

REVIEW

Open Access



The microbiome of the human lower airways: a next generation sequencing perspective

Velma T. E. Aho^{1*†}, Pedro A. B. Pereira^{1†}, Tari Haahtela², Ruby Pawankar³, Petri Auvinen¹ and Kaisa Koskinen^{1,4}

Abstract

For a long time, the human lower airways were considered a sterile environment where the presence of microorganisms, typically revealed by culturing, was interpreted as an abnormal health state. More recently, high-throughput sequencing-based studies have led to a shift in this perception towards the notion that even in healthy conditions the lower airways show either transient presence or even permanent colonization by microorganisms. However, challenges related to low biomass and contamination in samples still remain, and the composition, structure and dynamics of such putative microbial communities are unclear. Here, we review the evidence for the presence of microbial communities in the human lower airways, in healthy subjects and within the context of medical conditions of interest. We also provide an overview of the methodology pertinent to high-throughput sequencing studies, specifically those based on amplicon sequencing, including a discussion of good practices and common pitfalls.

Introduction

The human body harbours a varied, complex, and dynamic community of microorganisms known as the human microbiome. The term “microbiome” was originally coined by Joshua Lederberg [1]. Lederberg was specifically referring to the *human* microbiome, but since then the term has been extended to other environments as well. The original definition does not explicitly state whether the physicochemical characteristics of the environment in question are to be included, although not including them would make “microbiome” synonymous with “microbiota”. We prefer to make an explicit distinction that takes into account the ecological concept of “biome”. Therefore, the human microbiome consists of bacteria, archaea, protists, fungi, their respective viruses, and human viruses, as well as the surrounding host environment in which they exist. In practice, the terms are often used interchangeably, including in the present paper, but the meaning is nevertheless readily understood from context.

It has been estimated that the number of bacterial cells in the human body may exceed our own cells by an

order of magnitude ([2], but see discussion in [3]). Some of these resident organisms live in a state of commensalism with their human host, making out a living from the refuse of the host and/or from the metabolic by-products of other microbes, while others are part of a mutualistic relationship, in which both host and microbe benefit. Opportunistic pathogens are microorganisms that are commonly present in healthy individuals in low levels and that may become problematic due to abnormal outgrowth promoted by microbial community disruption and/or due to a compromised immune system.

Culture-independent methods have enabled the detection of bacteria in some rather unexpected niches in the human body. The lower respiratory tract of healthy individuals has been considered a sterile environment where the presence of any bacteria, typically revealed by culturing, represents an abnormal, unhealthy state [4, 5]. As with many other environments, this may have been caused by the fact that the microbes present are difficult to culture under standard laboratory conditions. Recently, numerous studies, particularly those based on high-throughput sequencing of the gene coding for the 16S ribosomal RNA, have provided tantalizing evidence that bacteria may be present in the lower respiratory tract even in healthy subjects. There are challenges in such studies, particularly related to the low biomass of the starting material and contamination of the samples

* Correspondence: velma.aho@helsinki.fi

†Equal contributors

¹DNA Sequencing and Genomics Laboratory, Institute of Biotechnology, University of Helsinki, P.O. Box 56 (Viikinkaari 4), 00014 Helsinki, Finland
Full list of author information is available at the end of the article

[6, 7]. Nevertheless, a growing number of studies are documenting the microbiome in healthy airways and in relation to medical conditions such as cystic fibrosis [8, 9], chronic obstructive pulmonary disease [10], asthma [11] and lung transplantation [12]. In some medical conditions, the sequencing data is supported by evidence from culturing methods [13, 14], but many studies rely entirely on a culture-independent approach.

Here, we provide an overview of DNA sequencing-based methodology and the microbiome of the human lower airways, mainly focusing on bacteria, from a 16S rRNA marker gene high-throughput amplicon sequencing perspective. The reasons for this taxonomic and technical focus are mostly practical, and include: 1) the vast majority of published studies deal specifically with bacteria; 2) phylogenetic marker gene studies that take into account other taxa require their own specific primers and downstream processing, increasing the cost and duration of the studies; 3) lack of standardized protocols and easy implementation and analysis; 4) less is known about other taxonomic groups within a human microbiome context, leading to weaker theoretical evidence from which to derive mechanistic hypotheses regarding relationships with the human host; 5) many of those groups have somewhat less relevance to human disease when compared to bacteria; and 6) shotgun metagenomics is significantly more expensive and requires much more complex and time-consuming analysis.

A brief overview of methods in microbiome studies

Next generation sequencing in microbial ecology

Next generation sequencing technologies have been widely applied in microbial ecology research. The most utilized approaches include shotgun metagenomics and amplicon sequencing. The term metagenomics was first used by Handelsman and colleagues [15], and it denotes examining the structure and function of the entire (microbial) community based on data derived from bulk genetic material extracted directly from samples. Therefore, the produced sequence data contains a fragmented puzzle of genomes of bacteria, archaea, viruses, fungi, protists, and even plants and animals, depending on the studied environment [16].

The first large-scale metagenomics study was conducted by Venter and colleagues in the Sargasso Sea [17]. They showed that by applying whole-genome shotgun sequencing to microbial populations it is possible to assess the diversity and community structure, and to discover and at least partially reconstruct previously unknown microbial genomes. Their results showed that a considerable share of the detected bacteria were indeed unfamiliar and resistant to cultivation by common laboratory methods, and therefore emphasized the

importance of DNA-based techniques in studying diverse and still largely unknown microbial populations.

Amplicon sequencing can be defined as a targeted version of metagenomics: a specific genetic region shared by the community members of interest is amplified using universal primers. The produced fragments of similar length are then sequenced. Amplification success is assumed equal across taxa, and hence each sequence read randomly represents the genetic diversity in the studied sample. Typically, the amplified fragment is part of a phylogenetically or functionally informative gene, such as the 16S ribosomal RNA gene or other marker genes [18–20]. Therefore, 16S rRNA amplicon sequencing is, strictly speaking, not metagenomics, although it is common to describe it as such. The choice of which variable region or regions of the 16S rRNA gene to target for sequencing is not trivial, since it affects what taxa are recovered and how accurately they can be classified [21, 22].

Human microbiome studies are a branch of microbial ecology. The goal is to study the variation of microbial communities in regard to, for example, population, genotype, age, nutrition, medication, biotic and abiotic environment, and/or its influence on disease. Figure 1 summarizes the typical workflow of an amplicon-based microbiome study. The first step in any such project is study design, where many details such as selection criteria for subjects, type of samples collected, and the methods used in laboratory and data analysis must be considered. Sample collection has its specific challenges, such as avoiding contamination. Once the samples have been obtained, the laboratory workflow includes DNA extraction, PCR amplification, purification, and sequencing. For sequencing, there are several different platforms, each with their pros and cons. The resulting sequence data is analysed bioinformatically, and finally, compared with statistical tools. All these steps and their potential obstacles will be briefly reviewed below. For further discussion on conducting a microbiome study, see [23].

Sampling and sample processing

Sampling of the lower airways is typically based on bronchoalveolar lavage (BAL) or collection of induced sputum (Table 1). Contamination of samples taken from low biomass environments is a pervasive problem in DNA sequence-based studies [7]. It can occur during sample collection and/or during sample processing at the pre-sequencing stage. Completely aseptic methods of sample collection, storage, and processing are a practical impossibility, but the risk of contamination-derived problems is exacerbated when the biomass of the targeted microbiota is low [24]. This is often the case in human microbiome studies that rely on samples where few microbes are present, such as skin swabs or BAL fluid. The lower the microbial biomass of the sample, the greater the possibility that due

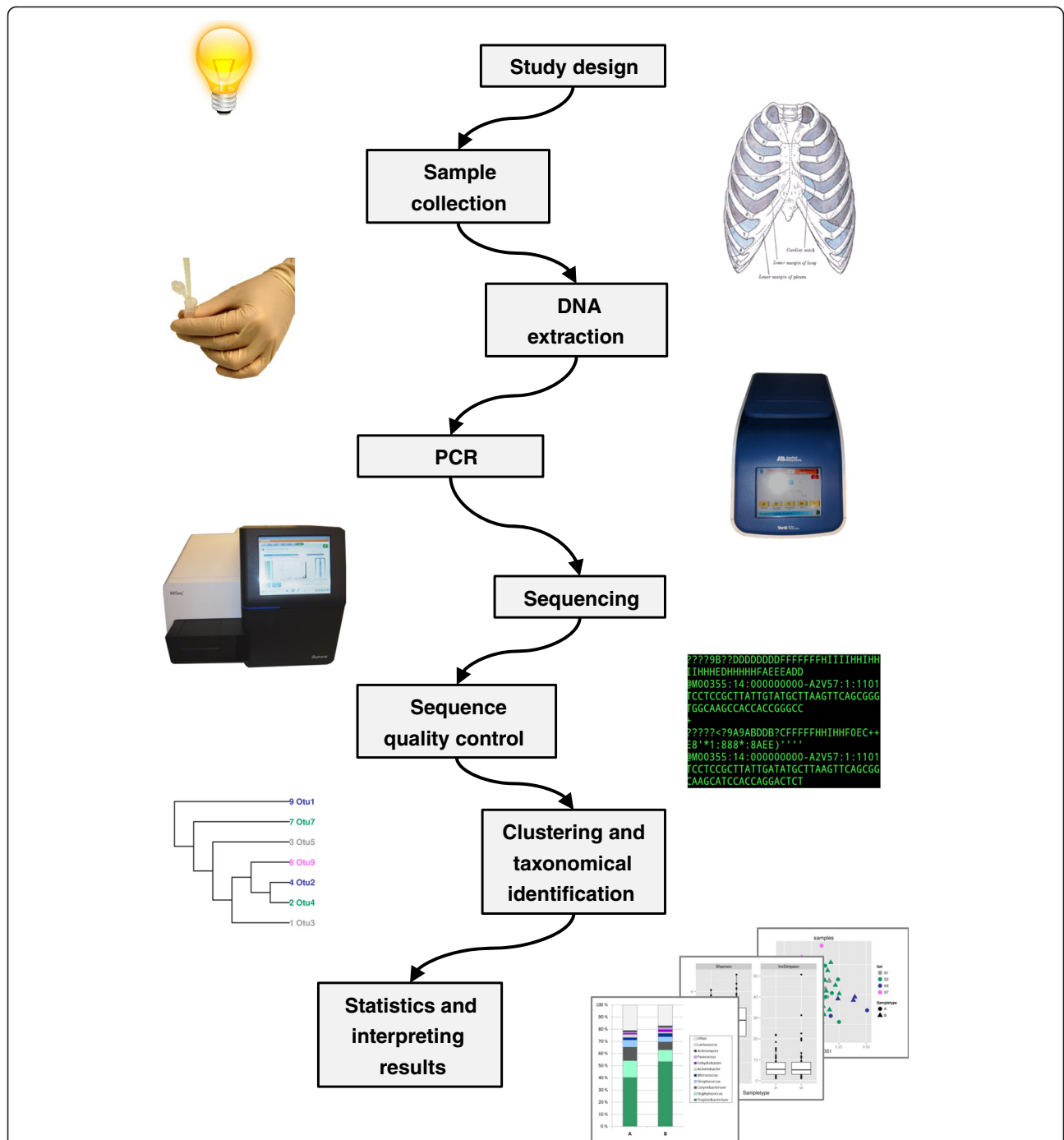


Fig. 1 General workflow of a microbiome study, from design to data analysis. This schematic is specific for target gene sequencing-based studies (e.g. 16S rRNA gene) and is not representative of studies using other approaches, e.g. shotgun metagenomics studies

to low concentration or quality of the obtained DNA, contaminants will swamp the real signal being sought. Potential sources of contamination include swabs, biopsy needles, collector, storage and reaction vessels, carryover by bronchoscopes, as well as the individual performing the sampling, either by accident or mishandling. To our knowledge, no thorough, systematic studies have been

conducted to evaluate the potential for contamination in these particular cases.

Contamination issues are not confined to sample collection; the problem is present in sample preparation at all stages. Different DNA extraction kits have been shown to contain different bacterial taxa as contaminants ([7], unpublished personal observations). Even different lots

Table 1 Next generation 16S rRNA gene amplicon based sequencing studies of the lower airway microbiome

Reference ^a	Medical condition	n (cases + controls)	Lower respiratory tract sample type	Sequencing platform	Targeted 16S rRNA regions	Sequence analysis software ^b
[108]	arsenic exposure	10 + 10	sputum	IonTorrent	V6	QIIME
[91]	asthma	10 + 10	sputum	454	V6	RDP
[92]	asthma	39 + 12	BAL	454	V1V2	BARTAB, RDP
[56]	CF	23	sputum	454	V1V2	AbundantOTU, Lucy, mothur, RDP
[59]	CF	10	lung explant secretions, sputum	454	V1V3, V2V3	KrakenBLAST, QIIME
[57]	CF	17	sputum	454	V4V6	custom in-house pipeline
[60]	CF	25	sputum	454	V1V3	RDP
[58]	CF	23	sputum	454	V6V7	QIIME
[109]	CF	19 + 6	sputum, lung tissue	454	V4V5	Matlab, mothur, RDP
[110]	CF	21	sputum	454	V1V2	BARTAB, RDP
[111]	CF	30	sputum	454	V1V3	custom software, Kraken BLAST
[68]	COPD	11	sputum	MiSeq	V4	QIIME
[42]	COPD, smoking	19 + 3	BAL, lung tissue	454	V1V3	mothur, RDP
[66]	COPD, smoking	16 + 16	lung tissue	454	V1V3	mothur
[41]	healthy	6	BAL, lower airway protected brush	454	V1V2	QIIME
[49]	healthy	28	BAL	454	V3V5	mothur
[44]	healthy smokers	19 + 45	BAL	454	V1V3, V3V5	mothur
[52]	HIV	82 + 77	BAL	454	V1V3	QIIME
[107]	ILD	24 + 9	BAL	454	V3V5	PyroTagger
[6]	intubation	5	ETA	454	V1V3	MG-RAST, mothur
[104]	IPF	55	BAL	454	V3V5	mothur
[105]	IPF	65 + 44	BAL	454	V3V5	QIIME
[96]	lung transplantation	4 + 2	BAL	454	V3	mothur, RDP
[95]	lung transplantation	21	BAL	454	V1V2	QIIME
[12]	lung transplantation	33 + 26	BAL	454	V3V5	mothur
[97]	lung transplantation	57 + 8	BAL	454	V7V8	QIIME
[102]	non-CF bronchiectasis	41	BAL, sputum	454	V1V3	custom software, Kraken BLAST
[103]	non-CF bronchiectasis	42	sputum	454	V1V3	custom software, Kraken BLAST
[101]	non-CF bronchiectasis	40	sputum	454	V1V3	AbundantOTU, Lucy, mothur, RDP
[45]	smoking, pulmonary inflammation	20 + 9	BAL	454	V1V2	QIIME
[106]	tuberculosis	22 + 14	sputum	454	V1V2	QIIME
[43]	various	6	BAL	454	V1V2	QIIME
[112]	various	56 + 4	BAL, sputum	454	V1V3	mothur

^aIn the case of several publications using the same sequence data, only the first one is included

^bSoftware for diversity calculation, statistical comparisons etc. is not included. RDP refers to the standalone implementation. If RDP was used via mothur or QIIME, it is not listed

Of special interest are the data on number of cases and controls, sample type, and 16S rRNA gene regions targeted for sequencing

from the same kit can show variability in taxa and amount of contaminant biomass. Yet another potential source of contamination are the laboratory reagents, such as enzymes, lab-grade water, and buffers. Additionally, the problem can be exacerbated during PCR amplification itself, in which with increasing PCR cycle number there is a corresponding increase in the production of contaminant sequences. Although the amount of the targeted amplicons will also increase and under ideal circumstances the proportions between both should be maintained, there may be biased amplification of the undesirable sequences depending on the primer sequences and DNA sequencing technology used (unpublished personal observations). This is not a significant issue with high biomass samples (e.g. stool), because the starting biomass is so high that there is no need for a large number of PCR cycles, and therefore the potential contaminant DNA will be suppressed by the real signal.

There are various approaches that can be used to minimize the effects of contamination and that are easy to implement in most lab settings. Table 2 summarises

some of those approaches, as well as other general considerations that are important for effective sample handling and processing.

There have been suggestions for treating reagents from extraction kits and PCR mixes using, for example, different types of radiation, restriction enzymes, or DNases [7]. All of these decontamination procedures have advantages and disadvantages, and their use needs to be evaluated within the context of the study in mind. Additionally, sequencing the complete genomes of common contaminants to identify and computationally subtract them from the data, will in the future help to distinguish between contaminants and real signal [24]. For a comprehensive review of these issues, as well as suggestions for dealing with them, see [7] and references therein.

Sequencing, bioinformatics and data analysis

The majority of published 16S rRNA gene sequencing based lower airway microbiome studies have used Roche's 454 pyrosequencing platform (Table 1). With

Table 2 Good practices in microbiome studies

Design considerations	<ul style="list-style-type: none"> The bacterial "universal" primers show bias in PCR amplification of certain taxa, and the use of different regions of the 16S rRNA gene in different studies (as seen for the lower airways in Table 1) affects inter-study comparisons (e.g. [21, 22])
Sample collection	<ul style="list-style-type: none"> Make sure that collecting and storage vessels are not needlessly subjected to potential contamination either by body contact or air exposure.
DNA extraction	<ul style="list-style-type: none"> Perform all activities (as long as it is practical to do so) under a hood with air filtering. Samples should be randomized before DNA extraction so that batch effects are minimized when groups of interest (case vs. control, age, sex, treatment, etc.) are compared. Varying contamination in kit lots and laboratory reagents can create artificial differences between groups during statistical analysis if the samples are not handled in randomized fashion (see [7] for an excellent and clear example). Be generous with controls. Every batch of samples being isolated at the same time should include one kit "blank" (control) to which no sample material is added but which undergoes the same process of DNA extraction and sequencing as the "real" samples. This serves the purpose of controlling for both contamination present in the kit and contamination introduced during the extraction process. Keep records of which kit lot was used for DNA extraction of which sample, and don't mix reagents from different kit boxes, even if from the same lot. Use kits with bead-beating to increase the chances that taxa with thicker cell walls will be properly lysed and that taxonomic representation biases will be avoided as much as possible. Ensure that all samples from the same project are handled in the same way following a common protocol, each individual step preferably executed by the same person.
PCR	<ul style="list-style-type: none"> When possible, work in a PCR clean room. Sequence a PCR master mix "blank" (control) for each different master mix aliquot. Do not add any template. Master mix controls serve the purpose of detecting potential contamination in PCR reagents (already present or accidentally introduced during preparation) every time a new master mix is prepared. Use PCR replicates to minimize PCR bias (uneven PCR amplification, lack of reaction effectiveness) and to detect the diversity present in samples as thoroughly as possible [113]. A minimum of two replicate reactions should be prepared per sample.
Sequencing	<ul style="list-style-type: none"> If possible, sequence a mock community prepared from genomic DNA from known isolates. Since the sequence composition of this community is known, it can be used to identify contamination effects and sequencing errors in the target samples.

This list is provided as an example of technical considerations that must be taken into account in studies involving DNA sequencing of environmental samples. The second and third sections are important for sequence-based studies in general, while the first, fourth, and fifth sections are of special interest for studies requiring target DNA amplification prior to sequencing

the forthcoming withdrawal of 454 from the market, Illumina's sequencers are quickly becoming the standard for both amplicon and shotgun metagenomics studies. The Illumina MiSeq platform in particular will probably eventually dominate amplicon-based microbial surveys. The various existing sequencing platforms have marked differences in read lengths, throughput, run time, error profiles, and cost [25, 26].

Regardless of the sequencing platform used, the raw sequence data requires several quality control steps to avoid bias and artefacts caused by the PCR amplification and sequencing [27]. Low quality reads are removed based on criteria such as quality scores, mismatches to barcodes and primers, sequence length, and the presence of ambiguous bases. Chimeric reads, which are the result of PCR products where two or more different template sequences have accidentally been joined together, are also removed [27]. After quality control, the sequences are typically clustered to operational taxonomic units (OTUs). Clustering can be performed based on only the sequences themselves ("de novo"), using a reference database of known sequences as cluster centroids ("closed-reference OTU picking"), or with a combination of the two approaches ("open-reference OTU picking") [28]. Regardless of the method, clustering is usually based on a 3% dissimilarity cut-off, which is used as a proxy for species [29]. It is also possible to omit the clustering step entirely and to simply use a reference database to identify each sequence read, binning together those that have the same taxonomic classification ("phylotype" approach). All methods that rely on a reference database are highly dependent on the quality of the database, while the database-independent "de novo" approach is computationally heavier than the other alternatives. It is particularly important to note what approach has been used when comparing different studies, since different clustering methods can produce widely varying results [30, 31]. After clustering, the resulting OTUs are identified taxonomically using a suitable reference database. This is another step where different databases may lead to major differences in the results.

An additional quality control consideration particularly for studies using low biomass samples is that the data may include contaminants [7]. The data from samples should be compared to sequenced negative controls and to existing publications, to look for unexpected discrepancies that could suggest contamination.

There is a large selection of software available for each step in the analysis of 16S rRNA gene sequence data, from quality control to clustering and taxonomical classification. Instead of using a different tool for each step, many researchers prefer software packages that cover

the complete workflow from raw sequence data to final results. The most commonly used ones are mothur [32] and QIIME [33], and the choice between the two is more of a question of personal preference, since both can be used to perform nearly the same tasks. The bioinformatic workflow should always be documented as meticulously as the laboratory work, to ascertain that it is replicable.

The results of the bioinformatic sequence analysis typically include representative sequences for the OTUs and a table that summarizes how many sequences of each OTU are present in each sample. The total amount of sequences per sample can vary significantly, and it is important to take this library size variation into account in statistical comparisons. Traditionally, this has been done by rarefaction or by converting the counts into relative abundances, but neither approach is optimal [34], and there are alternatives, such as the normalization methods offered by the metagenomeSeq [35] and phyloseq [34] R packages. Depending on the study design, the microbial community data can be used for a wide variety of statistical tests, which include, but are not limited to, searching for OTUs that are over or underrepresented in some predefined group (e.g. test subjects with a specific medical condition), comparisons of microbial diversity within and between samples or groups, and multivariate statistics for exploring the possible associations between clinical data and community composition. If OTUs of particular interest are discovered, the sequence data can be used for further analyses of the OTUs in question, both in continued bioinformatic and statistical analysis as well as in future experiments that focus on specific taxa instead of the entire microbiome.

As an end note, by microbial diversity we mean a measurement that takes into account both species richness (number of different taxa) and evenness (how abundant the taxa are). For any given sample, this would be represented by alpha diversity using such indices as Shannon or Inverse Simpson. Beta diversity consists of a distance measure between samples that represents the compositional dissimilarity or heterogeneity between those same samples. Typically, the higher the value of the beta diversity index for any given comparison the higher the dissimilarity between the samples. Notice that different indices have different meanings from an ecological point of view, and their interpretation is not always straightforward. In this review, we try to avoid ambiguity by identifying the meaning of "diversity" as used by the authors of specific articles mentioned in the text. For more information on this topic we suggest the articles by Jost [36, 37], Tuomisto [38], and Anderson et al. [39], and references therein.

The future

The development of DNA-based methods has already changed our knowledge of the structure and function of our genome as well as our second, microbial genome. Although DNA sequencing is going to become even cheaper in the future, easing the financial constraints on human microbiome studies, many problems remain that are related to study design and analysis of the results rather than just producing data. Bioinformatic tools are still being developed, and comparing results from report to report can be challenging due to the lack of consensus in analysis and presentation of data. Since new methods may require tailor-made analysis, this problem is likely to remain unsolved in the foreseeable future.

Shotgun metagenome projects, which avoid many of the limitations of the 16S rRNA gene approach, are gaining in popularity. They require no selected regions, just brute force sequencing. However, they also add new obstacles to the data analysis, making it more complex and significantly heavier computationally. On the other hand, the results will be more comprehensive and more to the point as reference databases are being filled with well annotated genomes and metagenomes. More functional insight can be obtained, especially as more RNA analyses are performed. Clinical diagnostics will be driven to sequencing-based approaches in the near future. Single-molecule sequencers, such as PacBio and the upcoming nanopore solutions, will be used to rapidly evaluate samples from unknown infections. Mass spectrometry is developing nearly as fast as DNA sequencing, opening novel opportunities in research and diagnostics. Still, sampling is the key to success. When samples are small or of limited quality, contamination problems will remain serious when these very sensitive assays are used.

The microbiome in healthy lower airways

The first studies using molecular methods to characterize the microbiome of the lower airways concluded that while the lower airways of healthy subjects are not sterile, the amount of bacteria is low, and the community is mainly made up of bacterial genera that are also common in the upper respiratory tract, including *Prevotella*, *Streptococcus* and *Veillonella* [40–42]. *Haemophilus* spp. [40] and *Tropheryma whipplei* [41] were suggested as bacteria specific to the lower airways. Later analysis of healthy subjects from an earlier study [41] suggested that three out of six subjects had some OTUs that were enriched in BAL fluid compared to oral wash samples, but all of them were found to be common in either negative controls or peri-glottic samples [43].

In more recent publications, several approaches have been used to eliminate possible contamination and to define statistically which bacteria are typical of the

healthy lower respiratory tract. One study listed OTUs classified as *Ralstonia*, *Bosea*, *Haemophilus*, *Methylobacterium*, *Tropheryma* and an unclassified *Enterobacteriaceae* OTU as potentially lung-specific [44]. Although samples with communities similar to negative controls were removed from the analysis, many of the genera on this list (particularly *Ralstonia*, *Bosea* and *Methylobacterium*, as well as *Enterobacter*, a genus of *Enterobacteriaceae*) correspond to a published list of common contaminants [7], raising questions regarding their actual origin. Another fairly recent study suggested two “pneumotypes”, one of them including genera similar to “background” or negative control samples, the other with a community resembling that of supraglottic samples, with high levels of *Prevotella* and *Veillonella* [45]. The negative controls and the samples similar to them were shown to contain a *Streptococcus* species different from the one typically found in supraglottic samples and supraglottic-like BAL samples. However, it is possible that the background-like “pneumotype” does not represent a true bacterial community, especially since the BAL samples were centrifuged before DNA isolation and only the supernatant was used, which is likely to have an impact on the resulting bacterial communities.

When viewed critically, the existing publications that strive to characterize the microbiome of the healthy lower respiratory tract are not without problems. Many of the bacteria common to the upper airways that have also been detected in the lower airways are likely to originate from microaspiration (subclinical aspiration of small droplets) of oropharyngeal secretions (e.g. [46]), which could mean that their presence is transient. It has been suggested that there is no lung “core” microbiome that is similar across study subjects and stable over time, but that the oral microbiome influences the microbial content in lungs [47]. Another possible source could be gastric reflux [48]. Contamination during sampling may also play a role. On the other hand, recent studies have shown that the BAL microbiome can be sampled without upper respiratory tract contamination [45] and that the lung microbes differ significantly from oral and gastric bacterial community [49]. Additionally, it seems that the sampling route (via nose or mouth) does not significantly affect the BAL microbiome [12].

Finally, it is also important to keep in mind that the standard 16S rRNA gene amplicon sequencing approach does not differentiate living, active cells from dead ones, raising the possibility that, at least in part, the recovered DNA has its origins in non-viable bacteria. Supporting this view is a study in healthy young pigs in which DNase was added to porcine BAL fluid and lung tissue samples [50]. While total bacterial DNA was abundant in both sample types, treatment with DNase substantially reduced the amount of cultivable cells, and showed

a total DNA reduction of 63% in BAL samples. On the other hand, Venkataraman et al. [51] showed evidence that 61% of 16S rRNA gene sequences recovered from healthy lung BAL could be matched to species recovered by a diverse assortment of culturing techniques. Of further interest, the study supports a neutral model of community ecology in which most of the bacteria in healthy (as opposed to diseased) lungs are the result of dispersal sourced mainly from the oral cavity, with dispersal and ecological drift dominating over selective forces in shaping healthy lung communities.

Out of the genera suggested to be enriched in the lower airways, some, such as *Haemophilus* [40, 44] and *Tropheryma* [41, 44, 52], have been reported in multiple studies, and are not known contaminants [7]. Further research is needed to ascertain whether there truly is a specific, stable lower airway microbiome in healthy subjects, or if the bacteria that have been discovered from such subjects are only transient visitors [48].

The lower airway microbiome and specific medical conditions

Cystic fibrosis

Cystic fibrosis (CF) is a recessive genetic disease particularly common in Caucasian populations [53]. It is caused by mutations that lead to the loss of function or dysfunction of the CF transmembrane conductance regulator protein (CFTR), which among other things causes abnormal mucous secretions and low airway surface pH, facilitating bacterial colonization of the lung [8]. Lung failure associated with chronic airway infections is the most common cause of morbidity and mortality in CF [9]. Based on the amount of publications (Table 1), CF is the lower respiratory tract related condition most commonly studied with next generation 16S rRNA gene sequencing.

CF airway infection is caused by a complex community of microbes [54], but only a handful of bacterial species have been established as clinically relevant. Particularly important pathogens are diverse strains of *Pseudomonas aeruginosa*, and the *Burkholderia cepacia* complex (Bcc) [8]. Other notable species include *Haemophilus influenzae* and *Staphylococcus aureus*, both common in paediatric CF patients, and emergent pathogens such as *Stenotrophomonas maltophilia*, *Achromobacter xylosoxidans*, *Mycobacterium abscessus* and the *Streptococcus milleri* group [8, 9]. In addition to recognized pathogens, many genera typical to the upper airways and to healthy lower airways, such as *Prevotella* and *Veillonella*, are also present in CF patients' lower airways [8, 9]. A study where the taxa found in CF sputum samples were partitioned statistically to satellite and core groups suggested that the core genera include *Porphyromonas*, *Prevotella*,

Streptococcus, *Catonella*, *Veillonella*, *Neisseria* and *Pseudomonas* [55].

Several studies have shown that the microbiome of the CF airways is very stable: neither clinical exacerbation nor antibiotic treatment have a significant effect on the patients' microbiota [56, 57]. A recent study reported that although antibiotic treatment of *P. aeruginosa* infected CF patients with pulmonary exacerbations initially caused a drop in the relative abundance of *Pseudomonas* and an overall increase in alpha diversity (measured with the Shannon diversity index), the microbial communities reverted to their pre-treatment state within 8–10 days of starting the treatment [58].

A majority of published CF studies have used sputum samples (Table 1), which are thought to represent the lower airway microbiota [59]. One study reports that the communities in sputum and in mouthwash samples are extremely similar, except for a small number of taxa, but suggests that this is probably due to microaspiration, referring to earlier studies as proof that sputum and oral microbiota are different [56]. However, another study using secretions obtained from lung explants concludes that sputum samples are not good surrogates for the lung microbiome, which is mainly dominated by typical CF pathogens and less diverse than the upper airway microbiome [59]. A recently suggested additional source of lower airway microbiota in CF patients is aspirated bile, which seems to affect the community composition significantly [60].

Chronic obstructive pulmonary disease and smoking

Chronic obstructive pulmonary disease (COPD) is a progressive lung disease that is mainly defined by irreversible airflow limitation, small airway fibrosis, mucus hypersecretion, and tissue destruction [61–63]. The leading cause of COPD in the developed world is exposure to tobacco smoke, and in developing countries indoor air problems from biomass fuel combustion also increase the occurrence [64]. It is not known why some of the exposed develop the disease and others do not, but a microbial component in the etiology is possible, as the lung microbiome seems to differ between healthy and affected individuals [40, 47, 65–67].

Several studies suggest that COPD patients carry bacteria in their lungs that are not present in healthy individuals. These bacteria include for example *Moraxella*, *Curvibacter*, *Corynebacterium* [65], and dominance of *Pseudomonas* [42]. However, a general consensus about the significance of individual taxa and their involvement in the pathogenesis of COPD is still lacking. Attempts to define the association between COPD and microbial diversity in the lungs have given rather contradictory results: it has been suggested that the COPD lung microbiome is richer [67] and more diverse (Shannon and inverse Simpson) [47], that there is

no difference in richness and evenness between healthy and affected lungs [10, 65], and that healthy lungs carry a more diverse microbiome (non-parametric Shannon) [42]. This inconclusiveness is probably linked to several factors: the interpersonal variation between lung microbiota is large and influenced heavily by the oral microbiome [47]; different lung compartments are dominated by different bacteria [42, 67]; and different levels of disease severity and medication affect the microbial community structure and diversity [47]. For instance, inhaled corticosteroids that are commonly used to treat COPD and asthma change the microbiome, advance enrichment of many community members, and may promote bacterial colonization in the airways [68]. In addition, the number of study subjects, sampling, selected laboratory protocols and analysis methods play a crucial role in the reliability of results.

Regarding smoking, in a recent thorough study comprised of sixty-four participants, the authors conclude that, although the data supports the existence of differences between the oral microbiome of “healthy” smokers and non-smokers, this does not seem to be the case regarding the lung communities, suggesting that the latter are not significantly altered by smoking [44]. An earlier study suggests that there is lower bacterial diversity (non-parametric Shannon) in smokers with moderate to severe COPD when compared to “healthy” smokers and non-smokers (the last two groups appearing to have no significant differences overall) [42]. Other studies have supported the hypothesis that lung bacterial communities don’t differ significantly between smokers and “healthy” smokers, while also adding to the evidence for differences in patients with severe COPD [45, 66].

Asthma

Asthma is a complex, chronic disorder where inflammation-related changes in the airways lead to air-flow limitation, causing recurring symptoms such as coughing, wheezing and breathlessness [69]. While the actual causes of asthma remain obscure, work conducted in recent years has unravelled a possible role for microbiota in either promotion of or protection from the disease [11, 70, 71]. As in other medical conditions discussed in this review, the data is strongly suggestive of a distinct pulmonary microbial community in healthy versus asthmatic people. Several studies suggest associations between environmental microbial exposure and the risk of developing asthma [72–75], or a role for antibiotic-induced gastrointestinal tract community effects during the first years of life [76–79]. The latter hypothesis is supported by the concept of a compartmentalized “common mucosal immune system”, in which microbe-induced differences in immunity at one mucosal site (e.g. gut) can affect immune function at other sites, including the mucosal surfaces of the airways [80–82].

Diverse studies, based on methods such as culturing, PCR, and/or DNA sequencing, have shown an association between asthma and certain microbial taxa. Implicated taxonomic groups include *Mycoplasma pneumoniae* and/or *Chlamydophila pneumoniae* [83–87]; *Moraxella catarrhalis*, *Haemophilus influenzae*, and/or *Streptococcus pneumoniae* [88, 89]; significantly more Proteobacteria, particularly *Haemophilus* spp., as well as *Staphylococcus* spp., but less Bacteroidetes, especially *Prevotella* spp. [40]; increased abundance of ~100 bacterial taxa distributed into 31 families, mostly Proteobacteria [90]; and higher levels of Proteobacteria, especially Gammaproteobacteria, in asthma cases [91]. Some studies report higher bacterial diversity in asthma [90, 91], while one study found no difference in diversity between controls and asthmatics [92]. Fungi have also been investigated in the context of asthma, with 90 fungal species showing higher relative abundance in sputum samples from asthma cases (particularly *Psathyrella candolleana*, *Malassezia pachydermatis*, *Termitomyces clypeatus* and *Grifola sordulenta*) while 46 were more abundant in controls (particularly *Eremothecium sincaudum*, *Systenostrema alba*, *Cladosporium cladosporioides* and *Vanderwaltozyma polyspora*) [93]. *Malassezia pachydermatis*, which has been associated with atopic dermatitis as well as other atopic conditions, was present in asthma patients but not in the control group. Viral infections at early ages have also been associated with asthma [75].

Whether the aforementioned associations represent a causative or mediating role for microbiota in asthma, or reflect the result of community disturbances due to other causes is unclear [71]. A potential confounding factor that may affect conclusions derived from some of these studies is the use of patients that are under medical treatment (e.g. [40, 90]) which may produce anomalous microbial communities as a consequence. This is a potential problem for clinical studies in general, and extra care and consideration should be given to disentangle the effects of medication (e.g. antibiotics) from the real signal. A particular confounder in asthma studies could be the use of inhaled corticosteroids, which are immunosuppressive and could affect the lower airway microbiome. However, the results of a study where most of the asthmatic subjects were not using corticosteroids [91] were in line with older studies where corticosteroid use was not controlled for [40, 90]. See also [92, 94].

Other conditions

It has been reported that the bacterial and fungal communities in the lower airways of lung transplant recipients differ from those of their upper airways, as well as from the upper and lower airways of healthy subjects [95, 96]. Although one small study suggests that the lower airways of transplant recipients have higher

microbial alpha diversity (Shannon and inverse Simpson indices) than those of healthy subjects [96], most studies have concluded that the diversity is lower in transplant recipients' airways (species richness and Shannon index) [12, 95, 97]. According to one study, the risk of bronchiolitis obliterans syndrome (BOS) is higher when the post-transplantation microbial community is distinct from the pre-transplantation state, while reestablishment of the pre-transplantation community has a protective effect, even when the subjects have CF and the community includes the potentially pathogenic *Pseudomonas* [97]. In a recent study, lung transplant recipients were found to have communities dominated by either *P. aeruginosa*, identified by both culturing and sequencing and associated with worse clinical status, or by *P. fluorescens*, which was only detectable by molecular methods, and typically found in asymptomatic subjects [12].

Since HIV patients have an increased risk of recurrent pneumonia, their lower airway microbiome has been studied, and there is a large multicenter research project, the Lung HIV Microbiome Project (LHMP), dedicated to it [44]. A high abundance of *Tropheryma whippelii* has been suggested as characteristic to HIV patients [52]. Antiretroviral therapy appears to have an effect on the microbiota [52, 98]. A metabolomics study using liquid chromatography-high-resolution mass spectrometry found that BAL samples from healthy subjects and HIV-1 infected patients had different metabolite profiles, and since many metabolites remained unidentified while one was identified as a siderophore of *P. aeruginosa*, the difference might be due to different microbiota [99]. While most of the airway microbiome publications cited in this review have studied Western populations, a microarray-based study compared BAL samples from HIV-infected pneumonia patients in Uganda [100] to an earlier study of a similar cohort from San Francisco [98], concluding that the Ugandan population had more diverse airway microbiota, and that several pathogenesis-related pathways were enriched in their predicted metagenome [100].

Several recent publications have addressed non-cystic fibrosis bronchiectasis. One study suggested that the complex microbial community present in the patients' lower airways remains unchanged during exacerbations [101]. Another study found that community composition and diversity, particularly species richness, are correlated with lung function [102], leading to the proposal of a microbiome-based stratification system for predicting exacerbations, where patients with high relative abundance of *P. aeruginosa* or *Veillonella* are more likely to suffer frequent exacerbations than those with high abundance of *H. influenzae* [103].

Recently, an association between the lung microbiome, idiopathic pulmonary fibrosis (IPF) and disease

progression has been identified [104, 105]. IPF is a progressive lung disease with unknown etiology that leads to respiratory failure and death typically within three to five years of diagnosis. It has been shown that patients with IPF have more than twofold higher pulmonary bacterial load compared to healthy individuals, and increased abundances of *Haemophilus*, *Neisseria*, and *Veillonella* [105]. Furthermore, the disease progression is associated with increased relative abundances of *Streptococcus* and *Staphylococcus*, and the overall bacterial burden: patients with highest bacterial load had more rapidly progressing disease, and a substantially increased risk of death [104, 105]. Further research is needed to determine if the changed microbiome is a cause or the result of IPF, but the findings provide support for testing antimicrobial therapy in IPF.

Other conditions that have been studied in relation to the microbiome of the lower airways so far include tuberculosis [106], interstitial lung disease [107], arsenic exposure [108] and intubation [6].

Conclusion

The study of the human lower airway microbiome is still in its infancy. The research suggests that, contrary to what was previously thought, the lower airways are inhabited by a diverse bacterial community. Although the evidence is not conclusive, it has strong support from a variety of studies performed throughout the history of the field, using different and complementary methods. Regardless of what the true scenario is, there is little consensus on the actual impact of specific bacterial taxa, or of the community structure and diversity on human health and disease. At present, published studies suggest that healthy and affected (e.g. CF, COPD, asthma) lungs carry partly distinct bacterial communities. Still, it is not known whether this difference is a cause or a consequence of the disease, and whether the varying bacterial communities reflect the disease severity, or if it is actually the medication that alters the community structure. As for healthy lung communities, it is at present not possible to differentiate between a mutualistic role for bacteria in promoting health as opposed to an essentially neutral, purely commensal role. Additionally, the amounts of study subjects in published papers are relatively small, the used methods rather dissimilar, and the differences between studies also suggest broad interindividual variation. More controlled studies are needed to solve the role of the lower airway microbiome in health and disease. Based on current data, larger study cohorts, standardised sampling protocols and laboratory procedures to avoid any misinterpretation caused by contamination, as well as carefully selected sequencing and detection methods are essential in future projects.

Abbreviations

BAL: Bronchoalveolar lavage; BOS: Bronchiolitis obliterans syndrome; CF: Cystic fibrosis; COPD: Chronic obstructive pulmonary disease; ILD: Interstitial lung disease; IPF: Idiopathic pulmonary fibrosis; NGS: Next generation sequencing; OTU: Operational taxonomic unit; PCR: Polymerase chain reaction; rRNA: Ribosomal ribonucleic acid.

Competing interests

The authors declared that they have no competing interests.

Authors' contributions

VA, PP, PA and KK read the relevant literature, and participated in designing, writing and revising the manuscript. TH and RP reviewed and edited the content of the manuscript in development. All authors reviewed and approved the final document.

Acknowledgments

The members of the Committee on Barrier Disease Issues and the Microbiome and the members of the Board of Directors of the World Allergy Organization have reviewed and approved this manuscript. VA is funded by the University of Helsinki Doctoral Programme in Biomedicine. TH has received funding from the Helsinki University Hospital (HUS) (grant no. 8361), and the European Union's Seventh Framework Programme FP7/2007-2013 under grant agreement 261357 (MeDALL). PA has received funding from The Michael J. Fox Foundation for Parkinson's Research. The funders had no role in the writing of the manuscript or the decision to submit the manuscript for publication.

Author details

¹DNA Sequencing and Genomics Laboratory, Institute of Biotechnology, University of Helsinki, P.O. Box 56 (Viikinkaari 4), 00014 Helsinki, Finland. ²Skin and Allergy Hospital, Helsinki University Hospital, Helsinki, Finland. ³Division of Allergy, Department of Pediatrics, Nippon Medical School, Tokyo, Japan. ⁴Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland.

Received: 20 November 2014 Accepted: 1 June 2015

Published online: 16 June 2015

References

- Lederberg J, McCray AT. 'Ome Sweet 'Ome—a genealogical treasury of words. *Scientist*. 2001;15:8.
- Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JL. The Human Microbiome Project. *Nature*. 2007;449:804–10.
- Torres VJ. Ten times more microbial cells than body cells in humans? *Microbe Magazine*. 2014;9.
- Baughman RP, Thorpe JE, Staneck J, Rashkin M, Frame PT. Use of the protected specimen brush in patients with endotracheal or tracheostomy tubes. *Chest*. 1987;91:233–6.
- Monso E, Ruiz J, Rosell A, Manterola J, Fiz J, Morera J, et al. Bacterial infection in chronic obstructive pulmonary disease: A study of stable and exacerbated outpatients using the protected specimen brush. *Am J Respir Crit Care Med*. 1995;152:1316–20.
- Lazarevic V, Gaia N, Emonet S, Girard M, Renzi G, Despres L, et al. Challenges in the culture-independent analysis of oral and respiratory samples from intubated patients. *Front Cell Infect Microbiol*. 2014;4:65.
- Salter S, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biol*. 2014;12:87.
- McGuigan L, Callaghan M: The evolving dynamics of the microbial community in the cystic fibrosis lung. *Environ Microbiol* 2014. Epub.
- Surette MG. The cystic fibrosis lung microbiome. *Ann Am Thorac Soc*. 2014;11 Suppl 1:61–5.
- Sze MA, Hogg JC, Sin DD. Bacterial microbiome of lungs in COPD. *Int J Chron Obstruct Pulmon Dis*. 2014;9:229–38.
- Beigelman A, Weinstock GM, Bacharier LB. The relationships between environmental bacterial exposure, airway bacterial colonization, and asthma. *Curr Opin Allergy Clin Immunol*. 2014;14:137–42.
- Dickson RP, Erb-Downward JR, Freeman CM, Walker N, Scales BS, Beck JM, et al. Changes in the lung microbiome following lung transplantation include the emergence of two distinct *Pseudomonas* species with distinct clinical associations. *PLoS One*. 2014;9:e97214.
- Sibley CD, Grinwis ME, Field TR, Eshaghurshan CS, Faria MM, Dowd SE, et al. Culture enriched molecular profiling of the cystic fibrosis airway microbiome. *PLoS One*. 2011;6:e22702.
- Huang YJ, Kim E, Cox MJ, Brodie EL, Brown R, Wiener-Kronish J, et al. A persistent and diverse airway microbiota present during chronic obstructive pulmonary disease exacerbations. *OMICS : a Journal of Integrative Biology*. 2010;14:9–59.
- Handelsman J, Rondon M, Brady S, Clardy J, Goodman R. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol*. 1998;5:245.
- Xu J. Microbial ecology in the age of metagenomics. In: de Bruijn FJ, editor. *Handbook of Molecular Microbial Ecology I*. Hoboken, NJ, USA: John Wiley & Sons, Inc; 2011.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA, et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science*. 2004;304:66–74.
- Cole JR, Konstantinidis KT, Farris RJ, Tiedje JM. Microbial diversity and phylogeny: Extending from rRNAs to genomes. In: Liu W-T, Jansson JK, editors. *Environmental Molecular Microbiology*. Norfolk, UK: Caister Academic Press; 2010.
- Huber JA, Welch DBM, Morrison HG, Huse SM, Neal PR, Butterfield DA, et al. Microbial population structures in the deep marine biosphere. *Science*. 2007;318:97–100.
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, et al. Microbial diversity in the deep sea and the underexplored “rare biosphere”. *Proc Natl Acad Sci U S A*. 2006;103:12115–20.
- Yarza P, Yilmaz P, Priesse E, Glockner FO, Ludwig W, Schleifer K, et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Micro*. 2014;12:635–45.
- Guo F, Ju F, Cai L, Zhang T. Taxonomic precision of different hypervariable regions of 16S rRNA gene and annotation methods for functional bacterial groups in biological wastewater treatment. *PLoS One*. 2013;8:e76185.
- Goodrich JK, Di Rienzi SC, Poole AC, Koren O, Walters WA, Caporaso JG, et al. Conducting a microbiome study. *Cell*. 2014;158:250–62.
- Laurence M, Hatzis C, Brash DE. Common contaminants in next-generation sequencing that hinder discovery of low-abundance microbes. *PLoS One*. 2014;9:e97876.
- van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. *Trends Genet*. 2014;30:418–26.
- Frey KG, Herrera-Galeano JE, Redden CL, Luu TV, Svetas SL, Mateczun AJ, et al. Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood. *BMC Genomics*. 2014;15:1–30.
- Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One*. 2011;6:1–14.
- Rideout JR, He Y, Navas-Molina JA, Walters WA, Ursell LK, Gibbons SM, et al. Subsampled open-reference clustering creates consistent, comprehensive OTU definitions and scales to billions of sequences. *PeerJ*. 2014;2:e545.
- Schloss P, Westcott S. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol*. 2011;77:3219–26.
- Sun Y, Cai Y, Huse SM, Knight R, Farmerie WG, Wang X, et al. A large-scale benchmark study of existing algorithms for taxonomy-independent microbial community analysis. *Brief Bioinform*. 2011;13:107–21.
- Koskinen K, Auvinen P, Björkroth KJ, Hultman J. Inconsistent denoising and clustering algorithms for amplicon sequence data. *J Comput Biol*. 2014. doi:10.1089/cmb.2014.0268.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol*. 2009;75:7537–41.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335–6.
- McMurdie PJ, Holmes S. Waste not, want not: Why rarefying microbiome data is inadmissible. *PLoS Comput Biol*. 2014;10:e1003531.
- Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nat Meth*. 2013;10:1200–2.
- Jost L. Entropy and diversity. *Oikos*. 2006;113:363–75.

37. Jost L. Partitioning diversity into independent alpha and beta components. *Ecology*. 2007;88:2427–39.
38. Tuomisto H. Commentary: do we have a consistent terminology for species diversity? Yes, if we choose to use it. *Oecologia*. 2011;167:903–11.
39. Anderson M, Crist T, Chase J, Vellend M, Inouye B, Freestone A, et al. Navigating the multiple meanings of beta diversity: a roadmap for the practicing ecologist. *Ecol Lett*. 2011;14:19–28.
40. Hilty M, Burke C, Pedro H, Cardenas P, Bush A, Bossley C, et al. Disordered microbial communities in asthmatic airways. *PLoS One*. 2010;5:e8578.
41. Charlson ES, Bittinger K, Haas AR, Fitzgerald AS, Frank J, Yadav A, et al. Topographical continuity of bacterial populations in the healthy human respiratory tract. *Am J Respir Crit Care Med*. 2011;184:957–63.
42. Erb-Downward JR, Thompson DL, Han MK, Freeman CM, McCloskey L, Schmidt LA, et al. Analysis of the lung microbiome in the “healthy” smoker and in COPD. *PLoS One*. 2011;6:e16384.
43. Charlson ES, Bittinger K, Chen J, Diamond JM, Li H, Collman RG, et al. Assessing bacterial populations in the lung by replicate analysis of samples from the upper and lower respiratory tracts. *PLoS One*. 2012;7:e42786.
44. Morris A, Beck JM, Schloss PD, Campbell TB, Crothers K, Curtis JL, et al. Comparison of the respiratory microbiome in healthy nonsmokers and smokers. *Am J Respir Crit Care Med*. 2013;187:1067–75.
45. Segal LN, Alekseyenko AV, Clemente JC, Kulkarni R, Wu B, Gao Z, et al. Enrichment of lung microbiome with supraglottic taxa is associated with increased pulmonary inflammation. *Microbiome*. 2013;1:19.
46. Gleeson K, Eggli D, Maxwell S. Quantitative aspiration during sleep in normal subjects. *Chest*. 1997;111:1266–72.
47. Pragman AA, Kim HB, Reilly CS, Wendt C, Isaacson RE. The lung microbiome in moderate and severe chronic obstructive pulmonary disease. *PLoS One*. 2012;7:e47305.
48. Marsland BJ. Influences of the microbiome on the early origins of allergic asthma. *Ann Am Thorac Soc*. 2013;10(Suppl):165–9.
49. Bassis CM, Erb-Downward JR, Dickson RP, Freeman CM, Schmidt TM, Young VB, et al. Analysis of the upper respiratory tract microbiotas as the source of the lung and gastric microbiotas in healthy individuals. *mBio*. 2015;6:e00037.
50. Pezzulo AA, Kelly PH, Nassar BS, Rutland CJ, Gansemer ND, Dohrn CL, et al. Abundant DNase I-sensitive bacterial DNA in healthy porcine lungs and its implications for the lung microbiome. *Appl Environ Microbiol*. 2013;79:5936–41.
51. Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, Huffnagle GB, et al. Application of a neutral community model to assess structuring of the human lung microbiome. *mBio*. 2015;6:e02284–14.
52. Lozupone C, Cota-Gomez A, Palmer BE, Linderman DJ, Charlson ES, Sodergren E, et al. Widespread colonization of the lung by *Tropheryma whipplei* in HIV infection. *Am J Respir Crit Care Med*. 2013;187:1110–7.
53. Lynch SV, Bruce KD. The cystic fibrosis airway microbiome. *Cold Spring Harb Perspect Med*. 2013;3:a009738.
54. Conrad D, Haynes M, Salamon P, Rainey PB, Youle M, Rohwer F. Cystic fibrosis therapy: A community ecology perspective. *Am J Respir Cell Mol Biol*. 2013;48:150–6.
55. van der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW, et al. Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J*. 2011;5:780–91.
56. Fodor AA, Klem ER, Gilpin DF, Elborn JS, Boucher RC, Tunney MM, et al. The adult cystic fibrosis airway microbiota is stable over time and infection type, and highly resilient to antibiotic treatment of exacerbations. *PLoS One*. 2012;7:e45001.
57. Price KE, Hampton TH, Gifford AH, Dolben EL, Hogan DA, Morrison HG, et al. Unique microbial communities persist in individual cystic fibrosis patients throughout a clinical exacerbation. *Microbiome*. 2013;1:27.
58. Smith DJ, Badrick AC, Zakrzewski M, Krause L, Bell SC, Anderson GJ, et al. Pyrosequencing reveals transient cystic fibrosis lung microbiome changes with intravenous antibiotics. *Eur Respir J*. 2014;44:922–30.
59. Goddard AF, Staudinger BJ, Dowd SE, Joshi-Datar A, Wolcott RD, Aitken ML, et al. Direct sampling of cystic fibrosis lungs indicates that DNA-based analyses of upper-airway specimens can misrepresent lung microbiota. *Proc Natl Acad Sci U S A*. 2012;109:13769–74.
60. Reen FJ, Woods DF, Mooij MJ, Chróinin MN, Mullane D, Zhou L, et al. Aspirated bile: A major host trigger modulating respiratory pathogen colonisation in cystic fibrosis patients. *Eur J Clin Microbiol Infect Dis*. 2014;33:1763–71.
61. Murray CJ, Lopez AD. Evidence-based health policy—lessons from the Global Burden of Disease Study. *Science*. 1996;274:740–3.
62. Lopez AD, Shibuya K, Rao C, Mathers CD, Hansell AL, Held LS, et al. Chronic obstructive pulmonary disease: current burden and future projections. *Eur Respir J*. 2006;27:397–412.
63. Barnes P, Drazen J, Rennard S, Thomson N. *Asthma and COPD: Basic Mechanisms and Clinical Management*. London: Academic; 2008.
64. Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. *Lancet*. 2009;374:733–43.
65. Zakharkina T, Heinzel E, Koczulla RA, Greulich T, Rentz K, Pauling JK, et al. Analysis of the airway microbiota of healthy individuals and patients with chronic obstructive pulmonary disease by T-RFLP and clone sequencing. *PLoS One*. 2013;8:e68302.
66. Sze MA, Dimitriu PA, Hayashi S, Elliott WM, McDonough JE, Gosselink JV, et al. The lung tissue microbiome in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2012;185:1073–80.
67. Cabrera-Rubio R, Garcia-Nunez M, Seto L, Anto JM, Moya A, Monso E, et al. Microbiome diversity in the bronchial tracts of patients with chronic obstructive pulmonary disease. *J Clin Microbiol*. 2012;50:3562–8.
68. 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;52:2813–23.
69. NHLBI (National Heart, Lung, and Blood Institute). National Asthma Education and Prevention Program Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma—Summary Report 2007. *J Allergy Clin Immunol*. 2007;120(5):94–138.
70. Huang YJ. Asthma microbiome studies and the potential for new therapeutic strategies. *Curr Allergy Asthma Rep*. 2013;13:453–61.
71. Huang YJ, Boushey HA. The microbiome and asthma. *Ann Am Thorac Soc*. 2014;11 Suppl 1:148–51.
72. Thavagnanam S, Fleming J, Bromley A, Shields MD, Cardwell CR. A meta-analysis of the association between Caesarean section and childhood asthma. *Clin Exp Allergy*. 2008;38:629–33.
73. von Mutius E, Vercelli D. Farm living: Effects on childhood asthma and allergy. *Nat Rev Immunol*. 2010;10:861–8.
74. von Mutius E. Environmental microorganisms and lung health. *Ann Am Thorac Soc*. 2014;11 Suppl 1:13–5.
75. Beigelman A, Bacharier LB. The role of early life viral bronchiolitis in the inception of asthma. *Curr Opin Allergy Clin Immunol*. 2013;13:211–6.
76. Chen Y, Blaser MJ. *Helicobacter pylori* colonization is inversely associated with childhood asthma. *J Infect Dis*. 2008;198:553–60.
77. Blaser MJ, Chen Y, Reibman J. Does *Helicobacter pylori* protect against asthma and allergy? *Gut*. 2008;57:561–7.
78. Marra F, Marra CA, Richardson K, Lynd LD, Kozyrskiy A, Patrick DM, et al. Antibiotic use in children is associated with increased risk of asthma. *Pediatrics*. 2009;123:1003–10.
79. Murk W, Risnes KR, Bracken MB. Prenatal or early-life exposure to antibiotics and risk of childhood asthma: A systematic review. *Pediatrics*. 2011;127:1125–38.
80. Chen K, Cerutti A. Vaccination strategies to promote mucosal antibody responses. *Immunity*. 2010;33:479–91.
81. Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science*. 2012;336:1268–73.
82. Hörmansperger G, Clavel T, Haller D. Gut matters: Microbe-host interactions in allergic diseases. *J Allergy Clin Immunol*. 2012;129:1452–9.
83. Kraft M, Cassell GH, Henson JE, Watson H, Williamson J, Marmion BP, et al. Detection of *Mycoplasma pneumoniae* in the airways of adults with chronic asthma. *Am J Respir Crit Care Med*. 1998;158:998–1001.
84. Kraft M, Cassell GH, Pak J, Martin J. *Mycoplasma pneumoniae* and *Chlamydia pneumoniae* in asthma: Effect of clarithromycin. *Chest*. 2002;121:1782–8.
85. ten Brinke A, van Dissel JT, Sterk PJ, Zwinderman AH, Rabe KF, Bel EH. Persistent airflow limitation in adult-onset nonatopic asthma is associated with serologic evidence of *Chlamydia pneumoniae* infection. *J Allergy Clin Immunol*. 2001;107:449–54.
86. Martin RJ, Kraft M, Chu HW, Berns EA, Cassell GH. A link between chronic asthma and chronic infection. *J Allergy Clin Immunol*. 2001;107:595–601.
87. Specjalski K, Jassem E. *Chlamydia pneumoniae*, *Mycoplasma pneumoniae* infections, and asthma control. *Allergy Asthma Proc*. 2011;32:9–17.
88. Bisgaard H, Hermansen MN, Buchvald F, Loland L, Halkjaer LB, Bønnelykke K, et al. Childhood asthma after bacterial colonization of the airway in neonates. *N Engl J Med*. 2007;357:1487–95.

89. Bisgaard H, Hermansen MN, Bønnelykke K, Stokholm J, Batty F, Skjott NL, et al. Association of bacteria and viruses with wheezy episodes in young children: prospective birth cohort study. *BMJ*. 2010;341:c4978.
90. Huang YJ, Nelson CE, Brodie EL, Desantis TZ, Baek MS, Liu J, et al. Airway microbiota and bronchial hyperresponsiveness in patients with suboptimally controlled asthma. *J Allergy Clin Immunol*. 2011;127:372–81.
91. Marri PR, Stern DA, Wright AL, Billheimer D, Martinez FD. Asthma associated differences in microbial composition of induced sputum. *J Allergy Clin Immunol*. 2013;131:346–52.
92. Goleva E, Jackson LP, Harris JK, Robertson CE, Sutherland ER, Hall CF, et al. The effects of airway microbiome on corticosteroid responsiveness in asthma. *Am J Respir Crit Care Med*. 2013;188:1193–201.
93. van Woerden HC, Gregory C, Brown R, Marchesi JR, Hoogendoorn B, Matthews IP. Differences in fungi present in induced sputum samples from asthma patients and non-atopic controls: a community based case control study. *BMC Infect Dis*. 2013;13:69.
94. Slater M, Rivett DW, Williams L, Martin M, Harrison T, Sayers I, et al. The impact of azithromycin therapy on the airway microbiota in asthma. *Thorax*. 2014;69:673–4.
95. Charlson ES, Diamond JM, Bittinger K, Fitzgerald AS, Yadav A, Haas AR, et al. Lung-enriched organisms and aberrant bacterial and fungal respiratory microbiota after lung transplant. *Am J Respir Crit Care Med*. 2012;186:536–45.
96. Borewicz K, Pragman AA, Kim HB, Hertz M, Wendt C, Isaacson RE. Longitudinal analysis of the lung microbiome in lung transplantation. *FEMS Microbiol Lett*. 2013;339:57–65.
97. Willner DL, Hugenholtz P, Yerkovich ST, Tan ME, Daly JN, Lachner N, et al. Re-establishment of recipient-associated microbiota in the lung allograft is linked to reduced risk of bronchiolitis obliterans syndrome. *Am J Respir Crit Care Med*. 2013;187:640–7.
98. Iwai S, Fei M, Huang D, Fong S, Subramanian A, Grieco K, et al. Oral and airway microbiota in HIV-infected pneumonia patients. *J Clin Microbiol*. 2012;50:2995–3002.
99. Cribbs SK, Park Y, Guidot DM, Martin GS, Brown LA, Lennox J, et al. Metabolomics of bronchoalveolar lavage differentiate healthy HIV-1-infected subjects from controls. *AIDS Res Hum Retroviruses*. 2014;30:579–85.
100. Iwai S, Huang D, Fong S, Jarlsberg LG, Worodria W, Yoo S, et al. The lung microbiome of Ugandan HIV-infected pneumonia patients is compositionally and functionally distinct from that of San Franciscan patients. *PLoS One*. 2014;9:e95726.
101. Tunney MM, Einarsson GG, Wei L, Drain M, Klem ER, Cardwell C, et al. Lung microbiota and bacterial abundance in patients with bronchiectasis when clinically stable and during exacerbation. *Am J Respir Crit Care Med*. 2013;187:1118–26.
102. Rogers GB, van der Gast CJ, Cuthbertson L, Thomson SK, Bruce KD, Martin ML, et al. Clinical measures of disease in adult non-CF bronchiectasis correlate with airway microbiota composition. *Thorax*. 2013;68:731–7.
103. Rogers GB, Zain NMM, Bruce KD, Burr LD, Chen AC, Rivett DW, et al. A novel microbiota stratification system predicts future exacerbations in bronchiectasis. *Ann Am Thorac Soc*. 2014;11:496–503.
104. Han M, Zhou Y, Murray S, Tayob N, Noth I, Lama V, et al. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med*. 2014;2:548–56.
105. Molyneaux P, Cox M, Willis-Owen S, Mallia P, Russell K, Russel A, et al. The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2014;190:906–13.
106. Cheung MK, Lam WY, Fung WYW, Law PTW, Au CH, Nong W, et al. Sputum microbiota in tuberculosis as revealed by 16S rRNA pyrosequencing. *PLoS One*. 2013;8:e54574.
107. Garzoni C, Brugger SD, Qi W, Wasmer S, Cusini A, Dumont P, et al. Microbial communities in the respiratory tract of patients with interstitial lung disease. *Thorax*. 2013;68:1150–6.
108. White AG, Watts GS, Lu Z, Meza-Montenegro MM, Lutz EA, Harber P, et al. Environmental arsenic exposure and microbiota in induced sputum. *Int J Environ Res Public Health*. 2014;11:2299–313.
109. Blainey PC, Milla CE, Cornfield DN, Quake SR. Quantitative analysis of the human airway microbial ecology reveals a pervasive signature for cystic fibrosis. *Sci Transl Med*. 2012;4:53ra130.
110. Zemanick ET, Harris JK, Wagner BD, Robertson CE, Sagel SD, Stevens MJ, et al. Inflammation and airway microbiota during cystic fibrosis pulmonary exacerbations. *PLoS One*. 2013;8:e62917.
111. Rogers GB, Cuthbertson L, Hoffman LR, Wing PAC, Pope C, Hoofman DAP, et al. Reducing bias in bacterial community analysis of lower respiratory infections. *ISME J*. 2013;7:697–706.
112. van der Gast CJ, Cuthbertson L, Rogers GB, Pope C, Marsh RL, Redding GJ, et al. Three clinically distinct chronic pediatric airway infections share a common core microbiota. *Ann Am Thorac Soc*. 2014;11:1039–48.
113. Lahr DJ, Katz LA. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. *Biotechniques*. 2009;47:857–66.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

