

## Microbiome Scoping Review Data extraction form

*Adapted from an original Qualtrics survey*

Who extracted information from this article?

(select one)

N/A<sup>1</sup>

### ARTICLE INFORMATION

1. What is the covidence citation number?

(free text)

covidence\_number

2. What is the article DOI?

(free text)

doi

3. What is the article title?

(free text)

title

4. Who were the authors of the article?

(free text)

author\_list

5. What was the primary purpose of the combined analysis? (Why did they combine data)

(select all)

- Replication or validation of targeted findings from a single cohort
- To describe the core microbiome of an environment. (The core microbiome is a shared set of bacteria found across all samples)
- To compare synthesize across multiple studies to describe a population
- The study described didn't actually involve combining multiple data sets from multiple prior publications or sequencing data sets

**[End Survey]**

- Not described

**[End Survey]**

- Other

(free text)

meta\_purpose[replication\_validation]

meta\_purpose[core\_microbiome]

meta\_purpose[population\_synthesis]

meta\_purpose[not\_combo]<sup>2</sup>

meta\_purpose[not\_described]<sup>2</sup>

meta\_purpose[other]

meta\_purpose\_other

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

6. What was the primary goal of this meta-analysis (what was the hypothesis they were testing)?

*(select all)*

- To characterize a microbiome-outcome association
- To characterize a microbiome-exposure association
- To characterize assembly, development, or temporal variation in the microbial community
- To characterize a unique microbiome (microbiome environment, population)
- Other *(free text)*
- Not described

meta\_goal[microbiome\_outcome]  
meta\_goal[microbiome\_exposure]  
meta\_goal[assembly]  
meta\_goal[unique\_microbiome]  
meta\_goal[other]  
meta\_goal\_other  
meta\_goal[not\_described]<sup>2</sup>

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

### SYSTEMATIC APPROACH TO DATA COLLECTION

7. Was there a clear systematic approach for gathering reference data sets? *(select one)* systematic approach
- No **[Skip 8-10]**
  - Yes **[Answer 8-10]**
- [If 7 was "yes"]**
8. Was the search strategy for this meta-analysis described? *(select one)* search\_described
- No
  - Yes
- [If 7 was "yes"]**
9. Was a PRISMA diagram describing how papers were selected included in the paper? *(select one)* prisma\_included
- No
  - Yes
- [If 7 was "yes"]**
10. Was there a clear set of inclusion/exclusion criteria for the final studies? *(select one)* study\_inclusion
- No
  - Yes

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

## DATA SOURCES

- |  |  |
|--|--|
| <p>11. How many studies were analyzed in the primary meta analysis, after the data was processed? (If this is not listed, leave it blank) <span style="float: right;"><i>(free text)</i></span></p>  | <p>num_studies_1o</p>  |
| <p>12. How many samples were analyzed in the primary meta-analysis, after the data was processed? (If not listed, leave blank) <span style="float: right;"><i>(free text)</i></span></p>   | <p>num_samples_1o</p>  |
| <p>13. What are the sources for sequencing and metadata for data included in the primary meta-analysis? <span style="float: right;"><i>(select all)</i></span></p> <ul style="list-style-type: none"> <li>• The authors of the meta-analysis or consortium</li> <li>• SRA/GenBank</li> <li>• ENA</li> <li>• Request to the authors of the original data set</li> <li>• MG-RAST</li> <li>• QIIME DB or Qiita</li> <li>• Other <span style="float: right;"><i>(free text)</i></span></li> <li>• Not described</li> </ul> | <p>sources_1o[consortium_authors]<br/>sources_1o[SRA]<br/>sources_1o[ENA]<br/>sources_1o[request_authors]<br/>sources_1o[mg_rast]<br/>sources_1o[qitta]<br/>sources_1o[other]<br/>sources_1o_other<br/>sources_1o[not_described]</p> |
| <p>14. Did the study include any independent validation cohorts? (not included in the original analysis) <span style="float: right;"><i>(select one)</i></span></p> <ul style="list-style-type: none"> <li>• No <span style="float: right;"><b>[Skip 15-17]</b></span></li> <li>• Yes <span style="float: right;"><b>[Answer 15-17]</b></span></li> </ul>  | <p>validation</p>  |
| <p><b>[If 14 was "yes"]</b></p>  |  |
| <p>15. How many studies were analyzed for validation (final number after processing)? (If not known, leave blank) <span style="float: right;"><i>(free text)</i></span></p>  | <p>num_studies_valid</p>   |

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

**[If 14 was "yes"]**

16. How many samples were analyzed for validation (final number after processing)? (if not listed, leave blank) *(free text)*

num\_sample\_validation

**[If 14 was "yes"]**

17. What were the sources for sequences and metadata for the validation analysis? *(select all)*

- The authors of the meta-analysis or consortium
- SRA/GenBank
- ENA
- Request to the authors of the original data set
- MG-RAST
- QIIME DB or Qiita
- Not described
- Other

*(free text)*

valid\_source[consortium\_authors]  
valid\_source[SRA]  
valid\_source[ENA]  
valid\_source[request\_authors]  
valid\_source[mg\_rast]  
valid\_source[qiita]  
valid\_source[not\_described]  
valid\_source[other]  
valid\_source\_other

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

## ENVIROMENT

18. Which environments were included? *(select all)*

- Built environment (i.e. house walls, office floor, etc)
- Host associated (non-human animals)
- Human-associated
- Not described
- Other

[Answer 20, 21, 22]

[Answer 20]

*(free text)*

envo[built\_envo]  
envo[host]  
envo[human]  
envo[not\_described]<sup>2</sup>  
envo[other]  
envo\_other

**[If " Host associated (non-human animals)" or " Human-associated" were selected in 18]**

19. If human or non-human animals were included, which body sites were analyzed? *(select all)*

- Gut (Feces, rectal swab, biopsy)
- Oral
- Urogenital
- Skin
- Airway
- Other

*(free text)*

envo\_bodysite[gut]  
envo\_bodysite[oral]  
envo\_bodysite[urogenital]  
envo\_bodysite[skin]  
envo\_bodysite[airway]  
envo\_bodysite[other]  
envo\_bodysite\_other

**[If " Host associated (non-human animals)" was selected in 18]**

20. What non-human animal species were included? *(free text)*

species

**[If " Host associated (non-human animals)" was selected in 18]**

21. If non-human animals were included, was this analysis a comparison between host species?  
*(select one)*

species\_compare

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

## STUDY DESCRIPTION

22. Which of the following are reported for the studies that were combined? (select all)

- Population description
- Experimental design (randomization, variable matching)
- Sampling method(s)/collection kit(s)
- Collection kit(s) used
- Extraction kit(s) used
- Hypervariable region(s)
- Sequencing platform

[Answer 24]

[Answer 23]

```
design_info[population]
design_info[experimental_design]
design_info[sampling_method]
design_info[collection_kit]
design_info[extraction_kit]
design_info[hypervariable_region]
design_info[sequencer]
```

[If "Sequencing platform" was selected in 12]

23. Which sequencing platforms were used? (select all)

- 454 pyrosequencing
- Illumina (MiSeq, HiSeq, NovoSeq)
- Ion Torrent
- Pac Bio
- Oxford Nanopore
- Other

(free text)

```
seq_platforms[454]
seq_platforms[illumina]
seq_platforms[ion_torrent]
seq_platforms[pac_bio]2
seq_platforms[nanopore]2
seq_platforms[other]
seq_platforms_other
```

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

**[If "Hypervariable region(s)" was selected in 12]**

24. Which hypervariable regions were included

(select all)

- V12
- V13
- V2
- V23
- V3
- V34
- V35
- V4
- V45
- V46
- V68
- V69
- Ion torrent or multiple region kit (TSLR)
- Other

(free text)

hypervar\_regions[V12]  
hypervar\_regions[V13]  
hypervar\_regions[V2]  
hypervar\_regions[V23]  
hypervar\_regions[V3]  
hypervar\_regions[V34]  
hypervar\_regions[V35]  
hypervar\_regions[V4]  
hypervar\_regions[V45]  
hypervar\_regions[V46]  
hypervar\_regions[V68]  
N/A<sup>2</sup>  
N/A<sup>2</sup>  
hypervar\_regions[other]  
hypervar\_regions\_other

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set



## TABLE CONSTRUCTION

25. Were all the data processed using a similar pipeline?

*(select one)*

- Yes
- No
- Not described

**[Answer 26-33]**

**[skip 26-33]**

**[skip 26-33]**

common\_pipeline

**[If 25 is "yes"]**

26. Did the authors perform taxonomic profiling and analyze the data? (i.e. denoise to ASVs, OTU clustering)?

*(select one)*

- Yes
- No

taxa\_profile

**[If 25 is "yes"]**

27. Did the authors perform functional profiling? (i.e. tools like PICRUSt, Tax4Fun, mention of KEGGs in results)

*(select one)*

- Yes
- No

fun\_profile

**[If 25 is "yes"]**

28. Were sequences denoised? (This might be indicated by the used of DADA2, Deblur, Unoise3, or the description of features as ASVs)?

*(select one)*

- No
- Yes
- Unclear/Not described

**[Skip 31; Answer 32]**

asvs

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

**[If 25 is "yes"]**

29. Were the sequences clustered into operational taxonomic units (OTUs)? (This might be indicated by the use of algorithms like mothur, usearch, or vsearch or the mention of clustering)(*select one*)

- No
- Yes
- Unclear/Not described

**[skip to 30]**

**[Answer 30; skip 32]**

**[skip 30]**

otus

**[If 29 is "yes"]**

30. What type of OTU clustering was performed and used for primary analysis (if multiple types of clustering were performed, or validation clustering was perform)? (*select all*)

- de novo clustering
- open reference clustering
- closed reference clustering
- unclear
- not described

otu\_type[de\_novo]  
otu\_type[open\_ref]  
otu\_type[closed\_ref]  
otu\_type[unclear]  
otu\_type[not\_described]

**[If 28 is not "yes" AND 29 is not "yes"]**

31. If the data was not denoised or clustered into OTUs, was the data collapsed to a taxonomic level without denoising or clustering? (*select one*)

- No
- Yes
- Unclear/Not described

qual\_filter

**[If 28 is "yes" OR 30 is " de novo clustering" OR 30 is " open reference clustering "]**

32. If the authors performed denoising without clustering, open reference clustering or de novo clustering, how did they describe their phylogenetic tree construction? (*select one*)

- They did not construct a phylogenetic tree (no UniFrac or Faith's diversity)
- They used MAFFT, Fast Tree, or another de novo approach
- They used fragment insertion (SEPP, epa-ng)
- Unclear/Not described

tree\_construction

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

**[If 29 is "yes"]**

33. Which tools were used to generate the feature table?

(select all)

- DADA2 in R or QIIME 2
- Deblur
- Unoise
- Userach
- Uclust
- Mothur
- CD-Hit
- SortMeRNA
- Sumacrust
- QIIME 1
- QIIME 2
- QIIME DB or Qiita
- Vsearch
- Naive Bayesian Classifier
- Cutadapt
- Clawback
- PICRUST 1
- PICRUST 2
- PICRUST (version not listed)
- Tax4fun
- Kraken
- Other

(free text)

```
table_tools[dada2]
table_tools[deblur]
table_tools[unoise]
table_tools[userach]
table_tools[uclust]
table_tools[mothur]
table_tools[cd_hit]2
table_tools[sortmerna]2
table_tools[sumacrust]2
table_tools[qiime1]
table_tools[qiime2]
table_tools[qiita]
table_tools[vsearch]
table_tools[nb_classifier]
table_tools[cutadapt]
table_tools[cutadapt]2
table_tools[picrust1]
table_tools[picrust2]
table_tools[picrust_tbd]
table_tools[tax4fun]
table_tools[kraken]
table_tools[other]
table_tools_other
```

34. Are there any other notes you'd like to add for the feature table generation?

(free text)

```
table_notes
```

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

## ANALYSIS

35. Which aspects of the microbiome were analyzed?

(select all)

- Descriptive analysis including taxonomy plots **[Answer to 36]**
- Alpha Diversity **[Answer 37, 38]**
- Beta Diversity **[Answer 39, 40]**
- Differential abundance **[Answer to 41-44]**
- Sample classification (i.e. random forest) **[skip to 26-31]**
- Core microbiome analysis (looking for a set of features common across all the samples)
- Co-occurrence networking
- Not described
- Other

(free text)

```
analyses_perf[descriptive]
analyses_perf[alpha]
analyses_perf[beta]
analyses_perf[differenital_abundance]
analyses_perf[sample_classification]
analyses_perf[core_microbiome]
analyses_perf[co_occurance]
analyses_perf[not_described]2
analyses_perf[other]
analyses_perf_other
```

**[If " Descriptive analysis including taxonomy plots" in 35]**

36. Which taxonomic levels were used for descriptive analysis?

(select all)

- OTU/ASV
- Species
- Genus
- Family
- Order
- Class
- Phylum
- Not described

```
taxa_level[otu_asv]
taxa_level[species]2
taxa_level[genus]
taxa_level[family]
taxa_level[order]2
taxa_level[class]2
taxa_level[phylum]
taxa_level[not_described]
```

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

**[If "Alpha Diversity" in 35]**

37. At which taxonomic levels was alpha diversity analyzed? (Faith's PD must be analyzed at the OTU/ASV level) *(select all)*

- OTU/ASV
- Species
- Genus
- Family
- Order
- Class
- Phylum
- Not described

alpha\_level[otu\_asv]  
alpha\_level[species]  
alpha\_level[genus]  
alpha\_level[family]  
alpha\_level[order]  
alpha\_level[class]  
alpha\_level[phylum]  
alpha\_level[not\_described]

**[If "Alpha Diversity" in 35]**

38. How did the authors handle differences between studies in their alpha diversity analyzes?*(select all)*

- Pooled all samples within a study
- Not adjustment for study effects (i.e. kruskal wallis testing)
- Adjusted for study as a fixed term (linear regression)
- Linear mixed effects model with study as a random effect
- Meta-analysis framework (i.e. forest plot, effect pooling)
- Not described
- Other

*(free text)*

alpha\_study[pooled]  
alpha\_study[no\_adjustment]  
alpha\_study[fixed]  
alpha\_study[random]  
alpha\_study[meta\_analysis]  
alpha\_study[not\_described]  
alpha\_study[other]  
alpha\_study\_other

**[If "Beta Diversity" in 35]**

39. At which taxonomic levels was beta diversity analyzed? (UniFrac distance must be analyzed at the OTU/ASV level) *(select all)*

- OTU/ASV
- Species
- Genus
- Family
- Order
- Class
- Phylum
- Not described

beta\_level[otu\_asv]  
beta\_level[species]  
beta\_level[genus]  
beta\_level[family]  
beta\_level[order]  
beta\_level[class]  
beta\_level[phylum]  
beta\_level[not\_described]

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

**[If "Beta Diversity" in 35]**

40. How did beta diversity analysis handle study effects?

(select all)

- Descriptive PCoA that did not show study effects
- PCoA showing study effect
- Adonis permanova showing or adjusting for study effect size
- Other test adjusted for study effect
- Not adjustment or acknowledgment of study effect
- Not described
- Other

(free text)

beta\_study[no\_study\_pcoa]  
beta\_study[pcoa\_study\_effects]  
beta\_study[adonis\_adj]  
beta\_study[other\_test\_adj]  
beta\_study[no\_adj]  
beta\_study[not\_described]  
beta\_study[other]  
beta\_study\_other

**[If "Differential abundance" in 35]**

41. At which taxonomic levels was differential abundance analyzed?

(select all)

- OTU/ASV
- Species
- Genus
- Family
- Order
- Class
- Phylum
- Not described

diff\_abund\_level[otu\_asv]  
diff\_abund\_level[species]  
diff\_abund\_level[genus]  
diff\_abund\_level[family]  
diff\_abund\_level[order]  
diff\_abund\_level[class]  
diff\_abund\_level[phylum]  
diff\_abund\_level[not\_described]

**[If "Differential abundance" in 35]**

42. What types of microbiome differential abundance testing was performed?

(select all)

- Differential abundance of a targeted taxa (IE is F. nucleatum higher in all the studies)
- Untargeted differential abundance

diff\_target[targeted]  
diff\_target[untargeted]

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

**[If "Differential abundance" in 35]**

43. How was the data filtered before differential abundance? *(select all)*

- The features were not filtered
- The filter was applied individually in each study
- The filter was applied across all studies
- the features were filtered based on prevalence
- The features were filtered based on abundance
- The features were filtered based on taxonomic assignment or other information

diff\_filter[no\_filtered]  
diff\_filter[per\_study]  
biofilter[consistent]  
diff\_filter[prevelance]  
diff\_filter[abundance]  
diff\_filter[taxonomy\_meta]

**[If "Differential abundance" in 35]**

44. How were multiple studies handled in differential abundance testing? *(select all)*

- The studies were pooled with no effect adjustment
- Pooled analysis adjusted for study or technical effect (fixed effect)
- Pooled analysis with study as a random effect
- A comparison of results across individual cohorts (i.e. forest plot, effect pooling)
- Not described
- Other

*(free text)*

diff\_type[pool\_no\_adj]  
diff\_type[fixed\_effect]  
diff\_type[random\_effect]  
diff\_type[meta\_analysis]  
diff\_type[not\_described]  
diff\_type[other]  
diff\_type\_other

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

**[If "Differential abundance" in 35]**

45. Which differential abundance algorithm(s) were used?

(select all)

- t-test or ordinary least squares regression on rarefied data
- kruskal wallis test on rarefied data
- LefSe
- DeSeq2
- CornCob
- edgeR
- MaAslin
- metagenomSeq
- limma voom
- kruskal wallis, t-test, or other parametric model with on CLR-transformed data
- ALDEx2
- ANCOM I or ANCOM II
- ANCOM-BC
- PhILR or PhyloFactor
- Gneiss
- Differential ranking (songbird, Birdman, Bayesian DR)
- Not described
- Other

(free text)

```
diff_algorithm[rare_kw]2
diff_algorithm[LefSe]
diff_algorithm[DeSeq2]
diff_algorithm[corncob]2
diff_algorithm[edgeR]2
diff_algorithm[MaAslin]
diff_algorithm[metagenomSeq]
diff_algorithm[limma_voom]
diff_algorithm[paremtric_clr]
diff_algorithm[aldex2]
diff_algorithm[ancom]
diff_algorithm[ancom_bc]
diff_algorithm[philr]2
diff_algorithm[nphilr]2
diff_algorithm[dr]
diff_algorithm[not_described]
diff_algorithm[other]
diff_algorithm_other
```

**[If " Sample classification (i.e. random forest)" in 35]**

46. What taxonomic level(s) were used to build a sample classifier?

(select all)

- OTU/ASV
- Species
- Genus
- Family
- Order
- Class
- Phylum
- Not described

```
class_level[otu_asv]
class_level[species]
class_level[genus]
class_level[family]
class_level[order]
class_level[class]
class_level[phylum]
class_level[not_described]
```

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set



**[If " Sample classification (i.e. random forest)" in 35]**

47. How was the classifier trained? *(select one)*

- It was trained on a single study
- It was trained on multiple studies with no consideration of study effects
- Leave a study out or cross validation training
- Not described
- Other

class\_train

*(free text)*

class\_train\_other

**[If " Sample classification (i.e. random forest)" in 35]**

48. Did the authors validate their classifier on other cohorts not originally used to train the classifier? *(select one)*

- Yes
- No
- Not described

class\_validate

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set

**OTHER NOTES**

49. Is there anything else you'd like to say about the study?

(free text)

final\_notes

<sup>1</sup>Internal Use only

<sup>2</sup>Not included in data set