Bacteria in the airways of patients with cystic fibrosis are genetically capable of producing VOCs in breath.
Bacteria in the airways of patients with cystic fibrosis are genetically capable of producing VOCs in breath

Lieuwe D J Bos, Simone Meinardi, Donald Blake and Katrine Whiteson

Department of Respiratory Medicine and Intensive Care, Universiteit van Amsterdam—Academic Medical Center, Amsterdam, the Netherlands
Department of Chemistry, University of California, Irvine, CA, USA
Department of Molecular Biology and Biochemistry, University of California, Irvine, USA
E-mail: katrine@uci.edu

Keywords: breath analysis, bacteria, microbiome, lung, cystic fibrosis

Abstract
Breath contains hundreds of volatile organic compounds (VOCs), the composition of which is altered in a wide variety of diseases. Bacteria are implicated in the formation of VOCs, but the biochemical mechanisms that lead to the formation of breath VOCs remain largely hypothetical. We hypothesized that bacterial DNA fragments in sputum of CF patients could be sequenced to identify whether the bacteria present were capable of producing VOCs found in the breath of these patients.

Breath from seven patients with cystic fibrosis was sampled and analyzed by gas-chromatography and mass-spectrometry. Sputum samples were also collected and microbial DNA was isolated. Metagenomic sequencing was performed and the DNA fragments were compared to a reference database with genes that are linked to the metabolism of acetaldehyde, ethanol and methanol in the KEGG database.

Bacteria in the genera Escherichia, Lactococcus, Pseudomonas, Rothia and Streptococcus were found to have the genetic potential to produce acetaldehyde and ethanol. Only DNA sequences from Lactococcus were implicated in the formation of acetaldehyde from acetate through aldehyde dehydrogenase family 9 member A1 (K00149). Escherichia was found to be genetically capable of producing ethanol in all patients, whilst there was considerable heterogeneity between patients for the other genera. The ethanol concentration in breath positively correlated with the amount of Escherichia found in sputum (Spearman rho = 0.85, P = 0.015). Rothia showed the most versatile genetic potential for producing methanol.

To conclude, bacterial DNA fragments in sputum of CF patients can be linked to enzymes implicated in the production of ethanol, acetaldehyde and methanol, which are VOCs that are predictive of respiratory tract colonization and/or infection. This supports that the lung microbiome can produce VOCs directly.

Introduction
Breath contains hundreds of volatile organic compounds (VOCs), the composition of which is altered in a wide variety of diseases [1]. The biochemical mechanisms that lead to the formation of breath VOCs remain largely hypothetical [2–6]. Although a mechanistic understanding of VOC origin is not required to use VOCs in the exhaled breath as a diagnostic tool [1], it would enable identification of the most appropriate patient populations to target with the test and would facilitate the acceptance of breath tests by clinicians.

Bacteria have been implicated in the production of VOCs for a long period of time. For centuries, lung abscesses have been reported to have a strong odor; breath odors were described as a diagnostic test in the ancient writings of Hippocrates. VOC profiles can distinguish between particular pathogens in vitro and in animal models of pneumonia [7–9]. This suggests that bacteria create the volatile compounds responsible for the odor associated with infection [10], and that particular bacterial species produce characteristic VOC profiles.

Recent developments in genomics allow for unbiased assessment of bacteria present in clinical samples.
The corresponding gene function and taxonomic origin can be predicted by comparing the sequenced DNA fragments to a reference database containing sequences of known taxonomy and/or function [12]. This ‘metagenomic’ approach has been especially useful for characterizing the polymicrobial communities that exist in human-associated environments, e.g. lungs of individuals with cystic fibrosis (CF), where there is a high abundance of different types of bacteria due to decreased mucociliary clearance [13].

In a related study, 2,3-butanedione was implicated as a marker for microbial fermentation processes in the lungs [12]. In this study, we focused on the presence and quantification of ethanol, acetaldehyde and methanol in breath from CF patients. Ethanol and acetaldehyde were described as an important marker of positive sputum and broncho-alveolar lavage fluid cultures with any bacterium in several studies [19–21], and they are found in the breath of healthy people [22]. Ethanol and acetaldehyde were both described as general markers of bacterial growth in a systematic review of all in vitro studies [9]. Methanol is found in the breath of healthy people [23], and could have microbial or human origins. Methanol has been shown to regulate the production of detoxification genes in mice [24].

We hypothesized that bacterial DNA fragments in sputum of CF patients could be sequenced to identify whether the bacteria present were capable of producing VOCs found in the breath of these patients. For this purpose, we used gas-chromatography and mass-spectrometry to detect VOCs in breath and genomics and metagenomics to characterize bacterial genes in sputum.

**Methods**

**Inclusion and obtained samples**

This was an observational cohort study that included 7 patients with CF recruited at the University of California San Diego Adult CF clinic as described in previous studies [12, 14]. IRB approval was obtained from the University of California Institutional Review Board (HRPP 081500) and San Diego State University Institutional Review Board (SDSU IRB#2121). After informed consent was obtained we collected sputum and breath samples. Species identification (with 16S) and per species genetic potential (with metagenomic shotgun analysis) were studied on sputum samples. The 16S data have been published previously, along with some of the metagenomic data [12, 15, 16].

**Breath analysis**

From each CF patient and a healthy volunteer (in the same room), three breath samples were collected within 5 min of each other. Volunteers were instructed to eat and drink normally until 1 h before sample collection; food and drink in the hours before sampling were recorded. Prior to sample collection, volunteers used saline to rinse their mouths, and volunteers were instructed to discard the first third of their breath (approximately 2 s), collecting the last two-thirds of the breath sample which is less influenced by the contents of the oral cavity. Samples were collected in 1.9L stainless steel canisters [17], along with a simultaneous background room air sample. Sample collection and analysis is described in Whiteson [12], using the approach detailed in Colman [18]. Briefly, breath samples were concentrated in a stainless steel loop submerged in liquid nitrogen, and then heated to ~80°C and split into six different columns. The data shown here were obtained from the DB-5 ms column (J&W; 60 m, 0.25 mm I.D., 0.5 mm film thickness) output to a quadrupole mass selective detector (MSD, HP5973). The absolute intensity information is not the major objective of the results shown here, but rather the relative amounts of the VOCs of interest in relation to the microbial sequence information obtained from parallel samples.

**KEGG**

The KEGG compound database was searched for ethanol (C00469), acetaldehyde (C00084) and methanol (C00132). The enzyme numbers (ec:x.x.x.x) associated with these compounds were extracted and coupled to the orthology database (K-numbers). For each of these K-numbers, the corresponding gene sequences in bacterial species that were previously described in the CF respiratory tract were extracted and then used to create a database.

**16S rDNA analysis**

The 16S rRNA gene amplicon data from these samples have been published, along with the associated methods [15]. Briefly, frozen sputum samples were thawed in the presence of Trizol reagent, DNA was extracted using a Life Technologies protocol, and amplicon sequencing of the V4 region (515F/860R) of the bacterial 16S rRNA gene was carried out at the RTSF Genomics Core at Michigan State University using the MiSeq v2 reagent kit for paired-end 250bp sequencing. Sequence data quality control with Prinseq [25] and a Mothur SOP for MiSeq data [26] were employed to obtain 97% operational taxonomic units (OTUs), a proxy for bacterial species. Data processing and OTU clustering are described in further detail in Quinn et al [15].

**Metagenomic analysis**

Fresh sputum samples were processed using hypotonic lysis and washing to remove eukaryotic cells, damaged bacterial cells and extracellular DNA, as described previously [12, 14, 27]. Ion Torrent sequencing yielded 19.8 million reads, and 13.6 million reads were retained...
after removal of low quality reads or human sequences [25, 28]. Some of these data were presented in previous publications that did not focus on breath analysis [15]. These metagenomic data were then compared to the previously constructed database (using BLASTn searches) to assist with assigning function to these data from CF patients. Metagenomic sequences were selected if they matched ethanol, acetaldehyde or methanol pathway genes from the KEGG with a minimum length of 40 bp, sequence identity of 40% and a BLAST e-value cutoff of \(1 \times 10^{-10}\).

**Data analysis**

The number of sequences in metagenomic data from CF patients that matched sequences in the database was calculated per patient (normalized to the number of sequence reads for that sample), per microbe (on the genus level) and per orthological unit (K-number from KEGG database). Heatmaps were constructed using the ggplot2 package within the R environment via the R-studio interface [29].

**Results**

**Patients**

Seven patients with CF were included from the UCSD Adult CF Clinic. Their ages ranged from 27 to 52 years. All patients were on chronic antibiotic therapy. Three patients received additional antibiotics in the week prior to sampling. The predominant microorganism cultured in the clinical microbiology lab did not often reflect the dominant microbe in the sequence data [12, 15, 30].

**Acetaldehyde**

Acetaldehyde was present in the breath of all included patients. Figure 1 shows the enzymes that are involved in the metabolism of acetaldehyde. The enzymes that convert between acetate and acetaldehyde, and between ethanol and acetaldehyde are bi-directional. Bacteria in the genera *Escherichia*, *Lactococcus*, *Pseudomonas*, *Rothia* and *Streptococcus* were found to have the genetic potential to produce acetaldehyde (figure 2). Only DNA sequences from *Lactococcus* were implicated in the formation of acetaldehyde from acetate through aldehyde dehydrogenase family 9 member A1 (K00149). Most reads associated with threonine aldolase were associated with *Rothia*.

**Ethanol**

Ethanol was also present in the breath of all patients. Figure 1 also shows the enzymes that are associated with ethanol (and vice versa), most enzymes are similar to those described in the section on acetaldehyde. *Escherichia, Pseudomonas, Rothia* and *Streptococcus* were found to have the genetic potential to produce ethanol (figure 3). *Escherichia* was found to be genetically capable of producing ethanol in all patients, whilst there was considerable heterogeneity between patients for the other genera. The ethanol concentration in breath positively correlated with the amount of 16S rRNA...
Figure 2. Genetic capabilities for ethanol production by different bacteria in the sputum samples of CF patients. Heatmaps of VOC concentrations and DNA fragment abundance. The intensity of blue represents the relative concentration/abundance. Upper panel: ethanol concentration in breath. Middle panel: relative abundance of Streptococcus species, Escherichia species, Rothia species and Pseudomonas species. Lower panel: number of DNA fragment copies per microbe, per enzyme (see figure 1), normalized to the number of reads for that sample.

Figure 3. Genetic capabilities for acetaldehyde production by different bacteria in the sputum samples of CF patients. Heatmaps of VOC concentrations and DNA fragment abundance. The intensity of blue represents the relative concentration/abundance. Upper panel: acetaldehyde concentration in breath. Middle panel: relative abundance of Streptococcus species, Escherichia species, Rothia species and Pseudomonas species. Lower panel: number of DNA fragment copies per microbe, per enzyme (see figure 1), normalized to the number of reads for that sample.
copies of *Escherichia* found in sputum (Spearman rho = 0.85, P = 0.015).

**Methanol**

Figure 4 displays the enzymes that are linked to the production of methanol in the KEGG database for which matching sequences were found in the CF sputum samples. *Rothia* showed the most versatile genetic potential for producing methanol. DNA sequences that matched to *Pseudomonas*, *Streptococcus* and *Escherichia* were also linked to methanol production (figure 5).

**Discussion**

The reported data support the hypothesis that bacterial DNA fragments in sputum of CF patients can be linked to gene functionality that would result in the production of VOCs that were also found in the breath of these patients. We found that the microbes present in these samples have the potential to produce ethanol, acetaldehyde and methanol, three VOCs that have been previously linked with airway colonization and pneumonia [19–21]. As a breath research community we should look at the pulmonary microbiome as a potential source for the VOCs we find in exhaled breath.

To our best knowledge, this is the second study to link VOCs in exhaled breath with their microbial producers via metagenomic analysis of the airway microbiome. In an earlier study drawing from the same samples, 2,3-butadione in breath was linked to anaerobic metabolism of *Streptococcus* spp., *Rothia mucilaginosa*, and other acetoin metabolizers. 2,3-butadione was more abundant during exacerbation, and decreased dramatically after antibiotic treatment; it may be a useful marker of exacerbation and of successful treatment in patients with CF [12]. This study further extends those findings to include other VOCs that have been studied more widely as markers of bacterial colonization and infection of the respiratory tract. The role of microbes in the production of these compounds is not new; it has been studied particularly well in the gut and in biogas systems [31, 32]. Excess production of ethanol by dysbiosis of the microbiome in the gut is even recognized as a clinical syndrome called auto-brewery syndrome, which is highly disabling as patients are constantly intoxicated [33]. Local acetaldehyde production by bacteria in the gut has been implicated in the pathogenesis of colon carcinoma [34]. However, no research has been performed on the potential harmful effects of ethanol and acetaldehyde locally in the lung.

This study has several limitations. First, we focused only on three VOCs; acetaldehyde, ethanol and methanol, as a lot is known about their metabolism and therefore we could obtain gene sequences from the bacteria in the respiratory microbiome. These VOCs are of great interest as they are implicated as markers of colonization and/or infection of the respiratory tract [19–21]. However, other markers may be just as, or even more, clinically relevant to study in the future. For that, more knowledge on the metabolic pathways leading to these markers is required. There are also likely to be other, unknown, metabolic pathways that lead to the production of acetaldehyde, ethanol and methanol that are currently not listed in the KEGG database.

Another limitation is our use of breath and sputum from seven patients with CF. The generalizability of the results has to be tested in patients with other respiratory diseases and the validity has to be replicated in larger cohorts of patients. With increasing numbers of included patients, more quantitative analyses might be possible. Furthermore, we focused on bacterial production of VOCs. We cannot exclude that fungi are also
capable of producing the studied metabolites. Indeed, it is suggested in the literature that ethanol and acetaldehyde is produced in vitro by Candida albicans [35–38]. We have also not studied the influence of substrates for the production of ethanol, methanol and acetaldehyde. Ingested ethanol is an obvious source; but this bias was excluded as the subjects did not drink alcohol around sample collection. However, non-alcohol beverages may also contribute [39].

The strengths of these studies are the prospective data collection and the meticulous analysis of the samples. The metagenomic analysis of the intact bacterial DNA allowed for a comprehensive view on the pathways that are involved in the production of the VOCs. Furthermore, while sequencing DNA from a sputum sample in itself does not support the viability of the bacteria identified, as the DNA could be extracellular or come from dead cells, most bacteria identified in the metagenomes from the same samples were found to be viable in a capillary culture model, supporting the possibility that they are capable of producing the VOCs identified here [16].

While Rothia and Streptococcus may be converting ethanol to acetaldehyde, the presented data suggest that Lactococcus, from the order of Lactobacilales and a close cousin to Streptococcus, was the only bacterium in the lung of CF patients that contains annotated genes to metabolize acetaldehyde from acetate, and vice versa. Acetaldehyde leads to a pro-inflammatory, oxidative stress and carcinogenic response. The conversion to acetate protects against this [32, 40]. Acetate is a short chain fatty acid whose production from mucins by oral anaerobes may enable Pseudomonas to derive nutrients from mucins that it otherwise cannot access [47]. Lactococcus is also part of the healthy, normal oral and airways microbiome; the conversion of a toxic molecule (acetaldehyde) to an anti-inflammatory short chain fatty acid (acetate) could be an important role, complicated by the fact that acetate is a nutrient for Pseudomonas. The fact that volatile molecules produced throughout the airways and oropharyngeal cavity can travel means that they can impact the physiology of host and bacterial cells throughout the airways, even coordinating physiological events the way that hormones do in multicellular organisms, with interesting implications for both microbiology and human health.

An important caveat to determining the origin of a particular VOC found in breath is that many molecules can be produced by humans or bacteria, therefore it is not possible to determine whether the molecule has human or bacterial origin. Furthermore, bacterial physiology is heavily influenced by interactions with the host immune system [41]. For example, Pseudomonas and some other anaerobes can use nitrate as an alternative electron acceptor [42], and inflammation results in the recruitment of immune cells that produce nitrate [43–45]. In addition, neutrophils are thought to consume a large fraction of the oxygen available in the CF lung [46], forcing the colonizing microbes to rely on fermentation or anaerobic respiration, which is likely to be a large part of why growth rates in the CF lung are very slow—Staphylococcus aureus doubling times are slow and heterogeneous, averaging about two weeks in a recent study using the incorporation of stable isotope labels into bacteria isolated from CF sputum samples [18]. VOC production profiles of slow growing microbes is likely to differ from typical culture conditions, and emphasizes the importance of considering...
physiologically relevant conditions and combining in vitro and in vivo approaches.

To conclude, bacterial DNA fragments in sputum of CF patients can be linked to enzymes implicated in the production of ethanol, acetaldehyde and methanol, which are VOCs that are predictive of respiratory tract colonization and/or infection. This confirms that the lung microbiome can produce VOCs directly.

Acknowledgement

LB would like to thank the European Respiratory Society for funding the short term fellowship that made this project possible. This independent research was also supported by the Gilead Sciences Research Scholars Program in Cystic Fibrosis to KW. We are also grateful to Dr Doug Conrad, Director of the UCSD Adult CF Clinic, Dr Robert Quinn, Mike Furlan, Dr Yan Wei Lim, and Prof. Forest Rohwer at San Diego State University during the era of the CF Expedition team.

References

[27] Lim Y W, Haynes M, Furlan M, Robertson C E, Harris J K and Rohwer F 2014 Purifying the impure: sequencing metagenomes and metatranscriptomes from complex animal–associated samples J. Vis. Exp. 94 e52117
[34] Homann N, Tillonen J and Salaspuro M 2000 Microbially produced acetaldehyde from ethanol may increase the risk of colon cancer via folate deficiency Int. J. Cancer 86 169–73


[37] Nieminen M T et al 2009 Acetaldehyde production from ethanol and glucose by non-*Candida* albicans yeasts in vitro Oral Oncol. 45 e245–8


[40] Sapkota M and Wyatt T 2015 Alcohol, aldehydes, adducts and airways Biomolecules 5 2987–3008

[41] Scales B S, Dickson R P and Huffnagle G B 2016 A tale of two sites: how inflammation can reshape the microbiomes of the gut and lungs J. Leukoc. Biol. 100 1–8

[42] Line L et al 2014 Physiological levels of nitrate support anoxic growth by denitrification of *Pseudomonas aeruginosa* at growth rates reported in cystic fibrosis lungs and sputum Front. Microbiol. 5 554


