*ERJ Express*. Published on November 28, 2018 as doi: 10.1183/13993003.01711-2018

EUROPEAN RESPIRATORY *journal*

FLAGSHIP SCIENTIFIC JOURNAL OF ERS

Early View

ERJ Methods

# Characterising the respiratory microbiome

Rebecca L. Watson, Emma M. de Koff, Debby Bogaert

Please cite this article as: Watson RL, de Koff EM, Bogaert D. Characterising the respiratory microbiome. *Eur Respir J* 2018; in press (https://doi.org/10.1183/13993003.01711-2018).

# Characterising the respiratory microbiome

**Authors: Rebecca L. Watson[*1], Emma M. de Koff[*2,3], Debby Bogaert[1,2]**

**Affiliations:**

**1: Center for Inflammation Research, Queens Medical Research Institute, University of Edinburgh, Edinburgh, United Kingdom**

**2: Department of Pediatrics, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, the Netherlands**

**3: Spaarne Academy, Spaarne Gasthuis, Hoofddorp, The Netherlands**

*** These authors contributed equally**

**Corresponding author:** Prof. Debby Bogaert, University of Edinburgh, 47 Little France Crescent, Edinburgh, EH16 4TJ, United Kingdom
d.bogaert@ed.ac.uk

Studying the respiratory microbiome provides critical novel insight into respiratory disease pathogenesis which may improve clinical management, and we should strive to standardise study design, laboratory procedures and statistical methods in the field.

Over the last decade, researchers have begun to unravel the causes and consequences of variation within the respiratory microbiota, developing a more profound understanding of its role in the pathogenesis of pulmonary disease to improve clinical management. Developments in culture-independent identification of bacterial species have provided faster and more cost effective methods to characterise niche-specifc microbial ecosystems. Historically, the gut has been the niche of focus for human microbiome research, but recent studies have revealed an unexpected diversity of bacteria in both the upper and lower airways, linking community composition to a number of respiratory diseases, including cystic fibrosis (CF), chronic obstructive pulmonary disease (COPD) and acute infections [1]. Monitoring temporal changes in community composition of the respiratory microbiota can reveal the influence of host and environmental drivers on ecosystem behaviour, as well as the consequences of infection susceptibility or severity, and treatment effects. Here we outline current best practices and upcoming developments for respiratory microbiome research and potential clinical applications.

**Study Design**

So far, in respiratory microbiome study design, we have learned that crucial elements in generating valid, useful results include clear research questions, power calculations, enrolment of sufficient numbers of subjects and controls, robust sampling and exhaustive patient information collection. Although this applies to any well-designed population-based or clinical study, we also need to carefully consider possible confounding effects of a broad range of environmental and host characteristics on microbiome composition [1, 2]. The first pioneering cross-sectional studies linked altered microbial community structure and composition to disease state [3, 4], but longitudinal sampling is needed to fully understand the causes and long-term clinical outcomes of variation in respiratory microbiota. For example, recent well-characterised healthy birth cohorts have shown the dynamics of nasopharyngeal microbiota development in relation to lifestyle factors [5, 6], and have revealed marked shifts in microbial community composition associated with acute respiratory infections [7, 8]. Intensive follow-up of CF [9, 10] and COPD [11] patients demonstrated changes in the airway microbiome composition preceding symptom onset, suggesting that dysbiosis coupled to a dysregulated host immune response could be at the basis of disease progression [12]. Support for the potential role for the respiratory microbiome in early disease pathogenesis is evident in early childhood, as microbial communities with fewer commensals and more potential pathogens are associated with consecutive wheeze and asthma [8]. So far, every study of respiratory microbiota in relation to any lung disease, has

revealed clear aberrations of microbial community composition from the healthy state, redefining commonly accepted pathophysiological concepts in respiratory disease pathogenesis [12].

## Sample Collection

With respect to anatomy and site of sampling, the respiratory tract is not a single uniform system, but consists of interconnected niches harbouring distinct microbial communities that depend highly on local microenvironmental conditions. Therefore, when designing a new microbiome study, the appropriate sampling niche will largely depend on research question, hypothesis and target population. Key procedural practicalities also require consideration; for example, sampling the lower respiratory tract (LRT) requires invasive bronchoscopic procedures, limiting sample size, age-groups to be studied, and frequency of repeated sampling. To overcome this lack of access to the lungs, many studies use the easily accessible upper respiratory tract (URT) which is considered the likely source community of the lungs as well as a reservoir for most respiratory pathogens [12, 13]. In healthy adults, microbial colonisation of the LRT is assumed to originate from micro-aspiration of the oropharyngeal 'flora', and hence, the oropharynx can be used, albeit imperfect, as a proxy for the lungs. In children, however, both the nasopharynx and oropharynx are likely sources of microbial seeding to the LRT, probably resulting from anatomical differences, nasal breathing, and higher production of nasal secretions by children [14, 15], further limiting result extrapolation. In chronic lung diseases such as CF and COPD, the URT and LRT communities appear to become segregated with increasing disease duration. This is probably due to chronic inflammation, failure of lung clearance mechanisms, and repeated antimicrobial treatment resulting in localised selection and evolution of independent communities, the latter rendering LRT sampling from multiple sites mandatory to obtain meaningful results [16, 17].

## Sample Processing

An important aspect to consider throughout the design and execution of a respiratory microbiome study is the risk of and control for contamination. The respiratory tract harbours low-density bacterial communities, with microbial densities dropping along the way from the URT to the LRT [14, 18]. As a result, environmental DNA introduction during sample collection and processing becomes a likely threat, and can entirely overrule the true

microbial signal [18]. Sampling of the LRT particularly carries a high risk of microbial carryover from the URT, and so accurate sampling should be undertaken by well-trained and consistent personnel to reduce the risk of contamination. During transportation, samples should be kept cooled in appropriate storage media, and then processed and stored at -80 °C as soon as possible to prevent selective bacterial outgrowth. Additionally, contamination from the laboratory environment and the reagents used for sample processing can significantly influence results from low-biomass microbial communities [19]. Implementing proper 'negative' controls for all sampling, storage and laboratory procedures allows for later comparison and identification of potentially confounding environmental signals (for more details see [20]). Variations in methodology and batches can also affect results, highlighting the importance of clean working during DNA extraction and using fully optimised methods for the specific sample type. In addition to contamination, the extraction method can also affect the quality of the data and care should be taken to use methods which do not bias the bacteria extracted from the samples [18]. Including 'positive' controls in the form of mock communities, will allow for adequate control and comparison between sequencing runs, laboratories and institutes [13].

**Sequencing Platforms**

Regarding sequencing platforms, amplicon sequencing is currently the most commonly used method for determining the microbial community composition and targets the bacterial 16S ribosomal RNA (rRNA) gene, containing highly conserved as well as hypervariable regions. This targeted approach has revealed a wealth of information regarding community compostion and dynamics. However, the taxonomic resolution provided by 16S rRNA sequencing is limited due to the short target region length, complicating accurate species- and strain-level identification. In comparison, metagenomic sequencing captures the entire microbial genomic content, including bacteria, viruses and eukaryotes, and allows for microbial characterisation at the deepest taxonomic levels as well as functional potential profiling. However, applying this technique to low-biomass respiratory samples is challenging, as genome assembly requires high numbers of sequencing reads per sample, which makes detection of low-abundant species difficult, and increases the risk of contamination [21].

## Data Handling

Once data is generated, the bioinformatics and statistical methods required to analyse the large amounts of raw DNA reads generated by sequencing can be daunting. Initially, raw reads are filtered to remove sequencing errors and are assembled into complete sequences, after which the sequences are grouped based on similarity and assigned taxonomic names to reveal their identities. Several bioinformatics pipelines are freely available for data pre-processing, including Qiime [22] and mothur [23]. Each resulting microbial profile shows the abundance of individual species relative to the entire bacterial population within a sample, and contains many zero abundances, demanding nonparametric statistical methods developed specifically for handling microbiome data [24]. Characterising microbial development over time requires multiple measurements of the same individual, further complicating data analysis, but several approaches have been proposed to correct for repeated measures [25, 26]. The increasing application of machine-learning techniques that perform predictive modelling of clinical outcomes from microbial profiles combined with host and environmental characteristics, is a promising development [27]. However, the study of temporal microbiome dynamics, especially while accounting for confounding factors, remains in its infancy [24].

## Clinical Applications

In the era of the 100,000 Genome Project and the launch of the NHS Genomic Medicine Service, it is clear that sequencing techniques are not only more accessible but are also becoming more integral to the clinical environment. In the clinic, identification by culture still dominates pathogen detection, and although quantitative methods such as qPCR are increasingly available, applications of sequencing technologies are lacking. Cost effectiveness and efficiency of sample and data processing are currently being improved to enable clinical implementation of sequencing methods. Single-use sequencing applications are being developed, as are faster methods of DNA extraction and library preparation [28]. Technological and bioinformatic advances are in the pipeline to improve detection of subtle strain-specific variation within the target region [24]. For applying sequencing at the point of care, the portable, low cost, real-time DNA sequencer Oxford Nanopore MinION has real potential with its ability to rapidly sequence the bacterial 16S gene, even up to strain-specific resolution [29]. The emergence of real-time sequencing technologies could dramatically influence diagnostic methods through accurate species identification and quantification within a clinically relevant time frame.

**Research Priorities**

To move closer towards clinical applications, comparative and meta-analyses must combine results from different cohorts to define actionable thresholds of microbial abundance. Current methodological heterogeneity restricts comparability across institutes, and so by underlining essential aspects of study design including consistent sample collection and processing, adequate contamination controls, and longitudinal sampling (summarised in the Figure and Box 1), we hope to encourage reaching a consensus on solid, robust methodology for respiratory microbiota research. Our increased understanding of respiratory disease pathogenesis will contribute to reshaping clinical diagnostic, preventative and therapeutic strategies. Important challenges remain to integrate the advances within microbiota research into everyday medical practice, and future efforts should prioritise standardisation of protocols and analysis, adaptation of technology for application in the field including remote settings, and collaboration across countries and disciplines (Box 2). However, current progress in respiratory microbiota research certainly provides a promising platform for the clinical application of culture-independent techniques in the future.

Box 1 | **Essentials for respiratory microbiome studies:**

- Longitudinal study design
- Appropriate power calculations
- Consistent sampling
- Appropriate niche (proxy)
- Minimise contamination at all stages
- Contamination controls at all stages
- Robust quality checks
- Consistent bioinformatic processing

Box 2 | **Research priorities for future studies:**

- International platforms for communication
- Uniform sampling and transport protocols
- Standardised controls across laboratories
- Agreement on handling complex data
- Adapt technology for remote settings
- Collaboration between research disciplines (clinics, microbiology, molecular biology, ecology, bioinformatics)

**REFERENCES**

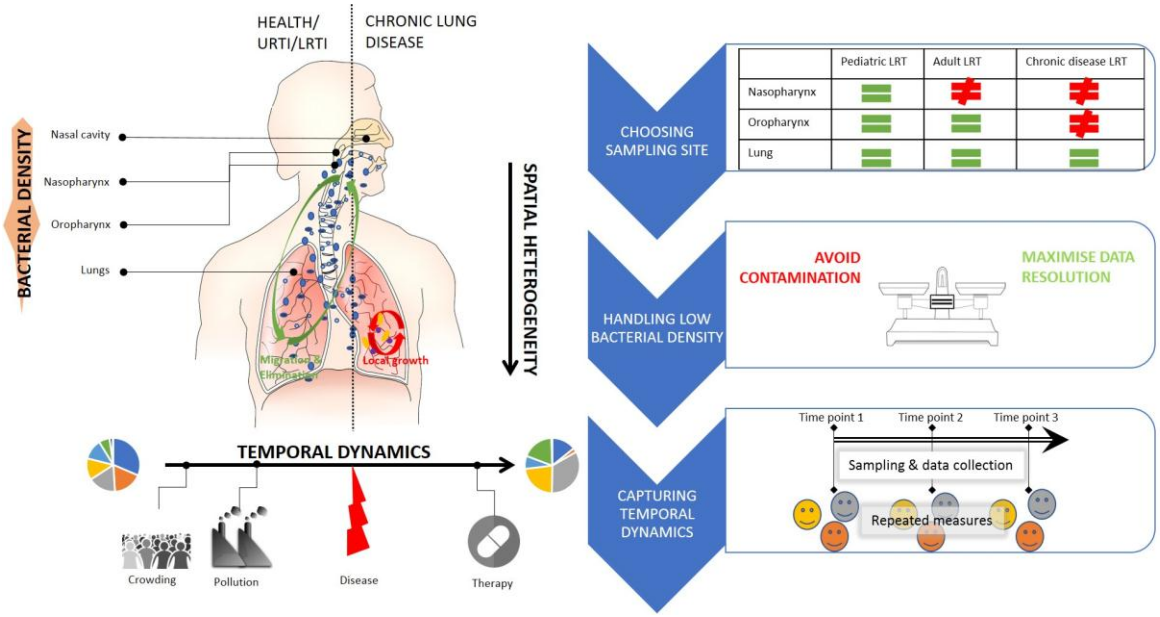

1. Hakansson AP, Orihuela CJ, Bogaert D. Bacterial-host interactions: Physiology and pathophysiology of respiratory infection. Physiol Rev. 2018;98(2):781-811.
2. Mattiello F, Verbist B, Faust K, Raes J, Shannon WD, Bijnens L, Thas O. A web application for sample size and power calculation in case-control microbiome studies. Bioinformatics. 2016;32(13):2038-40.
3. de Steenhuijsen Piters W, Huijskens E, Wyllie A, Biesbroek G, van den Bergh M, Veenhoven R, Wang X, Trzciński K, Bonten M, Rossen J, Sanders E, Bogaert D. Dysbiosis of upper respiratory tract microbiota in elderly pneumonia patients. ISME J. 2016;10(1):97-108.
4. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, Davies J, Ervine A, Poulter L, Pachter L, Moffatt M, Cookson W. Disordered microbial communities in asthmatic airways. PLoS One. 2010;5(1):e8578.
5. Biesbroek G, Bosch A, Wang X, Keijser B, Veenhoven R, Sanders E, Bogaert D. The impact of breastfeeding on nasopharyngeal microbial communities in infants. Am J Respir Crit Care Med. 2014;190(3):298-308.
6. Bosch AA, Levin E, van Houten MA, Hasrat R, Kalkman G, Biesbroek G, de Steenhuijsen Piters WA, de Groot PC, Pernet P, Keijser BJ, Sanders EA, Bogaert D. Development of upper respiratory tract microbiota in infancy is affected by mode of delivery. EBioMedicine. 2016.
7. Bosch A, de Steenhuijsen Piters WAA, van Houten MA, Chu M, Biesbroek G, Kool J, Pernet P, de Groot PCM, Eijkemans MJC, Keijser BJF, Sanders EAM, Bogaert D. Maturation of the infant respiratory microbiota, environmental drivers, and health consequences. A prospective cohort study. Am J Respir Crit Care Med. 2017;196(12):1582-90.
8. Teo S, Mok D, Pham K, Kusel M, Serralha M, Troy N, Holt B, Hales B, Walker M, Hollams E, Bochkov Y, Grindle K, Johnston S, Gern J, Sly P, Holt P, Holt K, Inouye M. The infant nasopharyngeal microbiome impacts severity of lower respiratory infection and risk of asthma development. Cell Host Microbe. 2015;17(5):704-15.
9. Carmody LA, Zhao J, Kalikin LM, LeBar W, Simon RH, Venkataraman A, Schmidt TM, Abdo Z, Schloss PD, LiPuma JJ. The daily dynamics of cystic fibrosis airway microbiota during clinical stability and at exacerbation. Microbiome. 2015;3:12.
10. Prevaes SM, de Winter-de Groot KM, Janssens HM, de Steenhuijsen Piters WA, Tramper-Stranders GA, Wyllie AL, Hasrat R, Tiddens HA, van Westreenen M, van der Ent CK, Sanders EA, Bogaert D. Development of the nasopharyngeal microbiota in infants with cystic fibrosis. Am J Respir Crit Care Med. 2016;193(5):504-15.
11. Huang YJ, Sethi S, Murphy T, Nariya S, Boushey HA, Lynch SV. Airway microbiome dynamics in exacerbations of chronic obstructive pulmonary disease. J Clin Microbiol. 2014;52(8):2813-23.
12. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. The microbiome and the respiratory tract. Annu Rev Physiol. 2016;78:481-504.
13. Faner R, Sibila O, Agusti A, Bernasconi E, Chalmers JD, Huffnagle GB, Manichanh C, Molyneaux PL, Paredes R, Perez Brocal V, Ponomarenko J, Sethi S, Dorca J, Monso E. The microbiome in respiratory medicine: Current challenges and future perspectives. Eur Respir J. 2017;49(4).
14. Charlson E, Bittinger K, Haas A, Fitzgerald A, Frank I, Yadav A, Bushman F, Collman R. Topographical continuity of bacterial populations in the healthy human respiratory tract. Am J Respir Crit Care Med. 2011;184(8):957-63.
15. Marsh RL, Kaestli M, Chang AB, Binks MJ, Pope CE, Hoffman LR, Smith-Vaughan HC. The microbiota in bronchoalveolar lavage from young children with chronic lung disease includes taxa present in both the oropharynx and nasopharynx. Microbiome. 2016;4(1):37.

16. Dickson R, Martinez F, Huffnagle G. The role of the microbiome in exacerbations of chronic lung diseases. Lancet. 2014;384(9944):691-702.
17. Boutin S, Dalpke AH. Acquisition and adaptation of the airway microbiota in the early life of cystic fibrosis patients. Mol Cell Pediatr. 2017;4(1):1.
18. Biesbroek G, Sanders EA, Roeselers G, Wang X, Caspers MP, Trzcinski K, Bogaert D, Keijser BJ. Deep sequencing analyses of low density microbial communities: Working at the boundary of accurate microbiota detection. PLoS One. 2012;7(3):e32942.
19. Salter SC, 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.
20. Marsh RL, 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 (Nathan). 2018;10:7.
21. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol. 2017;35(9):833-44.
22. Caporaso JK, J; Stombaugh, J; Bittinger, K; Bushman, FD; Costello, EK; Fierer, N; Gonzalez Peña, A; Goodrich, JK; Gordon, JI; Huttley, GA; Kelley, ST; Knights, D; Koenig, JE; Ley, RE; Lozupone, CA; McDonald, D; Muegge, BD; Pirrung, M; Reeder, J; Sevinsky, JR; Turnbaugh, PJ; Walters, WA; Widmann, J; Yatsunenko, T; Zaneveld, J; Knight, R. Qiime allows analysis of high-throughput community sequencing data. Nature Methods. 2010;7(5):335-6.
23. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol. 2009;75(23):7537-41.
24. Mallick H, Ma S, Franzosa EA, Vatanen T, Morgan XC, Huttenhower C. Experimental design and quantitative analysis of microbial community multiomics. Genome Biol. 2017;18(1):228.
25. Chen EZ, Li H. A two-part mixed-effects model for analyzing longitudinal microbiome compositional data. Bioinformatics. 2016;32(17):2611-7.
26. Subramanian S, Huq S, Yatsunenko T, Haque R, Mahfuz M, Alam MA, Benezra A, DeStefano J, Meier MF, Muegge BD, Barratt MJ, VanArendonk LG, Zhang Q, Province MA, Petri WA, Jr., Ahmed T, Gordon JI. Persistent gut microbiota immaturity in malnourished bangladeshi children. Nature. 2014;510(7505):417-21.
27. Pasolli E, Truong DT, Malik F, Waldron L, Segata N. Machine learning meta-analysis of large metagenomic datasets: Tools and biological insights. PLoS Comput Biol. 2016;12(7):e1004977.
28. Thakore N, Norville R, Franke M, Calderon R, Lecca L, Villanueva M, Murray MB, Cooney CG, Chandler DP, Holmberg RC. Automated trutip nucleic acid extraction and purification from raw sputum. PLoS One. 2018;13(7):e0199869.
29. Kerkhof LJ, Dillon KP, Haggblom MM, McGuinness LR. Profiling bacterial communities by minion sequencing of ribosomal operons. Microbiome. 2017;5(1):116.

**Figure. Challenges in characterising the respiratory microbiome.**

Niche-specific communities reside in the different parts of the URT and LRT, and therefore sampling site should depend on the research question and population studied. During health and acute URTI or LRTI, the LRT is transiently colonised with microbes from the URT (oropharynx for adults, naso- and oropharynx for children), while in chronic lung diseases, over time local selection and community assembly leads to differences between URT and LRT assemblages. In general, the local bacterial density in the respiratory tract is low, further decreasing when descending towards the LRT. Therefore, working with low-biomass samples requires careful sampling procedures and laboratory handling, including appropriate negative and positive controls to acquire reliable results. Microbial development over time is affected by environmental stimuli including crowding factors and pollution, and is altered in various acute and chronic diseases, so repeated sampling and exhaustive data collection of the same subjects over time is required to study cause-consequence relationships and estimate environment-induced variation. Appropriate bioinformatic processing of the sequencing data is required before robust statistical analysis is executed, which preferably accounts for covariates and repeated measures where relevant.

Abbreviations: URT = upper respiratory tract; LRT = lower respiratory tract; URTI = upper respiratory tract infection; LRTI = lower respiratory tract infection.


HEALTH/
URTI/LRTI
CHRONIC LUNG
DISEASE
BACTERIAL DENSITY
Nasal cavity
Nasopharynx
Oropharynx
Lungs
SPATIAL HETEROGENEITY
Local growth
TEMPORAL DYNAMICS
Crowding
Pollution
Disease
Therapy
CHOOSING
SAMPLING SITE
Pediatric LRT
Adult LRT
Chronic disease LRT
Nasopharynx
Oropharynx
Lung
HANDLING LOW
BACTERIAL DENSITY
AVOID
CONTAMINATION
MAXIMISE DATA
RESOLUTION
CAPTURING
TEMPORAL
DYNAMICS
Time point 1
Time point 2
Time point 3
Sampling & data collection
Repeated measures