of unconsolidated sediments, many GFS are actually exposed to much more severe impacts as listed here, for instance debris flows, which I guess will have severe impacts on the microbial communities.

AUTHORS: This is an interesting comment. Indeed, debris flow (if bisecting the stream) could have an impact on the resident community in the event's immediate aftermath. Yet, no evidence of recent debris flows were observed within streams at the time of sampling (this would definitely be noticeable on the streambed), and in fact, glaciers with a high chance of having such debris flows were avoided for safety reasons, and this is now mentioned in the Methods section of the manuscript (see our response to the reviewer's comment #6).

We would also like to point out that streamflow will eventually be re-routed through debris (unless it creates a lake), and the GFS community will return. Sediments are routinely delivered to the stream from the surrounding area through precipitation anyway, and therefore this pulse event would not likely introduce microbial cells that have not already been introduced before. What is really amazing is that despite debris flows, etc, the benthic GFS microbial community has a specific composition that differs from the community suspended in the stream water, for instance (see doi.org/10.1128/aem.00421-). Our understanding, shown in the present manuscript in particular, is that the environmental conditions (present in GFS habitats generally, but particularly within the benthos which necessitates biofilm formation) are sufficiently selective such that they overshadow discrete events such as debris flows. Of course, the latter can import bacteria to the GFS ecosystem, but these bacteria would then have to pass through environmental filters to persist and ultimately establish in the stream.

(5) Paragraph following L 289. See above (3-4) on the environmental effects beyond glacier size and %.

AUTHORS: This is certainly an interesting remark and we are grateful for it. We have already included some other environmental variables (e.g., streamwater temperature, electrical conductivity, ions, nutrients) that reflect glacier influence beyond their size and %. Yet, as explained above, to adequately capture more temporal glacial elements to incorporate into the analyses, it would require a timeseries of glaciological and biological data, the acquisition and curation of which is out of the scope of this present study.

(6) I could imagine that the glacier type, even within a region, could play an important role for microbes, beyond glacier and runoff climatology (3). E.g. debris-covered glacier, glacier with frontal lake etc.

AUTHORS: We are grateful for this comment. We are aware of these potential influences of glacier type and have therefore omitted from our sampling design systems with a proglacial lake or glaciers that are heavily debris-covered, as well as rock glaciers. Indeed, these factors can affect GFS hydrochemistry and biological communities, and in order to minimize the number of factors in our analyses, we have abstained from collecting samples from such systems. To address this comment, as well as those addressing the temporal element above, we have included the following in the Methods section: “Sampling was predominantly performed in spring or fall during hydrological ‘windows of opportunities' to avoid scouring conditions associated with the summer ice melt peak. Heavily debris covered glaciers and rock glaciers were furthermore avoided, and care was taken to not sample from GFSs with proglacial lakes, debris flows, or non-glacial tributaries in the reaches above the sampling site.”

(7) L 371-373. As commented above under (1) and (2), glacier change is quite diverse and not just a linear decline with warming atmosphere. So, the statement about ASV diversity and evolutionary lines being at risk should be formulated more careful and precise.

AUTHORS: Thank you for this comment. As aforementioned, we are aware of the non-linearity of glacier melt. With evolutionary lines, we refer to the bacterial clades as inferred from the phylogenetic analyses. In this context, it indeed makes sense to argue that GFS biodiversity is at risk of being lost as glaciers shrink — clearly it is a matter of timescales. Phylogenetic signatures of a highly-conserved gene (i.e. the 16S ribosomal RNA gene), reflect evolutionary relevant time-scales (albeit bacterial evolutionary rates are high compared to other organisms due to their large population size and short generation length). Given that both biogeographic and biodiversity patterns bear phylogenetic signatures, points to

processes that require somewhat stable conditions over sufficiently long times. These conditions are now changing rapidly. Moreover, our analyses highlight the role of spatial isolation. Dispersal could “rescue” biodiversity, but dispersal limitation, which we attribute to the topographic location and spatial isolation of GFS, suggests that such rescue is highly unlikely for GFS. Essentially, the underlying approach to our study is one of space-for-time substitution as it is generally used in GFS ecology studies (see e.g., doi.org/10.1038/NCLIMATE1435, doi.org/10.1038/s41559-017-0426-x, doi.org/10.1038/s41558-021-01004-x) . The respective sentence has been changed accordingly.

(8) L 375: glacier shrinkage is actually in principle not irreversible. Colder/wetter climate would within a few years to decades be able to restore a glacier. For this reason there are actually glaciers growing in a few regions. (How realistic it is that mankind will manage to reach a climate trajectory where glaciers can be stable or even recover, is a different topic, though).

AUTHORS: We agree with what is said here and are happy to share with the reviewer that we just started a project in the Pamir region where some glaciers are still having a neutral, even a positive mass balance. The irreversibility, (also a matter of timescales) is subject of a great debate on tipping points and these are about to be crossed as shown in doi.org/10.1126/science.abn7950. These discussions aside, we have changed that sentence accordingly.

(9) In sum, the final paragraph is quite surficial and speculative (see e.g. (7) and (8)).

AUTHORS: The final paragraph of this manuscript has been fully rewritten to address this comment and others.

(10) Extended data Fig 2a: What is the glacier area parameter? Average glacier size of the GFS investigated? Else? Same for the glazierisation %. Average? Else?

AUTHORS: We included a new table in the Supplementary Material (Supplementary Table 1) which provides further details on the median values, lower and upper quartiles of environmental and climatic variables, as well as the glacier metrics (i.e., glacier surface area as km2; relative glacier coverage of the catchment as %) used in our study. These glacier metrics are widely used in our field to quantify glacier influence on GFS ecology (e.g., doi.org/10.1038/s41559-017-0426-x).

Referee #2 (Remarks to the Author):

The manuscript of Ezzat et al reports 16S rRNA gene sequencing from a relatively high number of glacier fed stream ecosystems, and attempts to correlate these data with geographic and geochemical data to draw conclusions regarding global biogeographic patterns of diversity. The authors targeted specifically the benthic component of the streams. This is an important topic because, as the authors point out, glaciers are melting at an anomalously fast rate and their disappearance will have important implications for the ecology of the cryosphere. The conclusions regarding microdiversity are quite similar to a recent study published by many of the same authors, and therefore not necessarily a novel result (https://doi.org/10.1038/s41396-021-01106-6). What is new here is the high number of samples processed from a global dataset, and it is impressive and I think with the correct controls it could be a very nice study. Unfortunately however, there are fundamental controls that were missing (see below) that in my opinion preclude publication.

AUTHORS: We are grateful to the referee for her/his overall positive opinion on our work. First, we would like to apologize for not having reported the blank and mock communities serving as controls that we run through the lab procedures. This has now been addressed in the revised version of the manuscript — please see a detailed reply to this further below.

Second, we would like to stress that the present study goes far beyond the results and conclusions of the microdiversity work that we published in *The ISME Journal*, which focused on GFSs in New Zealand. The present manuscript focuses on the distinctiveness of the GFS microbiome from other cryospheric communities, global biogeographical patterns and their drivers, identifies high levels of regionality and endemicity, and extends our previous findings of the eco-evolutionary role of microdiversity substantially to 149 glacier-fed streams across 11 mountain ranges around the world (encompassing all continents except Antarctica). This is novel in many respects. Our work suggests, for instance, that both, selective environmental conditions and spatial isolation across spatial scales can jointly shape the GFS microbiomes. This extends findings from other global surveys, such as marine (doi.org/10.1126/science.12613) and soil (doi.org/10.1038/s41586-018-0386) microbiomes, which are predominantly shaped by environmental factors (e.g.,. temperature and pH, respectively) within these respective habitats. Microdiversity and phylogenetic clustering is an important feature of GFS microbiomes; however, we emphasize that the microdiversity element is not central to our manuscript and that we have well acknowledged how it differs from that previous study.

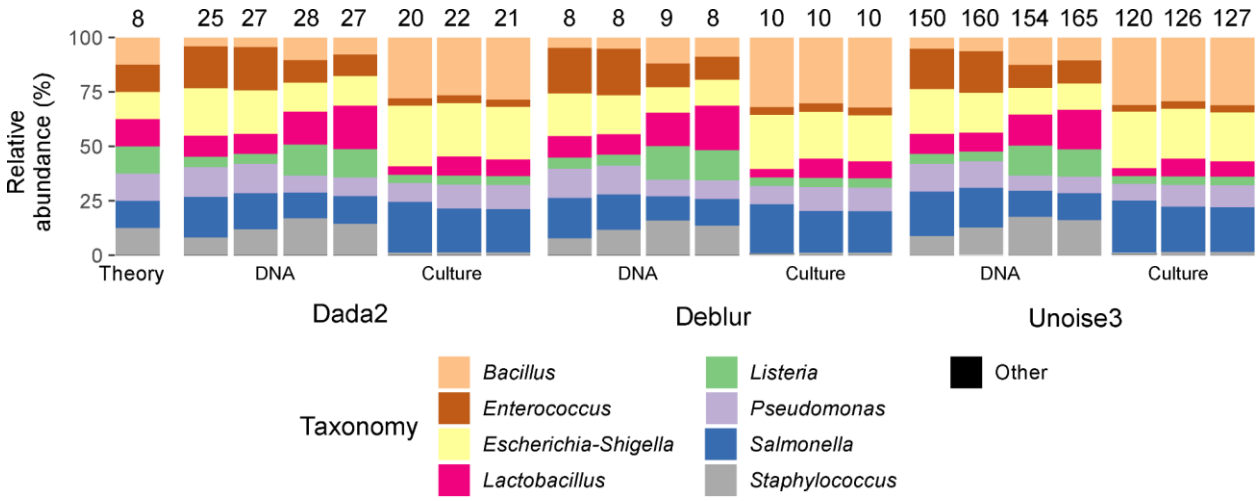
The authors main conclusions are that (1) a large microbiome fraction is endemic to single mountain ranges, and (2) a very small portion is disproportionately abundant and forms a cosmopolitan core. Because of the extreme dispersal potential of microbes, such questions

regarding biogeography are very difficult questions to answer in microbial ecology, and require careful controls. Unfortunately, there were many controls for contamination and checking ASV false positive rates (using mock communities) that were not done and the absence of these controls prevents the conclusions regarding core communities. Moreover, several taxa of their "core community" are well known as ubiquitous contaminants in molecular biology reagent kits (Salter et al 2014). This is concerning and the authors need to sequence extraction blanks to make sure that those taxa (e.g., Polaromonas) are indeed coming from the samples and not coming from the kits. The authors use a PCR primer combination that is fairly unconventional (why did they not use the Earth Microbiome Project primers?) and do not provide any information on whether their protocol was calibrated using mock communities and extraction blanks for contaminations. Some samples had a very low disproportionate sequencing depth compared to others, and this precludes conclusions regarding endemic taxa (they could be in the other samples just not detected due to a low sequencing depth.

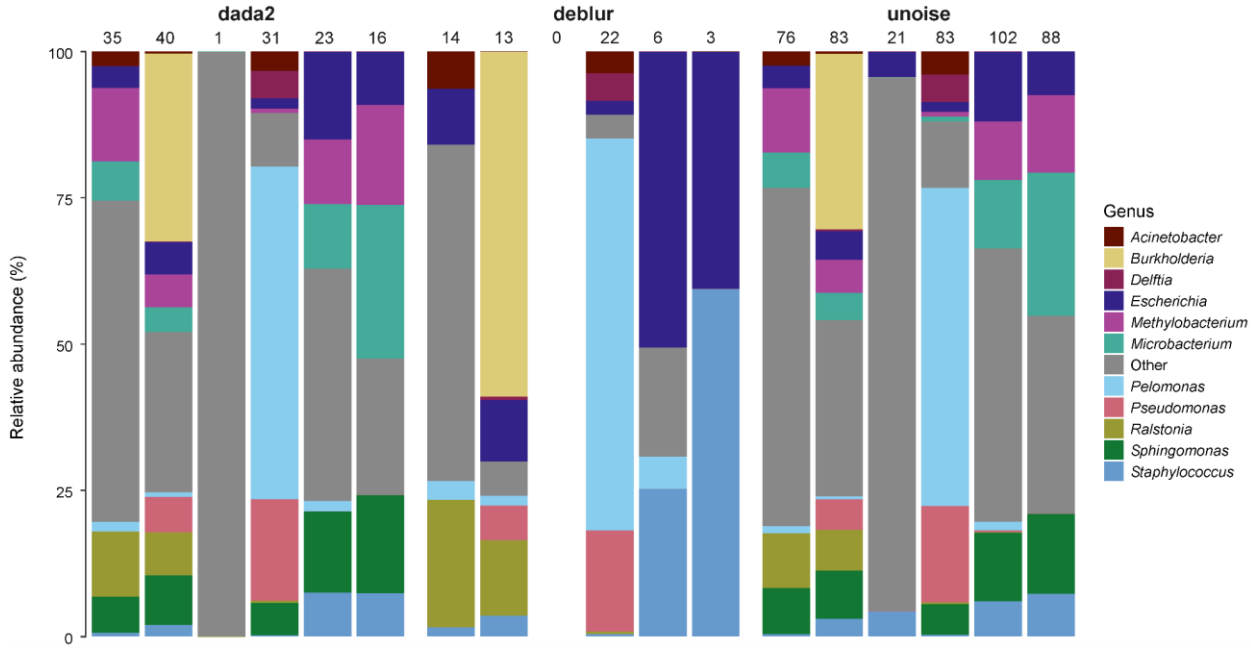
AUTHORS: We are grateful for these comments and acknowledge these concerns. We indeed missed adding the relevant information on controls and mock communities in the initial submission, for which we apologize again. In fact, we have previously published a paper (doi.org/10.7717/peerj.9973) describing an up-scaled phenol-chloroform based nucleic acid extraction method. In this paper, we benchmark the performance of our standard operational protocol using mock communities and blanks. However, we completely agree that it is pivotal to report our quality controls in this paper. Please, see our detailed replies to the points below.

Mock communities: Besides the mock communities reported in Busi et al., we now provide results from additional mock communities (ZymoBIOMICS, DNA-based mock community Cat Nr. 6305, 4 samples; cell-based mock communities, 3 samples, Cat Nr. 6300) as well as controls (7 extraction and sequencing blanks) to support our analyses. Cell-based mock communities were extracted using the same phenol-chloroform based extraction protocol as we used for all our samples. Importantly, and as suggested by reviewer #3, we benchmarked our analyses using three different sequence denoising pipelines (DADA2, Deblur, and UNOISE3) and found consistent results for beta-diversity and patterns of endemism across the three denoising approaches - taking into account blanks and mock communities (see Supplementary Information and Extended Data for further details). After comparing the three methods, we decided to proceed with Deblur, as this best resolved the taxonomy of the mock communities. For instance, Deblur accurately identified the expected bacterial taxa and effectively resolved the expected relative contribution of each strain of the mock communities. Moreover, compared to DADA2 and UNOISE3, Deblur resulted in few additional ASVs (e.g.,10 instead of a theoretical 8 bacteria that are present in the culture-based mock community) and an absence of contamination compared to the other methods (see Supplementary Material for details). Furthermore, Deblur has been demonstrated as one of

the most robust denoising approaches, a good compromise between UNOISE3 and DADA2 (doi.org/10.7717/peerj.5364).



Extended Data Figure 13b: Mock community composition obtained after using different denoising methods (DADA2, Deblur and UNOISE3). The mock communities were obtained both as DNA (DNA, Zymo, Cat Nr. 6305) and cell cultures (Culture, Zymo Cat Nr. 6300). The theoretical composition is shown on the left for comparison purposes and is composed of 8 distinct bacteria harboring similar abundances. Numbers above the bar plots correspond to the number of ASVs found in each community.



Extended Data Figure 13a: Taxonomic composition of the blank assemblages. Taxonomic compositions were computed using the three denoising approaches (DADA2, Deblur and UNOISE3). Numbers above the bar plots correspond to the number of ASVs found in each assemblage.

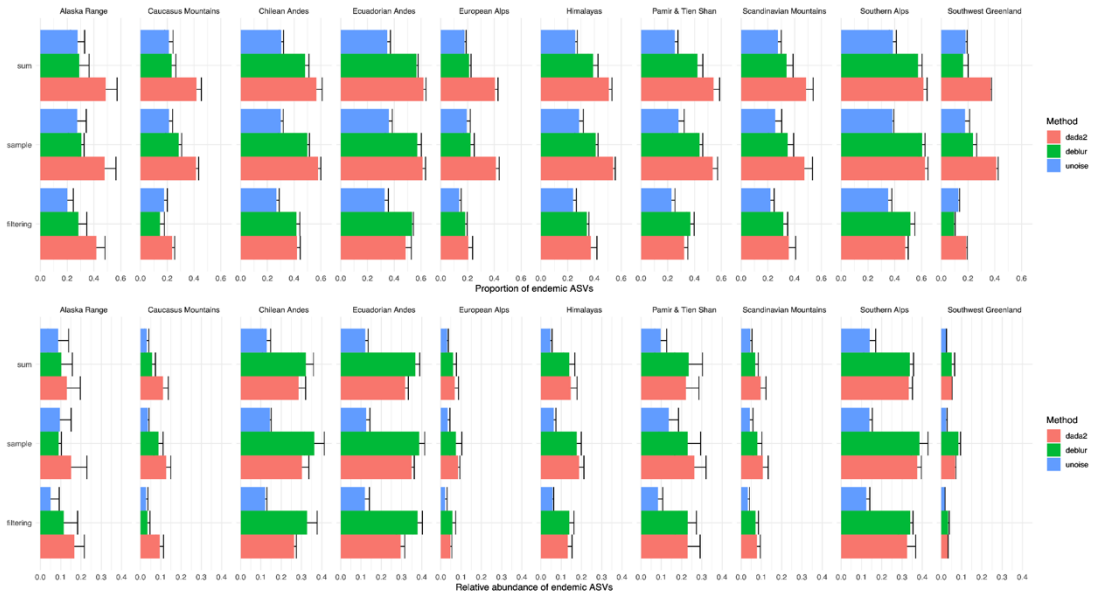
Blanks: We regularly included extraction blanks in different sequencing runs to aid QC. Moreover, as part of this revision, we re-extracted and re-sequenced samples with low read numbers and included additional blanks in this sequencing run as well. Overall, we found on average 486 reads in blanks, which Deblur clustered into 43 ASVs. This represents less than 0.6 % (relative abundance) of the overall community. These ASVs were removed prior to downstream analyses. The ASVs represent typical known contaminants such as *Escherichia*, *Pelomonas*, *Pseudomonas, and Staphylococcus*. These taxa do not naturally occur in GFSs and removal does not affect our conclusions. However, there was also a single ASV classified as *Polaromonas*, which we removed from downstream analyses. Please see also our specific response below.

Conclusions: Based on both, the results from the mock communities and blanks, in combination with our stringent filtering (i.e., retaining ASVs present in at least 50% of the biological replicates within a reach), we are confident that our findings, including the detection of the core community are not the results of false positives or cross-sample contamination but are valid estimates of the observed microbial communities.

Primers: We chose these specific primers (341F/785R) for the following reasons: First, they were designed to maximize environmental coverage and thus are widely used for the investigation of bacterial diversity in various environments (doi.org/10.1093/nar/gks808). Secondly, they were used in the majority of studies conducted on cryospheric samples (as found in doi.org/10.1038/s41467-022-30816-4) enabling us to maximally compare our results with those from other cold environments. Indeed, a literature search showed that 18 articles appear when searching for "cryosphere" in the papers that cite the Earth Microbiome Project (EMP) reference paper (doi:10.1038/nature24621) while 51 articles appeared when we proceed the same way as with doi.org/10.1093/nar/gks808. Finally, the 341F/785R primer pair allows for longer ASVs (444 bp) than that of the EMP primers (291 bp), which may increase taxonomic and phylogenetic resolution.

Sequencing depth: Differences in sequencing depth across samples can influence diversity estimates. To allow for comparability across samples we used rarefaction. However, we fully agree that rarefaction removes information, which is why we re-sequenced 34 samples and added 17 samples that allowed us to increase the number of samples with 3 replicates (see attached Table). We also provide endemism estimates based on the fully unfiltered and unrarefied dataset. Moreover, we performed a sensitivity analysis of our estimates of

endemicity. First, the three different denoising strategies (DADA2, Deblur and UNOISE3) could influence estimates of endemicity by differentially merging or keeping apart ASVs. Moreover, our filtering strategy could influence estimates of endemicity. In this context, we want to stress that we retained ASVs that were found in at least 50% of the biological replicates for a given GFS. However, if the same ASV was detected in only 1 out of three replicates of another sample, the ASV was retained in this sample as further removal would wrongly inflate endemicity. To account for this, we averaged the replicates counts, and subsequently multiplied them by three (i.e. number of replicated patches per sample) before applying the rarefaction (n=76,580 after multiplication by three of the counts, or 25,526.67). An alternative to averaging could be to sum read counts across replicates. Large inter-replicate variation in read counts could contribute to differences between averaged or summed read counts. This is important because of subsequent rarefaction. Rarefaction has the potential to influence endemicity as well. Rarefaction removes low-abundance taxa, and averaging or summing read counts across replicate samples can influence the distribution of reads. To assess the sensitivity of our approach (which is averaging), we compared the effect of summing, averaging and randomly sampling of reads on the estimate of endemicity (See Extended Data Figure 8).



Extended Figure 8. Comparison of the proportion and relative abundance of ASVs (top and bottom panel, respectively) across mountain ranges for the three different denoising approaches (DADA2, Deblur, UNOISE3). For method 'sum', replicates counts were summed, for method 'sample' one replicate was randomly sampled each time, and for method 'filtering' we applied the filtering as described in the methods (i.e., retaining ASVs only if

they were found in at least 50% of biological replicates for a given GFS), and averaged the replicates. For all methods, the dataset was subsequently rarefied (n=76,580 after multiplication by three of the counts), and 6 GFSs were randomly subset for each mountain range to account for uneven sampling across mountain ranges.

Notably, all combinations of denoising strategy and replicate treatment result in large proportions of endemic ASVs with overall large relative abundance. Furthermore, differences in endemicity between regions are generally larger than differences associated with methodological choices. While UNOISE3 tends to result in the lowest estimates of endemicity across regions, DADA2 tends to yield to the highest estimates. Deblur, in most cases, results in intermediate estimates of endemicity. This difference among denoising algorithms is particularly large when considering the contribution of endemic ASVs to relative abundance. UNOISE3, for instance, identifies a similar proportion of endemic ASVs; however, compared to the other denoising algorithms, these ASVs contribute less to relative abundance. This suggests that UNOISE3 tends to merge more abundant endemic ASVs across regions (therefore retaining a similar number but lower relative abundance of endemic ASVs).

The treatment of replicate patches within a given GFS reach (i.e., summing across replicate patches, randomly sampling or filtering followed by averaging), in contrast, had little effect on estimates of endemicity across most regions.

In conclusion, based on these considerations, we find that Deblur provides the most consistent numbers across methods, with our filtering representing a slightly more conservative approach that is bracketed by estimates using alternative denoising and replicate treatments).

Here are some more specific comments regarding my concerns on this matter:

line 123-126: The "rare biosphere" (doi: 10.1073/pnas.0605127103) is a controversial idea in microbial ecology that has been criticized (doi: 10.1111/j.1462-2920.2009.02051.x) to be biased by ASV clustering algorithms and sequencing errors that lead to high rates of false positive ASVs present at low abundance (https://www.nature.com/articles/nmeth.2604). It was bad with 454 sequencing, and has improved with more accurate illumina sequencing but the problem remains depending on how the ASVs are clustered and the degree of cross sample contamination, non-sample contamination, sample bleeding, and systemic errors in particular ASV picking algorithms. That is why all 16S based methods need to be ground truthed with empirical data from mock communities to figure out which parameters give an ASV richness that is correct and not overestimated. I cannot find any indication in the methods or manuscript that this was done for the authors 16S protocol reported here.

AUTHORS: While the rare biosphere concept has been challenged in its earliest days, we respectfully disagree with the reviewer that it remains a controversial idea. This is perhaps best supported by the following statement from a highly influential paper (Shoemaker et al. Nature Ecology and Evolution; doi.org/10.1038/s41559-017-0107) in the field: "One of the most frequently documented patterns of microbial diversity in recent years is the ‘rare biosphere’, which describes how the majority of taxa in an environmental sample are represented by few gene sequences." Furthermore, a plethora of papers in top outlets highlight the existence and relevance of rareness in global environments (e.g., Sogin et al. 2006 PNAS, 10.1073/pnas.0605127103; Pascoal et al., 2020. FEMS Microb Ecol, doi.org/10.1093/femsec/fiaa227; Jousset et al., 2017. ISMEJ, doi.org/0.1038/ismej.2016.174; Lynch & Neufeld, 2015. Nat Rev Microbiol, doi.org/10.1038/nrmicro3400; Saw, 2021, mSytems, doi.org/10.1128/msystems.00773-21).

We do however agree that ASV clustering algorithms and sequencing errors can inflate the numbers and richness of ASVs present in this category. We addressed these potential caveats in several ways. First, we note that our laboratory published a paper (Busi et al 2020; PeerJ, doi.org/10.7717/peerj.9973) where we compared and benchmarked methods for the extraction of DNA from GFS samples, mock communities and blanks. Essentially, that paper showed that the DNA extraction and library preparation that we applied in the present study and others (e.g., Ezzat et al., 2022 App Environ Microbiol, doi.org/10.1128/aem.00421-22; Busi et al., 2022 Nat Comm, doi.org/10.1038/s41467-022-29914-0) are not biased towards or against rare community members. This is, for instance, also reflected in our analysis of rank-abundance distributions (Extended Data Fig. 2,3), which follows theoretical expectations.

Nevertheless, we have included further blanks and mock samples (now included in the revised manuscript). As described above, the 43 ASVs detected in the blanks represented less than 0.6 % of the overall community and were removed. Furthermore, all the taxa expected in the mock communities were found in 8 to 10 ASVs compared to the theoretical 8 expected using Deblur.
A comparison with the other denoising methods showed that an average of 25 and 142 ASVs were found in DADA2 and UNOISE3, respectively, hence why we chose Deblur as our denoising method. Finally, our stringent filtering of ASVs (see above) further removes potential sequencing errors and extraction or PCR contaminations to only account for bacteria that can grow in these systems (i.e. are not contaminants from other sources, such as from humans, animals, spores, etc).

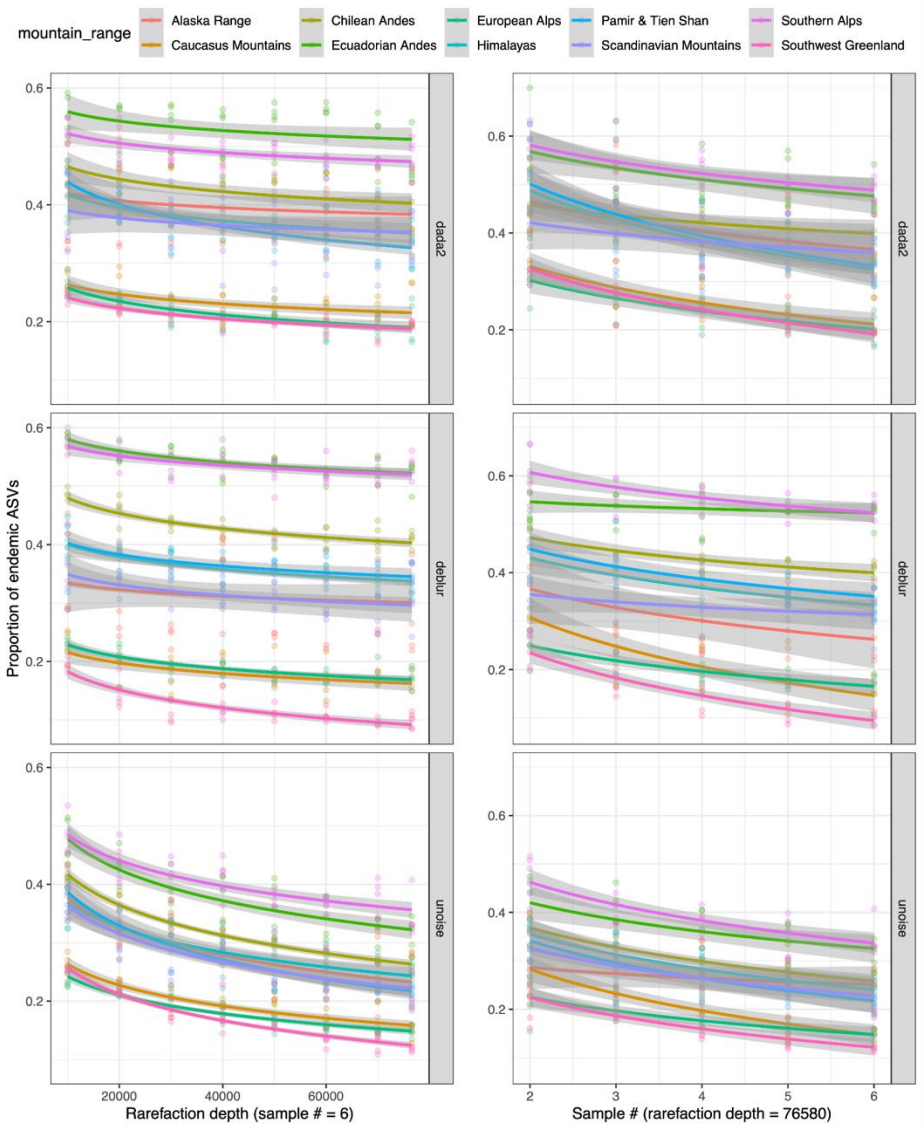
lines 168-169: If you define "endemic" as being present in a specific location, and not present in any others, then in my opinion, it is not possible to test this using 16S amplicon data unless

a very high sequencing depth was applied to all samples. But, looking at the methods some samples had as low as 10,800 reads and therefore many ASVs could easily been missed.

AUTHORS: Many thanks for this comment. We define endemic ASVs as those that we only find in a single mountain range. Indeed, the reviewer is right in the sense that “unique ASVs” (those present only in a single GFS) make up a substantial fraction of endemic ASVs. Given our filtering criteria, these unique ASVs are found in at least 50% of biological replicates of a single GFS - and in no other patch. Because of this, spurious ASVs (found sporadically in few samples) will not inflate our estimates. This data pre-treatment was performed prior to rarefaction, hence generally working with a much larger number of reads (i.e., the 10,800 reads refers to the number of sequences retained after rarefaction). Nevertheless, the reviewer is right that some samples had such relatively low numbers of reads (based on which the rarefaction depth was chosen). This decision was based on our ambition to retain as many GFSs as possible to maximize biogeographic information in our dataset. Moreover, the fact that our estimates suggested that we covered 73.6 % of the expected regional diversity with our sampling and sequencing effort (i.e., iNEXT estimates, frequency) reinforced the estimates of endemism.

Nevertheless, we agree that sequencing effort has the potential to influence estimates of endemism and hence we used the opportunity to re-sequence low-depth samples. By resequencing samples (see attached Table), we were able to more than double the number of reads in low-read samples (effectively increasing rarefaction depth from 10,800 to 25,526 reads). This additional sequencing effort also increased the percentage of observed (compared to estimated) gamma diversity from 56.7% to 73.6%. Despite this increase in coverage of regional diversity our initial estimates of endemism changed only marginally (from 58.3% of all ASVs to 58.4%), suggesting that our estimates of endemism are robust and relatively well constrained.

Additionally, we performed a sensitivity analysis (Extended Data Figures 8-10) by varying the rarefaction depth (per sample) and the number of samples randomly selected in a given mountain range, and computing endemicity on all three denoising methods. As expected, increasing rarefaction depth and/or sample number per region indeed reduced endemicity. Interestingly, rarefaction depth had more pronounced effects when the dataset was denoised using UNOISE3 as compared to DADA2 and Deblur.



Extended Data Figure 10. Sensitivity analysis of the proportion of endemic ASVs (within each mountain range) performed using all three denoising methods (rows), and shown for varying levels of rarefaction thresholds (left panel) and sample number per region (right panel). Generally, and as expected, estimates of endemicity are sensitive to both low-rarefaction depth and a low number of samples per region. Interestingly there are substantial differences between denoising algorithms, however, the lower bound estimates of the proportion of rarefaction range between approximately 10 and 55% (using Deblur). Geographically isolated GFSs (e.g, New Zealand's Southern Alps, Ecuadorian Andes) harbor consistently (and with high confidence) higher numbers of endemicity than other regions.

Please note that for comparability a maximum number of 6 samples was chosen and rarefaction was varied between 10,000 and 76,580 (i.e., summing the three replicates of 25,526 ASVs). For most regions we included many more samples and the final rarefaction was adjusted to 25,526 ASVs, leading to overall conservative estimates of endemicity.

lines 190-195: Some of the "core microbiome" is possibly the result of cross-sample contamination if all the samples were extracted and PCR amplified in the same lab. Particularly if you are including ASVs that are as lows 0.1% of total reads. Did the authors perform checks for cross sample contamination? Or laboratory, or kit contamination? I don't see any indication in the methods.

AUTHORS: Thank you for this comment. As stated earlier, we have removed 43 ASVs that we found in the blanks from our ASV table, which has not affected any of the biodiversity and biogeography patterns of the core or endemic ASVs here reported. However, given the low biomass and hypothesized low diversity, we were particularly careful to avoid contamination across samples. In fact, we routinely extracted only batches of samples from the same GFS. This is also because of logistic constraints with regard to large input quantities (requiring large centrifugation and limiting the number of samples that can be handled in parallel). Between each batch, laboratory surfaces and material were treated with bleach. Finally, we highlight 164 core ASVs based on thresholds of >0.1% for relative abundance and >50% for prevalence across mountain ranges. This is a relatively low number compared to the 39,434 ASVs in our entire dataset and (given the abundance threshold) is not influenced by low abundance (reads). Our estimate of a global GFS core at the ASV level is supported by taxonomic groups which we find at each site and by phylogeographic analyses which suggest that globally similar environmental constraints (i.e. homogeneous selection) shape the microbiome.

lines 199-204: Have you checked your data against contamination from the "kitome"? e.g., Salter et al 2014. Many of the group you write here are well represented as ubiquitous contaminants in commonly used molecular biology kits and DNA extraction kits.

AUTHORS: This is a valid comment, obviously linked to some of the previous and following comments as well. As described in the Methods section of our initial submission, and entitled ‘DNA extraction, library preparation and sequencing of the 16S rRNA gene’, we did not use any commercial kit for DNA extraction. Instead, we used an adapted phenol-chloroform extraction procedure that we had rigorously tested on GFS sediments and against commercial kits, as described in Busi et al. (doi.org/10.7717/peerj.9973) and accordingly referred to in our initial version of the manuscript. In that publication, which was centered on metagenomics, we used a mock community (ZymoBIOMICS, Cat.No. D6300) and did not find any signature

of contamination. As described in more detail above, we have included mock communities specifically for 16S sequencing.

line 253: Polaromonas is one of the most commonly found contaminants of molecular biology kits (Salter et al., 2014 BMC Biology). The fact that this is found to be an abundant cosmopolitan taxa of the "core community" is concerning, the authors need to sequence extraction blanks and make sure that polaromonas ASVs are not a contaminant of their PCR and DNA extraction reagents.

AUTHORS: Thank you for this point. As a follow-up to our previous comment, we want to reiterate that we did not use any commercial DNA extraction kit and that our blanks contained only one ASV related to *Polaromonas* compared to the 252 *Polaromonas* ASVs found across all our samples. This single ASV, which we removed, represented only 0.14% of the relative abundance of the all *Polaromonas* ASVs and did not have an effect on our results. Furthermore, the marked biogeographic signatures, notably of *Polaromonas* (see Extended Data Figure 12), refute concerns of contamination (at least partially) because contamination would be very unlikely to impart biogeographic patterns. We would also like to point out that, despite being regarded as a kit contaminant, *Polaromonas* has been repeatedly demonstrated to be an abundant and prevalent member of various cryospheric microbiomes

lines 286-287: Or perhaps it is an indication that there is some noise in the data (contamination, or ASV false positives?).

AUTHORS: May we kindly refer to our detailed responses above as this comment clearly is a continuation of some of the previous ones.

General comment on the 16S amplicon methods: Have the authors used control samples to determine the extent of kit or laboratory contamination (see Salter et al 2014), sample bleeding, contamination, ASV false positive rate, or cross sample contamination during library preparation? What is the false positive rate of ASVs using this approach? Many of the ASV clustering algorithms over estimate microbial richness, and the ASV false positive rate needs to be determined empirically using a mock community (https://www.nature.com/articles/nmeth.2604) for any 16S sequencing and analysis protocol. The authors do not report on this, or describe any of these controls, therefore I have to assume that they were not done.

AUTHORS: May we kindly refer to our responses to the previous points raised by this reviewer.

lines 473-475: Why did you use these primers, as opposed to the more conventional 515F/806R primers that are used by the Earth Microbiome Project (https://earthmicrobiome.org/protocols-and-standards/16s/) ? Have these 341f/785r primers been tested on a mock community by the authors and the ASV cutoff empirically determined?

AUTHORS: As stated above, we chose these specific primers (341F/785R) for two main reasons, first they were designed to maximize environmental coverage and thus are widely used for the investigation of bacterial diversity in various environments (Klindworth et al., 2013). Secondly they were used in the majority of studies performed on cryospheric samples, which allowed us to be able to compare our results with those of other cold environments. Indeed, a literature search showed that 18 articles appear when we search for "cryosphere" in the papers that cite the Earth Microbiome Project reference paper (doi.org/10.1038/nature24621) while 51 articles appear when we do the same with the Klindworth paper (doi.org/10.1093/nar/gks808).

Please, also see our detailed comment on the mock community, including the results in the respective figure.

lines 496-498: What was the identity cutoff used to classify ASVs? 97% , 98%, or 99% ?

AUTHORS: We assume this is a misunderstanding. In contrast to OTUs, ASVs are not clustered and are distinguished by a single nucleotide variant. The different pipelines, such as DADA2, Deblur and UNOISE3, thus allow investigators to resolve differences in amplicon sequence data at the level of single nucleotides, therefore avoiding potential pitfalls of clustering at a specific identity cutoff. Our results suggest that both biogeography (e.g. regional differences in GFS community composition) and biodiversity (e.g. microdiversity) of GFS microbiomes can be largely attributed to processes occurring at very fine scales of the phylogeny. Notably, these processes are non-random. Clustering ASVs into OTUs would blur much of these signals, rendering GFS microbiomes similar across spatial scales, and having important ramifications for our appreciation of microbial biodiversity.

line 507: It seems like a stretch to make conclusions about endemicity for samples that had as low as 10812 reads. With such low sequencing depth many ASVs at lower relative abundance could easily have been missed in the sequencing libraries. The authors allude to this , indicating they themselves are aware of the problem, on lines 157-159. Yet, they still draw very strong conclusions about endemicity in the manuscript. In my opinion such strong conclusions about endemicity are not support by the authors data for reasons described above.

AUTHORS: Thanks for this comment, which is clearly linked to our previous replies related to sequencing depth, blanks and mock communities, the various denoising pipelines benchmarked and our sensitivity analysis. With the results of the aforementioned analyses, we do hope that the reviewer is now convinced that our estimates of endemism are well constrained.

line 508: By "at least 50%" do you mean that the ASV had to be detected in at least 2 out of 3 replicates?

AUTHORS: Yes, this is the case. Following the preprocessing of our sequencing data, we decided to use a stringent approach to filter the ASV table. Given that benthic sediments were collected as 3 biological replicates from the same reach for a given GFS, we made sure to keep ASVs found in at least 50% of the ecological replicates (i.e., patch) for a given GFS, which means in 2 out of 3 replicates or 1 out of 2 when the third replicate was missing (3rd replicate missing for 16 GFS out of 149 GFS). We choose this analytical approach because previous work (Ezzat et al. Applied and Environmental Microbiology) showed that benthic communities differ from the suspended community in the streamwater (please, see also next comment). This suspended community reflects various upstream source communities, including supra-, en- and subglacial communities. Some ASVs of the suspended community are likely sampled alongside the benthic community. Leveraging our sampling design, we are able to identify and exclude these ASVs.

Referee #3 (Remarks to the Author):

Climate warming has caused significant loss of the world's glaciers, with great impacts on glacier-fed streams (GFS). However, our understanding of the biodiversity and biogeography on the biofilm microbiome of GFS is limited. Thus, the authors collected samples from about 150 GFS from world's major mountain ranges and determined bacterial distribution and biogeography using 16S amplicon-sequencing. Their results suggested that the global GFS benthic microbiome is quite different from other cryospheric microbiomes with very high diversity. About 60% of the total amplicon sequence variants are endemic, with very small core (0.4%). Further analysis indicated that the community diversity and changes are mainly shaped by dispersal limitation and environmental selection. Although the ecosystem examined is interesting, I have several major concerns on the novelty, significance, and conclusions of this study.

AUTHORS: We are grateful to the referee for their comments on our manuscript. Please, see our detailed replies to their three major concerns below.

First, it appears that the experimental results on high microbial diversity in this system just represent incremental improvements compared to previous studies by Wilhelm et al (2013, ISME J) and Ezzat et al (2022, AEM), and are not novel enough to warrant a publication in Nature, which generally publishes groundbreaking through, transformative studies in various fields.

AUTHORS: Many thanks for this comment, which we are happy to reply to in some detail. For the sake of clarity, we respectfully make the point that we do not report on experiments in our study as understood by the reviewer; rather, we report findings from a unique global survey on the GFS microbiome. As for the two papers mentioned by this reviewer: The Wilhelm *et al*. paper in *The ISME Journal* covers 24 GFS in Austria and is limited to OTU diversity and community composition. The Ezzat *et al*. paper in *Applied and Environmental Microbiology* is an important precursor study to the present study (*nota bene*, it has been designed as such) because it shows that the GFS benthic biofilm community is not random; rather it differs from the assemblage suspended in the water column of GFSs. This is critical for the present study, as otherwise it could be argued that we are looking at a microbiome that is not 'specific' to the benthic habitat in GFSs; for instance, it could be argued that it simply is a wash-out product from glaciers or debris flow (please also see above comments from referee #1).

Having this said, what is novel about our study?

This is the first global, and therefore most comprehensive study (149 GFS from eleven of the largest mountain ranges on this planet, encompassing all continents with the exception of Antarctica), of a microbiome dwelling in an ecosystem that is now rapidly vanishing because of climate change. The revised version (now improved thanks to your comments below) and leveraging both metabarcoding and metagenomics now shows that the GFS microbiome:

- differs from other cryospheric microbiomes in terms of its structure and function (based on 189 metagenomes)
- hosts a remarkable diversity that yet contains many unknown taxonomies
- demonstrates clear biogeographic patterns:
  i) high level of community dissimilarities among and within mountain ranges (also reflected by a distance decay pattern)
  ii) shows a remarkable level of endemism, even uniqueness (we provide a wealth of evidence supporting the robustness of these findings)
- harbors, as revealed by phylogeography, microdiverse clades that contribute to endemism and beta diversity
- shows signatures of functional redundancy as a response to the GFS environment

With this, we provide unprecedented insights into a microbiome at risk and present an important and valuable resource for future GFS microbiome studies.

It may be helpful to put this work into context. At the time when the first results of the Tara Oceans were published in *Nature* and *Science*, our knowledge on the microbial life in the world's oceans was far more advanced than our current understanding of the microbial life in the streams draining the roof of our planet. We all understand that mountain glaciers are melting at an unprecedented pace owing to climate change, and yet we do not know what actually lives in these ecosystems. In analogy to the Tara Oceans, we were cruising the major mountain ranges of our planet for four years, and now report for the first time, using transformative data indeed, on the biodiversity and biogeography of the bacterial communities in the GFSs. Therefore, novelty comes with first-of-its-kind insights into and understanding of the global GFS microbiome.

'Glacier ecology' (and cryospheric ecology, in general) is a burgeoning field for obvious reasons (please see doi.org/10.1038/s41559-019-1042-8, doi:10.1038/s41559-017-0426-x, doi.org/10.1038/s41586-023-06302-2, for instance). The GFS part of this field remains largely limited to meta-analyses of invertebrate ecology (e..g., doi.org/10.1038/s41559-019-1042-8, doi.org/10.1038/s41559-023-02061-5) - not on the microbiome. We do fill this gap (finally) — further adding relevance and novelty to our study.

We have changed the introduction accordingly to better emphasize how our work advances the field.

Second, the significance of the experimental results is not clear. The importance of dispersal limitation and environmental selection in shaping the community diversity and distribution is kind of expected given the large differences of continental level geographic locations and environmental conditions.

AUTHORS: This point is linked to the previous one, and we therefore take advantage to further argue why we believe that our results are significant and novel indeed. **(i)** The spatial structure (i.e., beta-diversity) of the global GFSs across scales — that is, within and between mountain ranges — is unexpected indeed, as well as its distinctiveness from other cryospheric ecosystems, in terms of taxonomic and functional composition. The GFS microbiome is shaped by both environmental filtering and dispersal limitation, which appears to be different from other ecosystems even at continental and global scales (e.g., ocean, soil). One could expect that the putatively high dispersal capacity of bacteria would homogenize spatial patterns. Our results show that these expectations are not met. As we hopefully made clear in the manuscript, the GFS environment by its very nature shares conditions independent of geographical location. However, GFSs, located at the tips of nascent river

networks, are spatially isolated such that spatial processes become relevant. At the same time, environmental processes are important in shaping the phylogenetic structure of the GFS microbiome. This is relevant beyond the realm of GFS microbiomes, as other globally resolved microbiomes tend to associate turnover predominantly with environmental factors (such as temperature in the global ocean doi.org/10.1126/science.1261359 or pH in soils doi.org/10.1073/pnas.0507535103 and doi.org/10.1038/s41586-018-0386-6). **(ii)** The high degree of endemism (now confirmed by our additional bioinformatic analyses, mock communities, etc; please, see our reply to the comment by this referee below as well as to comments from reviewer #2) is unexpected and adds to the ongoing debate of microbial endemism and biogeography. It highlights GFS ecosystems as 'islands' of endemicity. **(iii)** Our phylogeographic analyses provide fine-scale mechanistic understanding of the GFS microbiome beta-diversity, endemicity, and microdiversity. This is a noticeable advancement of microbial community ecology overall.

Third, there are serious flaws on methodologies and analyses (see below for details) and thus the results/conclusions are not reliable.

AUTHORS: May we kindly refer to our replies to the respective comments below.

In addition, the experimental analyses are preliminary, just based on 16S amplicon sequencing. No functional structure differences among these systems are presented. Other omics approaches such as shotgun sequencing, genechip, metabolomics beyond compositional profiling to assess the changes of functional traits could be considered. Moreover, it is not clear whether the community composition changes affect the functions of the glacier ecosystems, which is a critical question in ecology and climate change biology. There is no attempt to link community structure to the functions of these ecosystems.

AUTHORS: Most respectfully, we do not agree with the statement that 16S amplicon sequencing can just provide preliminary data. It is well known that 16S amplicon sequencing data and resulting ASVs (please see discussion on sequencing depth, mock communities, etc) are most appropriate to address questions of bacterial biodiversity, biogeography and phylogeography. Indeed, numerous are the studies on microbial diversity and biogeography (e.g., doi.org/10.1073/pnas.121242411, doi.org/10.1038/s41396-020-0657-8, doi:10.1038/ismej.2009.106, doi.org/10.1038/s43705-022-00099-3) that entirely rest on amplicon sequencing. We believe that this approach makes particular sense to lay the baseline of a hitherto unstudied microbiome on a global scale.

Without any doubt, we do agree with the referee that functional aspects of the GFS microbiome are interesting. In that context, we have now shared with the Editor (at her discretion) a synopsis for a follow-up paper that covers the functional aspects of the GFS

microbiome in great detail — based on 2,868 bacterial MAGs, as well as eukaryotic MAGs and novel insights into the GFS virome. It is clear that our manuscript on the biodiversity and biogeography of GFS bacteria fundamentally differs from that MAG-centric paper. It is also clear that they cannot be merged into one single, focused and coherent paper.

However, encouraged by the referee's suggestion, we have now included a functional element based on 24 million genes summarized into ca. 18,000 KOs from 97 GFS shotgun metagenomes that we compare to 92 metagenomes published from cryoconites, ice, and polar desert sand, for instance, (doi.org/10.1038/s41467-022-30816-4). With this analysis, we are now able to show that the GFS microbiome is functionally distinct from other key cryospheric microbiomes and that pathways associated with biofilm formation, cell cycles, energy and metabolism indeed makes the GFS microbiome distinct. We have added two panels on these findings to the new Figure 1 of our revision and an additional figure in the Extended Data.

We do hope that this extra work strengthens the conclusions of our study and adds novelty, while telling the previously untold story of the GFS microbiome biodiversity and biogeography.

(Below text split into separate comments for rebuttal purposes)
I have various more detailed questions on the methodologies. First, the experimental design and sampling is problematic. The authors collected three samples from each GFS with distance of average 300m as ecological replicates. Because of spatial heterogeneity, many more biological replicates are needed to represent the microbial community diversity of the GFS examined. Although 148 GFS were selected, only 418 sequencing libraries were constructed, which ended up only 1 or 2 libraries for some GFS. Rigorously speaking, the authors should spend more efforts in constructing sequencing libraries for all GFS with equal amount of sampling efforts, which is critical to the subsequent endemic analysis.

AUTHORS: These are important comments and we are grateful for them. Having worked in streams, particularly GFSs, for more than twenty years, we are well aware of the heterogeneity of these ecosystems. GFSs change downstream owing to chronosequences and related geomorphic and ecological processes. To avoid this gradient, we focused on the GFSs as close as possible to the glacier snout; please, see also response to reviewer #1. This approach greatly improves comparability between GFSs and provides insights into the microbiology of the nascent GFSs mostly affected by climate change (at the interface between the cryosphere and hydrosphere). To further ensure comparability across sites, we used standardized methods to retain the coarse sandy fraction of the benthic sediments; this is the most important fraction for microbial colonization on a surface-to-volume-ratio basis.

‘Experimental design’: We used a spatially nested sampling design typical for stream studies as they reflect the spatial organization and heterogeneity of stream ecosystems and their networks.

What may not have been clear from our manuscript is that each GFS was sampled as close as possible to the glacier snout (respecting safety issues, accessibility, etc). We have clarified this is the text now (median distance: 79 m). Within each reach, we sampled three sediment patches (i.e., the ecological replicates) within a perimeter of ca. 10 m — not 300 m (!) as understood by the referee. However, there is indeed turnover (dissimilarity) in benthic community composition across these very small scales. To address this, we quantified among patch dissimilarity using the dataset prior to merging (please, see the Figure below). Mean Bray-Curtis dissimilarity across patches from the same GFS was relatively low (0.38), particularly considering that mean Bray-Curtis dissimilarity of GFS communities from the same mountain range ranged between 0.70 (in the Caucasus Mountains) and 0.91 (in the Chilean Andes). Next, we sampled up to three GFSs within the same cluster (yet draining individual glaciers), and between three to four of such clusters per mountain range. This nested design helped us indeed to cover various spatial scales as well as environmental, geological and geographic features. This is particularly relevant for the analysis of distance decay patterns and variance partitioning models.

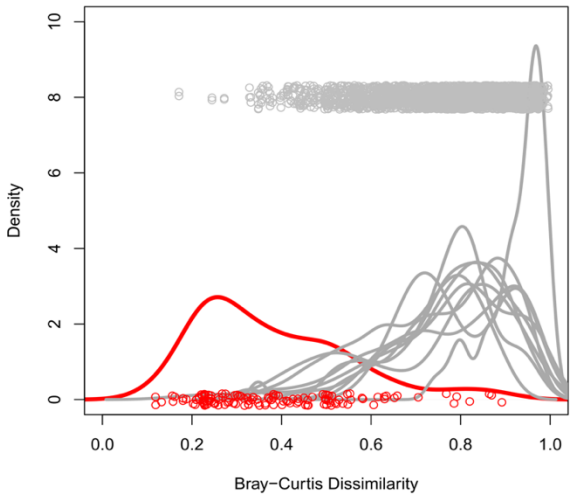


Figure. Density distribution of Bray-Curtis dissimilarity across 3 replicate patches sampled from the same GFS (red). Grey symbols and lines represent pairwise Bray-Curtis dissimilarities among GFSs from the same region. Turnover (dissimilarity) was generally larger across nearby GFSs than within independent biological replicates sampled from the same GFS.

149 GFS/418 libraries: We have sequenced an additional 51 samples of which 34 previously had low numbers of reads. This extra work on the bench has allowed us to increase the rarefaction threshold to 25,526 and include 14 replicates that failed previously. We need to emphasize how difficult it can be to extract sufficient high-quality DNA from GFS sediment samples. Given the complex mineral matrix, changing between mountain ranges, we consider our accomplishment fortunate. Nevertheless, we are most grateful to this reviewer for having encouraged us to return to the bench.

Second, the sequence preprocessing is also problematic. The authors used the average of three (two or one) ecological replicates to represent the diversity of each GFS. Since GFS is a big area, it will make more sense to use them as independent biological replicates (rather than as technical replicates) so that rigorous statistical tests can be performed to assess whether GFS X is significantly different from GFS Y.

AUTHORS: Thanks for this comment. Indeed, we have rigorously explored the pro's and con's of different data-pretreatment steps. Please also see our responses to referee # 2. Previous work by Ezzat et al. 2022 showed that the benthic biofilm and suspended assemblages of GFSs represent distinct entities. The suspended assemblage appears to be largely stochastically assembled from various upstream sources (including supra-, en- and subglacial microbial communities). This is different for the benthic community, for which the ability to form biofilms is a key trait. As shown in the above Figure 1, benthic communities from the same GFS are similar, particularly contrasted to the large dissimilarity observed even among nearby GFSs. This is important, because it provides the conceptual foundation of our data pretreatment. Given the focus of the present work on global-scale biogeographic signatures, we indeed opted for averaging biological replicates (thus not resolving patterns such as GFS X is different from GFS Y) yet using the per-replicate information to filter ASVs. Averaging reduces heterogeneity in relative abundance across replicates, which compared to summing or randomly sampling reads from replicated samples reduces the influence of outliers (e.g. individual ASVs with high number of reads in a single patch). However, we also provide a sensitivity analysis (please see above for details) of how averaging, compared to summing or randomly picking reads may affect endemicity estimates. This sensitivity analysis did not reveal marked differences between analyses considering replicates as independent or averaged. For statistical purposes especially in the context of the variance partitioning modeling, we therefore decided to further analyze our data by taking into account the average of replicates per patch.

Third, Data2 was used to preprocess the sequencing data. Since Data2 tends to remove many true ASVs, it is advisable to also use other sequencing preprocessing approaches such as dBlur, Unoise, and even UParse. Although the exact results obtained with different approaches will be different, the general trends should be consistent for the same datasets

with different preprocessing methods. It is important to ensure that the experimental results are not dependent on the methods used for sequence preprocessing.

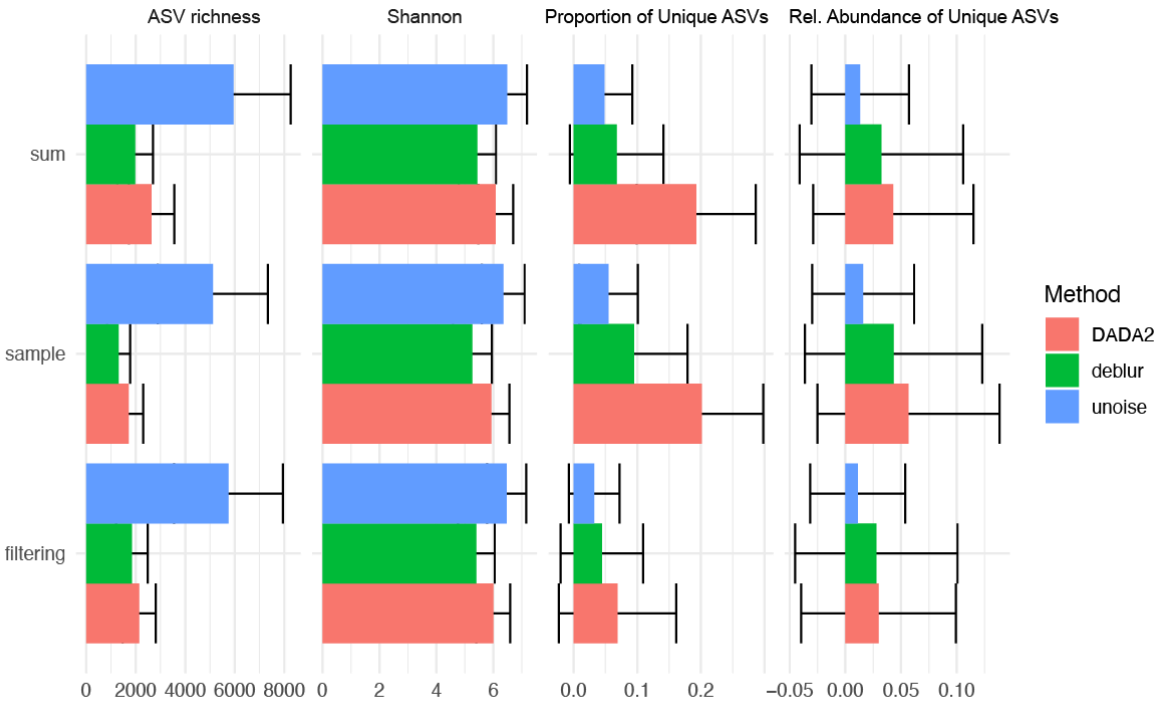
AUTHORS: We agree with the reviewer that it is critical to ensure that results are not dependent on the approaches used for sequence preprocessing. First, DADA2 is a well-accepted approach (or a gold standard) in 16S analysis for removing non-biological sequencing errors; it has been cited more than 15,000 *versus* other methods such as Deblur (about 1000 citations), UNOISE3 (UNOISE2 >1000 citations) and the OTU-picking approach UPARSE (12,272 citations). As stated in the review paper by Nearing et al. (2018, PeerJ, doi.org/10.7717/peerj.5364) "DADA2 tended to find more ASVs than the other two denoising pipelines when analyzing both the real soil and two other host-associated datasets, suggesting that it could be better at finding rare organisms, but at the expense of possible false positives". Further, as outlined in the research article by Prodan et al. (2020, PLoS One, doi.org/10.1371/journal.pone.0227434) "DADA2 offered the best sensitivity, at the expense of decreased specificity compared to USEARCH-UNOISE3 and Qiime2-Deblur."

Nevertheless, encouraged by this reviewer, we used two additional denoising approaches to pre-process our sequencing data, notably Deblur and UNOISE3 (please, see also our detailed comments and replies above). Our comparison can be found in the Supplementary Material. We here provide a brief summary:

- All methods show similar levels of diversity based on the Shannon index. However, compared to DADA2 and Deblur, UNOISE3 largely overestimates per-sample diversity when comparing the observed number of ASVs.
- Abundance-based beta-diversity estimates (e.g. Bray-Curtis dissimilarity) did not differ between denoising pipelines, suggesting that differences in alpha-diversity arise mainly because of ASVs with low relative abundance.
- The number of ASVs found in our mock samples, as stated above, was much lower using Deblur (8 to 10 ASVs) than DADA2 (average 25 ASVs) or UNOISE3 (average 142 ASVs), compared to the theoretical expectation of 8 ASVs.
- All denoising strategies show some level of uniqueness and endemicity. Deblur yields intermediate estimates of endemicity compared to DADA2 and UNOISE3.
- The regional differences in endemicity (which we attribute to different degrees of spatial isolation) are consistent across the three different denoising methods.

Based on these considerations, foremost because **(i)** Deblur best reproduced the mock communities and **(ii)** it provides more conservative estimates of endemic and unique AVSs than DADA2, we decided to use Deblur for our analyses. This decision entailed that we re-computed all our analyses that were initially performed based on DADA2.

We do hope that the reviewer appreciates this substantial benchmarking effort.



Extended Data Figure 9. Comparison of diversity metrics (Observed=number of ASVs per sample, Shannon=Shannon index, Unique prop.=proportion of unique ASVs in a sample, Unique rel. ab=relative abundance of unique ASVs) for the three different denoising approaches (DADA2, Deblur, UNOISE3). Values were computed with a rarefaction threshold set at 25,526 reads per sample after our filtering was applied and the replicates averaged.

We now present in the Supplementary Material our comparison and benchmarking of three pipelines for the mock communities, blanks, beta-diversity (weighted), as well as the number of endemic ASVs.

Fourth, defining endemism is very tricky and challenging. The conclusion that ~60% ASVs are endemic is problematic. An ASV which is not detected in other samples could be due to many other reasons beyond their “extinction” in these samples. One could be because sequencing depth is far not enough. Some OTUs which are not detected could be primarily due to shallow sequencing and/or random sampling processes. Thus, to determine “true”

endemism, much deeper sequencing (several magnitudes higher) is needed to ensure that an ASVs is present only in one GFS but not in other GFSs. Spatial heterogeneity will also have significant effects on the detection of endemic ASVs. If more samples are analyzed from each GFS, less endemism would be expected. Another reason could be due to the sequence preprocessing method (Data2) used. Data2 tends to remove many true ASVs, and rarefaction curve is easier to be saturated. I guess that the degree of the endemism will be greatly affected by the sequencing preprocessing methods.

AUTHORS: We are grateful for this comment, which was also raised by reviewer #2. Therefore, may we kindly direct you to our in-depth replies to the comments above. Please, also read in this context the previous comment where we refer to the effect of the various denoising pipelines on endemism.

We also want to emphasize that, although we provide evidence that our endemic signals are not flawed, we have now toned down (both in the abstract and main text) our statements on patterns of endemism.

Fifth, the authors estimated the community assembly processes for each mountain range. However, the authors used an out of dated approach to determine community assembly processes, which tends to overestimate selection and underestimate "drift". The recently developed more advanced approaches could be considered (Ning et al. 2020, Nature Communication, 11:4717).

AUTHORS: Thank you for raising this important point. While we kindly disagree that our initial approach is outdated, we followed the reviewer's suggestion and re-analyzed our dataset using iCAMP. Compared to our previous analyses (based on Stegen's QPEN; doi.org/10.1038/ismej.2013.93 combined with Fodelianakis' phyloscore analysis) in The ISME Journal, which used beta-MNTD, iCAMP uses beta-NRI. Beta-MNTD focuses on short phylogenetic distances, whereas beta-NRI focuses on mean (within-bin) phylogenetic distances, and hence reveals slightly deeper phylogenetic patterns (compared to beta-MNTD). As expected, and rightfully pointed out by the reviewer, this increases the contribution of "ecological drift" to overall community assembly. We are fully convinced that this is a more ecologically realistic representation compared to our previous analyses (also in view of the large unexplained variance in variance partitioning) and hence are very grateful for the reviewer's comment.

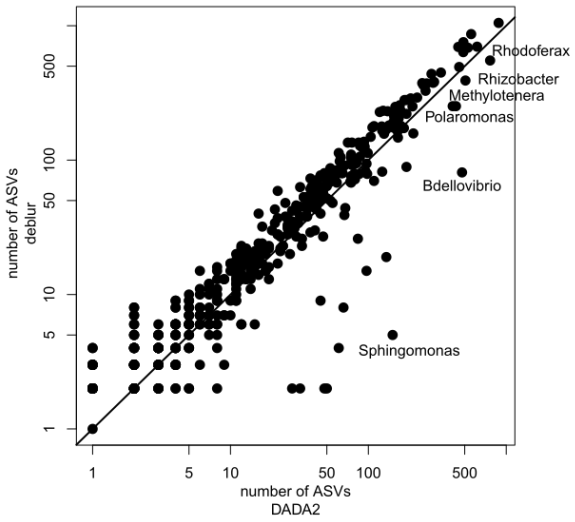


Figure. Shown are the number of ASVs for 481 genera with overlapping taxonomic assignment in datasets denoised using DADA2 and Deblur, respectively. Overall, Deblur identified slightly more ASVs (i.e. above the 1:1 line) compared to DADA2. However, for genera with particularly high contribution to relative abundance in GFS (i.e. *Rhodoferax, Rhizobacter, Methylotenera, Polaromonas*) DADA2 identified substantially more ASVs.

Moreover, as we re-analyzed our dataset using different ASV binning strategies (i.e. Deblur compared to DADA2), we noticed that phylogenetic clustering among certain genera was reduced. For instance, DADA2 resolved 764 and 506 ASVs within the genera *Rhodoferax* and *Polaromonas*, respectively. Deblur, resolved substantially fewer ASVs within these phylogenetically highly clustered genera (i.e. 551 and 391 ASVs, respectively). This effect was only apparent for genera with particularly short phylogenetic distances (including, besides the above-mentioned genera also the genera *Methylotenera*, *Rhizobacter* and *Bdellovibrio*). All other genera yielded slightly more ASVs using Deblur as compared to DADA2. Nevertheless, given the importance of these microdiverse clades (they are still highly phylogenetically clustered using Deblur) in terms of contribution to relative abundance, alpha and beta diversity in GFSs, we found that the choice of denoising algorithms can influence community-wide estimates of dominant assembly processes. We attribute the increased fraction of dispersal limitation (which was the second-most important assembly process in our initial analysis) to a combination of denoising algorithms and the differences in phylogenetic metric used by iCAMP and QPEN.

Sixth, the estimated drift is very low for most of the mountain ranges (undetermined in Extended Fig. 11), which is contradictory to db-RDAs and VPA analyses that showed at least

52% community variations could not be explained. Theoretically, the unexplained variations are generally due to unmeasured environmental variables and ecological drift.

AUTHORS: Thank you for this comment. Please also see our response above. We attribute this to the differences between QPEN and iCAMP - and the phylogenetic metric they are based on, respectively. Indeed, this contains interesting information - based on short phylogenetic distances among closest taxa across communities (i.e. beta-MNTD), communities often fall below -2 standard deviations of the null model expectations (hence indicative of HoS). This is different for mean pairwise phylogenetic distances within bins (i.e. beta-NRI), which more often fall between -2 and +2 compared to null model expectations (which in combination with an absence of taxonomic overdispersion/clustering is indicative of ecological drift). This is interesting because it highlights that across GFSs within mountain ranges there are indeed very similar (but not identical) community members (resulting in low beta-MNTD). This reflects one of the main findings of our work - taxonomically similar (and phylogenetically related) taxa dominate the global GFS microbiome, yet these ASVs are distinct, differing in their relative abundances across individual GFSs, therefore contributing to high dissimilarity. These taxonomically similar and phylogenetically closely related taxa also contribute to regional diversity and endemicity. Nevertheless, we agree that our initial estimates probably underestimated ecological drift, and that our revised analyses represent a more nuanced view of the assembly processes shaping the global GFS microbiome. We agree with the reviewer that unexplained variance from the variance partitioning models could result from environmental variables not included in the models and ecological drift, but also to variables not following linear trends and to biotic interactions between microorganisms (i.e. competition). This is now more congruent in the revised version.

In addition, the authors discussed the importance of horizontal gene transfer in shaping community structure. This is highly speculative because no experimental data are presented. To address this type of questions, as mentioned above, functional traits-based approaches are needed to determine the diversity and distributions of microbial functional genes.

AUTHORS: We agree with the reviewer that this is speculative indeed and it should have been conveyed as such. If not, please accept our apologies. This observation is purely based on theoretical considerations by Souza and colleagues (*Nature Review Microbiology;* doi.org/10.1038/nrmicro1917) and our data on resource stoichiometry, that suggest P limitation in our GFSs indeed. We have properly cited that paper and have now made clear that this represents theory. In fact, we find it interesting to discuss the potential link between resource stoichiometry and endemism as a concept (*sensu* Souza and colleagues) as we think that this could pave the way for exciting future research avenues.

Table. List of re-sequenced samples, samples from which DNA was re-extracted and sequenced, blanks and mock communities.

Glacier Number - Patch	Region	Status
GL161-1	Alaska	Newly Extracted and Sequenced
GL156-2	Alaska	Resequenced
GL156-3	Alaska	Resequenced
GL157-1	Alaska	Resequenced
GL158-1	Alaska	Resequenced
GL158-2	Alaska	Resequenced
GL158-3	Alaska	Resequenced
GL159-1	Alaska	Resequenced
GL159-2	Alaska	Resequenced
GL160-1	Alaska	Resequenced
GL160-2	Alaska	Resequenced
GL160-3	Alaska	Resequenced
GL161-3	Alaska	Resequenced
GL162-1	Alaska	Resequenced
GL162-2	Alaska	Resequenced
GL162-3	Alaska	Resequenced
GL163-1	Alaska	Resequenced
GL163-2	Alaska	Resequenced
GL163-3	Alaska	Resequenced
GL165-1	Alaska	Resequenced
GL165-3	Alaska	Resequenced
GL166-1	Alaska	Resequenced
GL166-2	Alaska	Resequenced
GL166-3	Alaska	Resequenced
GL167-1	Alaska	Resequenced
GL167-2	Alaska	Resequenced
GL167-3	Alaska	Resequenced
GL168-3	Alaska	Resequenced
GL169-1	Alaska	Resequenced
GL169-2	Alaska	Resequenced
GL101-1	Alps	Newly Extracted and Sequenced
GL31_1	Alps	Newly Extracted and Sequenced
GL96-3	Alps	Newly Extracted and Sequenced
GL99-3	Alps	Resequenced
GL146-3	Chile	Newly Extracted and Sequenced
GL66-3	Ecuador	Newly Extracted and Sequenced
GL35-2	Greenland	Newly Extracted and Sequenced
GL36-3	Greenland	Newly Extracted and Sequenced
GL35-1	Greenland	Resequenced
LM-1		Lab Mock community
LM-2		Lab Mock community
ZDNA-1		Zymo Mock community DNA
ZDNA-2		Zymo Mock community DNA
ZM-1		Zymo Mock community extraction
ZM-2		Zymo Mock community extraction
ZM-3		Zymo Mock community extraction
oldM-1		Zymo Mock community extraction
oldM-2		Zymo Mock community extraction

GL113-1	Nepal	Newly Extracted and Sequenced
GL113-2	Nepal	Newly Extracted and Sequenced
GL113-3	Nepal	Newly Extracted and Sequenced
GL6_2	New_Zealand	Newly Extracted and Sequenced
GL19_2	New_Zealand	Resequenced
GL19_1	New_Zealand	Resequenced
GL19_3	New_Zealand	Resequenced
GL87-1	Norway	Newly Extracted and Sequenced
GL87-2	Norway	Newly Extracted and Sequenced
GL87-3	Norway	Newly Extracted and Sequenced
GL88-3	Norway	Newly Extracted and Sequenced
GL89-3	Norway	Newly Extracted and Sequenced
oldBL-1		Blank
oldBL-2		Blank
BLEX-1		Blank
BLEX-2		Blank
BLEX-3		Blank

Revisions of Manuscript #2023-03-04207A-Z (Nature)
Ezzat et al. - Towards a global biogeography of the glacier-fed stream benthic microbiome

Dear Editor, dear reviewers,

Before we provide our detailed responses to the various constructive comments raised by the three referees below, we want to make three more generic points upfront.

First, the topic of glacier-fed stream heterogeneity has emerged several times during the review process, linked for instance to 'endemicity', representativeness of our data, etc. We, of course, acknowledge that heterogeneity is inherent to stream ecosystems in general. But we also acknowledge a tradeoff between increasing the number of replicate samples to better capture instream heterogeneity and samples to cover a broad geographical range - all of which come with more physical and analytical effort and resources. Initially, we focused our microbiome analyses on samples collected as close to the glacier snout as possible (referred to the 'upstream reach' hereafter). As written in the last version, the median distance between the upstream reach and the snout is 78 m. However, to better encompass the environmental gradient and related heterogeneity, we also sampled a 'downstream reach'. This latter reach was typically just upstream from the terminal moraine from the Little Ice Age, which we were mostly able to read from the landforms (n.b., our field team included experts trained to read high-alpine landforms). Accordingly, the median distance between the downstream reach and the snout is 773 m, while the median distance between the two reaches is 538 m.

To meet the criticisms related to capturing environmental heterogeneity and sample size, we have now included the sequencing (metabarcoding and shotgun metagenomics) data from all downstream sites as well. We note that the median dissimilarity between the two reaches is lower (Bray-Curtis Dissimilarity = 0.48) compared to median dissimilarity among GFSs from the same mountain range (Bray-Curtis Dissimilarity = 0.77). With this extra effort, we now present data from on average 5.5±1.4 sediment patches (i.e., ecological replicates) per GFS, almost doubling the previous dataset. Overall, this makes our data more representative (we cover more than 80% of expected regional diversity; please see below) and statistical analyses more robust. Notably, while some numbers have improved (e.g., sample size, number of ASVs), the overall patterns and main results reported throughout the manuscript remain unchanged.

The following table highlights how the increased sample size has affected some of the key numbers reported in the manuscript. All other changes can be seen from the submitted version with the changes highlighted (with the track change function).

Parameter	Initial	Revision	foldchange
Number of GFSs studied	149	152	1.02
Total ASVs retained	38983	54837	1.41
Projected proportion of observed ASV (based on asymptotic estamiates)	73.6	86.2	1.17
Alpha diversity	1801	2794	1.55
BC dissimilarity between mountain ranges	0.9	0.89	0.99
BC dissimilarity within mountain rangesranges	0.71	0.68	0.96
Number of indicator ASVs	12027	20522	1.71
AVSs specific to a mountain range (formerly endemic) (%)	58.4	62.2	1.07
Number of core ASVs	164	165	1.01
Variation partitioning, total variablity explained for all ASVs	50.5	54.9	1.09
Dispersal limitation as assembly process (%)	43.5	59.4	1.37
Ecological drift as assembly process (%)	39.8	30	0.75
Homogeneous selection as assembly process (%)	11.5	11.7	1.02
Beta diversity decreasing with phylogenetic resolution	0.395	0.383	0.97

Second, we were repeatedly asked to add data on the GFS microbiome function, which we had already done to some extent in the previous version. Given that the focus of this manuscript is on the biodiversity and biogeography of the GFS microbiome, and given both the complexity and length needed to address this research question alone, we purposely limit functional insights to a relatively high level.

We also want to make the point that we currently have a separate manuscript focusing on thousands of prokaryotic, eukaryotic, and viral metagenome assembled genomes from our GFSs that is currently revised for a major microbiology journal. We have decided to share with you this manuscript — understandably at your discretion. Hopefully, this will allow you to share with us our opinion that it does not make sense to integrate the two manuscripts into one. For your information, we are also currently preparing two additional MAG-centric manuscripts.

Third, it is in the nature of each review round that revisions reflect referees' criticisms, typically making manuscripts longer. To counter this, we have removed the short discussion on the potential role of horizontal gene transfer and its theoretical link to phosphorus limitation (btw, also criticized by referee #3). Furthermore, we have removed the variance partitioning of the core microbiome. Where we encouraged by referees to elaborate on some arguments, we have done so in the most concise way possible.

We remain committed to meeting any further requirements or suggestions to ensure the review process.

Referee #1 (Remarks to the Author):

My expertise lies more in glaciers and glacier runoff, and I thus comment mainly on these aspects.

My concerns regarding the initial version of the manuscript have been addressed to a large extent and the glacier-related aspects of the paper have improved.

For the main text there remains one question in my view. The authors agree with me on the highly dynamic nature of the glacial environment (daily + seasonal runoff, runoff peaks/floods, debris flows, sediment reworking, glacial action, etc.). To what extent can the (cumulative) history of such actions have influenced the microbial communities found? For instance, the timing of glacier retreat from the sample location. How long is the sample location free of glacier cover? Is there are history of heavy stream reworking, not necessarily easily visible on site? Etc. I assume such (cumulative) history can be very different from site to site. I recommend a sentence or two discussing the potential influence of such environmental history on the study results, or why it is expected to not play a role, resp.

AUTHORS: We thank the reviewer for her/his overall positive evaluation of our revision. We have tried to include several of the suggestions into the text without becoming too explicit for the sake of focus and word count. For instance, we now state in the last paragraph of the Introduction (and associated Methods) that we sampled two reaches per GFS — the upstream reach specifically was sampled as close to the glacier snout as possible (median distance 78 m across all studied GFSs). This reach therefore represents relatively recently deglaciated terrain. A downstream reach was sampled just upstream of the terminal moraine of the Little Ice Age (wherever we were able to detect it from landscape forms, acknowledging that the Little Ice Age was not a globally coherent phenomenon). Nevertheless, this description provides an implicit notion of the time brackets that we encompass.

Furthermore, the paragraph entitled ‘The glacier-fed stream environment’ now better illustrates that annually recurring meltwater peaks scour the unstable GFS streambed, etc. However, it is not possible to explicitly assess whether there were cumulative effects of such events and how they may vary from site to site. If adequate at all, geomorphological surveys would have been required to address this point, which would not have been feasible in the field due to logistical constraints and given the focus of our study.

This reviewer also mentions the potential for legacy effects from streambed reworking, for instance, on the microbial communities. This is a valid point, and our understanding is that any such effects would depend upon the predictability (along with the frequency and magnitude) of such events. For example, for discrete

hydrological/geomorphological events to act as environmental filters, even agents of evolution, they would have to happen frequently and predictably enough, which would run against their nature as discrete events. For example, debris flow events are typically unpredictable, discrete events in time, in which populations and communities typically cannot adapt to. Therefore, it is hard to imagine that discrete events in the history of a stream would imprint upon microbial communities. But, if predictable (and occurring frequently enough), then they would be linked to the intra- and inter-annually recurring ice melt inherent to GFSs (and thus part of the GFS flow regime, and hence ecosystem phenology), in which species and communities can adapt to. It is worth noting that our phylogeographic analyses implicitly encompass (besides dispersal, drift, and speciation, etc) some of the environmental selection, partially related to hydrological (e.g. flow regime) and geomorphological legacies. In fact, the GFS flow regime could be a structuring mechanism for the microbiome, and perhaps contributes to the observed homogenizing selection in phylogenetic analyses. Yet, we prefer not to discuss this in the manuscript in depth due to being rather speculative in nature.

What does appear interesting in this context, however, is that (i) benthic bacterial cell numbers in GFSs are relatively high compared to other extreme sedimentary environments (please see reviewer #2) and (ii) the benthic bacterial community composition differs from the assemblage transported in the streamwater. Thus, the benthic zone comprises a distinct niche across numerous GFSs, independent of 'disturbance history', but perhaps evolved to the specific hydrologic conditions of GFSs.

Supp. Inf.:

There is a parameter "permafrost" in several tables that I don't find explained. In the main text it is referred to as "permafrost soil". How is it defined? From which data? What would be its influence on microbial communities? You typically don't have permafrost under glaciers.

AUTHORS: These SI tables report statistics for comparisons between GFS and other cryospheric habitats, including permafrost soils. These other cryospheric habitats (glacier ice, snow, cryoconite, etc) were not sampled by our group; but refer to microbial community sequence data retrieved from repositories for use in these comparisons and utilized in a previous meta-analysis (Bourquin et al. 2022, Nature Communications). We have now changed "permafrost" to "permafrost soil", and added a clear explanation to these tables (see also comment below).

Tab2: what refers "glacier" to? Area? %? Similar for snow. I find Tab2 and parameters used too little explained.

AUTHORS: We believe that this comment refers to Supplementary Information (SI) Table 2 regarding the anova.manyglm analysis testing the effects of ecosystem on community structure. Similar to SI Table 4, these categories refer to various other cryospheric habitats. Specifically, “glacier” refers to glacial ice communities. We agree that these tables lacked a clear explanation, which we have improved in the revised version.

Referee #2 (Remarks to the Author):

The authors have done a lot more work in attempts to improve their manuscript, which includes new mock community analysis, analysis of blanks, and resequencing of many samples. This extra effort is appreciated, and I recognize the hard work that they have put in to make the manuscript better. However, there unfortunately still remain points that need to be clarified from the analysis of the blanks and the mock community which I provide below. I know this may seem like a moving target, but I assure the authors it is not. Rather, I am encouraging the authors to take necessary steps to ensure that the data are of highest quality (presumably the authors also want this as well). As a global dataset, I think it meets the high mark required for Nature. But, I do not think that the authors claims for endemicity can be proven with their data - and therefore in its current form I cannot support the manuscript as it stands for publication. If the authors are willing to change their stance on this, I would be more positive. I think it comes down to what the word "endemic" means, which I discuss in more detail below.

AUTHORS: We are most grateful for the overall positivity of this review. As detailed below, we have now done our best to fully address the points raised by this reviewer. May we therefore guide her/his attention directly to our detailed responses below.

Here is why the authors data do not support conclusions about endemic taxa: the authors write on lines 150-151 that their analysis predicts that they only sampled 73.6% of the total ASVs present in the mountain ranges. That means that about 25% of the taxa were not detected. It is therefore entirely possible that some ASVs are not actually endemic to a specific geographic region as the authors claim, but rather simply not detected in other geographically separated mountain ranges (because they are part of the undetected 25%). Surely the authors know this, and would agree with this point. I suggest to remove all claims of endemicity from the manuscript, because this is not possible to prove unless all taxa have been found in all samples (not the case here). If the authors publish this dataset in Nature, and make claims about endemicity when only 75% of ASVs are detected, this will surely result in a critical response from the scientific community. The authors define "endemic ASV" on lines 231-232 as "being specific to a single mountain range". I suggest changing the word "endemic" to "specific" throughout the entire manuscript, that might be an easy way to soften these strong conclusions.

AUTHORS: We agree with this point, and now abstain from using the word 'endemic' in the manuscript. Rather, as suggested by the reviewer, we instead refer to 'ASVs specific to a mountain range'. At the same time, we would like to point out that the inclusion of additional data increased the number of specific ASVs, along with the expected coverage of regional gamma diversity to 86.2%. Therefore, the underlying statistics used in identifying 'specific ASVs' are more robust now than before. We strongly believe that "specificity" is not merely a consequence of not detecting ca. 25% (or, now, 14%) of the expected regional diversity, as further highlighted by our sensitivity analyses (please, see Extended Data Figs. 12-14). While specific ASVs represent 62.2% of all ASVs, they contribute 9% (median) of the relative abundance across mountain ranges . Yet, they are consistently present among GFSs within the same region. Based on our extrapolations of expected regional diversity, we failed to sample (only) 14% of the ASVs. These ASVs are almost certainly of low relative abundance, and hence unlikely to be present across multiple regions (and thus reducing the degree of specificity). Notably, this extrapolation of gamma-diversity does not account for the effect of increased sampling of GFS. However, *in silico* varying sampling effort (i.e. number of GFS) revealed that sampling effort does not substantially affect estimates of mountain-range specificity. Moreover, the fact that nearly doubling the number of patches, and at the same time observing an increased coverage of regional diversity and a relatively stable number of specific ASVs, highlights the relative homogeneity of GFS communities.

Related to the endemicity topic, on lines 253-260 (and Figure 3a) If you only sampled each GFS 3 times (and some even less), I don't think you can statistically show that ASVs detected in only one GFS is truly specific to that GFS (more sampling is needed within the GFSs to show this).

AUTHORS: As described upfront to these detailed replies, we have now increased the number of ecological replicates (i.e., sediment patches) by including a maximum of three more replicates from a downstream reach, increasing statistical power. In the same spirit, we filtered out six GFSs, which had less than 3 total ecological replicates. Overall, we now have on average 5.5±1.14 ecological replicates per GFS. Furthermore, adding samples from a downstream reach allowed us to increase the number of studied GFSs from 149 to 152, where previously the amount of reads in the samples was either too low, or the DNA extraction, amplification, and/or sequencing had failed. In addition, our analyses, which leverage the occurrence of ASVs across biological replicates from the same GFS (i.e., an ASV is retained only if it is present at least in 2 out of 6 independent replicates), limit the number of spuriously "specific" ASVs (i.e., ASVs that are only detected in 1 or 2 ecological replicates of a given GFS). As for unique ASVs, please note that we have consistently found hundreds of them in all ecological replicates (i.e., 100% catch per unit effort) within a given GFS. Independent of this, we do not use the term 'endemicity' anymore.

I disagree with Reviewer 3 that the results are incremental, to my knowledge this is the first global dataset on GFSs with data spanning all of the major glacier covered mountain ranges on all continents. As a global dataset, I think it meets the high mark required for Nature. I agree with Reviewer 3 that functional, or ecological information is the end goal of any microbial ecology study, and I think that the added metagenomic data in figure 1 strengthen the study even more. I agree with Reviewer 3 original comments, that the original methodologies were standing on weak ground. However, the authors have reanalyzed their data with multiple pipelines and apparently obtained similar results. I don't think the main results are a spurious result of clustering algorithms. There are still some things to check however with the mock community and inadequate sequencing depth of the blanks (see below for details).

AUTHORS: We are appreciative of this comment and agree that our revision has produced a solid and robust (i.e., not spurious) data set based on rigorous methods - even more so now after including downstream samples. Please see more detailed replies below to the point related to novelty, mock communities, and sequencing depth.

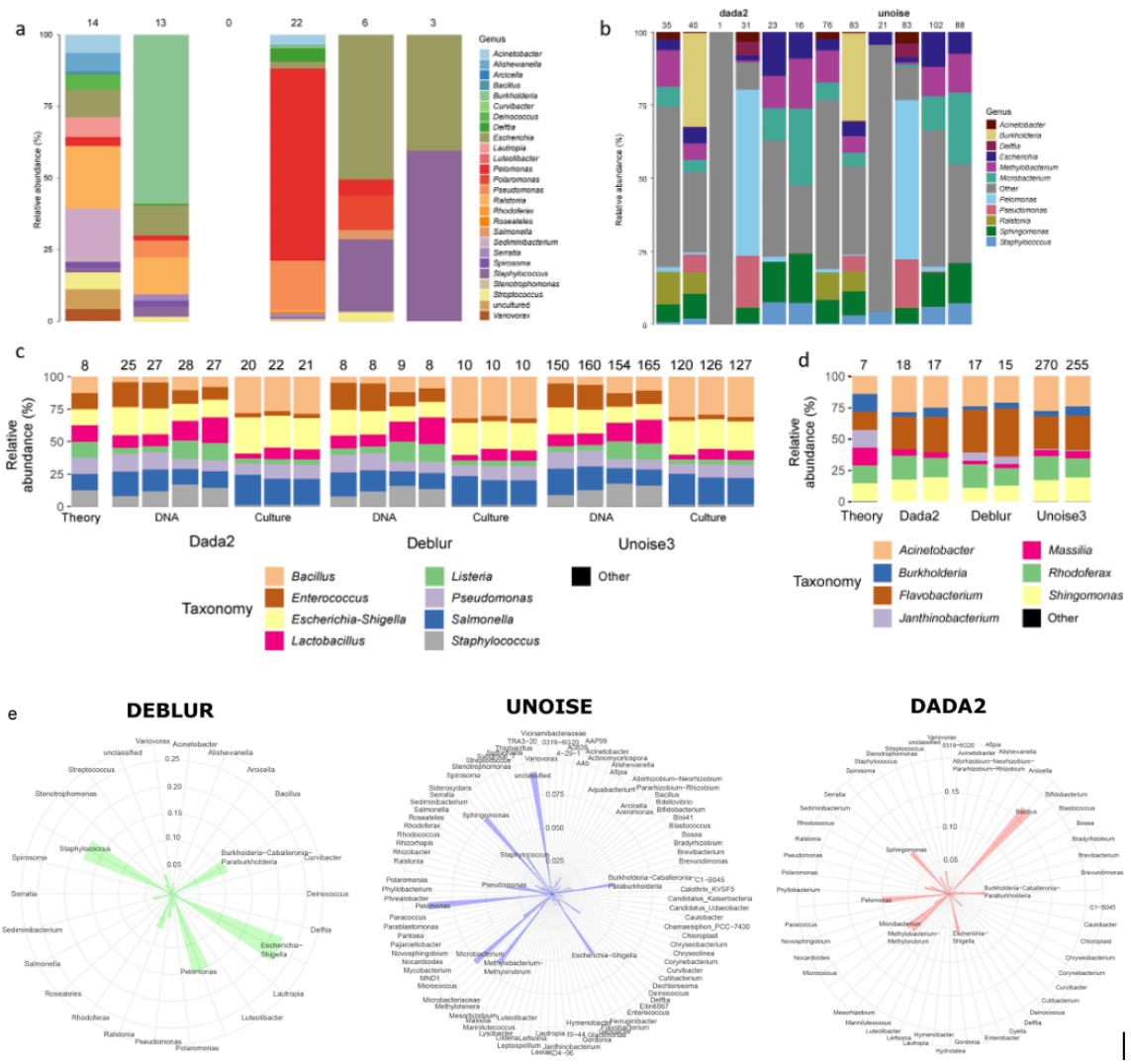
Where I agree most strongly with Reviewer 3, is in our shared critical opinion on the authors use of the word "endemism". I agree with Reviewer 3 that the GFS sampling could have been better (3 samples per GFS is a bit low and some GFSs had less than 3 samples). This is another reason, that precludes the authors from drawing conclusions about endemicity (maybe if you took more samples, you would have found the 'endemic' taxa in other mountain ranges) - I think Reviewer 3 and I are in agreement on that. However, I get the impression that the authors have a different understanding of the word "endemic" than I do (and apparently Reviewer 3 as well). Their results show that many GFS ASVs are highly specific to certain geographic areas, and it might be a matter of them just changing the word "endemic" to "specific" (or something similar). The authors cannot prove with their data that ASVs are not present in different mountain ranges because they only detected on average 75% of the ASVs (lines 150-151). The authors must know this and I think should therefore agree.

AUTHORS: We agree and thank the reviewer for this constructive suggestion. As discussed above, we replaced the term ‘endemic’ with ‘specific’ throughout the manuscript. Doubling the sample size makes the statistics underlying the identification of specific ASVs more robust. It is worth noting that the larger sample size increased coverage of gamma diversity and slightly higher number of specific ASVs, which occur at relatively low abundances.

Here I provide my comments on the new results of the mock community and the blanks:

Comments on the mock community results:
It is good that the authors finally included a mock community here, in Figure S13. However, the authors need to show the total number of ASVs detected using the different algorithms compared to the true number of species in the mock community. Does the number of species in the mock community equal the number of ASVs after sequencing ?
This is a critical point the authors need to address. The stacked histogram they presented does not show this.

AUTHORS: We agree with the reviewer that this is an important point, and have now included detailed results from the sequencing of mock communities in Extended Data Figures 17 and 19. In the commercial mock communities, we expected 8 ASVs, whereas for our in-house mock community we expected 7 ASVs (that is, 7 strains isolated from a glacier-fed stream in Switzerland). All of the expected genera were found in our datasets, with the proportion of relative abundance between taxa comparable between the three analysis algorithms. We are aware that those proportions deviated to some extent from the theoretical expected, especially between DNA and culture communities, probably due to biases introduced during purification and amplification. Such biases are unavoidable, and indeed prior to this analysis we invested largely in optimizing extraction protocols to reduce such biases as much as possible (10.7717/peerj.9973). Comparing the different algorithms using the commercial mocks we find that Deblur retained 8 to 10 ASVs, DADA 20 to 27, and UNOISE 120 to 165 ASVs. Based on the fact that we only expect 8 ASVs in the commercial communities (and we find only a few reads coming from contamination, see response below) we decided to utilize Deblur for analysis of our samples - as discussed in the first round of review. Similarly, for the in-house mock community, Deblur only retained 15 and 17, whereas DADA retained 17 and 18, and UNOISE 255 and 270 ASVs. Overall, all ASVs with a high number of reads (>100) belonged to the mock strains. This also underlines the conservative nature of the denoising algorithm (i.e., Deblur), which does not inflate ASV richness.



Extended Data Figure 17. Taxonomic composition of (a) blanks analysed with deblur with all the genera assigned and (b) with DADA2 and UNOISE with an "other" category. (c) Taxonomic composition of commercial mock communities obtained both as DNA (DNA, Zymo, Cat. Nr. 6305) and cell cultures (Culture, Zymo Cat. Nr. 6300). (d) Mock community created based on isolates obtained from Swiss GFS. Taxonomic compositions were computed using the three denoising approaches (DADA2, Deblur and UNOISE3). Numbers above the bar plots correspond to the number of ASVs found in each community. (e) Visual representation of ASVs detected in blanks and retained by the three different denoising algorithms (Deblur, UNOISE3 and DADA2). These circularized barcharts show the mean relative abundance of all genera detected in the blanks. Abundant genera are highlighted (moved inside). Note that for UNOISE, only the 100 most abundant genera could be visualized and that GFS taxonomies, including Polaromonas, Rhodoferax and Methylotenera, were either absent (Methylotenera) or of very low relative abundance in these blank samples.

The authors say that in the mock community they "did not find any signature of contamination" what does this mean exactly? Are the authors claiming that no other ASVs other than the mock community were detected? This is very hard to believe, since contaminating ASVs in mock communities are always detected at low relative abundance in sequencing runs (see papers by Robert Edgar).

AUTHORS: We thank the reviewer for this comment and apologize for the apparent confusion. Regarding the mock communities, we sequenced seven commercial (Zymo) and two lab-internal communities for mocks. Four of the commercial mocks were obtained as pure DNA (i.e., only requiring library preparation prior to sequencing) while three of them were in the form of cells for which the DNA had to be extracted following the same approach as our GFS samples (i.e., requiring DNA extraction and sequence library preparation prior to sequencing). The lab internal mock communities were prepared from pure cultures obtained from a Swiss GFS and culturing at 10°C in R2A liquid media. For taxonomic identification of these mock-strains, DNA was extracted, the 16S rRNA gene amplified) and Sanger-sequenced for identification (using the 27F and 1492R primer pair for amplification and Sanger sequencing in both directions). To obtain mock communities, strains were grown in R2A medium and optical density (OD600nm) was monitored. After reaching a plateau, the cultures were diluted to the same OD600 and mixed to obtain the mock community. DNA was then extracted from this mock community using the same protocol as for the GFS. DNA extracted from the different mock communities were included in our Illumina NovaSeq libraries and sequenced along with our samples. The results are shown in Extended Data Figures 17 and 19. Using Deblur, we detected between 8 to 9ASVs out of the expected eight prokaryotic strains in the four replicates of the DNA mock and 10 ASVs for all three replicates of the cultured mock. All ASVs belonged to the mock strains except one associated to the genus *Paenibacillus* and one unknown Gallionellaceae, although both contaminations had only very few reads of less than 1% compared to the lowest reads for an expected community member (Extended Data Figure 19 a). On the one hand, this underlines the conservative nature of the denoising algorithm (i.e., Deblur) which does not inflate richness. Regarding our lab “isolate mock” communities, the conclusion was similar, with 17 ASVs classified to the expected 7 taxa. We agree that this needs to be reported more explicitly and hence include the results of the mocks and blanks in the Extended Data Figures. 17-19.

		Zymo DNA								Zymo Culture					
		ZDNA1		ZDNA2		ZDNA3		ZDNA4		ZCult1		ZCult2		ZCult3	
Taxa	InMocks	Reads	ASVs	Reads	ASVs	Reads	ASVs	Reads	ASVs	Reads	ASVs	Reads	ASVs	Reads	ASVs
Bacillus	Zymo	8669	1	5815	1	3804	1	3277	1	22336	1	21144	1	26038	1
Enterococcus	Zymo	7902	1	6968	1	16683	1	13588	1	2484	1	2683	1	2869	1
Escherichia-Shigella	Zymo	8461	1	7870	1	15565	1	11307	1	17330	1	14927	1	17089	1
Lactobacillus	Zymo	11111	1	13370	1	7933	1	5986	1	2709	1	6158	1	6251	1
Listeria	Zymo	11239	1	9276	1	4091	1	3150	1	2788	1	2959	1	3511	1
Paenibacillus		0	0	0	0	0	0	0	0	6	1	5	1	11	1
Pseudomonas	Zymo	5459	1	5497	1	10723	1	8456	1	5780	1	7488	1	8740	1
Salmonella	Zymo	8074	2	8078	1	14774	1	10267	1	15672	2	13387	2	15391	2
Staphylococcus	Zymo	11497	1	8922	1	6127	1	7356	1	679	1	794	1	929	1
Unknown Gallionellaceae		0	0	2	1	0	0	0	0	0	0	0	0	0	0

Extended Data Figure 19. a - Table representing the genera found in the commercial mock community standards from Zymo. The reads and ASV columns correspond to the number of reads and ASVs assigned to the genera, respectively. The gradient of blue colors represents the number of reads assigned to each genera, while the gradient of color from purple to yellow indicates the number of ASVs found for each genera. White indicates the absence of (0) reads and ASVs in some samples.

		LM1		LM2	
Taxa	InMocks	Reads	ASVs	Reads	ASVs
Acinetobacter	Lab	20859	1	15294	1
Burkholderia-Caballeronia-Paraburkholderia	Lab	1934	1	2236	2
Flavobacterium	Lab	20129	2	18261	2
Janthinobacterium	Lab	3577	2	2866	2
Massilia	Lab	1847	2	1526	1
Rhodoferax	Lab	11307	2	7518	2
Sphingomonas	Lab	267	9	319	7

Extended Data Figure 19. B.Table representing the genera found in the mock samples from cultured bacteria obtained from a Swiss GFS. The reads and ASV columns correspond to the number of reads and ASVs assigned to the genera, respectively. The gradient of blue colors represents the number of reads assigned to

each genus, while the gradient of color from purple to yellow indicates the number of ASVs found for each genera.

	BL1		BL2		BL3		BL5		BL6		BL4	
Taxa	Reads	ASVs	Reads	ASVs	Reads	ASVs	Reads	ASVs	Reads	ASVs	Reads	ASVs
Acinetobacter	12	1	0	0	0	0	0	0	0	0	54	3
Alishewanella	12	1	0	0	0	0	0	0	0	0	0	0
Arcicella	0	0	0	0	0	0	0	0	0	0	3	1
Bacillus	2	1	0	0	0	0	0	0	0	0	0	0
Burkholderia-Caballeronia-Paraburkholderia	0	0	382	2	0	0	0	0	0	0	0	0
Curvibacter	0	0	0	0	0	0	0	0	0	0	11	1
Deinococcus	10	1	0	0	0	0	0	0	0	0	0	0
Delftia	0	0	4	1	0	0	0	0	0	0	70	2
Escherichia-Shigella	18	1	68	1	0	0	46	1	15	1	34	1
Lautropia	13	1	0	0	0	0	0	0	0	0	0	0
Luteolibacter	0	0	0	0	0	0	0	0	0	0	2	1
Pelomonas	6	1	11	1	0	0	5	1	0	0	983	1
Polaromonas	0	0	0	0	0	0	11	1	0	0	0	0
Pseudomonas	0	0	38	1	0	0	0	0	0	0	253	3
Ralstonia	41	1	84	2	0	0	0	0	0	0	7	1
Rhodoferax	0	0	0	0	0	0	0	0	0	0	2	1
Roseateles	0	0	0	0	0	0	0	0	0	0	8	1
Salmonella	0	0	0	0	0	0	3	1	0	0	0	0
Sediminibacterium	35	1	0	0	0	0	0	0	0	0	0	0
Serratia	0	0	14	1	0	0	0	0	0	0	19	2
Spirosoma	4	1	14	1	0	0	0	0	0	0	0	0
Staphylococcus	3	1	23	2	0	0	23	1	22	2	6	1
Stenotrophomonas	0	0	0	0	0	0	0	0	0	0	7	1
Streptococcus	11	1	10	1	0	0	3	1	0	0	4	1
Unknown Alcaligenaceae	0	0	27	1	0	0	0	0	0	0	0	0
Unknown Micrococcaceae	6	1	8	1	0	0	2	1	0	0	38	1
Variovorax	8	1	0	0	0	0	0	0	0	0	4	1
uncultured	13	1	0	0	0	0	0	0	0	0	0	0

Extended Data Figure 18. Table representing the genera found in the blank samples. The reads and ASV columns correspond to the number of reads and ASVs assigned to the genera, respectively. The gradient of blue colors represents the number of reads assigned to each genus, while the gradient of color from purple to yellow indicates the number of ASVs found for each genera.

Comments on the blank results:
It is a good start, that the authors included blanks to check for contaminants. However, they state in the rebuttal that they only obtained an average of 486 reads per blanks. This is very, very low, (the authors must know this) and the blanks need to be re-sequenced to a depth at least similar to their samples before it can be compared. Its entirely possible if the blanks are sequenced to the same depth as the samples (>20,000 reads) then more ASVs are found. The current extremely low sequencing depth of the blanks makes it impossible to reliably compare to the samples.

AUTHORS: We agree with the reviewer that controlling for contaminants is crucial in microbiome research, particularly when samples contain very low biomass (doi. 10.1038/s41564-018-0202-y). Similar to our explanation above for the mock communities, the number of reads coming from contamination was extremely low in our blanks (i.e., DNA-free water “no-input” DNA extraction). We explain that by the fact that all blanks obtained in different extraction and sequencing runs contain marginal DNA (< 0.05 ng/µl) and 16S PCR amplification failed to yield any visible bands on agarose gels (please see figure below with gels). Because of failed PCR amplification, hence the quasi absence of DNA to be sequenced, sequencing can simply not happen at the same depth as other samples, which contain more DNA — please see also the next comment below. Furthermore, resequencing the blanks may create a high number of spurious sequences and sequencing artifacts that will be removed from the preprocessing steps of our bioinformatic pipeline and not generate more “exploitable reads”. Thus, it is expected that the number of reads in the blanks (486 ± 587) was much lower than in the actual samples (46,167 ± 15,805) with a total of 43 ASVs representing 26 genera. Those ASVs represented only 0.6 % of the total relative abundance of the GFS microbiome. The fact that blanks contain very few reads with ASVs that are present in less than 1% of the total microbiome was also observed previously in multiple papers such as Reitmeier et al., (2021, ISMEJ com, 10.1038/s43705-021-00033-z), Kim et al., (2017, Microbiome 10.1186/s40168-017-0267-5), Hornung et al., (2019, FEMS Microb Ecol, https://doi.org/10.1093/femsec/fiz045) and Díaz et al., (2021, mSphere, 10.1128/msphere.00506-21). Therefore, our blanks indeed show that very little contaminants are introduced during sample preparation. This is further confirmed by the fact that three different Mock communities (Zymo DNA, Zymo Mock, in-house mock), with three replicates for each, contain almost exclusively the expected ASVs and very few contaminants. In fact, one can argue that these mock communities serve as a better contamination control since they were not only processed in parallel to other samples, but similar DNA quantities were added to the sequencing libraries and they were sequenced to similar depth.

In the blank sequencing results (SI figure 13), please update the stacked histogram plot currently with the "other" label by showing all the taxonomic groups that were detected. In their rebuttal the authors state they only found 43 different ASVs. These

should be possible to group into the different genera as colors and plot. Readers deserve to know what was found in each blank sequencing replicate and what the "others" actually are (Polaromonas for example). I am very curious what is in the "others" particularly since some of the blank samples have almost 100% "others", and the authors write in the rebuttal that they found one Polaromonas ASV but don't label it in this figure but rather group it with "others". This is a bit conspicuous.

AUTHORS: We agree - this is an important point and we apologize if our choice of representation was not clear. We revised this Figure to show all 26 genera detected in blanks using Deblur. Due to the high number of genera retained in the blanks by DADA2 (n=105) and UNOISE (n=383), we still use an “other” category in these stacked barcharts. Nevertheless, the most abundant taxonomies detected by Deblur in the blanks included *Escherichia-Shigella* (accounting on average for 22.3% of reads across all blanks), *Staphylococcus* (17.9%), *Pelomonas* (15.1%), *Burkholderia-Caballeronia-Paraburkholderia* (11.2%), Ralstonia (6.8%), *Pseudomonas* (4.4%), unclassified ASVs (3.9%), *Sediminibacterium* (3.6%) and *Polaromonas* (2.4%). However, while *Escherichia-Shigella* and *Staphylococcus* were found in all blanks, *Polaromonas* was present in a single blank (with 11 out of 93 reads, i.e. 11.8%). This was similar when using other denoising algorithms. To better illustrate this, we included novel Figures showing the relative number of reads recruited by all genera for blanks analyzed using Deblur (See Extended Data Figure.17e). For blanks processed with DADA2 and UNOISE, we still had to limit this to the 100 most abundant genera to somehow maintain readability. We further note that the ASVs present in the blanks represented only 0.6% of relative abundance of the total community as indicated in the previous response to reviewers.

Unoise retained 383 zOTUs across all blanks, which were assigned to 158 genus-level taxonomies. The most abundant genera included *Pelomonas* (9.3%), *Microbacterium* (7.7%), *Sphingomonas* (7.5%), *Methylobacterium-Methylorubrum* (6.5%), *Escherichia-Shigella* (5.3%), *Burkholderia-Caballeronia-Paraburkholderia* (4.7%) and *Pseudomonas* (3.5%). *Rhodoferax, Methylotenera, Thiobacillus, Polaromonas* and *Massilia* which were commonly detected in GFS around the world were also found in the blanks (using UNOISE), accounting on average for 1.4%, 1.3%, 0.6%, 0.6%, and 0.5% of reads, respectively.

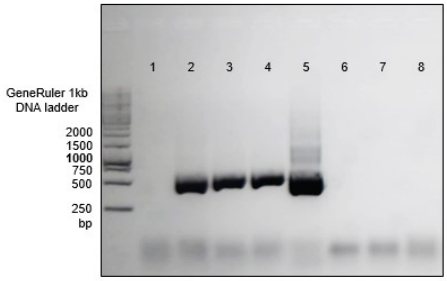
DADA2 retained 105 ASVs, with *Bacillus* (16.7%), *Pelomonas* (10.0%), *Sphingomonas* (8.3%), *Microbacterium* (7.8%), *Methylobacterium-Methylorubrum* (7.6%), *Escherichia-Shigella* (5.7%) and *Burkholderia-Caballeronia-Paraburkholderia* (5.1%) accounting for most of the reads in the blanks. *Polaromonas* accounted on average for 0.6% (detected in a single blank sample).

I understand, that it may be a possibility (but the authors didn't point it out) that the blanks are so clean, as to only produce a faint band on a gel resulting in the prepared libraries being extremely low concentration and did not sequence well. That

would explain the extremely low sequencing depth for the blanks. Can the authors explain this, or at least show a picture of the gel with the blank compared to the samples?

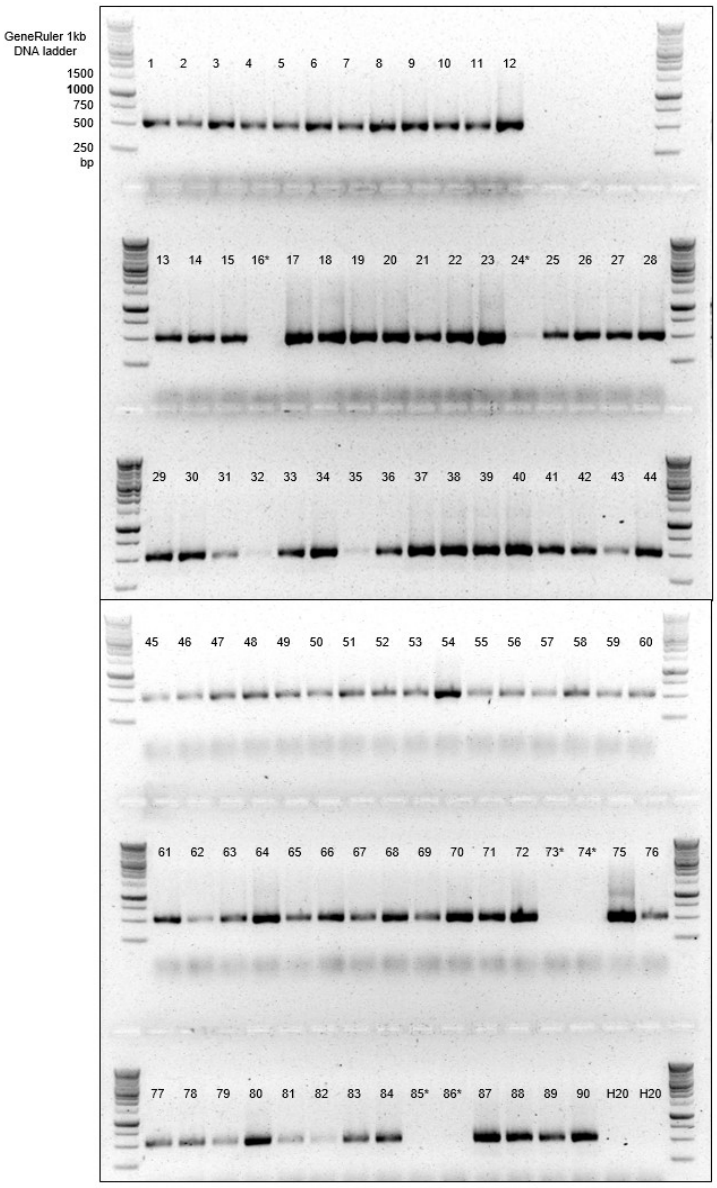
AUTHORS: We agree, this is an important point for which we are grateful. Please, see also our response above. Prior to the expeditions, we spent considerable time and resources on optimizing both field and laboratory procedures to maximize yields and minimize contamination (e.g., flame-sterilization of all field equipment *in situ*, large-volume kit-independent DNA extraction and purification, etc). These efforts are reflected in the very low DNA concentrations in our blanks (<0.05 ng/µL using Qubit High Sensitivity dsDNA assay Cat. Nr. Q32854) and the absence of visible bands on agarose electrophoresis gels after PCR amplification compared to the positive mock controls (commercial and in-house mock communities) and samples with sufficient quantity and quality of DNA. To substantiate this, we here included the gels for your appreciation from the Zymo mock DNA and mock communities, as well as respective blanks, and all 90 PCR amplifications of GFS sediment samples from our last expedition to Alaska. Upon request, we can provide gels from the other extractions (we would have to retrieve them from the lab journals) and include them to the Extended Data or Supplementary Information. As the reviewer concluded, blanks were of very good quality and contained very little DNA that was not seen on gels. Furthermore, the blanks were sequenced in the same lane as the actual samples, which contained on average 46,167 ± 15,805 reads, while blanks contained (as stated before) 486 ± 517 reads.

a



16S (V3-V4) PCR gel for controls:
1: negative H20 control for PCR
2: Zymo Mock DNA 1 (Cat. Nr. D6305)
3: Zymo Mock community 1 (Cat. Nr. D6300)
4: Zymo Mock community 2 (Cat. Nr. D6300)
5: Lab-made Mock community from isolates 1
6: Blank extraction 1
7: Blank extraction 2
8: Blank extraction 3

b



16S (V3-V4) PCR gel for Alaska samples:
1-3: Glacier **156** up1-3
4-6: Glacier **156** down1-3
7-9: Glacier **157** up1-3
10-12: Glacier **157** down1-3
13-15: Glacier **158** up1-3
16-18: Glacier **158** down1-3
19-21: Glacier **159** up1-3
22-24: Glacier **159** down1-3
25-27: Glacier **160** up1-3
28-30: Glacier **160** down1-3
31-33: Glacier **161** up1-3
34-36: Glacier **161** down1-3
37-39: Glacier **162** up1-3
40-42: Glacier **162** down1-3
43-45: Glacier **163** up1-3
46-48: Glacier **163** down1-3
49-51: Glacier **164** up1-3
52-54: Glacier **164** down1-3
55-57: Glacier **165** up1-3
58-60: Glacier **165** down1-3
61-63: Glacier **166** up1-3
64-66: Glacier **166** down1-3
67-69: Glacier **167** up1-3
70-72: Glacier **167** down1-3
73-75: Glacier **168** up1-3
76-78: Glacier **168** down1-3
79-81: Glacier **169** up1-3
82-84: Glacier **169** down1-3
85-87: Glacier **170** up1-3
88-90: Glacier **170** down1-3
H20: negative H20 control for PCR
*samples failed to amplify

Representative results of agarose gel electrophoresis analysis of 16S rRNA PCR amplifications. **(a)** Representative results of positive controls of commercial and lab-made Mock communities as well as of PCR negative controls (just water) and products of blank DNA extractions. **(b)** Results for the 90 GFS sediment samples from the Alaska Range as a representative example. PCR analysis from other

regions yielded similar results. Note that samples which did not amplify in this run (indicated with asterisk) were re-extracted and the PCR was repeated. The expected size of the V3-V4 of the 16S rRNA gene is 530 to 550 base pairs (bp).

More general comments on the role of (high) microbial biomass for the interpretation of the authors results. I think this is an important issue that the authors should discuss in more detail in regards to interpreting their data.

lines 113-114: I disagree with this statement. The biomass of the GFS (10^5 - 10^6) that is shown in Table S1 is actually quite high in my opinion, because it is comparable to seawater where you have about 1 million cells per mL. Compared to more extreme sedimentary environments like the deep subseafloor biosphere (where cell counts are as low as 10^2 per gram, (Kallmeyer et al., 2012 PNAS; Inagaki et al., 2015 Science), the GFSs cell counts the authors show in table S1 are very high. I would therefore not classify these GFSs as "low biomass". These high cell counts actually support the authenticity of the authors 16S data, because at these high cell concentrations contamination should not be a major issue. Indeed, the authors refer to stream "biofilms" throughout the manuscript. You cannot have a biofilm, and low biomass. Biofilms are by definition high biomass. The authors need to correct this in the manuscript.

AUTHORS: We are most grateful for this comment and have fully implemented the suggested changes. Indeed, not so low cell numbers are indicative of a bacterial niche, which is further supported by our previous observations that their community composition differs from the composition of the bacteria transported in the streamwater. We have now highlighted this upfront when introducing the GFS microbiome diversity. We apologize if our use of term ‘biofilm’ insinuated high-biomass systems; in fact, we follow the understanding of HC Flemming and S Wuertz of what a biofilm is (please see, http://doi.org/10.1038/s41579-019-0158-9) - also including few cells attached to sediments or inhabiting their porous space.

lines 519-521: Is this the first time you used this sonication and flow cytometry approach to count bacterial cells from sediments? If you've done it before please cite the protocol here. I would expect that it underestimates cell counts by at least an order of magnitude since not all cells will become detached from the sediment grains. It would be good to at least discuss this important point. e.g., that the sampled ecosystems are actually not that low biomass, 10^6 per gram is quite high in my opinion (its similar to what you have in seawater). This is actually a good thing for the authors, and makes concerns regarding contamination much less of an issue.

So, its a point that will strengthen the results quite substantially, and the authors should point this out in the manuscript.

AUTHORS: We are very grateful for this comment, obviously linked to the previous one. The protocol (including pyrophosphate as a detergent and sonication) that we used to detach bacterial cells from sediments is very well established in stream microbial ecology. This is an optimized protocol, based on former work by Tom Bott and Martin Pusch, for instance, and that we have used for more than 30 published studies. In these same studies, we used flow cytometry for cell counting, compared several times with counts from epifluorescence microscopy. In fact, the detachment step is repeated twice to improve yields. When we optimized our protocols prior to the expeditions we have also benchmarked several parameters of this step (duration and speed of shaking (15, 30 and 45 minutes at speed 2 (low), 5.5 (medium) and 7 (high)), duration (30 sec, 1 minute and 1.5 minutes) and intensity (duty cycles 40 and 60; output 2 (low), 4 (medium), 5 (high)) of sonication, and the number of repetitions). It is correct that cell-extraction efficiency from sediments can be improved by app. 20% by a second extraction step (using 15 minute shaking at speed 5.5, 1 minute sonication at DC 60, output 5; please see doi.10.3389/fmicb.2020.591465 for more details). However, we found that a third iteration typically does not increase yields any further. Therefore, this method is solid. Following the suggestion, we have now added the above reference in the *Methods*.

Some additional comments on the new meta genomic data:

lines 128-129: I would keep the ordination in Figure 1 showing the meta genomic data differences between the samples, but get rid of the heat map that only has two columns. Its misleading and gives the impression only two meta genomes are being compared (but in reality its many many more in the ordination).

AUTHORS: We agree with the reviewer and updated Figure 1 as suggested. The heatmap now appears in the Extended Data Figure. 5, with additional information to enhance clarity.

lines 129-136: This is all very speculative and a very general statement based on gene annotation data from meta genomes (not activity measurements were made). I suggest removing it, since its not needed to make the points. Rather, just keep the ordination showing that the GFS microbiome has a completely different functional profile compared to the other cryosphere environments and let the 16S tell the rest of the story about vanishing "specific" GFS microbiomes.

AUTHORS: Thank you for this comment and agree that this could be misleading. We have therefore removed this statement from the revised version of the manuscript.

Referee #3 (Remarks to the Author):

I appreciated the author's efforts with very long detailed responses to my previous comments. However, the authors failed to address most of my major concerns despite long rebuttal letter. In the following, I will describe them in details.

1. Novelty of this study

In the rebuttal letter, the authors listed five points for novelty of this study in terms of sampling, analyses, and conclusions. It seems that the authors misunderstood my previous comments. I have no doubt that some new information is presented. Actually, new knowledge or information has to present for any publications. Otherwise, there is no point to publish. But, the key question is whether this study is novel enough to warrant a publication in Nature, the most prestigious multidisciplinary scientific journal, which publishes top 1-2% groundbreaking, transformative discoveries in various fields. I searched Web of Science using the key words related to global microbial diversity and biogeography. There are hundreds, if not thousands, of publications on these topics. Thus, to my knowledge and standards, this study definitely does not meet the standards by Nature as a transformative groundbreaking study in microbial ecology.

AUTHORS: We do agree with the reviewer that numerous papers on microbial diversity and biogeography are being published. Many of them are still regional studies of thoroughly explored ecosystems (e.g., ocean, soil). Today, no single global study exists on the microbial diversity and biogeography of one of the least explored freshwater ecosystems on Earth — glacier-fed streams, 90% of which are predicted to disappear within 75 years due to climate-induced glacier shrinkage (doi/10.1126/science.abo1324).

Moreover, our findings are relevant for microbiomes beyond GFSs. In fact, we show that in spatially isolated and extreme environments (with comparable environmental conditions), diverse microbiomes assemble deterministically - with large turnover (i.e., dissimilarity across different systems) among closely related community members, which leads to biogeographic patterns. This is important for understanding how microbial communities form in general, but also because of its potential functional consequences in the face of climate change. On the one hand, such communities may be expected to respond similarly to environmental changes due to high similarity at deeper phylogenetic levels. On the other hand, we provide evidence for functional redundancy - and redundancy may preserve functioning of these communities in the face of change.

The authors also mentioned Tara Ocean as justification. I do not think it is comparable. The first study of Tara Ocean was published approximately 10 years ago. During last decade, microbiology has been developed so quickly. As time proceeds, the standards become much higher. For instance, 20 years ago, genome sequencing from a pure culture can be published in Nature, Science, Nature sister journals, and other decent journals. Now it can be only published typically as a genome announcement. Similarly, many of this type of community analyses can be published in broad impact multidisciplinary journals such as Nature and Science several years ago, but now only in typical professional journals.

AUTHORS: We respectfully disagree with this comment. In fact, global-scale studies (please see DOI: 10.1126/science.abb3717, DOI: 10.1038/s41586-021-04233-4, DOI: 10.1126/science.abn6358, for instance) are powerful because they do unravel underpinning mechanisms of biological variation along large spatial, climatic and environmental gradients, and allow investigators to answer research questions that are not possible to investigate at local/regional scales.

In the rebuttal letter, it seems that the authors misunderstood my previous comment on 16S-based studies. My point is that given the rapid advance in microbial ecology, 16S sequencing is so easy as a routine tool, advanced microbial ecology studies have to be moved beyond 16S sequencing, either with additional omics and other experimental approaches (i.e., isotope), and/or theoretical analyses. Also, the authors cited similar papers published 10 years ago in PNAS and ISME J, and current paper in Ecology and Evolution to justify this study in Nature. Yes, as I mentioned before, I have no doubt on the certain merits of this study, but they do not meet the high standards set up by Nature. As I know, many top professional journals like ISME J now have editorial policy to directly reject 16S-based diversity studies without external review. However, this does not mean that 16S-based approaches are not important. They are parts of the important toolboxes, but the research questions must be beyond diversity per se (either experimental or theoretical) to be justified for publications in prestigious journals, particularly for Nature.

AUTHORS: We respectfully disagree with the comment. Metabarcoding is appropriate for studies on biodiversity and fine-scale phylogenetics — the two main pillars of our study. Our data provide substantial and new insights into an endangered microbiome at a global scale, which has previously only been investigated in a patchwork of local/regional scales. We note that microbial ecology has long been method-driven (see the efforts by James Prosser and colleagues to redress this), and we believe that science is only as good as the questions, not necessarily how fancy the methods are. Nevertheless, we have included metagenomic data — to the point where they serve the purpose of the study. Having said that, and with all due respect, we do not want to further debate the utility of using large-scale metabarcoding.

2. Significance of the results

The authors found that the GFS are different from other studies with many unknown diversity, and with biogeographic patterns, this is not surprising. It is kind of expected based on numerous microbial studies from soils, ocean, freshwater, groundwater, wastewater treatment plants, as well as human microbiomes with high throughout omics technologies. More important and interesting questions are whether the diversity and distributions are important to the functions of GFS ecosystems. It seems that no efforts were attempted to explore the importance of the microbial diversity to the functions of the GFS ecosystems.

AUTHORS: We do acknowledge such microbial studies from various ecosystems. The significance of our study may be best appreciated in an Humboldtian context, that is, it discloses for the first time the structure, diversity, and distribution of a hitherto largely unseen microbiome at a large scale. The fact that glacier-fed streams are vanishing ecosystems makes our study even more timely and appealing for a very broad readership. With this, we show for the first time what else besides water we may be losing as glaciers shrink.

We do agree with the reviewer that the functionality of the GFS ecosystem is important (and which we partially address). However, this is an entirely different research question — one that we are currently addressing in other manuscripts and that we have just published in *Nature Geoscience* (doi.org/10.1038/s41561-024-01393-6).

As I mentioned previously, the importance of dispersal limitation and environmental selection in shaping the community diversity and distribution is also expected and not surprising at all from theoretical perspectives. Because the samples are from different continentals, dispersal has to be limited. Since environmental conditions are quite different across samples, environmental filtering has to play roles. Numerous global studies on various habitats have supported such theoretical predictions. From this angle, this study does not provide new transformative data/information/thoughts as the authors claimed in the rebuttal letter.

AUTHORS: We are grateful to the reviewer for sharing her/his opinion on this point. Indeed, dispersal limitation and environmental selection are understood as major drivers for community assembly — equally true for microbes, plants and animals. And yet, the relative contributions of these various drivers are still debated for microbial communities, as witnessed by the various analytical frameworks published over the last few years (10.1038/ismej.2013.93; 10.1038/s41396-021-01106-6; doi.org/10.1038/nrmicro2795; 10.1038/s41467-020-18560-z). Given the ease with which numerous bacteria seem to disperse, and given the fact that environmental conditions are not ‘quite different across samples’, one would expect high similarity

within and between mountain ranges. Please note that as ice melts, water is equally cold independent of geography. Similarly, high runoff, steep slopes and unconsolidated sediments translate into highly turbulent flow and high turbidity. Furthermore, given that most mountain glaciers are above the tree line, glacier-fed streams are devoid of allochthonous organic energy and therefore ultra-oligotrophic. What can vary is streamwater geochemistry because of parent geology, etc. All this has been described upfront in the text, referring to Figure 1a. Having said that, despite these assumptions (dispersal) and little environmental variation, we still observe distinct patterns in community composition. Moreover, these findings bring new insights into the field of microbial ecology as not that many ecosystems are similarly spatially isolated as GFS - exclusively located at the tips of stream networks.

In addition, in the rebuttal letter, the authors highlighted the importance of the results related to endemism and functional redundancy. Honestly, I still have serious concerns on this type of analyses and robustness of the conclusions as before (see below). Even if we assume that the analyses on endemism are reliable, do the results represent groundbreaking, transformative discoveries in microbial ecology? The answer is not. Given extremely high microbial diversity, continental scale, and dramatic differences in environmental conditions, high degree of endemism is theoretically expected, which are also demonstrated by numerous previous studies in various habitats with soils, groundwater, freshwater, wastewater, and so on. (e.g., doi: 10.1002/mbo3.644, doi.org/10.1186/s40168-023-01572-4, doi.org/10.1073/pnas.1919139117, doi: 10.1038/ncomms12083, doi.org/10.1038/s41564-019-0426-5, doi: 10.1128/mSystems.00803-19, https://doi.org/10.1093/femsec/fiz128). Furthermore, given the extremely high microbial diversity, existence of functional redundancy in the GFS microbial communities and its importance to functional resilience is not surprising at all. It is kind of expected from both ecological theories and other empirical studies.

AUTHORS: Thank you for these comments. As previously mentioned, we have now abstained from using the concept of ‘endemism’. Following the advice of reviewer #2, we now refer to these ASVs as specific to a mountain range; given the substantially larger sample size that we now present, underlying statistical analyses are even more robust than in the previous version. However, we also stress the difference between our work and other studies that report high levels of specificity. First, our sampling effort allowed us to cover more than 80% of the expected regional diversity - the fact that we find high levels of regional specificity despite such large coverage reduces the likelihood that specificity arises because of undersampling (i.e. that ASVs may appear specific to a single mountain range because they were present but not detected in another mountain range). Moreover, we leverage more than 850 individual samples and remove approximately ⅓ of the ASVs detected during filtering to yield conservative estimates of specificity (i.e. ASVs that were not found in at least 2 out of 6 patches in any GFS were removed). This

high level of specificity and even uniqueness is further illustrated by 499 ASVs that were present in all 6 patches from a single GFS - but in no other of the 848 GFS samples. This equals a probability of 1:5088 for being false positive.
High degrees of regional specificity and uniqueness may also be unexpected. It runs counter to the fact that environmental conditions are rather similar across GFSs (nota bene, there are no 'dramatic differences in environmental conditions' as erroneously and repeatedly claimed by this reviewer) and the dispersal ability of many microorganisms. The latter is still a subject of debate as highlighted again by two most recent publications (doi.org/10.3389/fmicb.2022.855859, doi.org/10.1111/1462-2920.16270).

The reviewer argues that 'endemism' is to be expected because of 'continental scale', among others. We note that uniqueness (see above) arises at the scale of nearby GFS within the same mountain ranges. This can be illustrated by ASVs 57eb8f5618f1144c279b27ee2a3a2de5 (classified as *Methylotenera*) which recruits between 275 and 1391 reads in all 6 patches sequenced from the same GFS in Nepal, but no single read in the 6 patches sampled from the next GFS only 4.3 km away and no other GFS worldwide. We are grateful to the reviewer for listing studies reporting 'endemicity' from various ecosystems and stress the fact that our sampling effort identifying specific ASVs is well beyond those reported in these studies.

3. Experimental design problem of this study

The authors argued that 149 GFS were sampled, representing global most comprehensive study. This is an over-statement, and it is not a comprehensive study. Actually, as I mentioned before, this study has some serious flaws in sampling design. It seems that Reviewer 1 had similar concerns. It is not clear whether the 149 GFS samples represent the global diversity of GFS. Given the huge area of GFS, more sites need to be selected to cover the habitat diversity of a given GFS. For instance, the sites selected from Alaska are very close based on the map. Since Arctic is huge area and highly heterogeneous, it is hard to justify that the selected few adjacent sites can fully represent the Arctic GFS. How and what sites are selected are critical to addressing the research questions, particularly related to endemism. I do understand the logistic difficulties to sample more sites, but the data collected are hard to address the type of questions related to "true" endemism. More importantly, given the high spatial heterogeneity, three replicates from each site are far not enough to reflect the microbial community diversity of the GFS examined. In their rebuttal letter, the authors mentioned that they focused on the GFSs as close as possible to the glacier snout to avoid environmental gradient. Yes, this will help minimize the variations among replicates, but jeopardize the sample representation of a particular GFS. Thus, to consider spatial heterogeneity, many more samples (not three in this study) from each GFS should have been collected. In addition, this is just a snapshot study. To determine "true" endemism, temporal sampling of

representative sites are needed because ecosystem dynamics are not linear and hence time is a critical factor in controlling microbial diversity and distribution.

AUTHORS: We respectfully disagree with the reviewer claiming that our study is not the 'most comprehensive' effort studying the GFS microbiome today. We are not aware of any similar study or any laboratory involved in a similar effort. We wished we were able to sample 1000 GFSs but even that number would not suffice - and would be akin to filtering thousands of liters of water from the world's ocean to infer conclusions on the 'global ocean microbiome'. We all know what the limits to these studies are and respectfully acknowledge them — respectfully also for the gigantic and unprecedented resources and efforts spent to sample 152 GFSs across 11 mountain ranges.

As for the sampling design, first, we identified major glacierized mountain ranges with glaciers with GFSs (not those calving into the sea). Within each mountain range, we identified 'clusters' to account for differences in geology, aspect etc, and further within each cluster, up to four or five GFSs. This nested design allows us to cover geographic distances from 94 m to more than 18'817 km, and altitudinal differences from 93 to 5093 m above sea level. Equally true, this design allowed us to maximize the potential gradient of environmental variation, which proved to be low obviously (please, see comment on this apparent misunderstanding above and elsewhere). Moreover, as aforementioned, we are happy to report that we were able to include 3 more patches from most GFS - nearly doubling the number of samples underlying this study. Strikingly, doubling the number of ecological replicates did not change our main findings - effectively eliminating the concern that within-GFS heterogeneity could jeopardize representativity of our samples. We agree that temporal sampling would be a desirable goal - however, clearly beyond what is currently feasible at large spatial scale. Furthermore, it would not answer questions of biogeography as communities are likely to reorganize themselves rather than fundamentally turning over.

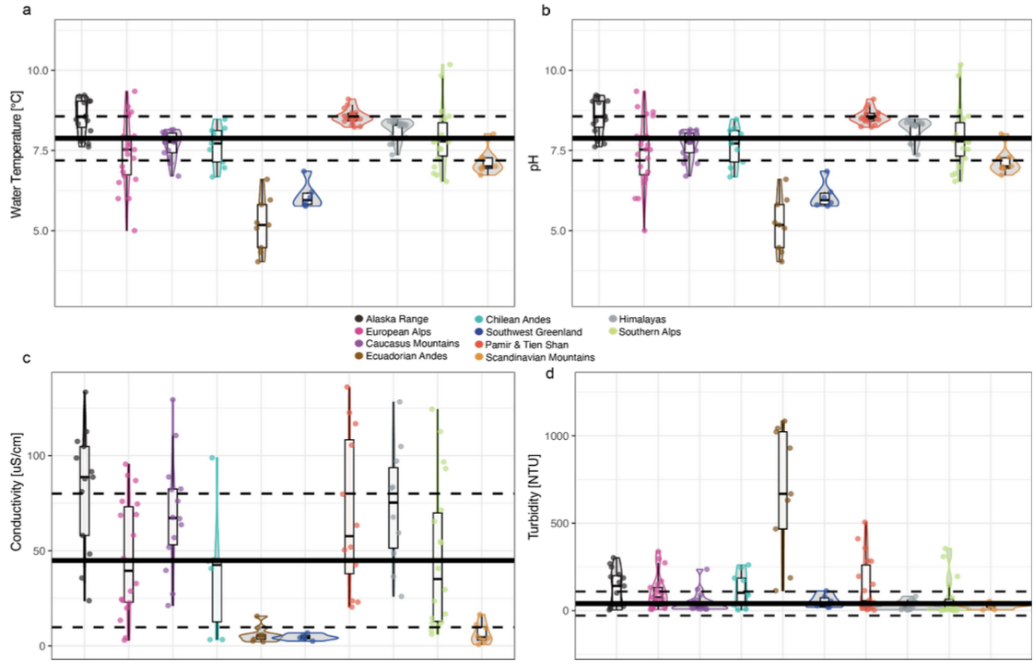
4. Methodological problems for quantifying endemism

The author claimed that one novel aspect of this study is high degree of endemism. As I mentioned above, this is also expected because of dispersal limitation due to continental sampling scales, and environmental selection owing to huge differences of environmental conditions. More importantly, quantifying endemicity is very tricky, which depends on sample numbers, sample representation, sequencing approaches, sequencing depth and bioinformatic processing approaches, as well as taxonomic resolutions of the molecular markers used for analysis. Reviewer 2 also had similar concerns, particularly related to contaminations. Beyond contaminations, I would like to highlight some issues again. First, spatial heterogeneity will have significant effects on the detection of endemic ASVs. If more samples are analyzed from each GFS, less endemism would be expected. As I mentioned above, it seems

that the 149 GFS examined may not cover the global habitat diversity of GFS, and three replicates from each GFS do not reflect the spatial heterogeneity. Also, microbial community abundances in the GFS ecosystems could be quite dynamic. If time is considered and more spatial samples are analyzed, the picture of endemism could be quite different. In fact, the experimental designs and the current availability of samples do not allow the authors to address questions related to endemism although it is an interesting topic. Second, as I mentioned previously, the sequencing depth is shallow. Although the authors resequenced some samples and increased sequencing depth to 25,526 sequences per sample for analyses. nIt is still not clear whether the sequencing depth enough to represent the diversity of the communities examined, which is critical to address the questions related to endemism. Instead of using DATA2, the authors used Deblur to preprocess the data, which is also very conservative. Many sequences could be removed prior to analyses, which could greatly overestimate the estimation of endemism because such removed sequences are in the samples, but they are simply not counted for estimation. Third, sequencing approaches are problematic in terms of reproducibility as demonstrated by numerous studies since last decade (e.g., doi:10.1111/vop.12854, doi:10.1093/bioinformatics/btaa926, doi:10.2306/scienceasia1513-1874.2021.073, doi:10.1186/s40168-020-00998-4, doi:10.1016/j.ebiom.2021.103649, doi:10.1038/nbt.3780, doi:10.1371/journal.pone.0099414, doi: 10.1016/j.mimet.2013.07.015, doi: 10.1016/j.mimet.2013.07.015, doi:10.1128/mBio.02288-14. https://doi.org/10.1371/journal.pone.0043093). For instance less than 40% overlaps were detected among technical replicates (i.e., the same DNA samples are sequenced multiple times) even with very simple communities (doi.org/10.1371/journal.pone.0043093). The low reproducibility is an inherent problem of sequencing approaches due to random sampling processes (doi:10.1128/mBio.00324-13). Although reproducibility is less problematic in addressing questions related to community structure differences, it is critical to determining endemicity because information on presence/absence of OTUs (or ASVs) is needed for estimating endemicity. For instance, some missed ASVs in a sample could be due to low reproducibility problem of the sequencing approaches. They are in the sample, but are simply not sampled. In addition, taxonomic resolution of the molecular markers used is critical to understanding endemism. As an example, if one focuses on the taxonomic resolution at kingdom (e.g. bacteria, archaea, fungi) or phylum (e.g., acidobacteria, proteobacteria) levels, there could be no or very little endemism. It is well known that the sequenced region of 16S gene generally provides the resolution at genus or family levels. Ideally, to define “true” endemism, molecular markers with the taxonomic resolutions at species/strain levels are needed. In a word, given all sampling and methodological problems, I really doubt the reliability and robustness of the conclusions on endemism in GFS ecosystems.

AUTHORS: With all due respect, we would like to draw the reviewer’s attention, after she/he has brought this argument up a couple of times, to the fact that there are no

'huge differences in environmental conditions' across GFSs. To counter this potential misunderstanding, we have now added three new Extended Data Figures (2,3,4) showing box plots for all environmental parameters as also used in Figure 1a. These parameters are resolved at the level of mountain ranges and compared to the median and IQRs for all GFSs. Please, see here below a sample (Extended Data Figure 2) showing the variation of GFS streamwater temperature, pH, conductivity and turbidity. Overall, it becomes obvious that at the level of mountain ranges, variation is well constrained. However, there are regional variabilities for some few mountain ranges (e.g., Ecuador, SW Greenland), which we do acknowledge in the text (e.g., in the subsection describing the GFS environment). We refer to this as well, when discussing the spatially structured environment among mountain ranges in the variance partitioning section. We do sincerely hope that with this extra information, we can convince this referee that **(a)** overall GFS environmental conditions are relatively well constrained globally, but **(b)** yes indeed, there is a weak level of regionality.



Extended Data Figure 2. Violin plots of environmental data. **(a)** Water temperature [°C], **(b)** pH, **(c)** Conductivity [uS/cm], **(d)** Turbidity [NTU] across the different mountain ranges. Each dot represents a GFS. Solid and dashed lines are IQR values.

Furthermore, and as mentioned previously, we have significantly increased the number of samples, allowing us to better account for the spatial heterogeneity within each GFS. It is worth noting that within-GFS microbial dissimilarity is smaller compared to between-GFS microbial dissimilarity within the same and across mountain ranges. Furthermore, we do not use the word 'endemic' anymore.

We agree that sequencing depth can impact the observed proportions of endemic species, however, we disagree that our sequencing depth is shallow. First, rarefaction curves (or collectors curves) saturate for nearly all samples. We also

estimated Good's coverage, which ranged between 99.28 and 99.995, suggesting that our sequencing effort covered the vast majority of ASV present in these communities. Conceptually, we attribute this to the selective environmental conditions of GFS, which, to some extent, limit biodiversity. Moreover, we performed sensitivity analyses to benchmark the effects of sequencing depth (and number of GFS sampled per mountain range) and observed that the proportion of specific ASVs only marginally decreased with increasing rarefaction depth (from 5,000 to 20,000 reads, please, see Extended Data Fig. 14). This suggests that we have sufficient sequencing depth to accurately estimate the proportion of specific ASVs in GFS microbial communities.

We also agree that Deblur is the most conservative method for denoising in terms of the number of resulting ASVs (please see, Supplementary Figure 9 and our response to reviewer #2 in terms of blank and mock communities). However, when we compare the proportion of specific ASVs among all three investigated methods, Deblur does not consistently result in higher proportions (please, see Extended Data Fig. 12). In contrast, mountain ranges such as the Southern Alps and the Ecuadorian Andes consistently display the highest proportions of specific/endemic ASVs independent of the denoising method (please, see Extended Data Fig. 12).

Furthermore, Deblur, while conservative in terms of richness, gave the best estimates when looking at the mock communities compared to the other methods. Indeed, as more extensively described in our response to reviewer 2, we obtained 8 to 10 ASVs when analyzing a commercially available mock community (Zymo) that contains 8 specific strains, while DADA2 and UNOISE3 respectively gave 20-28 and 120-160 ASVs for the same community. As most of the ASVs in those two methods corresponded to the expected species, we assumed that, in accordance to a paper by Schloss (2021, mSpere, doi.org/10.1128/mSphere.00191-21), both DADA2 and UNOISE3 tend to separate the same genome into different ASVs based on the number of 16S rRNA gene present and thus overestimate diversity and potentially endemism.

Indeed amplicon sequencing has been under discussion to have low reproducibility. In particular, occurrence-based overlaps can be very low. However, taking into account abundances, overlap between replicates is much higher. Additionally alpha diversity is typically very consistent between replicates and beta-diversity between replicates is much smaller than between samples. (doi.org/10.1128/spectrum.04048-22, https://doi.org/10.1371/journal.pone.0176716). We are further countering these uncertainties by averaging up to six ecological replicates for each GFS. Additionally, we are comparing mountain ranges to identify specific ASVs and not single samples, thus further compensating possible shortfalls of the sequencing approach.

5. Problems related to shotgun sequencing data and functional redundancy

I mentioned that examining functional genes-based traits could be more interesting. The authors provided a very limited shotgun sequencing information in this study. I appreciated such efforts, but I also have some concerns. It seems that the majority of such data will be presented in another manuscript. To be honest, I do not think this helps to strengthen this current study because based on my experiences, unlike other typical microbiological journals such as Applied and Environmental Microbiology, Environmental Microbiology, mSystems, and etc, Nature only publishes original, first hand data for original research papers (could be different from meta-analysis papers). Such data should be merged with this paper to really strengthen this type of studies.

AUTHORS: We have used 'first-hand' shotgun metagenomics for comparing the potential functional profile of the GFS microbiome to those of other cryospheric microbiomes. This highlights its uniqueness among cryospheric systems and suggests functional properties linked to the biofilm mode of life underlying this basic difference. Shotgun metagenomics has also revealed functional redundancy of the GFS microbiome. Given the focus of our manuscript on the GFS microbiome diversity and biogeography, further functional data would be beyond the scope of this work. A companion paper on functional aspects only is currently in revision with a leading outlet in the field.

Second, no information on shotgun sequencing methods is provided in M&M except KEGG functional analyses. I could not find the data mentioned in Supplementary Table 8 with this link, doi:10.5281/zenodo.10029840. Thus, I could not judge the appropriateness of the methodologies and the reliability and robustness of the conclusion with respect to functional redundancy.

AUTHORS: We apologize for this. The data are now available at the following link: https://zenodo.org/records/10985735; doi.org/10.5281/zenodo.10985735. In addition, we updated the method section and included additional information regarding the methodologies such as the statistical tools (i.e., Maaslin2) that we used to compare the data from GFSs and the other cryospheric samples, alongside the metagenomic analyses methods. The methods in the revised manuscript, now included, are as below:

"Shotgun metagenomics

DNA samples obtained from the benthic sediments also underwent whole genome shotgun sequencing as described previously[58]. Briefly, the libraries were prepared using the NEBNext Ultra II FS library kit, where 50 ng of DNA was used. Library preparation included six PCR amplification cycles after a 12.5-min enzymatical fragmentation of the input DNA. On average, an insert size of 450 bp was obtained, where the libraries were quantified by Qubit (Invitrogen) coupled with quality and size estimation using a Bioanalyzer (Agilent).

Sequencing was performed at the Functional Genomics Centre Zurich on as S4 flowcell (150 bp paired-end) (NovaSeq).

For the metagenomic sequence data, the Integrated Meta-omic Pipeline (IMP; v3.0, commit# 9672c874; available at https://git-r3lab.uni.lu/IMP/imp3)[59], was used to process the paired forward and reverse reads from 97 GFS metagenomes. The IMP workflow for GFS analyses has previously been described[58]. Briefly, IMP employs MEGAHIT (v1.2.9)[59,60] for the assembly, following an adapter and primer trimming step with cutadapt[55]. Subsequently, the assemblies were used for functional gene calling through Prodigal[61], yielding 24,946,385 non-redundant gene clusters after clustering with mmseqs2 with the following parameters: --cov-mode 0 c 0.8 --min-seq-id 0.3[62,63]. FeatureCounts was then used to estimate gene abundances, while the KEGG annotations from the gene annotations were generated using Mantis[64], mapping to 17,536 KEGG Orthologs (KOs). We filtered KEGG data so that only the 8,518 KOs associated with prokaryotes were retained as determined by the KofamScan profiles[65]. For comparison with other cryospheric systems, we obtained 92 metagenomes from ref.[20] and processed them via the IMP workflow using the same parameters."

In the rebuttal letter, the authors indicated that this study "shows signatures of functional redundancy as a response to the GFS environment". Because functional redundancy is critical to ecological stability, it is a central but challenging topic in ecology. Thus, I am very interested in this part and searched all three files (main text, extended figs and supplementary). Unfortunately, I could find such signatures as the authors mentioned in the rebuttal letter. Based on my own experiences, defining and quantifying functional redundancy are very difficult, particularly for microbial communities. Identifying signatures for functional redundancy is even more difficult in microbial ecology.

AUTHORS: We agree with the reviewer that functional redundancy is an interesting area in microbial ecology because of its inherent link to stability. We also agree that it is not trivial to infer functional redundancy from large environmental data sets. Therefore, we present signatures of functional redundancy at community level with care. Before walking through them one-by-one, we here consider functional redundancy when community structure (from ASVs) varies independently from community potential function (from KOs) across scales. This is analogous to the approach used by doi.org/10.1016/j.ecolind.2019.03.029, for instance, introduced by doi.org/10.1016/j.ecolind.2012.02.004, and resting on the notion that community structure and function are decoupled (at least to a certain extent) (e.g., doi.org/10.1016/j.mib.2022.102263, http://doi.org/10.1038/s41559-018-0519-1 opted for a multifaceted approach to address community-level functional redundancy. First, comparing altitudinal gradients of ASV *versus* KO diversity, we notice an absence of covariation between structure and function with altitude, which we interpret as a signature of functional redundancy. Next, exploring beta-diversity patterns, we show that dissimilarities of the functional potential are overall lower and vary less than dissimilarities in the composition (i.e., ASVs). We complement that analysis with

estimates on niche breadth (Levin's), showing that common KOs essentially explain only marginal differences in the functional potential. To highlight this, we have added a new Extended Data Figure 11. Not unexpectedly perhaps, distance decay pattern (DDP) analyses do not show much of a biogeographic trend for the functional potential, in stark contrast to composition. We have now added a panel to Figure 4 with the KO DDP. These levels of decoupling between microbiome composition and potential functions are further signatures of functional redundancy. Finally, on completely different grounds, we also provide phylogenetic evidence, namely the microdiversity contained within clades, towards functional redundancy.

To conclude, we want to bring forward that these lines are in agreement with our conclusions based on the analyses of more than 2800 genomes assembled from metagenomes, as described in the manuscript that we share at the discretion of the reviewers.

In conclusion, the revised manuscript still has serious issues in terms of novelty, significance, experimental design, experimental analyses, data analyses, and data interpretation. More importantly, the experimental design and the data generated by the authors could not allow the authors to address the research questions in mind (e.g. endemism, functional redundancy). The results presented in this study does definitely not represent groundbreaking advance in microbial ecology to warrant a publication in Nature, one of the most two prestigious multidisciplinary scientific journals, along with Science. I have no doubt that this study can be published in other professional journals, but not suitable to Nature.

AUTHORS: We respectfully acknowledge the conclusions of this reviewer.

2023-03-04207B : Global diversity and biogeography of the glacier-fed stream bacterial microbiome

Responses to Reviewers

Referee #1 (Remarks to the Author):

I am satisfied with the responses of the authors to my comments and the respective changes made in the manuscript. As final thought, the authors might want to consider to at least include one sentence that summarizes the discussion that the authors give in their response in the paragraph "This reviewer also mentions the potential for legacy effects from streambed ... rather speculative in nature.". In order to at least roughly indicate that there are potential influences that the study could - well understood - not consider. The fact that the referee wonders might indicate that other readers could well wonder the same.

Authors: We are pleased to read that our revisions have addressed the comments and suggestions raised by the referee. We would like to thank the referee for her/his constructive comments, which greatly improved the quality and robustness of our study. We agree with the above comment and as suggested, the section *The glacier-fed stream environment* now includes an additional sentence mentioning the potential for legacy effects**.**

Referee #2 (Remarks to the Author):

The revised manuscript of Ezzat et al has been substantially improved and I was pleased to read and see that the authors took many of my suggestions into consideration. Namely, there is now a very extensive analysis of mock communities and ASV calibrations and the authors did a good job using their mock communities to pick the correct algorithm for their data (Deblur). I think that their work in this area sets the bar for quality control for future 16S based microbial ecology studies, I m not aware of any study that did a more extensive checking on this than the current study. Moreover, I was furthermore pleased to see that the authors took my comments about the word "endemic" seriously, and have changed the word to "specific" or "unique". That being said, I think that the discussion and introduction should have a stronger connection to the historical debate on microbial biogeography, namely that the authors results may shed new light on the original hypothesis proposed by Bass-Becking in 1934 that "everything is everywhere". This hypothesis has been debated ever since, with some arguing that endemic species may not exist because of limitless dispersal potential of microbes. The citation provided in reference 16 is a good recent review on the topic of microbial biogeography, but I think the last part of the Introduction of this manuscript (and also in the discussion where the results are presented) should provide a little bit more on the historical perspective and long standing debate. For example, here are some papers to familiarize with past examples of the debate

(doi:10.1111/j.1462-2920.2006.01017.x)
(doi: 10.1126/science.1070710)
(doi: 10.1126/science.1070710)

Perhaps, a nice example of the debate is well documented in this article

(DOI: 10.1641/0006-3568(2004)054[0885:RFFAF]2.0.CO;2)

Where those authors stated "We therefore remain unconvinced by the limited evidence for microbial biogeographies"

Its seems this new manuscript by Ezzat et al on global GFS biogeography is a very strong dataset that adds a lot of value to this historical debate and therefore I think the authors could improve the discussion a bit by adding it into the context. I know that the main focus is on the vanishing unique cryosphere biomes, but this is an additional aspect of the work that will be highly impactful. While the current manuscript does not prove endemicity, it provides very strong evidence for highly specific communities that have undergone strong selective pressures resulting in very clear patterns of global biogeography. This discovery in itself on a global scale in the cryosphere environment is very novel in my opinion.

In general, I think that this manuscript is vastly improved from the first submission and is well on its way to be published in Nature. It meets the high bar for novelty, impact, and quality in my opinion. Aside from the points above, I have a series of additional comments below where I identified a few incorrect statements in relation to new references that were added. I also point out a few regions of the text and figures that are unclear and provide suggestions for improvement.

Authors: We are pleased to hear that our revisions have successfully addressed the referee's concerns, and we thank the referee for her/his positive feedback and overall appreciation of our work. We stress the fact that her/his constructive comments have greatly improved our study.

We fully agree with the reviewer that this dataset perfectly adds to the ongoing conversation about microbial ubiquity and biogeography. Following her/his advice, we have now added a few sentences that better anchor our study within this debate of microbial biogeography, at the same time being mindful to not convolute the narrative and/or exceed the article length limits of *Nature*.

Please find below our detailed replies to the additional comments made by Referee #2.

Specific comments:

line 70: typo 'fulfil', change to "fulfill".

Authors: This has been changed accordingly.

lines 89-95: This section could benefit from a couple of sentences introducing the nearly 100 year old debate on 'everything is everywhere, but the environment selects' that was originally proposed by Bass-Becking and was reignited decades ago when new molecular techniques began to be applied to microbial communities (see comments above).

Authors: We agree, and have included one sentence on this topic in the introduction. Furthermore, we have changed the final paragraph of the '*High mountain range specificity*'

section to relate to this idea. However, for the sake of word counts, number of references, etc, we have kept this discussion concise and referred to doi:10.1038/nrmicro1341, which was already included in the reference.

line 125: What metric was used to conclude that these pathways "drive" the segregation? Was a statistic used in the ordination for this?

Authors: As indicated in the figure legend, analyses are based on PERMANOVA ($F_{1,111}$=92.5, R^2=0.46, p=0.001). We also included multivariate tests outside of the PERMANOVA to pull out pathways that differ between groups (See Supplementary table 5). We agree that the legend was not clear, and we have now altered it.

line 135: I think this statement could be made more precise, particularly what "The cryosphere was long thought to be devoid of microbial life...." means. I don't agree with the statement as written, because the existence of microbial life in the cryosphere goes back at least 70 years (https://doi.org/10.2307/2257289). Perhaps, instead of 'long thought to be devoid of life' the authors could rephrase to say the microbial cryosphere is gaining more attention in recent years.

Authors: We have changed the sentence accordingly.

line 145: Do you mean generalizable conclusions about global microbial biogeography?

Authors: Thank you for spotting this. Yes, indeed for global GFS microbial biodiversity and biogeography. We have added this precision.

lines 146-148: This acknowledgement is appreciated, as is the new tone of the manuscript in regards to topics of specificity and endemicity.

Authors: We agree and thank the referee for their comment.

line 151: Im not sure if "harshness" is the right word here, to be more specific it is a matter of low primary productivity and organic matter that keeps the microbial biomass low.

Authors: Thanks for the critical eye. We have changed this to ‘low resources’.

lines 157-159: This statement is inaccurate and needs a revision. The concentrations of the GFS cells reported here are actually lower (not higher, as is currently written) than what Ref 25 reports. Ref 25 reports cell concentrations of surface deep sea sediments in the range of 10^7 - 10^9 cells g-1. The authors manuscript reports orders of magnitude lower, around 10^6 - 10^7 cells g-1. What the authors measured in GFSs is however higher than the extreme subseafloor biosphere that exists down to several kilometers below the seafloor where cell concentrations drop to levels 10^5 - 10^2 cells g-1 (https://doi.org/10.1038/s41579-018-0046-8). Can the authors please revise this text for accuracy. I think its important to place the measured microbial biomass into context. Please explicitly state the statistics on concentrations of microbial cells is in the text (in addition to

citing Supplemental Table 1), and provide a comparison to other "high" and "low" biomass biomes.

Authors: The referee is correct, and we apologize for the inaccurate statement. We have revised the sentence and added the range of cell abundances across all mountain ranges (that is, IQR: 3 x 106 - 1.8 x 107 cells g-1sediment) and further referred to Supplementary Information Table 1. We have now added doi.org/10.1038/s41579-018-0046-8 as a reference for the lower bound. However, for the sake of space limitation and distraction from the actual topic, we abstain from explicitly adding the cell numbers reported in the two papers that we cite now.

line 212: Could you provide a brief explanation of what Levin's niche breath is, for the reader. Has this ever been applied to microbial communities before, in order to classify specialists and generalists as is depicted in Extended Data 11?

Authors: Levin's niche breadth is a relatively simple and commonly used index to identify (habitat) generalists and specialists based on abundance and occurrence frequencies. Briefly, abundant and widespread taxa are identified as generalists, while rare (low abundance) taxa found only in a few locations are considered specialists. An R function using ecological null model permutations for significance testing is implemented in the EcolUtils R package. Several publications on microbial metacommunities used Levin's niche index (e.g.https://academic.oup.com/femsec/article/87/1/102/505876) and the EcolUtils implementation (e.g. https://enviromicro-journals.onlinelibrary.wiley.com/doi/full/10.1111/1462-2920.16044). We added a brief description of Levin's index and the Yan et al. 2022 reference to the revised manuscript.

line 223: I suggest using "range specific ASVs", to make it clear which ASVs are specific to the ranges as opposed to those specific to GFSs.

Authors: Thank you for the suggestion. The manuscript now includes, where appropriate, the term "range specific ASVs" to distinguish ASVs that are specific to mountain ranges from those specific to GFSs.

lines 226-231: I suggest to state here that these calibrations were all made using comparisons to mock communities and therefore the decision to use Deblur is very robust.

Authors: This is a good idea, and we have altered the sentence to highlight that our analyses rely on comparisons made with mock communities, and thus our decision to use Deblur was rigorously made.

line 238-239: This is very interesting indeed!

Authors: Thank you. So do we think as well.

lines 243-246: This statement is appreciated, and provides a much more balanced view and interpretation of the results in my opinion. The dataset provides a fresh insight into this age

old debate, and may stimulate new research in this area.

Authors: We thank the referee for her/his positive feedback.

lines 273-275: Were any of these organisms isolated from these locations, or is this only based on 16S gene fragments in sequencing libraries? If they have been isolated that is additional evidence supporting their presence (as opposed to laboratory contaminants or cross sample contaminants from PCR products).

Authors: Thank you for your comment. These organisms were not isolated from such locations. Results are based on 16S gene fragments.

lines 317-323: I suggest that the authors also point out, that just because a KO is shared between different locations, does not mean that the exact same function is occurring. Recent evolutionary events such as point mutations, single amino acid changes in key functional parts of the protein, can have a significant effect on protein or enzyme function in different strains of bacterial species that colonize different habitats (doi: 10.1126/science.1218198). Therefore, its possible that recent evolution in the genomes of ASVs has taken place and that even though they share KOs, the encoded proteins are not exactly the same and might have different functional profiles between ASVs inhabiting different ranges or GFSs. I suppose this is something for a future project, but nevertheless I think its important to point out. That the conclusion of "functional redundancy" might not apply if this was the case.

Authors: We thank the referee for her/his suggestion and have now included an additional sentence to point out that although KOs are shared across mountain ranges, this is not necessarily proof of functional redundancy.

lines 368-370: Do these results also potentially reflect recent speciation events in range specific ASVs?

Authors: This is an interesting point. Speciation and extinction, but also dispersal and species sorting, contribute simultaneously to turnover and beta-diversity. We observed that beta-diversity is concentrated at the tips of the phylogeny (i.e. there is no evidence for turnover of larger phylogenetic clades). While speciation might be one of the mechanisms contributing to beta-diversity in GFS microbiomes, we are not currently able to unambiguously identify speciation based on 16S rRNA barcoding.

lines 380-381: What do you mean by "fine-scale phylogenetic structure"? Is this referring to "micro diversi

Authors: Yes, this refers to the phylogenetic structure within microdiverse genera, as illustrated by *Polaromonas, Rhodoferax, Methylotenera, and Rhizobacter* (Extended Data Fig. 11a). The phylogeography of *Polaromonas* ASVs is depicted in Ext. Data Fig. 11b. Within this microdiverse genus, there are wide-spread (i.e., occur in all/most mountain ranges) and abundant *Polaromonas* ASVs; however, there are also clades of closely-related *Polarmonas* ASVs that are specific to one or a few mountain ranges. Unfortunately, due to

space limitations, we can not further discuss this (although very interesting) part in the main text. But, we have revised the sentence accordingly to avoid misunderstandings and hope that it is clearer now.

Figure 3 panel d: It looks like the core microbial community is actually the most abundant part of the community compared to the specific and indicator ASVs. This result did not come out strongly to me in the text. Can the authors please highlight this important point a bit more in the discussion (if I am interpreting this correctly). Why do the specific ASVs tend to be less abundant than the core ASVs?

Authors: We are grateful for this comment. Indeed core ASVs were abundant, compared to range-specific and indicator ASVs. Elevated abundance of core taxa is relatively well documented in the literature. We argue that these ASVs are particularly well adapted to the selective environment of GFS overall and therefore competitively superior to others, potentially the result of homogenous environmental selection. We have now added a comment on this to the manuscript.

Figure 5: This is a really nice figure. However, for the reader I think it might be a good idea in the text to briefly define what "phylogenetic agglomeration" means (combining all ASVs within a certain larger hierarchical group?). Also, is it possible that DL and HoS colors (blue and red) are mis labeled in panel B? It looks like HoS increases with distance but DL decreases (which is different from what is written in the legend).

Authors: Thank you for the suggestions. We have now added further explanation of “phylogenetic agglomeration” in the figure legend. The term “phylogenetic agglomeration” describes the gradual combination of phylogenetic tips along a gradient from the tips towards the root. In this panel, we show that beta-diversity (within and across mountain ranges) changes most when phylogenies close to the tips are considered. Thank you for noting the error in the legend, indeed the colors for DL and HoS were switched. This was corrected.

Extended data 5: The top dendrogram can be removed since it is only connecting two sample groupings.

Authors: We agree with the reviewer and removed the top dendrogram from the figure.

Extended data 7: These individual plots are so small they are barely readable. Does this figure really add something critical to the story? Aside from a brief mention on lines 164-165 it doesn't feature in the story. Its unclear to me what the authors mean when they say that extended data 7 suggests truncation of the rare ASVs. What do you mean 'truncation' in this part of the text? Do you mean removing? What does each plot show? Are they different samples? Please provide the units and scale for X and Y axes. I could not find this information in the legend either.

Authors: Thank you for this suggestion. We agree that the individual Figures were too small to be meaningful. We revised the main text to be more clear about the scarcity of very rare taxa (i.e. the “truncated” rarity), which we attribute to the unstable and selective environment. We find this to be an important and interesting aspect of the GFS microbiome - and hence respectfully disagree that this is not featured in the story. Moreover the two panels of this Figure also provide a comparison between filtered and unfiltered datasets (showcasing that our filtering procedure did not bias against rare taxa) and quantitative fits of the log-linear and an ecological null model (i.e. broken stick model). We therefore decided to retain larger versions of examples of these panels in Extended Data. However, we now show only a subset of the panels in the Extended Data Fig. 4 - and deposit a zoomable version of both, the filtered and unfiltered dataset on Zenodo. A link to the Zenodo repository is provided in the “Data availability” section of the manuscript.

Extended data 10a (heat map): I think that this heat map tells a lot more the story about uniqueness and potential endemism compared to some of the current main text figures. I would consider making the heat map (or a version thereof) a main text figure. Because, it is the only figure I have seen so far that is actually showing the distribution of the "specific" ASVs across the samples.

Authors: Thank you for the suggestion. However, we would first like to clarify that this heatmap shows the distribution of the most abundant indicator ASVs across mountain ranges, not the specific ASVs. While we agree that it is an interesting figure, we prefer to keep this figure in the Extended data rather than include it in the main text. This is because, If we did wish to include it in the main text, it would be necessary to shrink its size (in order to meet the figure limit), which would make it hardly readable.

Referee #3 (Remarks to the Author):

I reviewed this MS twice and raised various critical concerns. The authors have addressed my comments to some degree. In the previous two rounds of reviews, I have been focused more on high level of questions, particularly novelty, significance, and potential impacts based on Nature standards. Although I still have concerns, I will leave these issues upon Editor’s decisions. In this round of review, I am more focused on details about experimental analyses, data analyses, and data interpretation and presentation for how to improve the quality of this study. After more seriously thinking of this study, although habitats are unique, but the analyses seem very shallow. I have various additional comments and suggestions for improvements. Please notice that my comments are not presented in a logical order.

Authors: We are grateful for this comment and refer to our replies below. Overall, we appreciate the suggestions, several of which have been an integral part of our manuscript from the initial submission onwards. We apologize if several of them were not clear.

1. Although habitats studied are interesting, this study is quite descriptive. Right now, it is a fish-expedition, and no hypothesis is presented. Based on numerous such types of studies plus their previous experiences with this habitat, the authors should be able to formulate some interesting and attractive hypotheses to test.

Authors: We agree with the reviewer that glacier-fed streams are interesting ecosystems. Given the Humboldian (not experimental) character of our study, we did not put forward any hypothesis. A hypothesis broad enough to encompass our entire study would be trivial by nature. Most similar such studies (e.g., DOI: 10.1126/science.1261359, DOI: 10.1126/science.abb3717) investigating for the first time a microbiome, do not present any major hypothesis. So yes, we do agree, these studies are rather descriptive. Yet, they form an important basis for future hypothetico-deductive studies. And so will ours, we are convinced.

Please, note that we do test, however, several assumptions (e.g., altitudinal gradients, distance decay patterns), without explicitly referring to them as hypotheses. We abstain from referring to these as explicit hypotheses; this as it would not be appropriate, a.o., as they would add to the length of the manuscript, including additional references.

2. Two of the main drawbacks of amplicon sequencing approaches are low reproducibility and poor quantification. Due to the randomness nature, quantitation problem should be less serious for shogun sequencing approaches. The authors could recover phylogenetic markers (e.g. 16S, 18S and etc) from shotgun sequencing data and then compare them with PCR-based amplicon sequence data to assess the consistency and reliability of the experimental datasets.

Authors: Thank you for this comment. We mitigate the putatively “low reproducibility” and “poor quantification” of 16S rRNA gene amplicon sequencing with our experimental design, including numerous ecological replicates, several blanks and mock communities, coupled with rigorous data filtering. Hence, we are convinced that random error is not an issue of our study.

Moreover, we believe that a comparison between amplicon sequencing and shotgun sequencing is clearly beyond the scope of this study. In fact, we dedicated a massive sequencing effort to capture most of the bacterial diversity resolved by the 16S rRNA gene. Specifically, our amplicon sequencing efforts yielded 4.0E+07 16S rRNA gene reads. While our total sequencing effort for shotgun metagenomes was even larger with 1.1E+08 reads, 16S rRNA reads only represent 4.3% of these (i.e., a total of 4.8E+06 16S rRNA reads were identified using Metaxa2). This nearly 10 fold difference in sequencing effort dedicated to 16S rRNA amplicons *versus* shotgun metagenomes would hamper meaningful comparisons.

3. Because this is a global analysis, it will be interesting to see whether the GFS microbial communities exhibit some general macroecological patterns such as latitudinal diversity patterns (LDGs), and species abundance patterns (SAD). But, no such analyses are presented.

Authors: We are most grateful for this comment, which we share with the referee. She/he may have overseen that indeed we did test for elevational patterns, corrected for latitude (Extended Data Figure 5). Moreover, Extended Data Figure 4 provides species abundance patterns and fits of the log-normal model and an ecological null (broken-stick) for comparison for all our samples. However, based on a comment of Reviewer #2, we merged this Figure and thus updated it into Extended Data Figure 5. We have also presented distance decay curves (Figure 4) as they are frequently used in macroecology. We do refer in the text to all

these analyses and understand if Extended Data sometimes are being overlooked during the review process.

4. Another very attractive theoretical question to address is how high is the microbial diversity across global GFSs, i.e., the extent of microbial biodiversity. This has recently become an interesting hot topic in global biodiversity studies. The authors can look at some representative studies (Locey et al. 2016. Proc. Natl Acad. Sci. 113, 5970–5975; Wu et al. 2019. Nature Microbiology, 4:1183–1195; Lennon and Locey 2020, Biology Direct, 15:5), and other similar studies which cited the 2016 PNAS paper).

Authors: We appreciate this comment. However, we argue that there are numerous and diverse approaches to quantify and analyze biodiversity, as demonstrated by the references provided, and one study simply cannot do everything. In our study, a major focus was on the three components of biodiversity: alpha-, beta- and gamma-diversity. As outlined in the manuscript, we present data on gamma-diversity (both observed and predicted) and amply analyze beta-diversity, using various approaches and techniques. While certainly more could be done to answer some of the theoretical questions presented in the references listed (a good idea for a future paper), we respectfully reply to this referee that our analyses on the GFS bacterial diversity is comprehensive for our outlined purposes, and doing more would exceed the paper's scope. Furthermore, given that this is the first time that GFS biodiversity is being studied at a global level, we decided to present the diversity metrics that makes our study easily comparable with others.

5. In global biodiversity pattern section, it will be interesting to perform real global biodiversity analysis by comparing the GFS microbial communities to representative microbial communities from other habitats such as lakes, ocean, sediments, and soils to experimentally demonstrate the uniqueness of the GFS microbial communities under the extreme conditions. This will also help to understand whether GFSs microbial communities are more closely related to lakes, oceans soils or they totally stand alone.

Authors: Many thanks for this comment. We are not convinced that it makes sense, in the context of our study, to compare the GFS bacterial microbiome to the sunlit or abyssal ocean, desert soils or tropical rivers, for instance, which are almost certain to be different from the GFSs.

However, may we kindly guide the reviewer's attention to our Figure 1, where we do compare the GFS community composition (ASV and KOs) to microbiomes from other systems, notably including glaciers, snow, permafrost soils, lakes and marine systems under the influence of glaciers. With these analyses, we have shown that the GFS bacterial microbiome differs indeed from these other cryospheric microbiomes — all of this discussed in our manuscript.

Therefore, with all due respect, we do not wish to potentially complicate and overload the current study with further analyses.

6. As I mentioned previously, the authors should address whether the GFS microbiome diversity is important to the functions of GFS ecosystem. The authors measured some basic geochemical variables in C, N and cycling, which can be used as proxy of community

functions. Thus, the authors could assess whether the biodiversity in the GFSs is important to the ecosystem functions to some degree by performing various statistical analyses.

Authors: Many thanks for this comment. However, this suggestion opens a completely new route, which is without any doubt beyond the scope of our study. As the referee is certainly aware of, establishing relationships between microbial diversity and ecosystem functioning is a daunting task — for numerous well-known reasons (one of them relates to functional redundancy as discussed in our manuscript). Please, note that we have not measured any elemental cycling processes *sensu stricto*, and even if we did include such analyses (i.e. determining function by comparing the community with environmental data), the results would be set on a very shallow foundation requiring large logical leaps.

We have shared with this referee as well our MAG-based study, which shall be published soon, and which does connect the GFS microbiome to functions. However, we want to stress the fact that it is still inherently difficult to bridge microbiome functions with ecosystem functioning. These are two different things.

7. The authors used iCAMP to assess the relative importance of different community assembly processes in controlling the GFS microbial community diversity and distribution. Their results (Fig. 5b) revealed that HoS increases with spatial distance, but dispersal limitation decreases with spatial distances. These results are counteractive to intuition. I expect that DL would increase with geographic distance, whereas HoS would decrease. The authors should check whether they perform iCAMP correctly, and if so, they should give some appropriate explanations. In addition, the authors could test how different community assembly processes (e.g. selection, drift) change with environmental conditions of the GFS systems (Ning et al. 2024. Nature Microbiology, 9:490-501).

Authors: We are most grateful for this comment, and we apologize for the mistake - the legend in this panel was wrong. In fact, and as expected (both by us and Reviewer #3), the relative importance of HoS remains invariant across spatial distance but dispersal limitation increases with increasing distance. We have corrected this mistake in Fig. 5. In the course of the revisions, we changed from QPEN and the phyloscore analysis to iCAMP - and extensively tested the iCAMP settings. We are certain that there are no issues with our final iCAMP analysis (besides the error in the legend mentioned above). Thank you for pointing to the recent work by Ning et al. 2024. The conceptual framework described there is certainly useful when assembly processes are studied along environmental gradients - such as in the stress gradient of the Oak Ridge Field groundwater. However, and importantly, GFSs share fundamental environmental constraints, such as low temperature, ultra-oligotrophy and unstable streambeds. Therefore, across GFS within regions environmental gradients may not bear sufficient variance that would warrant such an analysis such as performed in by Ninhg and colleagues.

8. The authors also use db-RDA to partition community variations. Their results indicated that 45.1% variations could not be explained, which could be due to unmeasured environmental variables, biotic interactions and/or drift. The authors should compare the results with the estimations of drift from iCAMP to assess the consistency across different approaches. But I failed to locate the data for drift's estimation.

Authors: Thanks for this comment. Results from VP indeed indicate that there is 45.1% of the total variance that could not be explained. This level of variance explained is very close to the 49.7% reported from other studies on microorganisms and well bracketed by studies on animals and plants (doi:10.1038/nrmicro2795). Indeed and not unexpected (as also shown and discussed in the aforementioned study), as pointed out by the reviewer, such a proportion of unexplained variance could result, among other, from drift, biotic interactions, and unmeasured environmental variables. Thus, one cannot evaluate/quantify how much of this unexplained variance is attributable to drift. In addition, db-RDA/VP and iCAMP are based on different methodological and conceptual approaches; therefore, it would not be appropriate to directly compare their respective results.

9. Line 258. The section title, Taxonomy of the GFS microbiome, is not very attractive. It could be more interesting and informative to change it to global core bacterial community or global core microbiomes if other microbial groups are included (see Point 12). Also, the authors could test whether the core microbial communities are important to the functions of GFS ecosystems by using various statistical analyses.

Authors: Thank you for this comment. We have now changed the sub-heading to better refer to the core microbiome and its small size. As mentioned previously, relating microbiome structure and functions to ecosystem-level functions (e.g. metabolism, elemental cycling) is a daunting task (it may work in reductionist lab-based settings), which would be clearly beyond the scope of this manuscript with its thematic focus on the GFS microbiome biodiversity and biogeography.

10. As I mentioned before as well as indicated by Reviewer 2, the experimental design and the data generated by the authors could not allow the authors to address the research questions in mind (e.g. endemism, functional redundancy). It is good to not use the words endemic or endemicity, which will help to avoid potential confusion and misunderstanding. Using specific or specificity also problematic because, to me, specific or specificity and endemic or endemicity means same within this context. The authors can easily get away from this problem without using them. For example, this sentence “ASV specificity was not uniformly distributed across mountain ranges” can be easily rewrote as ASV was not uniformly detected across mountain ranges.

Authors: We are grateful for this comment and happy to see that Reviewer #2 has now accepted the use of ‘specific ASVs’ as suggested by her/himself. Semantologically, the alternative suggested by this reviewer is not different from ASV specificity. Therefore, with all due respect, we will keep the latter.

11. The authors could perform structural equation modeling (SEM) analysis to assess how important various environmental factors/climate change factors are in shaping the diversity and distributions of the global GFS microbiomes.

Authors: We used db-RDAs and VP to explore how the GFS environment but also climatic factors shape beta-diversity of the GFS bacterial communities across spatial scales. Our focus was to investigate the proportion of variance explained by predictors within and among mountain ranges, not to model causal pathways and relationships among a set of variables.

Furthermore, we did not aim at exploring how climate change may affect the GFS microbiome as suggested by this reviewer. Please note that we have submitted a full study using models to forecast the GFS microbiome under various climate change scenarios.

db-RDAs and VP combined are widely used across ecological studies because this framework is well-suited for nested multi-scale data — such as ours — and requires far fewer assumptions compared to SEM. Accommodating numerous a priori assumptions between variables in SEM would increase model complexity, potentially leading to difficulties in achieving convergence and impeding result interpretation. Therefore, with all due respect, we do not see the superiority of SEM over our db-RDAs and VP framework for the questions that we have addressed.

12. The habitats are unique, but the analyses are not as comprehensive as it should have been. The authors examined just the diversity of bacteria. As a comprehensive study, it will be very interesting to determine the diversity and distributions of other groups of microbiota such as archaea, fungi, and protists. Archaea could be important in this habitat from biodiversity perspective. But, the currently used primers are not good to recover archaea sequences. To examine the diversity and distribution of archaea, archaea-specific primers must be used. The diversity of fungi and protists can be easily done by sequencing 18S genes. In fact, the shotgun sequences could include sequences from all these groups of microorganisms although the majority could be bacteria. Thus, it will be nice to use amplicon-sequencing approach to also examine the diversity of archaea, fungi, and protists in these important habitats to make the data more comparable and comprehensive.

Authors: This is a valuable idea, which may be the topic of future studies — each with specific questions. As has been clear from the first review round onwards, the focus of this manuscript is on bacteria. The main reason for this decision is the pivotal function of bacteria in GFS biofilms; we do acknowledge the role of algae and other eukaryotes in these systems — however, outweighed by bacteria for some of the most prominent ecosystem functions as in most aquatic ecosystems. A further argument is scope; accommodating the same /similar analyses multiple times for ‘other groups of microbiota such as archaea, fungi, and protists’ as we did for bacteria would generate a paper that is simply hard to read.

Again, we have shared our MAG-based study with this referee as well. This manuscript, currently in a most advanced stage of the review process, highlights the cross-domain complexity of the GFS biofilm, with bacteria outweighing other ‘microbiota’.

To avoid any further ambiguity and as suggested by the editor, we have now better emphasized throughout the text that our study is one on bacteria.

13. It is not easy to understand the meanings of the titles of some subsections, e.g. “Taxonomy of the GFS microbiome”, “The phylogenetic structure of beta-diversity and specificity”, “Drivers of GFS microbiome beta-diversity”. They are not straightforward, and it is hard to figure out what they really mean from ecological standpoints. Based on the text, I think they should simply use more common terms, as I mentioned above, e.g., global core bacterial community, community assembly mechanisms, environmental drivers of the GFS microbiomes.

AUTHORS: We are grateful for this comment. We have now adjusted several of the sub-headings accordingly.