January 14, 2021

Dear Dr. Veronique van den Berghe,

We would like to thank you for taking time to handle our manuscript (“Compilation of longitudinal microbiota data and hospitalome from hematopoietic cell transplantation patients”) and encouraging a revision of the manuscript for further consideration.

We also want to thank the two anonymous reviewers for their constructive comments they made. We are very glad that the reviewers are positive and found our compiled data potentially useful for the field.

We have made revisions to our manuscript in response to your and the reviewers’ comments. Our detailed, point-to-point responses to the specific comments are listed below.

**Editor’s Comments:**

**Comment 1: “***Dear Professor Xavier,*

*First of all, our best wishes for the new year and please accept our apologies for the delay in sending you a decision on your manuscript. Your manuscript entitled "Compilation of longitudinal microbiota data and hospitalome from hematopoietic cell transplantation patients" has now been seen by the referees, whose comments are appended below. As you will see, the referees were supportive, but they raised important points and made suggestions for modifications, which they felt must be addressed before this work would be acceptable for publication at Scientific Data.*

*In line with Reviewer #1's comments, we request that you make all SRR accessions public or at least provide the SRA metadata. Also, could project level identifiers be added to the tblASVsamples file?*

*Based on the recommendation from the handling Editorial Board member, Nicola Segata, we therefore invite you to revise and resubmit your manuscript, taking into account the points raised.***”**

**Author Response**: Thank you and the other editor Nicola Segata for handling our manuscript and providing us feedbacks from reviewers during the pandemic, when a delayed review process is totally understandable. Below we have addressed all their comments and made corresponding modifications in our manuscript.

In response to your specific concern, we have made all SRR accessions public and provided project-level identifies (column “BioProject”) in the tblASVsamples file (<https://figshare.com/articles/dataset/samples/12016983>). We described this new column in the section “Data Records” of our manuscript (Line 314).

**Comment 2: “***If you have not already done so, at this time, we ask you to use our metadata tool (https://scientificdata.metadata-creator.com) to submit key characteristics about the data described in SDATA-20-01150. Submitting your metadata at this point will result in a quicker publication process should your manuscript be accepted. Our in-house curation team uses the information submitted to finalise the machine-readable metadata for all accepted Data Descriptor manuscripts.***”**

**Author Response**: We have submitted our metadata using the online tool.

**Comment 3: “***At the same time, we ask that you ensure your manuscript complies with our format requirements explained in full in our Submission Guidelines:*

*https://www.nature.com/sdata/publish/submission-guidelines*

*\* Please ensure that your references, code and data citations conform fully to the Nature style. See the examples at the links below:*

*https://www.nature.com/sdata/publish/submission-guidelines#refs> and**https://www.nature.com/sdata/publish/submission-guidelines#data\_citations>   
  
For example, with the current details reference 36 should become:*

*Liao, C. samples. figshare* https://doi.org/10.6084/m9.figshare.12016983.v6 *(2020).*

*You may want to adapt the title and co-authors of your figshare files, so that the references become more informative as well.***”**

**Author Response**: We have checked the submission guideline thoroughly and ensured that our manuscript, including the reference style, conforms to the general guidelines.

Regarding data citations, we have put together all our datasets within a single figshare collection (<https://figshare.com/collections/Compilation_of_longitudinal_microbiota_data_and_hospitalome_from_hematopoietic_cell_transplantation_patients/5271128>) and generated the following informative citation format including title and author

Liao, C. & Xavier, J. B. Compilation of longitudinal microbiota data and hospitalome from hematopoietic cell transplantation patients. *figshare* <https://doi.org/10.6084/m9.figshare.c.5271128.v1> (2021).

This reference has been used in the manuscript when referring to our data.

**Comment 4: “***\* Please upload all figures and tables as separate files. Tables should preferably be in Excel format. Line-art figures should be provided in a suitable vector-graphic format, ideally by saving them directly in the EPS or PDF formats from the program used to create them.* *Please note that some of the coloured boxes in Figure 5 are difficult to distinguish.***”**

**Author Response**: We have uploaded all figures (no table was provided) as separate PDF files in vector-graphic format. For Figure 5, we created different patterns for boxes with similar colors to distinguish among them.

**Comment 5: “***\* For formatting reasons, please change the following section titles* *"Contributions" and "Acknowledgement" into "Author contributions" and "Acknowledgements" respectively. Please also replace "subjects" by "participants" at line 100.***”**

**Author Response**: We have implemented these changes (Line 152, 572, 585).

**Reviewer #1**

**Comment 1: “***The authors provide a collection of >10,000 fecal microbiota samples longitudinal collected from >1,000 patients, along with some analyses of those specimens. Given the scope of the journal this review focuses on the dataset.*

*First, this is a large and rich dataset with excellent potential for re-use. Metagenomes were sequenced on an Illumina MiSeq platform using a paired-end 250 × 250 reads. The dataset includes:*

* *Counts of mapped 16S sequences*
* *16S qPCR data for total microbiome estimation.*
* *Metadata about hematopoietic cell transplantation*
* *Timing and route of drug administration*
* *Dates of positive blood cultures*
* *Repeated body temperature measures*
* *Daily measurements of white blood cells, platelets and red blood cells*
* *NCBI SRA accession numbers for the sequencing data*
* *Silva taxonomy tables for OTUs*

*The sequencing data are distributed through NCBI’s Sequence Read Archive, and all other data through Figshare. I was able to easily download all these data as CSV files, load them into R, and join them based on sample ID and patient ID. It is a great improvement over the presentation of data and metadata in most publications that are not data-centric. I have only some moderate and minor comments.***”**

**Author Response**: We thank the reviewer for taking time to evaluate our data tables and manuscript. We are glad to know that the reviewer thinks our data is potentially useful.

**Comment 2: “***Moderate: "Did the authors provide all of the information needed for others to reuse this dataset, or integrate it with other data?": I selected No here only because the SRR accessions are not publicly visible on SRA, and I didn’t see a reviewer link to access them. I imagine these are just still not yet made public, but would want to make sure they are made public with at least minimal metadata on SRA with publication. The data in file tblASVsamples enables linking the sequence data to sampleID and patientID in the figshare metadata.***”**

**Author Response**: We are sorry if the reviewer cannot access some sequences via SRR accessions. All our sequences have been made publicly visible before initial submission. Some of the sequences, however, have been submitted to SRA more than once under different projects and the later submissions may or may not replace the earlier versions, leaving some SRR accessions obsolete. We apologize for the obsolete SRR accessions in the *tblASVSamples* file we provided in the initial submission.

In the revised *tblASVSamples* file, we have corrected the SRR accessions by using the accession number of the most recent submission (if multiple submissions exist) of the same sequencing data. We explained how we chose the SRR accession in the manuscript (Lines 311-312).

**Comment 3: “***The DADA2 taxonomic assignment is great because it doesn’t involve OTU picking. But it does use different “Amplicon Sequence Variants” (ASVs) for each sample, so conversion to wide format and naming of taxa is needed for most analysis. It would help some users to also provide a taxa x samples matrix version of the ASV table.***”**

**Author Response:** We agree with the reviewer that a taxa-by-sample table of relative abundance may be convenient for readers who will reuse our data. Therefore, we provide such tables at different taxonomic levels (ASV, genus, family, order, class, phylum) and described these tables in the manuscript (Lines 233-234).

**Comment 4: “***Minor: The second fig share URL has an error (*[*https://doi/org/10.6084/m9.figshare.12016989*](https://doi/org/10.6084/m9.figshare.12016989)*- should be doi.org, not doi/org)***”**

**Author Response**: Thanks for pointing out the typo. We have merged all four folders into one collection (<https://figshare.com/collections/Compilation_of_longitudinal_microbiota_data_and_hospitalome_from_hematopoietic_cell_transplantation_patients/5271128>) and, therefore, the four figshare URLs (including the incorrect one) has been replaced by a single URL for the entire data collection (<https://doi.org/10.6084/m9.figshare.c.5271128.v1>).

**Comment 5: “***Folder “samples” and Folder “taxonomy” – these are provided as files, not folders.***”**

**Author Response**: We created folders for “samples” and “taxonomy” on purpose. Although there is a single file in each folder at the moment, we expect to expand the contents of these folders by cleaning up more data in the future. For example, the taxonomic classification of fungal ITS sequences may be saved to a different file in the “taxonomy” folder. We explained our rationale in the manuscript (Line 220-222).

**Reviewer #2**

**Comment 1: “***Nice repository model.***”**

**Author Response**: We thank the reviewer for reading and commenting on our manuscript.

**Comment 2: “***Some comments: Have you added, in the repo, the episodes of GVHD? How many did you find? Have you identified some risk factors for them?***”**

**Author Response**: We appreciate the reviewer’s interest in GVHD outcomes. We have not added GVHD data as detailed GVHD outcomes are not available for the entire patient cohort without vetting by manual chart review. We regretfully cannot publicly post such patient-level outcomes, even de-identified.

We have previously reported that microbiome risk factors for GVHD-related mortality include low alpha-diversity, low abundance of Blautia [1] and domination by *Enterococcus* [2] and that chronic GVHD is associated with high circulating concentrations of microbiome-derived short-chain fatty acids [3]. We indicated these GVHD-microbiome links in the manuscript (Lines: 118-119).

**Comment 3: “***About your patients: have you registered other factors that influence gut microbiota composition (e.g. diet, delivery, breastfeeding, PPI use)?****”***

**Author Response**: We have previously reported drug exposure [4] and host-microbiome interaction [5] as two major drivers of gut microbiota dynamics in our allo-HSTC patients. Both drug administration records and counts of white and red blood cells have been released together with microbiome data in the current data collection.

We have not included other factors that might be influencing as well in our dataset. However, there are ongoing projects in our institute studying the associations of these factors such as diet with gut microbiota composition. In the revised manuscript, we indicated that it is not yet available to analyze the effects of these factors using our data (Line 525-529).

**Reference:**

[1] Jenq, Robert R., et al. "Intestinal Blautia is associated with reduced death from graft-versus-host disease." Biology of Blood and Marrow Transplantation 21.8 (2015): 1373-1383.

[2] Stein-Thoeringer, C. K., et al. "Lactose drives Enterococcus expansion to promote graft-versus-host disease." Science 366.6469 (2019): 1143-1149.

[3] Markey, Kate A., et al. "Microbe-derived short chain fatty acids butyrate and propionate are associated with protection from chronic GVHD." Blood Journal (2020): blood-2019003369.

[4] Morjaria, Sejal, et al. "Antibiotic-induced shifts in fecal microbiota density and composition during hematopoietic stem cell transplantation." Infection and immunity 87.9 (2019): e00206-19.

[5] Schluter, Jonas, et al. "The gut microbiota is associated with immune cell dynamics in humans." Nature (2020): 1-5.