

CORRESPONDENCE

Open Access



Realising respiratory microbiomic meta-analyses: time for a standardised framework

David Broderick^{1,2}, Robyn Marsh³, David Waite¹, Naveen Pillarisetti⁴, Anne B. Chang^{3,5,6} and Michael W. Taylor^{1*}

Abstract

In microbiome fields of study, meta-analyses have proven to be a valuable tool for identifying the technical drivers of variation among studies and results of investigations in several diseases, such as those of the gut and sinuses. Meta-analyses also represent a powerful and efficient approach to leverage existing scientific data to both reaffirm existing findings and generate new hypotheses within the field. However, there are currently limited data in other fields, such as the paediatric respiratory tract, where extension of original data becomes even more critical due to samples often being difficult to obtain and process for a range of both technical and ethical reasons. Performing such analyses in an evolving field comes with challenges related to data accessibility and heterogeneity. This is particularly the case in paediatric respiratory microbiomics — a field in which best microbiome-related practices are not yet firmly established, clinical heterogeneity abounds and ethical challenges can complicate sharing of patient data. Having recently conducted a large-scale, individual participant data meta-analysis of the paediatric respiratory microbiota ($n = 2624$ children from 20 studies), we discuss here some of the unique barriers facing these studies and open and invite a dialogue towards future opportunities.

Introduction

Meta-analyses represent a powerful approach to maximise the use of existing data to improve an area of scientific understanding [1]. Such studies often extend the usefulness of data beyond the original research question for which they were generated, allowing the data to

continue to provide novel and valuable insights. By combining data from multiple studies, one can dramatically increase statistical power and seek to identify overarching patterns that may not be evident from a single study. This is particularly so for sample types that are difficult to obtain, and therefore, sample numbers are constrained. One such environment is the paediatric lower airways where large-scale microbiota data are limited compared to other fields (e.g. the gut). Individual participant data (IPD) meta-analyses, which utilise more granular data, are even more powerful than meta-analysis of studies where only overall results are considered. As a form of secondary analysis, meta-analyses also allow for the reproduction of original results, a concern highlighted in recently published literature [2, 3]. Additionally, combining data in meta-analyses allows for broader overarching questions — such as determining if there are non-specific disease signatures in the respiratory microbiota — to be investigated in a manner that is largely unfeasible within

*Correspondence:

Michael W. Taylor
mw.taylor@auckland.ac.nz

¹ School of Biological Sciences, University of Auckland, Auckland, New Zealand

² Present Address: Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand

³ Child Health Division, Menzies School of Health Research, Charles Darwin University, Darwin, NT, Australia

⁴ Starship Children's Hospital, Auckland, New Zealand

⁵ Department of Respiratory and Sleep Medicine, Queensland Children's Hospital, Brisbane, QLD, Australia

⁶ Australian Centre for Health Services Innovation, Queensland University of Technology, Brisbane, QLD, Australia



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

a single study. Meta-analyses have been proven valuable in furthering understanding of the human microbiota, shining a light on microbiota evolution [4], technical drivers of variation in microbiota studies [5] and providing valuable insights into human disease in both the gut [6] and sinuses [7].

Although often regarded as the highest level of evidence in clinical research [8], the utility of meta-analyses becomes limited where the underlying data are complicated by clinical, technical and/or analytic heterogeneity among the primary studies [9, 10]. Meta-analysis guidelines have been established to address such limitations [11]; however, existing frameworks can be ill-suited to microbiomic datasets that have higher complexity. Standardised frameworks specifying minimum reporting criteria for microbiome studies have emerged (e.g. MlXS [12], STORMS [13] and STROBE metagenomics [14]); however, such frameworks are not yet widely or consistently applied [15], and their suitability for meta-analyses remains to be determined. Context-specific refinement of current frameworks is also needed for respiratory microbiome studies to address heterogeneity related to diagnostic inconsistencies and to account for the different methods used to sample the airways, each of which carries its own biases [16]. Given the complexity and ever-expanding nature of datasets emerging from omics-based studies, the emergence of new airway sampling methods (e.g. breath) [17] and the growing recognition of clinically relevant disease endotypes [18], there is a pressing need to define standardised reporting criteria for respiratory microbiome and multi-omic studies.

Having recently conducted a large-scale, 16S rRNA gene-based IPD meta-analysis focused on the paediatric respiratory microbiota that involved 2624 children from 20 studies [19], we identified several challenges around both data sharing and methodological heterogeneity within this research niche. Here, we examine limitations of respiratory microbiota datasets that emerged from our study [12] and propose strategies to achieve the higher inter-study standardisation needed to support future large-scale respiratory microbiome meta-analyses.

Data accessibility challenges

After determining a research question, the first challenge in performing any meta-analysis is obtaining access to usable data [20]. Access to published data varies greatly among studies. For our meta-analysis [19], just over half of the 20 included studies had publicly available sequence data, and, at the time of the initial literature search, only two contained sufficient metadata to enable meaningful linking with the sequence data. There was also variability in terms of the types of samples that had been uploaded to public repositories: some publicly uploaded

datasets included samples which were not explained or mentioned in the original article, potentially representing mock communities, sequencing controls, negative extraction controls, samples excluded from the original publication or samples from different studies that were uploaded under the same accession number. Key clinical metadata characteristics were frequently excluded, including the anatomical site from which the sample was taken, respiratory diagnosis of the individual providing the sample, and demographic variables such as age and sex, all of which are specified as minimum reporting criteria in multiple reporting frameworks [21, 22]. While contact with corresponding authors yielded access to the required metadata for all the 20 included studies, there is no guarantee that this would be the case for future investigations, with at least one potential study excluded due to a lack of author response. Similar challenges have been encountered in other contexts, including meta-analyses of the gut microbiome [20]. Heterogeneity in the uploading of sequence data to public repositories is also recognised as a systemic challenge in microbiomics that likely stems from the limited training of users who have variable levels of computational experience [15]. There remains a need for best-practice solutions and training to support uploading of sequence data and the associated metadata in consistent and useable formats, as highlighted by earlier standardisation efforts (e.g. MIGS) and exemplified by initiatives like the American National Microbiome Data Collaborative (<https://microbiome.data.org/>) [15, 21]. As there are currently no applicable metadata reporting criteria for respiratory contexts, we outline the key issues arising from our study, potential solutions and our recommendations in Table 1.

Overcoming barriers to accessing primary datasets

The recognised need for improved data-sharing processes has prompted development of frameworks that define current best practices. One recent example is the American National Institutes of Health's (NIH) Data Management and Sharing Policy that requires data handling according to FAIR (Findable, Accessible, Interoperable and Reusable) principles [23]. While such initiatives will undoubtedly improve data-sharing practices, current frameworks are not specifically designed to address the data access and heterogeneity issues affecting microbiomic meta-analyses. Additionally, time frames for data sharing are not specified beyond being made available "as soon as possible". Delays in accessing primary datasets, particularly those generated from human cohorts, are a widely acknowledged barrier to secondary analyses [20]. For our meta-analysis, it took ~18 months to access data from the 21 studies that were ultimately included (one of which was excluded following quality filtering of

Table 1 Issues and potential solutions to data cohesiveness in respiratory microbiota studies

Issue	Short-term solution	Long-term solution
Data accessibility	Consider deposition of data in original research proposals	Develop and support the use of guarded archives with established regulatory oversight. Develop patient communication plans that could be used when consenting individuals in order to inform permissions for future data use
Cohort heterogeneity	<p>Raw sequence data</p> <p>Clearly labelled controls</p> <p>Minimal metadata^a</p> <p>Clearly state diagnostic criteria for cases</p> <p>Where possible, report symptoms or tests used for decision-making</p>	<p>Determine how accurately current diagnostic criteria reflect pathobiological processes and determine if any can be combined from a microbiota perspective</p> <p>Further research is needed to determine key demographic variables that may confound meta-analyses</p>
Sample collection	<p>Report key demographic variables^b</p> <p>Try to recruit diverse cohorts</p> <p>Create a standardised (and granular) reporting metric for reporting recent antibiotic exposure</p> <p>Address knowledge gaps in lower airway sample literature and (where possible) collect paired upper- and lower-airway samples</p>	<p>Improve understanding of respiratory microbiota recovery from antibiotics</p> <p>Improve understanding as to how related the microbiotas of different respiratory anatomical sites are in different conditions or clinical contexts (e.g. during long-term antibiotic therapy)</p> <p>Determine consistency of different sampling approaches used for the same anatomical site (e.g. nasal swabs and nasal aspirates; sputum and induced sputum; BAL vs protected brushing)</p>
Sample processing	<p>Clearly state sample method used</p> <p>Include and report results of negative extraction controls and methods used to mitigate the presence of kitome among low bacterial density samples</p> <p>Laboratory protocols should be reported consistent with STORMS and STROBE-metagenomics requirements [11, 12]</p>	<p>Determine the effects of different extraction methods on perceived microbiota profiles for different anatomical sites</p>
Bioinformatics processing	<p>PCR bias in amplicon sequencing studies</p> <p>Clearly state software, version numbers, and where possible include analysis scripts as supplementary files when submitting primary data for publication</p>	<p>Determine the effects of different sequence regions on perceived microbiota profiles</p> <p>Make data available for secondary analyses which may use a new pipeline</p>

^a Minimal metadata, diagnosis, age, gender and anatomical site sampled. ^bKey demographics, age, gender and ethnicity

sequence data), while despite best efforts we were unable to access data for a further 10 studies that had met our inclusion criteria. There is broad consensus in the environmental microbiomic field that (sequence and meta)-data should be made publicly available as a condition of manuscript publication [24], thereby removing access barriers entirely; however, such an approach ignores the complex ethical considerations involved in sharing human datasets, particularly those derived from vulnerable or underserved populations [25, 26]. Two such groups are commonly included in paediatric respiratory microbiota studies: firstly, children are an inherently vulnerable population, whose original consent is typically obtained through the proxy of parents, and secondly, marginalised populations who may be overrepresented in respiratory microbiota studies that focus on high disease burdens among vulnerable populations [27–29]. An important example of this is First Nations populations, as emphasised in recent research examining Indigenous data governance frameworks [30]. Informed consent and data privacy are key concerns with respect to data access, as children may wish to remove their data when they reach maturity, a right to be forgotten which has already been acknowledged in other contexts [31, 32]. Reliance on ethics committees alone to determine whether secondary data use is reasonable risks ignoring important considerations related to the rights of First Nations and other vulnerable populations [30]; however, studies of vulnerable populations are essential to improving overall population health, particularly where early interventions may be needed to prevent paediatric conditions progressing into disease in adulthood [33, 34]. Indeed, the importance of diversity within medical research is now increasingly recognised and even demanded [30, 35]. Thus, solving these ethical challenges is imperative to furthering scientific understanding as these issues will only become more complex as datasets become increasingly comprehensive, potentially containing multiple types of microbial sequences (bacterial, fungal or archaeal) as well as human genome data, even if only present through a “bystander” effect in whole-genome shotgun sequencing.

Guarded archives offer a solution for balancing the ethical need to maintain participants’ agency in determining how their data are used against the scientific need for timely access to datasets that are commonly generated using public funds. Guarded archives (databases where sequence and metadata are uploaded but can only be accessed if a request is approved by a recognised data-oversight regulatory body) enable external verification that data presented in original research papers (a) exist and (b) are genuinely accessible to valid research requests (something not necessarily true when data are only made available “upon request” [36]). When paired

with data communication plans that are implemented at the time of sample collection, guarded archives provide an opportunity for meta-consent processes to be applied in which individuals can refuse participation in a secondary analysis or withdraw their data from the database at any time. Furthermore, guarded archives allow the validity of new research requests to be reviewed both in the context of the original consent and ongoing patient and community interest group consultation. The Atlas of Living Australia, The Indigenous Background Library and Aotearoa Variome (Genomics Aotearoa, 2019) represent some examples of smaller archives where appropriate Indigenous consultation has been undertaken during development to achieve a more bespoke and ethically sound repository than larger public repositories such as the Sequence Read Archive (SRA) that can be accessed without regulatory oversight. The use of smaller bespoke archives may also mean they are better equipped to deal with the unique challenges facing each individual study in comparison with the monolithic archives currently in common use. Funding and control of bespoke archives, however, bring new challenges as they should aim to be Nagoya Protocol compliant [37], particularly where the microbiota data are later used to develop novel treatments [38]. A limitation of this approach is that smaller organisations managing bespoke repositories may be less able to resist pressure from forensic and law enforcement agencies to release data, a problem already noted with human genome information [39]. It is worth noting that the NIH Data Management and Sharing Policy encourages consideration of controlled access repositories for “sensitive data” [23].

Heterogeneity in clinical metadata

Another substantial hurdle to IPD meta-analyses, or even “merely” comparing literature on the paediatric respiratory microbiota, is the considerable degree of heterogeneity in study methodology. The sources of this variability stretch from the bedside (where diagnoses, symptoms and treatments may be reported differently) through to sample collection and analytical processing with an overabundance of different laboratory protocols (Fig. 1). Benchmarking protocols for microbiome studies are emerging [40], but, as with reporting guidelines, these remain to be widely adopted. Resolving the variability in the published literature creates a dilemma for meta-analyses, whereby technical differences, which may contribute to inter-study differences, must either be ignored or standardised by data filtration. In our study, this resulted in substantial sample size reduction and therefore constrained our ability to generate findings (Fig. 1).

Assigning a paediatric respiratory diagnosis is a complex process which is influenced by many factors that

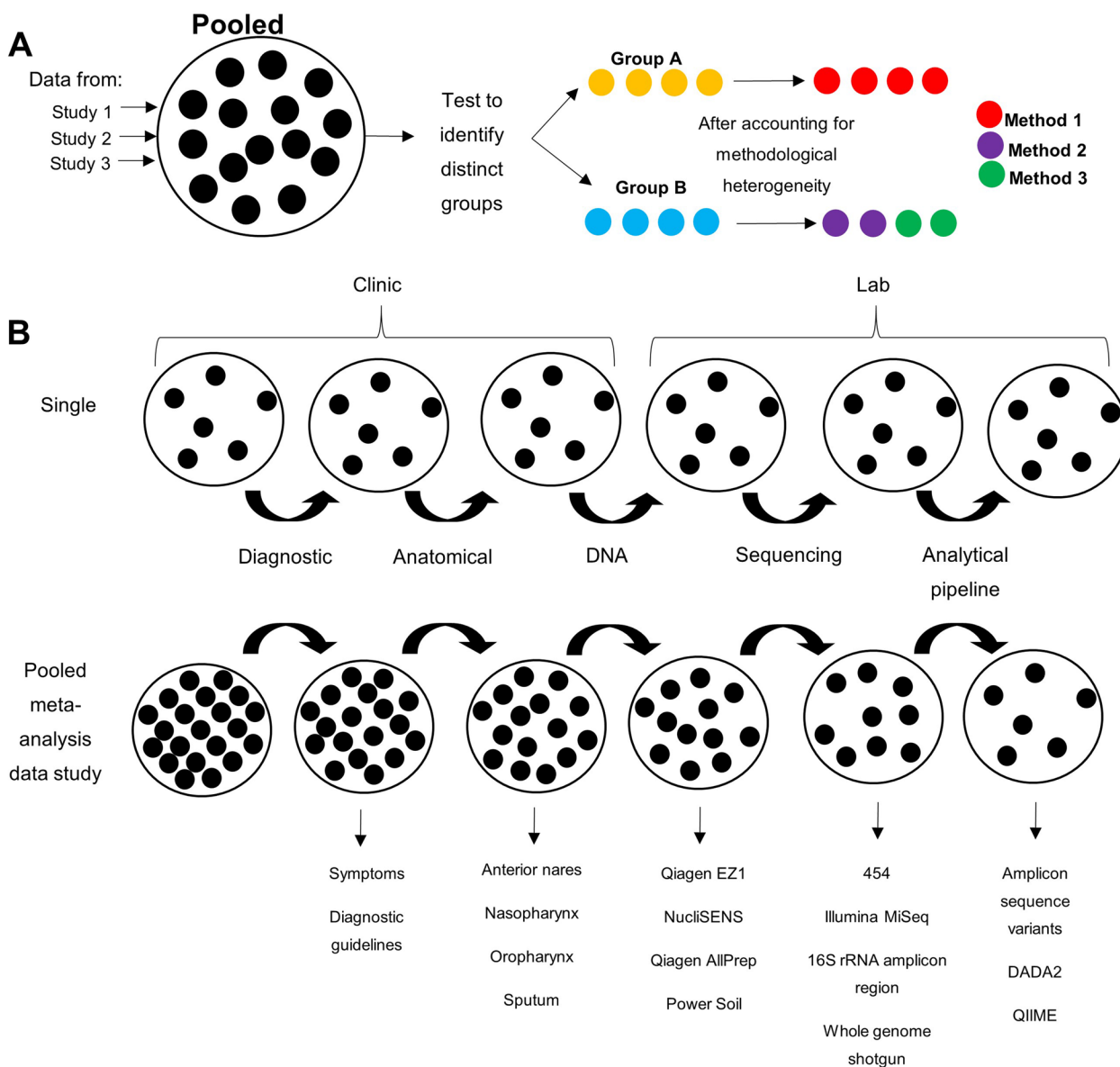


Fig. 1 Impacts of data heterogeneity in meta-analyses. A key strength of meta-analyses is the ability to take data from multiple studies to increase sample size. **A** *Methodological heterogeneity risks introducing confounding batch effects in pooled datasets.* In this figure, data from three studies have been pooled and analysed using a uniform pipeline that revealed two distinct clusters (group A and group B). However, after accounting for methodological heterogeneity, it becomes evident that all of the data in group A were generated using a single method that was distinct from those that generated the data in group B. This type of batch effect risks incorrect clinical interpretation of the analysis outputs. **B** *Selecting studies for inclusion in meta-analyses based on standardised methodologies can reduce batch effects outlined in A but risks loss of sample size.* In respiratory studies, heterogeneity emerges from a multitude of clinic and laboratory factors, from initial differences in diagnostic criteria through to variation in the pipeline used to analyse sequence data. Within a single study (example at top), a given method is applied consistently to all samples and therefore has no impact on sample size. By contrast, within IPD meta-analyses (example at bottom), correction for differences in the methods used at each successive stage progressively erodes the sample size. If the field contains a low level of methodological standardisation, then this reduction could potentially make the statistical power of a meta-analysis no more useful than that of a single study

include country-specific nuances, age of the child, local guidelines and level of expertise such as individual physicians considering factors differently when making clinical guidelines or diagnostic decisions [41, 42]. Even if all

diagnostic decisions were consistent, there are limitations as to how much traditional diagnoses may reflect underlying pathobiological processes, for example many diagnoses could contain multiple pathologies with shared

symptoms, while common pathologies could present differently in clinic [43]. A recent article highlighted this complexity, calling for the combination and elimination of several diagnoses to clarify this process, aiming to achieve better diagnostic accuracy [43]. Such clarity would also benefit meta-analyses, as diagnostic heterogeneity complicates which patient cohorts should be compared or combined; for instance, microbiota studies of infants have used both respiratory syncytial virus (RSV) infection [44] and bronchiolitis [45] as grouping factors. Despite bronchiolitis being most commonly caused by RSV infection [46], and the high probability that these studies may be discussing identical conditions, the cohorts cannot necessarily be conflated in meta-analyses without additional diagnostic information. Furthermore, while RSV infection in children aged under 2 years is usually clinically diagnosed as bronchiolitis, the same wheezing illness in an older child could be classified as viral-induced wheeze, a non-specific lower respiratory infection. Indeed, the PERCH study identified RSV as the most common cause of pneumonia [47]. Overcoming such challenges requires clearer symptom reporting in microbiomic studies, either at an individual level or by specifying clearly defined case criteria. To address these gaps, we recommend that the criteria in Table 1 be considered as minimum criteria for diagnostic reporting. An alternative approach to a specific clinical diagnosis which is gaining popularity is the concurrent use of endotypes (“distinct molecularly defined functional or pathobiological pathways that may be associated with distinct treatment responses” [48]) and/or phenotypes (clinical parameters of a patient’s condition) or a nuanced approach combining these. Endotypes have proven useful in asthma studies where inflammatory endotypes have been used to successfully identify specific anti-inflammatory therapies [49]. The benefit of using an endotype approach is also reflected in the bronchiectasis field where factors such as sputum production can have high prognostic predictive value [18]. The adoption of respiratory phenotypes and endotypes that are clearly defined and consistent across studies (both in the clinic and in research), rather than a clinical diagnosis per se, will likely be useful in future microbiome-based meta-analyses.

Accounting for antibiotic usage in respiratory microbiota studies

Respiratory microbiota studies often focus on populations with high antibiotic usage. With antibiotics likely to influence an individual’s microbiota profile [50], consideration of this factor when comparing studies is essential. The STORMS framework requires that authors report data “known about antibiotic usage before or during

sample collection” [13]; however, this definition may be too broad in the context of respiratory microbiota studies where different classes of drugs may be used for variable periods of time.

Our meta-analysis [19] identified two broad approaches to reporting antibiotic usage that are common among respiratory microbiota studies: either the time between a patient’s last antibiotic use and sampling is reported or, alternatively, no samples are taken from individuals with antibiotic exposure more recent than a specified cut-off date. The former can be challenging to accommodate in meta-analyses, as a standardised definition of “recent antibiotic use” is lacking. While it is generally believed that the gastrointestinal microbiota will recover in approximately 4 weeks after antibiotic treatment [51], it is unclear whether a comparable window exists for the respiratory microbiota and, if so, how long microbial community recovery might take. Current reporting approaches do provide a way to either standardise or examine the effects of antibiotics within an original study, though comparing across studies with different cut-off dates may mean that antibiotics rather than other biological differences are driving the microbiota differences between study populations. Reporting of patients’ health status at the time of sampling is also important, as the response of the respiratory microbiota to maintenance antibiotics during a period of relative stability and wellness may well be different to that observed during a severe exacerbation where the antibiotic treatment is more intense (e.g. with intravenous antibiotics) [52]. Finally, there remains a need for the effects of different antibiotics to be accounted for in respiratory microbiome analyses. Ideally, more granular data pertaining to the type(s) of antibiotics used, as well as their method of delivery (e.g. oral, inhaled or intravenous), coverage and duration, will be reported. Such data will be essential to achieving a more nuanced understanding of the effects of specific antibiotics than is currently possible when different antibiotic types and classes are analysed collectively under the umbrella of “recent antibiotics”.

Sample collection in respiratory microbiota studies

The method(s) used to collect respiratory samples is another point of variability among studies that has potential to confound inter-study comparisons and meta-analyses [16]. Despite the respiratory microbiota often being described as a single entity, there are distinct physicochemical niches present in, for instance, the proximal and distal lower airways (e.g. epithelial differences across the bronchi and lung parenchyma) and within the upper airways (oropharynx and nasopharynx) [50]. It remains unclear how similar or different microbial communities may be across the lower and upper airways, with

reports indicating both similar [45] and different [53] bacterial communities. It may well be that the connectedness of the communities, and therefore their overall similarity, is differentially altered by illness [54]. A paucity of detailed research on this topic makes it challenging to compare the findings of, say, a study investigating the sputum microbiota with those of a study investigating the nasopharyngeal microbiota, even if the study population has the same condition [55]. Furthermore, within a single anatomical site, there may be differences in the method used to collect the sample. For instance, the microbiota of nasopharyngeal aspirates differs somewhat from nasopharyngeal swabs within the same individual [56]. Similarly, the bronchial microbiota recovered from protected-specimen brushing may be different to that sampled by bronchoalveolar lavage (BAL) or sputum [54]. A nuanced approach to this issue is needed in respiratory microbiome studies, as determining from which anatomical sites samples should be collected, and how this should be performed, differs depending on clinical accessibility, patient tolerance and ethical considerations. While sampling needs to be tailored to the question at hand and the patient population being studied, standardised reporting frameworks would greatly facilitate inter-study comparisons.

Sample processing heterogeneity

In addition to the aforementioned lack of consistency in clinical diagnoses and sample collection, there is a panoply of ways in which respiratory microbiota studies can progress from a clinically collected sample to ultimately generating a microbiota profile. This includes differences in DNA extraction protocol and selection of sequencing technique (e.g. amplicon sequencing vs shotgun metagenomics). Most respiratory studies to date (and indeed all that were included in our meta-analysis) have employed amplicon sequencing of the 16S rRNA gene. Sources of heterogeneity among the 20 studies we analysed included the 16S rRNA gene region sequenced, PCR amplification and purification parameters, the sequencing platform used and bioinformatic analytical pipelines. While meta-analyses represent a powerful tool to address the variability in bioinformatic pipelines (by analysing all included data via a uniform pipeline), they cannot easily redress the methodological differences which generated the original sequence data. Realising a framework to address this issue in respiratory studies is critically needed as existing sources of heterogeneity will be compounded by the increasing use of new technologies generating additional omics-data types [7, 57] and the emergence of multi-omic applications. Lack of standardisation in laboratory methods used for microbiota studies is a long-standing problem well known in other areas of microbial ecology

research [58, 59], including the human gastrointestinal microbiome [60]. While the best protocol for any given study will always be determined by independent research groups, and there exists an argument that true biological signals should transcend differences in methodological approach, large-scale methodological reviews prompted by this heterogeneity in other areas have provided standardised frameworks on which future work to support meta-analyses could be based [60]. For the respiratory microbiota, there is a relative paucity of such studies, and while approaches used to investigate the gastrointestinal microbiota are often taken wholesale and transferred to the respiratory system, there is a lack of evidence that this is the best approach [61].

DNA extraction methods in respiratory microbiota studies

The DNA extraction method is possibly the factor that most influences observed microbiota profiles [40, 62]. Consideration of this methodology is important, particularly in respiratory studies given the variable bacterial density across the airways that is markedly lower than that observed for gut samples [63]. Failure to account for low bacterial density risks confounding due to reagent contamination or the “kitome”, as highlighted by placental microbiome studies [64]. DNA extraction methods can also result in substantial batch effects which may mask or imitate biological signals [65, 66]. For example, in a longitudinal study of nose microbiomes among infants, Salter and colleagues demonstrated that an apparent association between the composition of the nose microbiomes and the age of the infants was actually due to a DNA extraction batch effect [65]. The application of batch correlation tools, particularly in the limited examples for the respiratory microbiota where both case and control data are available [67], is also worthy of further consideration.

Sequence analysis in respiratory microbiota studies

Once DNA is obtained from a sample, many respiratory microbiota studies use the 16S rRNA gene amplicon approach to profile an individual's microbiota. The different regions of this gene which may be selected can lead to different PCR biases and may alter the accuracy of taxonomic assignment [68]. It is not yet clear which 16S rRNA gene region is best to use for respiratory microbiota studies. An analysis of a respiratory-specific database, eHOMD, indicated that the hypervariable V1–V3 regions may provide the greatest taxonomic resolution [69]; however, the length of this region (~490 bp) introduces new challenges related to the limited overlap of forward and reverse reads that occurs when $DNA > 460$ bp is

sequenced using most current technologies. As a result of this limitation, many studies instead target the shorter V3–V4 region (~460 bp) where sequencing is performed using the Illumina MiSeq platform. Sequencing of much longer amplicons, as achievable using PacBio technology, could circumvent this issue but may be cost prohibitive. Avoiding amplicon-based approaches altogether, such as in whole-genome shotgun sequencing [16], offers another alternative, although high amounts of contaminating host DNA can be problematic in analyses of respiratory specimens [70, 71], and it is unclear how or if such data could be fairly compared to the majority of current amplicon-based literature.

Bacterial quantification, diversity and contamination

Bacterial alpha diversity is often used as an indicator of overall respiratory health [34]; however, reporting alpha diversity without bacterial load information can result in errors due to sample contamination and limits our ability to fully interpret these data. The influence of contamination can be exacerbated in respiratory samples compared to, for example, gut samples, given the relatively low bacterial density in the former [63]. Statistical measures used to estimate alpha diversity may also be prone to biases increased by the common practice of rarefaction [72]. As such, direct quantification of total bacterial density (or, most commonly, 16S rRNA gene copies) and/or individual taxa may be more insightful [73], particularly if used in conjunction with alpha diversity metrics, as both overall bacterial community size and diversity can be considered [74]. While several studies use quantitative PCR (qPCR) to estimate bacterial density [53, 75], the myriad of methodologies and inherent variability present within qPCR [76] make inter-study comparisons of such data particularly challenging, especially given the potentially low bacterial density in paediatric respiratory samples [63]. Much like with amplicon-based taxonomic analysis, it is important that standardised reporting processes are encouraged (e.g. MIQE [77]), potentially incorporating new technologies such as droplet digital PCR, which has less inherent variability than standard real-time qPCR platforms [74, 78].

Concluding remarks

Ongoing research into the microbiota of the paediatric respiratory tract has already yielded promising findings, such as the role of the respiratory microbiota in respiratory tract development [79] and disease susceptibility [80]. This demonstrates the potential of this research avenue to improve understanding of respiratory disease in children. Despite this promise, there remains many more questions and avenues to explore in this complex

topic. By cutting through (some of) the noise, meta-analyses offer a promising way to provide new insights into this area while also addressing the issue of small sample sizes that often affects individual respiratory microbiome studies. Such analyses are particularly effective at identifying patterns and generating hypotheses for future investigation and “ground truthing”. However, variability in how studies are conducted and the ways that data are managed and reported can limit both the ability of meta-analyses to reach new conclusions as well as prevent the field from moving forward as a whole, rather than as several splintered research groups. As the microbiome field continues to move toward standardised methodologies [13, 14, 62, 81], these approaches may not necessarily meet the context-specific needs of respiratory microbiome studies, and as such, this work remains to be done. While some potential solutions to these long-standing and complex challenges are offered in Table 1, this manuscript alone cannot propose a complete solution and functions primarily as an invitation for an open dialogue where solutions can be built upon by the field as a whole.

Acknowledgements

Not applicable.

Authors' contributions

All authors were involved in developing the concepts and ideas for this article, and all authors contributed to the writing and editing of the manuscript. The authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 27 March 2022 Accepted: 20 February 2023

Published online: 22 March 2023

References

- Gonzalez A, Navas-Molina JA, Kosciolk T, McDonald D, Vázquez-Baeza Y, Ackermann G, DeReus J, Janssen S, Swafford AD, Orchanian SB, et al. Qiita: rapid, web-enabled microbiome meta-analysis. *Nat Methods*. 2018;15:796–8.
- Stuppel A, Singerman D, Celi LA. The reproducibility crisis in the age of digital medicine. *NPJ Digit Med*. 2019;2:1–3.
- Miyakawa T. No raw data, no science: another possible source of the reproducibility crisis. *Mol Brain*. 2020;13:1–6.

4. Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nat Rev Microbiol*. 2008;6:776–88.
5. Debelius J, Song SJ, Vazquez-Baeza Y, Xu ZZ, Gonzalez A, Knight R. Tiny microbes, enormous impacts: what matters in gut microbiome studies? *Genome Biol*. 2016;17:217.
6. Duvallet C, Gibbons SM, Gurry T, Irizarry RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun*. 2017;8:1784.
7. Wagner Mackenzie B, Waite DW, Hoggard M, Douglas RG, Taylor MW, Biswas K. Bacterial community collapse: a meta-analysis of the sinonasal microbiota in chronic rhinosinusitis. *Environ Microbiol*. 2017;19:381–92.
8. Lee YH. An overview of meta-analysis for clinicians. *Korean J Intern Med*. 2018;33:277.
9. Gurevitch J, Koricheva J, Nakagawa S, Stewart G. Meta-analysis and the science of research synthesis. *Nature*. 2018;555:175–82.
10. Imrey PB. Limitations of meta-analyses of studies with high heterogeneity. *JAMA Netw*. 2020;3:e1919325.
11. Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F, Niu Y, Du L. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. *J Evid-Based Med*. 2015;8:2–10.
12. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, et al. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nat Biotechnol*. 2011;29:415–20.
13. Mirzayi C, Renson A, Zohra F, Elsafoury S, Geistlinger L, Kasselmann LJ, Eckenrode K, van de Wijgert J, Loughman A, Marques FZ, et al. Reporting guidelines for human microbiome research: the STORMS checklist. *Nat Med*. 2021;27:1885–92.
14. Bharucha T, Oeser C, Balloux F, Brown JR, Carbo EC, Charlett A, Chiu CY, Claas EC, de Goffau MC, de Vries JJ, et al. STROBE-metagenomics: a STROBE extension statement to guide the reporting of metagenomics studies. *Lancet Infect Dis*. 2020;20:e251–60.
15. Vangay P, Burgin J, Johnston A, Beck KL, Berrios DC, Blumberg K, Canon S, Chain P, Chandonia J, Christianson D, et al. Microbiome metadata standards: report of the National Microbiome Data Collaborative's Workshop and follow-on activities. *mSystems*. 2021;6:1194.
16. Carney SM, Clemente JC, Cox MJ, Dickson RP, Huang YJ, Kitsios GD, Kloefer KM, Leung JM, LeVan TD, Molyneux PL, et al. Methods in lung microbiome research. *Am J Respir Cell Mol Biol*. 2020;62:283–99.
17. Bos LD, Sterk PJ, Fowler SJ. Breathomics in the setting of asthma and chronic obstructive pulmonary disease. *J Allergy Clin Immunol*. 2016;138:970–6.
18. Flume PA, Chalmers JD, Olivier KN. Advances in bronchiectasis: endotyping, genetics, microbiome, and disease heterogeneity. *Lancet*. 2018;392:880–90.
19. Broderick DT, Waite DW, Marsh RL, Camargo CA Jr, Cardenas P, Chang AB, Cookson WO, Cuthbertson L, Dai W, Everard ML, et al. Bacterial signatures of paediatric respiratory disease: an individual participant data meta-analysis. *Front Microbiol*. 2022;12:711134.
20. Schloss PD. Identifying and overcoming threats to reproducibility, replicability, robustness, and generalizability in microbiome research. *MBio*. 2018;9:525.
21. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol*. 2008;26:541–7.
22. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg*. 2014;12:1495–9.
23. The National Institutes of Health. Data management and sharing policy. <https://sharing.nih.gov/data-management-and-sharing-policy>. Accessed 19 Feb 2023.
24. Langille MG, Ravel J, Fricke WF. "Available upon request": not good enough for microbiome data! *Microbiome*. 2018;6:1–2.
25. Chuong KH, Hwang DM, Tullis DE, Waters VJ, Yau YC, Guttman DS, O'Doherty KC. Navigating social and ethical challenges of biobanking for human microbiome research. *BMC Med Ethics*. 2017;18:1–10.
26. Kling S. Ethics and the microbiome. *Curr Allergy Clin Immunol*. 2019;32:101–3.
27. Kovetski T. Respiratory disease in Canadian First Nations and Inuit children. *Paediatr Child Health*. 2012;17:376–80.
28. O'Grady KF, Hall KK, Bell A, Chang AB, Potter C. Review of respiratory disease among Aboriginal and Torres Strait Islander children. *Aust Indig Health Bull*. 2018;18:2.
29. Twiss J, Metcalfe R, Edwards E, Byrnes C. New Zealand national incidence of bronchiectasis "too high" for a developed country. *Arch Dis Child*. 2005;90:737–40.
30. Hudson M, Nanibaa'A G, Sterling R, Caron NR, Fox K, Yracheta J, Anderson J, Wilcox P, Arbour L, Brown A, et al. Rights, interests and expectations: indigenous perspectives on unrestricted access to genomic data. *Nat Rev Genet*. 2020;21:377–84.
31. Kaye J. The tension between data sharing and the protection of privacy in genomics research. *Ann Rev Genomics Hum Genet*. 2012;13:415–31.
32. Kim H, Kim SY, Joly Y. South Korea: in the midst of a privacy reform centered on data sharing. *Hum Genet*. 2018;137:627–35.
33. Cox MJ, Ege MJ, von Mutius E. The lung microbiome: European Respiratory Society. 2019.
34. Pillariseti N, Broderick D, Ainsworth A, Mulholland A, Wagner Mackenzie B, Middleton D, Byrnes CA, Taylor MW. The airway microbiota in children newly diagnosed with bronchiectasis largely retains its diversity. *Eur Respir J*. 2019;54:1900704.
35. Dresser R. Wanted single, white male for medical research. *Hastings Cent Rep*. 1992;22:24–9.
36. Krawczyk M, Reuben E. (Un) available upon request: field experiment on researchers' willingness to share supplementary materials. *Account Res*. 2012;19:175–86.
37. Ambler J, Diallo AA, Dearden PK, Wilcox P, Hudson M, Tiffin N. Including digital sequence data in the Nagoya Protocol can promote data sharing. *Trends Biotechnol*. 2020;39:116–25.
38. Yamanishi S, Pawankar R. Current advances on the microbiome and role of probiotics in upper airways disease. *Curr Opin Allergy Clin Immunol*. 2020;20:30–5.
39. Wolf LE, Hammack CM, Brown EF, Brelsford KM, Beskow LM. Protecting participants in genomic research: understanding the "Web of Protections" afforded by federal and state law. *J Law Med Ethics*. 2020;48:126–41.
40. Hasrat R, Kool J, de SteenhuijsenPeters WAA, Chu MLJ, Kuiling S, Groot JA, Van Logchem EM, Fuentes S, Franz E, Bogaert D, et al. Benchmarking laboratory processes to characterise low-biomass respiratory microbiota. *Sci Rep*. 2021;11:17148.
41. Liu LL, Gallaheer MM, Davis RL, Rutter CM, Lewis TC, Marcuse EK. Use of a respiratory clinical score among different providers. *Pediatr Pulmonol*. 2004;37:243–8.
42. de São José BP, Camargos P, Bateman ED, Botelho C, de Seixas Maciel J, Mancuzo EV, de Amorim Corrêa R. Primary care physicians' ability to diagnose the most prevalent respiratory diseases. *Int J Tuberc Lung Dis*. 2016;20:1392–8.
43. Douros K, Everard ML. Time to say goodbye to bronchiolitis, viral wheeze, reactive airways disease, wheeze bronchitis and all that. *Front Pediatr*. 2020;8:218.
44. de SteenhuijsenPeters WAA, Heinonen S, Hasrat R, Bunsow E, Smith B, Suarez-Arrabal M, Chaussabel D, Cohen DM, Sanders EAM, Ramilo O, Bogaert D, Mejias A. Nasopharyngeal microbiota, host transcriptome, and disease severity in children with respiratory syncytial virus infection. *Am J Respir Crit Care Med*. 2016;194:1104–15.
45. Luna PN, Hasegawa K, Ajami NJ, Espinola JA, Henke DM, Petrosino JF, Piedra PA, Sullivan AF, Camargo CA, Shaw CA, et al. The association between anterior nares and nasopharyngeal microbiota in infants hospitalized for bronchiolitis. *Microbiome*. 2018;6:2.
46. Hasegawa K, Mansbach JM, Teach SJ, Fisher ES, Hershey D, Koh JY, Clark S, Piedra PA, Sullivan AF, Camargo CA Jr. Multicenter study of viral etiology and relapse in hospitalized children with bronchiolitis. *Pediatr Infect Dis J*. 2014;33:809.
47. O'Brien KL, Baggett HC, Brooks WA, Feikin DR, Hammit LL, Higdon MM, Howie SR, Knoll MD, Kotloff KL, Levine OS, et al. Causes of severe pneumonia requiring hospital admission in children without HIV infection from Africa and Asia: the PERCH multi-country case-control study. *Lancet*. 2019;394:757–79.

48. Reddel HK, de Verdier MG, Agustí A, Anderson G, Beasley R, Bel EH, Janson C, Make B, Martin RJ, Pavord I, et al. Prospective observational study in patients with obstructive lung disease: NOVELTY design. *ERJ Open Res.* 2019;5:36.
49. Svenningsen S, Nair P. Asthma endotypes and an overview of targeted therapy for asthma. *Front Med.* 2017;4:158.
50. Man WH, de SteenhuijsenPeters WAA, Bogaert D. The microbiota of the respiratory tract: gatekeeper to respiratory health. *Nat Rev Microbiol.* 2017;15:259.
51. Lange K, Buerger M, Stallmach A, Bruns T. Effects of antibiotics on gut microbiota. *Dig Dis.* 2016;34:260–8.
52. Jubinville E, Veillette M, Milot J, Maltais F, Comeau AM, Levesque RC, Duchaine C. Exacerbation induces a microbiota shift in sputa of COPD patients. *PLoS One.* 2018;13:e0194355.
53. Marsh RL, Kaestli M, Chang AB, Binks MJ, Pope CE, Hoffman LR, Smith-Vaughan H. The microbiota in bronchoalveolar lavage from young children with chronic lung disease includes taxa present in both the oropharynx and nasopharynx. *Microbiome.* 2016;4:37.
54. Dickson RP, Erb-Downward JR, Freeman CM, McCloskey L, Beck JM, Huffnagle GB, Curtis JL. Spatial variation in the healthy human lung microbiome and the adapted island model of lung biogeography. *Ann Am Thorac Soc.* 2015;12:821–30.
55. Lu J, Carmody LA, Opron K, Simon RH, Kalikin LM, Caverly LJ, LiPuma JJ. Parallel analysis of cystic fibrosis sputum and saliva reveals overlapping communities and an opportunity for sample decontamination. *mSystems.* 2020;5:e00296–20.
56. Perez-Losada M, Crandall KA, Freishtat RJ. Two sampling methods yield distinct microbial signatures in the nasopharynges of asthmatic children. *Microbiome.* 2016;4:25.
57. Gurevitch J, Hedges LV. Statistical issues in ecological meta-analyses. *Ecology.* 1999;80:1142–9.
58. Anonymous: raising standards in microbiome research. *Nat Microbiol.* 2016;1:16112. <https://doi.org/10.1038/nmicriol.2016.112>.
59. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, et al. A communal catalogue reveals Earth's multiscale microbial diversity. *Nature.* 2017;551:457–63.
60. Sinha R, Abu-Ali G, Vogtmann E, Fodor AA, Ren B, Amir A, Schwager E, Crabtree J, Ma S, Abnet CC, et al. Assessment of variation in microbial community amplicon sequencing by the Microbiome Quality Control (MBQC) project consortium. *Nat Biotechnol.* 2017;35:1077.
61. Chang D, Dela Cruz CS, Sharma L. Challenges in understanding lung microbiome: it is NOT like the gut microbiome. *Respirology.* 2019;25:244–5.
62. Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Truitt M, Driessen M, Hercog R, Jung F, et al. Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol.* 2017;35:1069–76.
63. Marsh R, Nelson MT, Pope CE, Leach AJ, Hoffman LR, Chang AB, Smith-Vaughan HC. How low can we go? The implications of low bacterial load in respiratory microbiota studies. *Pneumonia.* 2018;10:7.
64. Sterpu I, Fransson E, Hugerth LW, Du J, Pereira M, Cheng L, Radu SA, Calderón-Pérez L, Zha Y, Angelidou P, et al. No evidence for a placental microbiome in human pregnancies at term. *Obstet Gynecol.* 2021;224:296.
65. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol.* 2014;12:87.
66. Hornung BV, Zwittink RD, Kuijper EJ. Issues and current standards of controls in microbiome research. *FEMS Microbiol Ecol.* 2019;95:fiz045.
67. Gibbons SM, Duvallet C, Alm EJ. Correcting for batch effects in case-control microbiome studies. *PLoS Comput Biol.* 2018;14:e1006102.
68. Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, Fabani MM, Seguritan V, Green J, Pride DT, et al. Library preparation methodology can influence genomic and functional predictions in human microbiome research. *Proc Natl Acad Sci USA.* 2015;112:14024–9.
69. Escapa IF, Chen T, Huang Y, Gajare P, Dewhirst FE, Lemon KP. New insights into human nostril microbiome from the expanded human oral microbiome database (eHOMD): a resource for the microbiome of the human aerodigestive tract. *mSystems.* 2018;3:e00187–18.
70. Marsh RL, Aho C, Beissbarth J, Bialasiewicz S, Binks M, Cervin A, Kirkham LS, Lemon KP, Slack MP, Smith-Vaughan HC. Panel 4: Recent advances in understanding the natural history of the otitis media microbiome and its response to environmental pressures. *Int J Pediatr Otorhinolaryngol.* 2020;130:109836.
71. Mackenzie BW, Waite DW, Biswas K, Douglas RG, Taylor MW. Assessment of microbial DNA enrichment techniques from sino-nasal swab samples for metagenomics. *Rhinology Online.* 2018;1:160–93.
72. Willis AD. Rarefaction, alpha diversity, and statistics. *Front Microbiol.* 2019;10:2407.
73. Reese AT, Dunn RR. Drivers of microbiome biodiversity: a review of general rules, feces, and ignorance. *mBio.* 2018;9:e01294–18.
74. Galazzo G, van Best N, Benedikter BJ, Janssen K, Bervoets L, Driessen C, Oomen M, Lucchesi M, Van Eijck PH, Becker HE, et al. How to count our microbes? The effect of different quantitative microbiome profiling approaches. *Front Cell Infect Microbiol.* 2020;10:403.
75. Pettigrew MM, Gent JF, Kong Y, Wade M, Ganseboom S, Bramley AM, Jain S, Arnold SLR, McCullers JA. Association of sputum microbiota profiles with severity of community-acquired pneumonia in children. *BMC Infect Dis.* 2016;16:317.
76. Smith CJ, Osborn AM. Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. *FEMS Microbiol Ecol.* 2009;67:6–20.
77. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem.* 2009;55:611–22.
78. Sillanpää S, Kramna L, Oikarinen S, Sipilä M, Rautiainen M, Aittoniemi J, Laranne J, Hyöty H, Cinek O. Next-generation sequencing combined with specific PCR assays to determine the bacterial 16S rRNA gene profiles of middle ear fluid collected from children with acute otitis media. *mSphere.* 2017;2:e00006–17.
79. Yun Y, Srinivas G, Kuenzel S, Linnenbrink M, Alnahs S, Bruce KD, Steinhoff U, Baines JF, Schaible UE. Environmentally determined differences in the murine lung microbiota and their relation to alveolar architecture. *PLoS One.* 2014;9:e113466.
80. Biesbroek G, Tsivtsivadze E, Sanders EA, Montijn R, Veenhoven RH, Keijser BJ, Bogaert D. Early respiratory microbiota composition determines bacterial succession patterns and respiratory health in children. *Am J Respir Crit Care Med.* 2014;190:1283–92.
81. Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, Gevers D, Petrosino JF, Abubucker S, Badger JH, et al. A framework for human microbiome research. *Nature.* 2012;486:215.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

