The personalized temporal dynamics of microbiome in the airways of cystic fibrosis patients

Giovanni Bacci[1], Giovanni Taccetti[2], Daniela Dolce[2], Federica Armanini[3], Nicola Segata[3], Francesca Di Cesare[1], Vincenzina Lucidi[4], Ersilia Fiscarelli[4], Patrizia Morelli[5], Rosaria Casciaro[5], Anna Negroni[6], Alessio Mengoni[1], and Annamaria Bevivino[6*]

[1]Department of Biology, University of Florence, Sesto Fiorentino, 50019, Florence, Italy,
[2]Cystic Fibrosis Center, Anna Meyer Children's University Hospital, Department of Pediatrics Medicine, Florence,
50139, Italy
[3]Centre for Integrative Biology, University of Trento, Trento, 38122, Italy
[4]Children's Hospital and Research Institute Bambino Gesù, Rome, 00165, Italy
[5]Cystic Fibrosis Center, IRCCS G. Gaslini Institute, Department of Pediatrics, Genoa, 16147, Italy
[6]Department for Sustainability, Italian National Agency for New Technologies, Energy and Sustainable Economic Development, ENEA, 00123 Rome, Italy

* Corresponding author
Annamaria Bevivino, Department for Sustainability, ENEA, Rome, Italy. Email: annamaria.bevivino@enea.it

Author contribution:
Conceived and designed the experiments: AB VL GT EVF AM. Performed the experiments: GB FDC. Analyzed the data: GB AM AB. Contributed reagents/materials/analysis tools: DD FA PM RS AN. Wrote the paper: GB AM AB. Provided comments and recommendations that improved the manuscript: NS GT VL. Supervised research: AB VL GT EVF AM.

Abstract

Although the taxonomical composition of the cystic fibrosis (CF) lung microbiome has been largely inspected, little is still known about the overall gene content and functional profiles of the resident microbiome. To understand the dynamics of the lung microbiome in relation with patient's disease status, a large cohort of CF subjects with moderate-severe lung disease was followed over a 15-month period. Longitudinal assessment of sputum microbiome by shotgun metagenomics revealed a patient-specific colonization of the primary and emerging CF pathogens. Even if patient genotype and exacerbation events impacted the microbiome diversity, CF microbiota rebounds to pre-treatment state. A core set of antibiotic resistance genes was found although their presence was not affected by antibiotic intake. The microbial resilience and persistence of antibiotic resistant genes support the growing consensus that the management of chronic CF infection may be improved by a more patient-specific personalization of clinical care and treatment.

Introduction

Cystic Fibrosis (CF) is the most common lethal autosomal recessive disease in Caucasians, caused by mutations in the gene coding for cystic fibrosis transmembrane conductance regulator (CFTR) channel[1]. Disruption of chloride anion transport, one of the key underlying features of CF, leads to altered physiological conditions at epithelial surfaces. In the airways, CFTR mutations result in a dehydrated viscous mucus that compromises mucociliary clearance and predisposes CF patients to repeated cycles of airway infection, mucous impaction, and bronchiectasis resulting in the majority of morbidity and mortality in the patient population[2]. In particular, bacterial lung infections reduce life expectancy in most CF patients[3]. The affected individuals consistently maintain high bacterial loads in their airways also during periods of clinical stability that are punctuated by episodes of

pulmonary exacerbation[4]. Such periodic episodes of acute pulmonary exacerbation strongly contribute to the irreversible decline of lung function. Though much is known about the composition of the microbial infections in CF (for a recent review see[5]), the factors leading to such exacerbations are still poorly understood. In the past years, studies employing DNA-based analyses of the airway microbiota of CF patients have shown somewhat discordant results. Indeed, some authors report a largely stable airway microbiota through periods of exacerbation and antibiotic treatment[6], while other indicate of a high inter-patients variability[5,7–9], but also suggested the possibility to identify some microbial taxa as biomarker of exacerbation[10], as well as a role of rare species in exacerbation[11]. Most of these works are targeted metagenomic surveys performed on a variable number of patients and focusing on the 16S rRNA gene sequence. However, this approach offers limited possibilities to infer strain-level and functional (meaning based on functional genes) insights[12]. These two last points are particularly relevant when host-microbiome interactions are studied. Indeed, the overall genetic repertoire of the microbiome (i.e. the entire set of genes in all the genomes of the community members) is the main responsible of the interaction with the host[13]. Recently, the functional interactions among members of a bacterial community have stirred the attention of investigators for relating microbiome functionality to human-microbe interaction[14] and as a perspective for understanding the airway microbiome dynamics in CF[15]. In several human diseases where the microbial infection is an important factor, such as CF, single patients harbor genomically different strains, which ultimately may lead to explain individual differences in clinical outcomes[16–18]. Until now, few longitudinal studies, with a limited number of patients, on CF airway microbiota have been performed[19,20]. Moreover, studies on CF microbiome are few and on a limited number of patients[9,21–23] or specific metabolic functions[24]. Moving away from taxonomic inventories towards a better understanding of the CF microbiome genes opens a new avenue for the identification of the microbial gene repertoire associated with CF lung disease. An ecological

perspective on multispecies and multi-strain colonization of CF airways will permit to understand the role of polymicrobial dynamics in lung disease progression[25] and provide the clinicians with new biomarkers of CF progression and targets for antibiotic therapy.

In this work. we tried to fill the gap of knowledge about the temporal dynamics of the airway microbiome in CF, paying special attention to the episodes of exacerbation, by using a shotgun metagenomic approach[26], that is targeting the entire genomic repertoire of the microbial community, down to the strain level[27,28]. A cohort of 22 patients with moderate-severe lung disease, grouped according to different genotypes ($F508_{hom,}$ homozygote F508; $F508_{het}$ heterozygote F508), was selected and followed over 15 months during which 8 patients underwent exacerbation events. This offered the opportunity to investigate the taxonomic and functional dynamics of the overall microbiome. The main outcome from this study is a highlight on a patient-specific temporal dynamic of the microbiome and a clear resilience, following exacerbation, of the microbiome fraction which includes the main CF pathogens.

Methods

Ethics Statement

The study was approved by the Ethics Committees of Children's Hospital and Research Institute Bambino Gesù (Rome, Italy), Cystic Fibrosis Center, Anna Meyer Children's University Hospital (Florence, Italy) and G. Gaslini Institute (University of Genoa, Genoa, Italy) [Prot. N. 681 CM of November 2, 2012; Prot. N. 85 of February 27, 2014; Prot. N. FCC 2012 Partner 4-IGG of September 18, 2012]. All participants provided written informed consent before the enrollment in the study. All sputum specimens were produced voluntarily. All procedures were performed in agreement with the "Guidelines of the European Convention on Human Rights and Biomedicine for Research in Children" and the Ethics Committee of the three CF Centers involved. All measures

were obtained and processed ensuring patient data protection and confidentiality.

Demographic and clinical characteristics of enrolled patients

Twenty-two adolescents and adults with CF were enrolled in the study between October 2014 and March 2015 (Table 1). The study subjects were selected based on eligibility criteria that included all of the following: (i) a diagnosis of CF, i.e., a sweat test showing sweat Cl > 60 mmol/l and two known CFTR mutations causing the disease with pancreatic insufficiency (elastase< 5 μg/g/feces)[29], (ii) aged more than six years, i.e., between 14 and 55 years, (iii) chronically infected with *Pseudomonas aeruginosa* and iv) decline in FEV_1 in the previous three years before enrollment[30]. Patients were excluded if they were chronically infected with *Burkholderia cepacia* complex. Using these criteria, 22 patients were included in the study for a total of 79 shotgun metagenomic samples. The cohort was enrolled in three Italian Hospital, namely: Bambino Gesù Children's Hospital (Rome, Italy), Giannina Gaslini Children's Hospital (Genoa, Italy) and Meyer Children's Hospital (Florence, Italy). Subjects were treated according to current standards of care with periodical microbiological controls[31] with at least four microbiological controls per year[4]. At each visit, clinical data collection and microbiological status (colonizing pathogens with available cultivation protocols) were performed according to the European CF Society standards of care[32]. Forced expiratory volume in 1 second as a percentage of predicted (%FEV1) is a key outcome of monitoring lung function in CF[33]. FEV1 values were measured according to the American Thoracic Society and European Respiratory Society standards[31]. CFTR genoptyping, sex, age, and antibiotic treatment for each patient were reported in (Table 1 and S1). During serial sampling, data (antibiotic usage and spirometry) were collected.

Sample collection, processing, DNA extraction and sequencing

Sputum samples were obtained by spontaneous expectoration at stable, exacerbation, and post-exacerbation state. Sampled were processed according to standard methods as previously

described[13,34]. Bacterial respiratory pathogens were identified using the conventional techniques reported in the Guidelines, as previously described[34,35]. The number of samples, microbiological status at sampling and samplings following exacerbation events are reported in Table 1. Sputum samples were washed in 5 mls PBS and then centrifuged (3,800 g) for 15 minutes. Resulting pellets were resuspended in 5-10 mls DNAse buffer (10 mM Tris-HCl pH 7.5; 2.5 mM MgCl2; 0.5 mM CaCl2, pH 6.5) with 7.5 ul of DNAse I (2000 Units/ml) per 1 ml of sample (15U/ml final), incubated for 2 hours at 37C, and washed twice by pelleting at 3,800 g for 15 minutes and resuspending in 10 ml SE buffer (75 mM NaCl, 25 mM EDTA, pH 7.5). Pellets were then resuspended in 0.5 ml lysis buffer (20 mM Tris-HCl pH 8.0; 2 mM EDTA pH 8.0; 1% (v/v) Triton; 20 mg/ml Lysozyme final concentration), incubated for 30 minutes at 37C before extracting DNA with the MoBio Powersoil DNA extraction kit as per manufacturer's instructions. Libraries were prepared with Nextera XT kit (Illumina) Sequencing was performed on an Illumina HiSeq2500 apparatus (Illumina). Raw sequence data reported in this study have been deposited in the NCBI "Sequence Read Archive" (SRA) under the project accession PRJNA516870.

Bioinformatic analyses

Sequence quality was ensured by trimming reads using StreamingTrim 1.0[36], with a quality cutoff of 20. Bowtie2[37] was used to screen out human-derived sequences from metagenomic data with the latest version of the human genome available in the NCBI database (GRCh38) as reference. Sequences displaying a concordant alignment (mate pair that aligns with the expected relative mate orientation and with the expected range of distances between mates) against the human genomes were then removed from all subsequent analyses. Metabolic and regulatory patterns were estimated using HUMAnN2[38] and considering only those pathways with a coverage value ≥ 80%, whereas the taxonomic microbial community composition was assessed using MetaPhlAn2[39]. Reads were assembled into contigs using the metaSPAdes microbial assembler[40] with automatic k-mer length

selection. To establish an airway microbiome gene catalog[12] we first removed contigs smaller than 500bp and then used prodigal in Anonymous mode[41], as suggested by the author of the tool, to predict open reading frames (ORFs). Translated protein sequences obtained from assembled contigs were classified using eggNOG mapper against the bactNOG database[42]. Each protein was classified according to its best hit with an e-value lower than 0.001 as suggested in[43]. The CARD database[44] was used in combination to the Resistance Gene Identifier (RGI, version 4.0.3) to inspect the distribution of antibiotic resistance gene (AR genes). Genes predicted within each metagenome were quantified using the number of reads that mapped against metagenomic contigs obtained for each sample. Reads were mapped back to contigs using Bowtie2[37] and the number of reads mapping each ORF was obtained with the bedtools command "multicov" (version 2.26.0). To quantify gene content across different samples, genes were collapsed using the bestOG given by eggNOG mapper by summing together the number of reads that mapped genes with the same annotation. The same approach was used to quantify AR genes predicted with RGI but this time the unique identifier provided by CARD was used to collapse counts.

Strain characterization was performed using StrainPhlAn[27]. Sequence variants for each organism detected were assessed against the MetaPhlAn2[39] marker genes and a tree has been generated including all samples in which the organism was found at least in one time point. One reference genome per organism was downloaded form the RefSeq database and added to the tree.

Taxonomic classification of metagenomic contigs

Assembled contigs were taxonomically classified using BLAST. First, all genomes available for each species detected with MetaPhlAn2 were downloaded from NCBI and used to build a database for each sample. All genomes reporting an identity higher than 90% and a coverage higher than 80% were collected and used for taxonomic classification. Contigs reporting hits with genomes coming

168 from a single species were assigned to that species whereas contigs reporting hits from multiple
169 species were flagged as unknown.

170 **Statistical analyses**

171 Statistical analyses were performed in R[45] version 3.4.4. The taxonomical and functional
172 composition on lung microbiome was explored using permutational multivariate analysis of variance
173 (PERMANOVA with 1000 permutations), 'adonis2' function of vegan package version 2.5-2;
174 whereas differences in bacterial diversity were tested using analysis of covariance (ANCOVA),
175 'aov' function. The model fitted for both analyses was:

176 $$X \sim Exacerbation + Genotype + Subject + FEV_1 + days$$

177 where, Exacerbation is the exacerbation event, Genotype is the CFTR genotype, Subject is the
178 patient, FEV_1 was the forced expiatory volume in 1 second, and days, was the number of days from
179 the enrollment in the study. For the ANCOVA analyses Tukey's post hoc tests were performed to
180 test for mean differences within each factor used to build the full model (excluding FEV_1 value and
181 days since they were not categorical variable). Ordination analyses were conducted on both taxa and
182 pathways using the function 'ordinate' of the phyloseq package (version 1.23.1) with principle
183 coordinate decomposition method (PCoA) and the Bray-Curtis dissimilarity index. The same index
184 was used to inspect the distribution of samples and compare beta diversity level in bot taxonomic
185 composition and pathways.

186 To test for differentially distributed pathways and taxa across exacerbation events and genotypes we
187 used a moderated t-test as implemented in the limma package[46], version 3.34.9. Data obtained with
188 MetaPhlAn2 (taxonomic composition) and HUMAnN2 (pathway composition) were fitted into
189 limma's model using subjects as blocking variable. Since both software quantify biological units

using relative counts (HUMAnN2 uses "copies per million" and MetaPhlAn2 uses percentages) we transformed this data into logarithmic values using the formula: $\log_2(x + 0.1)$, where x are the relative counts. Obtained p-values were corrected using the Benjamini-Hochberg correction method. A similar approach has been used for antibiotic genes detect along assembled contigs. Here the number of reads that mapped onto each gene was used to estimate differentially abundant gene. Since the number of reads for each sample was variable (the ratio of the largest library size to the smallest was more than 10-fold) we used limma's voom method[47] to fit our model, as suggested by the author of limma.

Results

Population and sampling

Twenty-two patients with CF were enrolled for a total of 15 females and seven males. The patients were chosen from a larger cohort of patients with moderate-severe lung disease ($30 < FEV_1 < 70$) and chronically infected by *Pseudomonas aeruginosa*. During the study period, they were treated with maintenance antibiotics (aerosol) and only a subset (n = 8) received clinical intervention in form of supplementary antibiotics (oral or/and intravenous) for a pulmonary exacerbation (CFPE) (Table S1). The bacterial microbiome was investigated on sputum samples obtained every 3-4 months from 22 individuals along a survey of 15 months. Within the 22 subjects monitored, 8 underwent episodes of exacerbations, which provided the opportunity to explore the microbiomes composition along the events. In total, 79 samples from these 22 subjects were collected and analyzed by a whole metagenomic sequencing approach.

Airway microbiomes are taxonomically distinct and show patient-specific strain colonization

The overall taxonomic representation of the microbiomes from the 79 samples is reported in Fig. 1a and 1b, whereas a summary of obtained reads per sample was reported in Table S2. Firmicutes,

Proteobacteria, Bacteroidetes, and Actinobacteria were the most represented phyla. A massive presence of the "classical" CF bacterial signatures (taxa), such as *Staphylococcus aureus*, *Rothia mucillaginosa*, *Pseudomonas aeruginosa*, and *Prevotella melaninogenica* (all present in the top-10 species within each phylum, Fig. 1b), was found. These species, indeed, represent the 49% of all detected taxa as reported in Table S3.

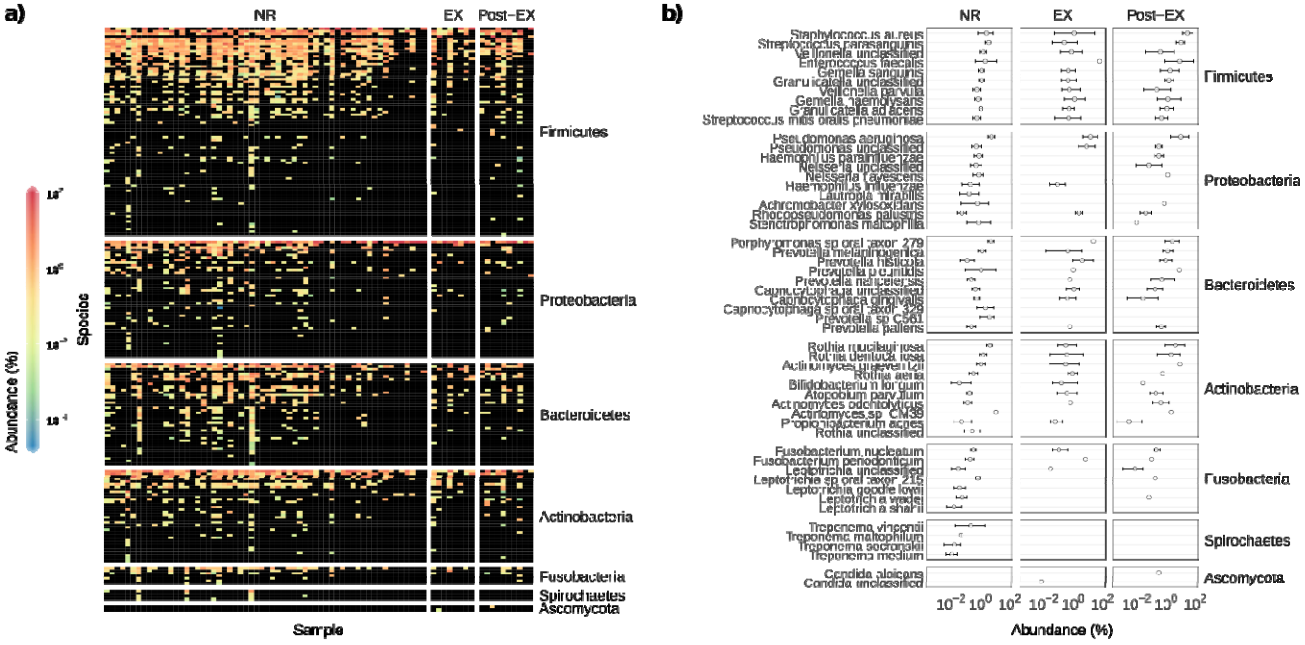


Figure 1: Taxonomic distribution in patients enrolled in the study. a) The taxonomic distribution of all species detected using MetaPhlAn2 was reported in each row of the matrix whereas columns represent samples collected during the study. Colors from dark blue to red were used to report "copies per million" (CPM) values as obtained from HUMAnN2 with black reporting a CPM value of zero. The plot was divided according to exacerbation event (EX), normal status (NR) or post-exacerbation samples (Post-EX) defined as the first samples collected after an exacerbation event has been occurred. Species were ordered according to their mean abundance and grouped according to their Phylum. b) The mean abundance value of the top-ten species (if available) detected within each Phylum was reported together with the standard error.

Although samples can be hardly clustered based on exacerbation event and/or genotype (Fig. 2a), the PERMANOVA analysis reported a significant effect (p-values < 0.05) of both factors. However,

the R^2 values, namely the proportion of variance explained by the factor considered, were very low (0.03 for both factors). The interaction effect between exacerbation event and genotype was not significant (p-value > 0.05), meaning that different genotypes did not influence the lung microbiome during exacerbation events and vice versa. The predominant effect observed was the subject effect (p-value < 0.05), reporting a R^2 value of 0.52, indicating that a high fraction (more than 50%) of the total variance can be explained by subject (patient) individuality. Both FEV_1 and time did not show any significant effect (p-values > 0.05, Table 2).

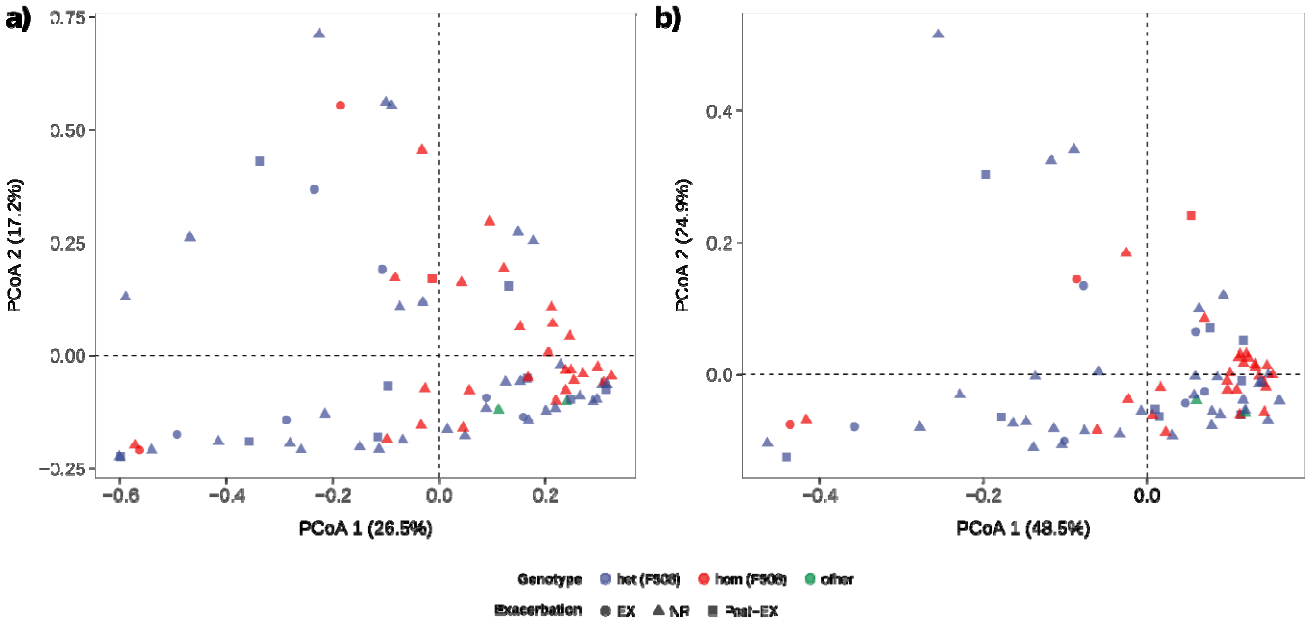


Figure 2: Ordination analyses based on a) taxonomic assignments and b) pathway distribution detected with MetaPhlAn2 and HUMAnN2, respectively. Ordination analyses were conducted using the Bray-Curtis dissimilarity index and ordered following the principle coordinate decomposition method (PCoA). The percentage of variance explained by each coordinate was reported between round brackets.

The strain-level analysis conducted on both the main CF signatures and on the overall biodiversity revealed that samples from the same patients tightly clustered together, confirming a high patient-specific colonization by strains of the above-mentioned species (Fig. 3 and Fig. S1).

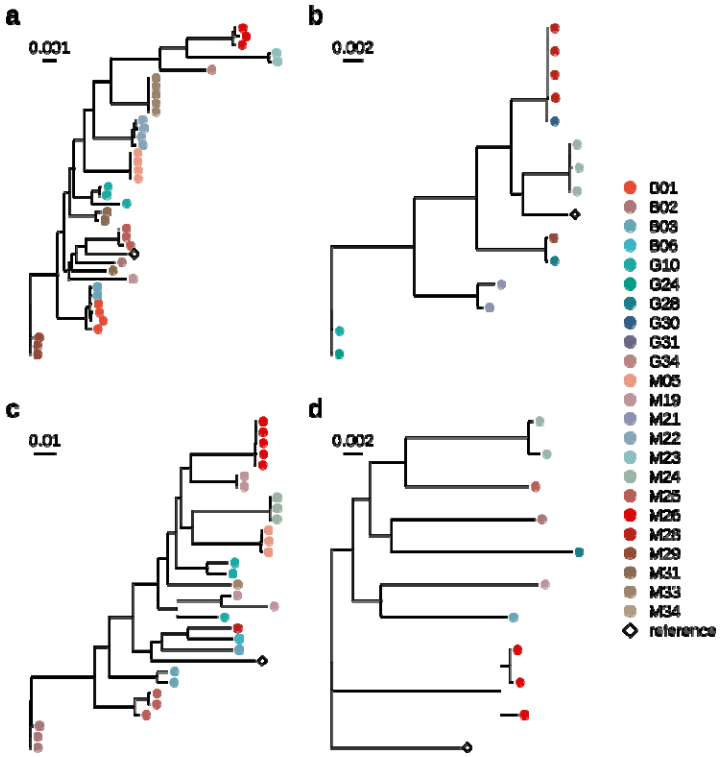


Figure 3: Strain-level phylogenetic trees of the main CF pathogens detected in the study. Phylogenetic trees obtained through StrainPhlAn pipeline were reported for the main pathogenic signatures of CF disease: a) *Pseudomonas aeruginosa*; b) *Staphylococcus aureus*; c) *Rothia mucilaginosa*; d) *Prevotella melaninogenica*. Patients were reported using different colors as specified in the figure legend.

Alpha diversity analysis confirmed the overall picture of results mentioned above. Different values of bacterial diversity were found according to exacerbation events, genotypes, and subjects (Fig. 4a, Fig. S2, and Table S4). Samples collected during exacerbation events reported a lower biodiversity than samples collected during normal visits, highlighting the role of clinical treatments in perturbing CF lung communities as confirmed by the Tukey's post hoc test (Table S5).

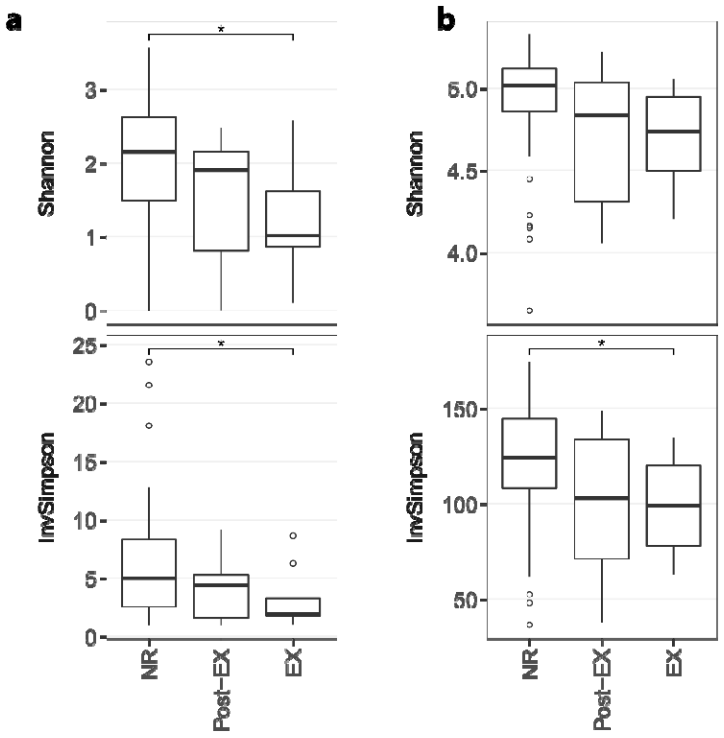


Figure 4: Differences across exacerbation events. The effect of an exacerbation event on alpha diversity was inspected using both the Shannon index and the inverse Simpson index. Diversity indexes were computed for both a) taxonomic signature and b) metabolic pathways.

Airway microbiomes are functionally consistent and show subject-specific distribution patterns

Similar results as those reported above were obtained considering the pathway distribution. Indeed, the PERMANOVA analysis (Table 2) confirmed the effect of exacerbation events and genotypes in shaping the pathway distribution of CF lung microbiome (R^2 values of 0.04 and 0.03 respectively), though less marked than the subject-specific effect (R^2 = 0.48). The sample distribution according to the ordination analysis (PCoA) was very heterogeneous with no sharp differences according to genotypes or exacerbation events. Even here, alpha diversity analyses reported a significant drop of diversity in samples collected during exacerbation events, but the drop was significant only considering the inverse Simpson index (Fig. 4b and Table S4). Overall, the pathway distribution was

more consistent with respect to the taxonomic one, with biosynthetic pathways being the most represented functional category (Fig. 5, Fig. S2, and Table S5). Pathways were mainly detected in members of Firmicutes and Proteobacteria phyla, though Bacteroidetes and Actinobacteria were quite well represented. Even if these results confirmed the results from the analysis of the taxonomic distribution, metabolic pathways showed a more consistent distribution across samples. Indeed, the beta-diversity analysis on both taxonomic and functional distribution showed a lower similarity based on taxonomy in respect with pathways (Table S6, Fig. 6a and 6b). These results were additionally confirmed by the differential abundance analysis. For contrasts made within each genotype, only 40 pathways reported significant differences across exacerbation statuses (p-values < 0.05 and |log(fold-change)| > 5) all in the homozygote group (Fig. S3, Table S7), whereas, considering all samples together, no pathway was found to be more abundant in one condition in respect with another (data not shown). These results confirmed the extraordinary resilience of the CF microbiome even from a functional perspective.

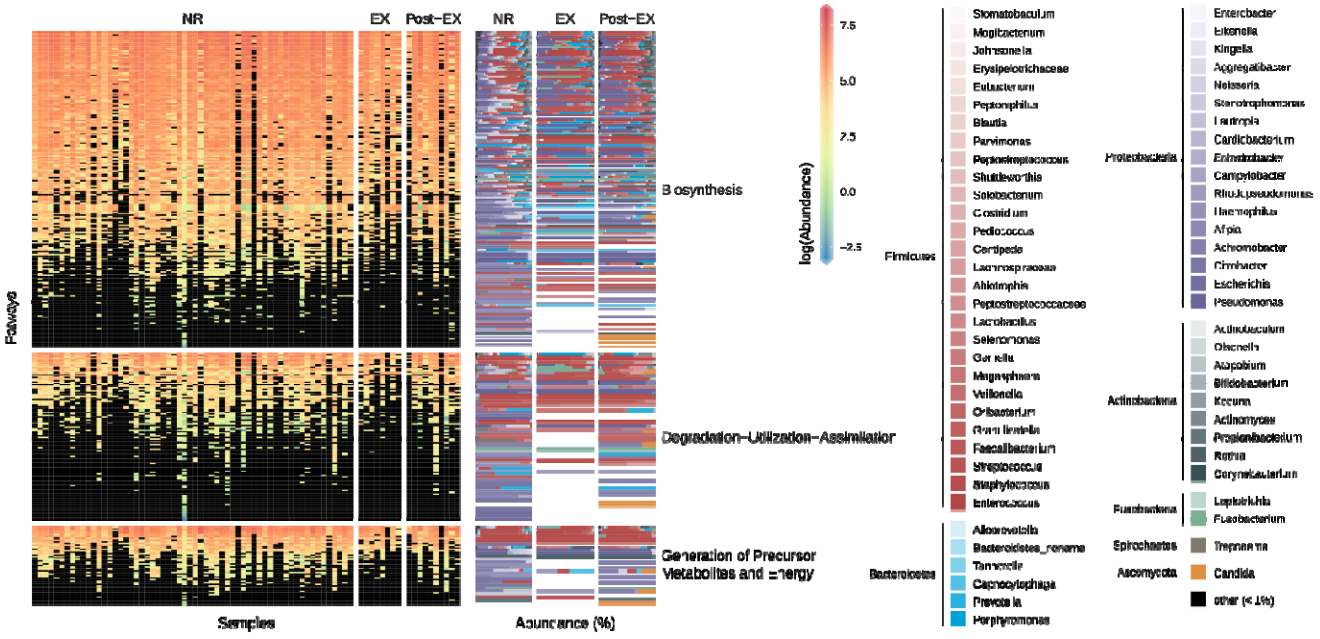


Figure 5: Pathway distribution according to exacerbation events. The pathway distribution was reported for each sample (columns) and for each pathway detected (rows). Colors from dark blue to

red were used to report "copies per million" (CPM) values as obtained from HUMAnN2 with black reporting a CPM value of zero. On the left, the percentage of taxa in which each pathway was detected was reported using different colors. The main colors correspond to the Phylum whereas the different shades correspond to the genus detected (if available).

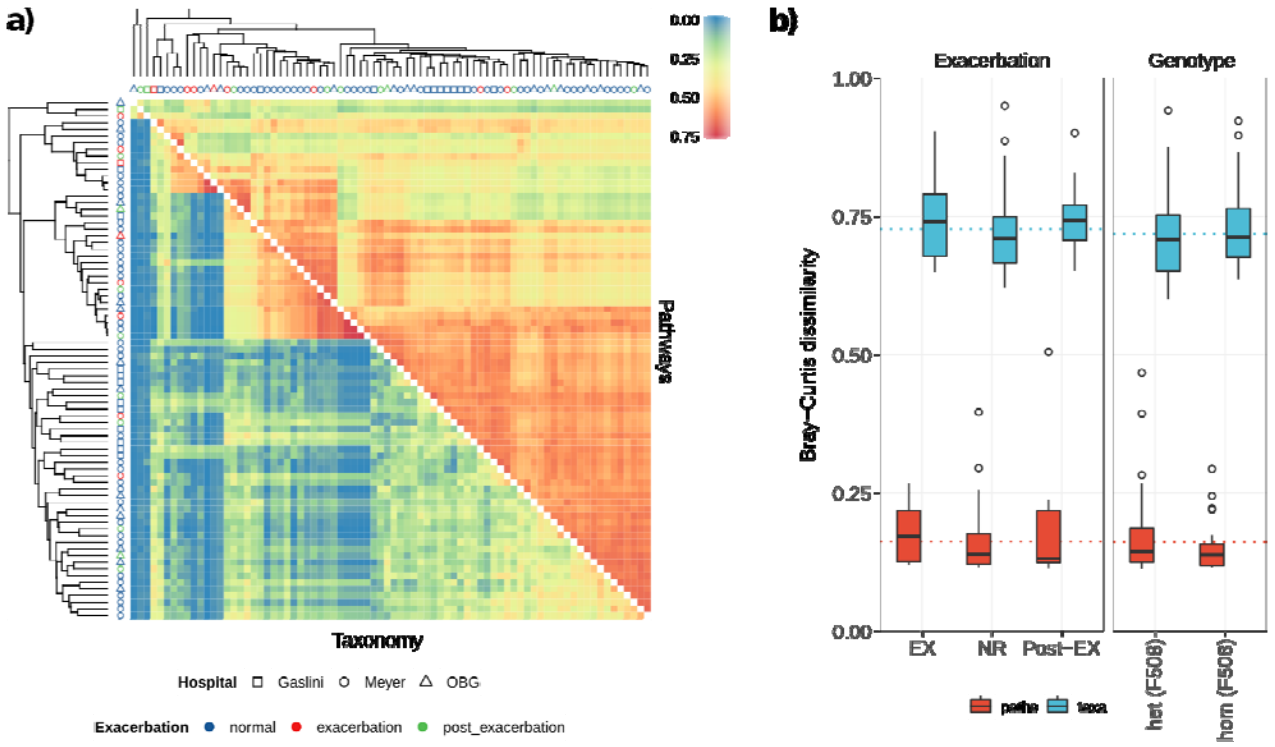


Figure 6: Beta diversity analysis on both taxonomic and functional distribution. a) Hierarchical clustering based on UPGMA method. Clustering was performed on both pathway distribution (the upped triangle) and taxonomic composition of samples (lower triangle). The Bray-Curtis distance was used to compute distances between samples, but it was transformed into similarity value by subtracting 1 before plotting. Thus, red colors represent high similarity values whereas blue colors represent low similarity values. The shape of the points on each tip of trees refers to the hospital whereas the colors refer to the exacerbation events. b) Results of Tukey's post hoc test on beta diversity values across patient genotypes and exacerbation events. Contrasts were computed even to test differences between taxonomic distribution and pathways with taxa reporting higher level of beta diversity.

Antibiotic resistance genes through exacerbation events and treatments

Similar to the pathway analysis reported above, antibiotic resistance genes (ARG) were inspected in relation to exacerbation events. Only six genes were found to be affected by an exacerbation condition, all regarding samples form F508 heterozygote patients whereas, as found for metabolic pathways, no gene was significantly impacted by antibiotic treatment when considering all samples at once (Fig. S4 and Table S8). A similar approach was used to inspect the effect of antibiotic treatment on ARG distribution. ARG were inspected in relation to the antibiotic treatments reported in Table S1. The class of each antibiotic was correlated to the presence (and the abundance) of genes that may, in principle, confer resistance to antibiotics from the corresponding class. Differential abundance analyses were performed for each classes of antibiotics that was used in this study and results obtained were reported in Fig. S5 and Table S9. Only 11 genes were found to be affected by antibiotic intake in different ways. Indeed, 8 out of 11 reported a reduction of abundance during the treatment whereas the remaining 3 reported an increased abundance in respect with antibiotic intake. Results obtained confirmed the high resilience of the gene composition of CF lung microbiome. A highly variable composition along time passing from patients to patients was found (Fig. S2). The presence of ARGs coupled with antibiotic intake was also explored. Results showed that the antibiotic resistance classes of each gene corresponded to the antibiotic treatment used in each sample reporting a big block of ARGs that were present in most of the sample considered (Fig. S6 and Fig. S7).

Discussions

Longitudinal studies allow to provide important clues on stability and dynamics of microbial ecosystems[48]. As all biotic communities, microbial communities tend to evolve towards a stable composition, either in natural environment and in association with host (as human-associated

microbiomes). Changes in the community can be triggered by external conditions, as changes in host physiology (e.g. inflammation status) and/or other perturbations (e.g. antibiotic treatment). Indeed, perturbation studies help to probe community dynamics and resilience and possibly discover new findings for accessing ways for modifying the microbiome[49,50]. Although patients with CF experiments repeated episodes of pulmonary exacerbations during their lives, a broadly accepted definition of these events is still missing[4]. Here, we have investigated the temporal dynamics of CF airway microbiome by using shotgun metagenomics posing attention on exacerbation events which usually bring to an acute decrease in lung function and an increase in respiratory symptoms (such as: increased cough, sputum production, and shortness of breath). Key questions were i) what was the composition and stability of the lung microbiome in patients with CF when longitudinally sampled at stable and exacerbation events; and ii) if the clinical status influenced the metabolic repertoire and the AR gene composition of lung bacterial community. Our results describe a unique examination of the dynamic of the lung microbiome in patients with moderate-severe lung disease carrying the F508del mutation and containing clinical measurements over a 15-month period.,

The lung microbiome of CF patients seems to be a highly patient-specific environment which can be directly conditioned by the host and its habits. Indeed, there was less variation within the same individual at different time points than between different individuals at the same time point, proving some degree of temporal stability of an individual's lung microbiome. This last point agrees with the lack of a time effect on the taxonomic distribution of microbiome. The predominant taxa that colonized the lung of CF patients showed an extraordinary resilience, as witnessed by the presence of the same strains during the whole period of infection. These results agree with previous observations based on 16S rRNA gene profiling, though these studies failed to report a strain-specific overview of the whole dynamic due to the limitations intrinsic to the approach[6,8,11]. Carmody and colleagues showed a relatively stable lung community that may be altered during

period of exacerbation even in the absence of viral infection or antibiotic only in a small group of patients[10]. Even in other pulmonary diseases, such as non-cystic fibrosis bronchiectasis, lung bacterial communities showed a conserved structure for long period of time, as showed in the work by Cox and colleagues where patients were followed for a six-month period[8]. A similar result was shown in the work from Fodor and colleagues[6] where, though occasional short-term compositional changes in the airway microbiota were found, the main taxonomic signatures of CF disease were highly stable.

The antibiotic treatment used did not seem to alter this micro-environment for long period of time since most of the main taxa linked to CF infection are still present even after exacerbation events that are usually handled by a massive amount of antibiotic. From a taxonomic perspective, samples coming from the same patient clustered together highlighting the role of the host in bacterial strain selection during the baseline but even during (and after) exacerbation events. Despite this patient-specific colonization, the taxonomic composition was very different from one subject to another event if sampled at the same time point.

On the other hand, pathways reported a more homogeneous distribution across patients. This high conservation could be related to the characteristic of the lung environment itself, such as mucus compositions, nutrient availability, and oxygen levels, which can be broadly similar across patients with a similar clinical status. This, is in line with the finding that the function of a biotic community is more conserved than the presence of single members[51]. In fact, though the lung microbiome in our study was populated with a relatively large set of microorganisms, the main functions detected are similar across all patients. From this point of view the airway microbiome can be considered as performing a similar “ecosystem service”, irrespective of the taxonomy present as pointed out by various authors in other environments[51–53]. The finding that CFTR genotypes a different representation in some pathways, may suggest that the airways microbiome is influenced by the type

of CFTR alteration. However, this hypothesis deserves further attention to clarify the specific role of microbial pathways with respect to CFTR genotype. Pathogenic bacteria, such as *Pseudomonas aeruginosa*, need to colonize human tissues to grow and in this sense, even pathway that could be related to a worsening of clinical conditions or that could be targeted by antibiotic molecules will be part of this core set of functions. Despite a clear effect of antibiotic treatment during (and after) exacerbation periods, the community structure is always recovered with the main pathogenic taxa emerging again. This effect is confirmed by the correlation of ARG distribution and antibiotic intake. Patients subjected to a given antibiotic treatment did not seem to select bacteria resistant to the antibiotic used but the detection of a particular mechanism seems to be distributed in almost all patients regardless of the treatment. An evidence of functional stability of the lung microbiota was previously reported in other works not concerning CF disease[54,55]. Both works focused their attention on the gut microbiome of obese and healthy individuals (human and mouse) reporting a considerable metabolic redundancy. This high degree of redundancy in the gut microbiome supports a more ecological view where subjects can be considered as different ecological niches all inhabited by unique collections of microbial phylotypes but all sharing the same set of genes. This concept can be extended to the lung microbiome where it is possible to define a core set of features only at the level of metabolic functions. This functional conservation may thus be needed by the whole community and patients can be seen as multiple micro-environments inhabited by a peculiar set of strains, which share the same functions. This work represents a step forward toward a patient-specific interpretation of CF microbiology.

Acknowledgments

This study was supported by Italian Grants funded by the Italian Cystic Fibrosis Research Foundation (FFC) (http://www.fibrosicisticaricerca.it/) to Annamaria Bevivino: Project n.

FFC#14/2015 with the contribution of “Delegazione FFC di Latina”, “Latteria Montello Nonno Nanni”, and “Gruppo di Sostegno FFC Valle Scrivia Alessandria”), and Project n. FFC#19/2017 with the contribution of “Delegazione FFC Lago di Garda con i Gruppi di Sostegno FFC di Chivasso, dell'Isola Bergamasca, di Arezzo”. Giovanni Bacci was supported by a postdoctoral fellowship by Italian Cystic Fibrosis Research Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors greatly acknowledge the Italian Cystic Fibrosis ResearchFoundation (FCC) for its support and administrative tasks, and Ricciotti Gabriella, Tuccio Vanessa and Campana Silvia for their technical support.

Conflict of Interest

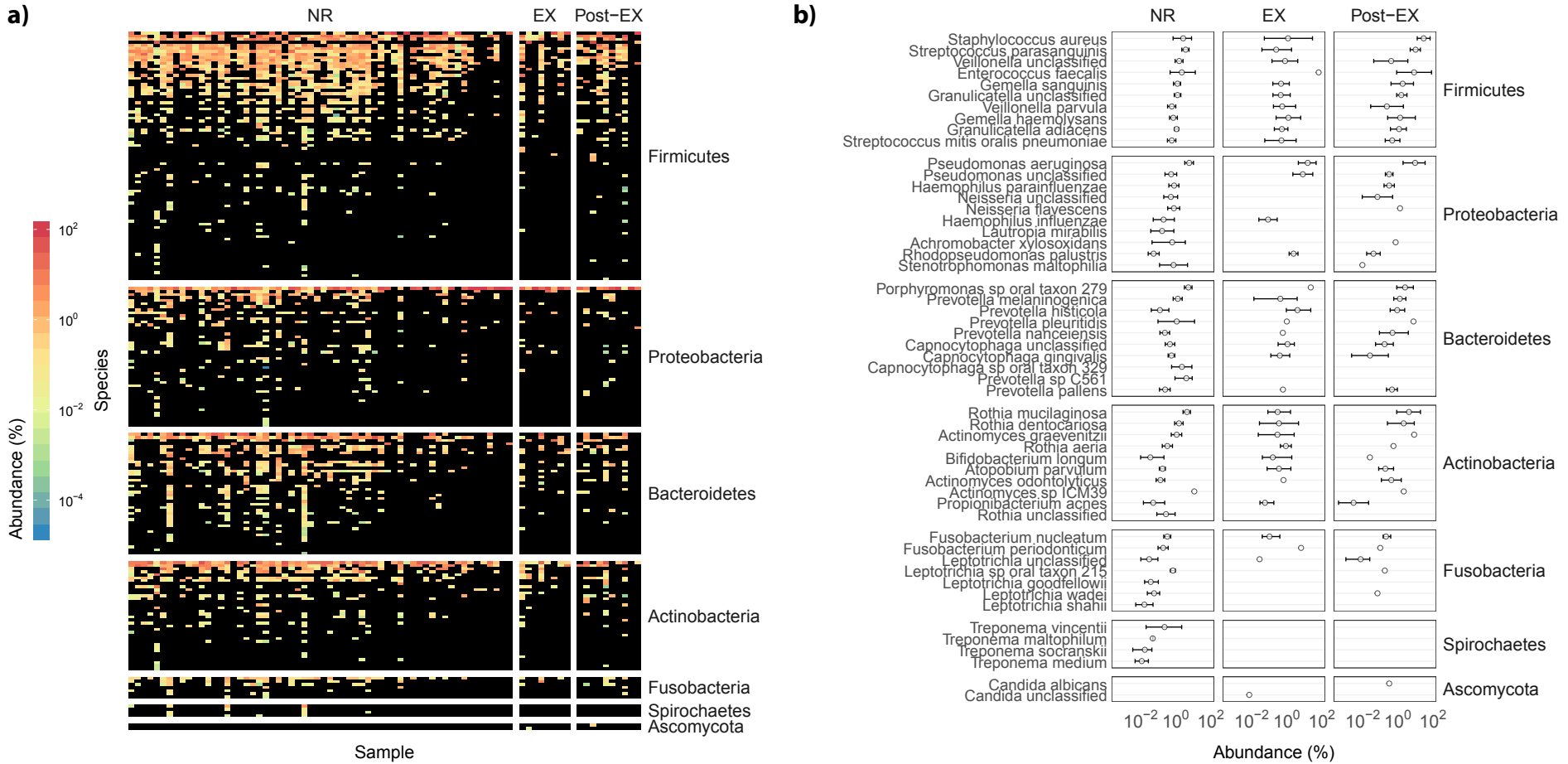
We have no conflict of interest to declare.

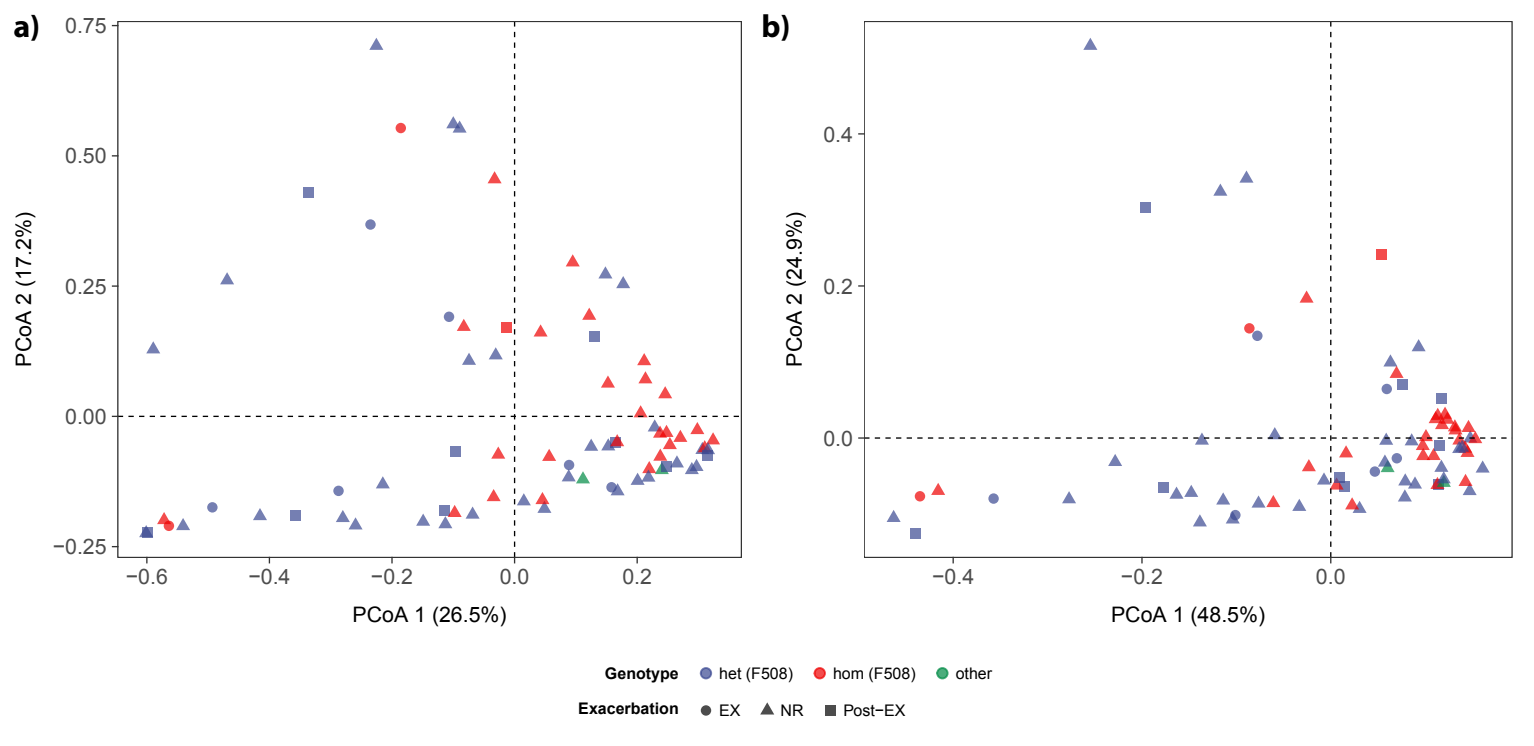
References

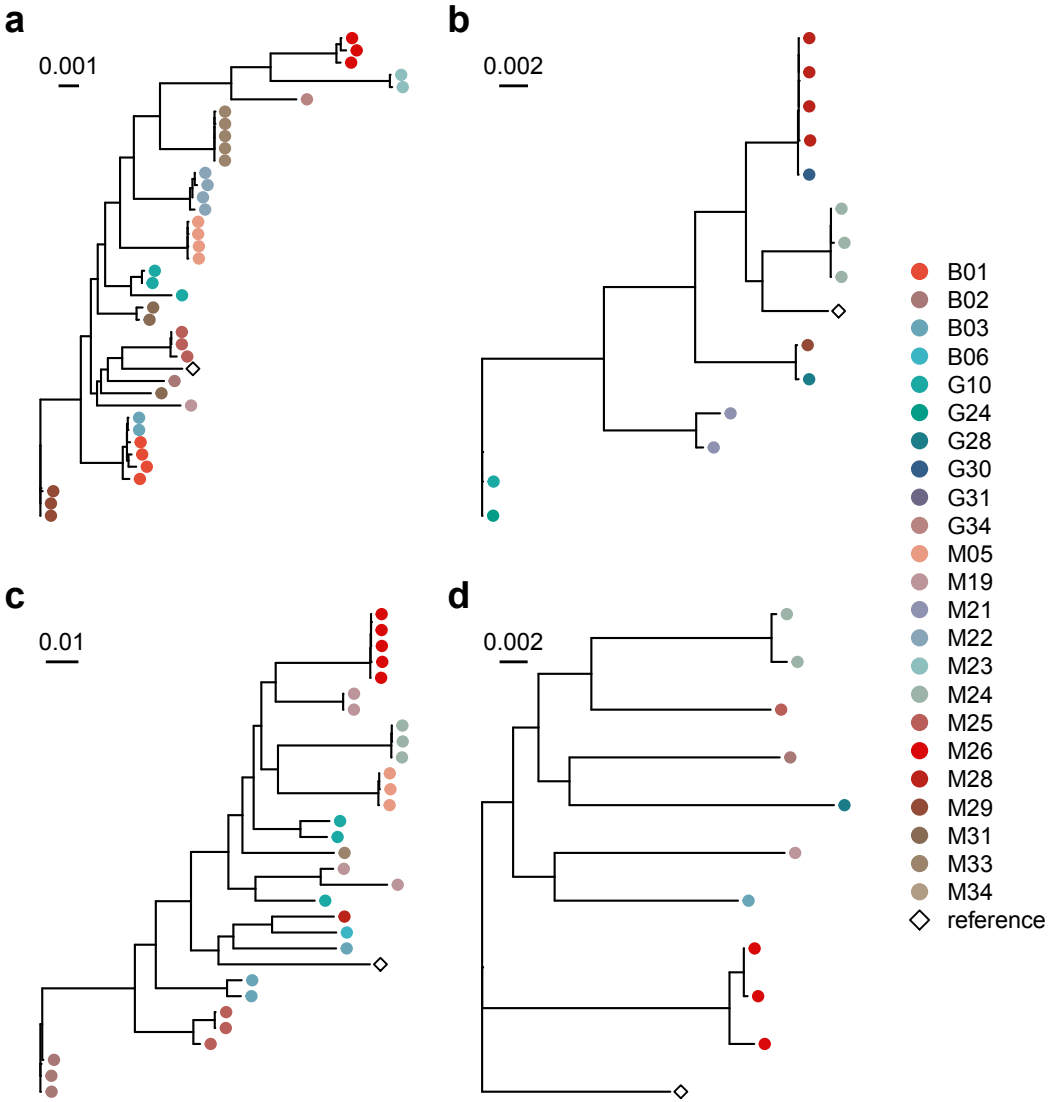
1. Ratjen, F. & Döring, G. Cystic fibrosis. *Lancet* **361,** 681–689 (2003).
2. Gibson, R. L., Burns, J. L. & Ramsey, B. W. Pathophysiology and Management of Pulmonary Infections in Cystic Fibrosis. *Am. J. Respir. Crit. Care Med.* **168,** 918–951 (2003).
3. Knapp, E. A. *et al.* The Cystic Fibrosis Foundation Patient Registry. Design and methods of a national observational disease registry. *Ann. Am. Thorac. Soc.* **13,** 1173–1179 (2016).
4. Flume, P. A. *et al.* Cystic fibrosis pulmonary guidelines: Treatment of pulmonary exacerbations. *Am. J. Respir. Crit. Care Med.* **180,** 802–808 (2009).
5. Huang, Y. J. & LiPuma, J. J. The Microbiome in Cystic Fibrosis. *Clin. Chest Med.* **37,** 59–67 (2016).
6. Fodor, A. A. *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* **7,** e45001 (2012).
7. Price, K. E. *et al.* Unique microbial communities persist in individual cystic fibrosis patients throughout a clinical exacerbation. *Microbiome* **1,** 27 (2013).
8. Cox, M. J. *et al.* Longitudinal assessment of sputum microbiome by sequencing of the 16S rRNA gene in non-cystic fibrosis bronchiectasis patients. *PLoS One* **12,** 1–17 (2017).
9. Whelan, F. J. *et al.* Longitudinal sampling of the lung microbiota in individuals with cystic fibrosis. *PLoS One* **12,** e0172811 (2017).
10. Carmody, L. A. *et al.* Changes in Cystic Fibrosis Airway Microbiota at Pulmonary Exacerbation. *Ann. Am. Thorac. Soc.* **10,** 179–187 (2013).
11. Carmody, L. A. *et al.* The daily dynamics of cystic fibrosis airway microbiota during clinical stability and at exacerbation. *Microbiome* **3,** 12 (2015).
12. Qin, J. *et al.* A human gut microbial gene catalogue established by metagenomic sequencing.

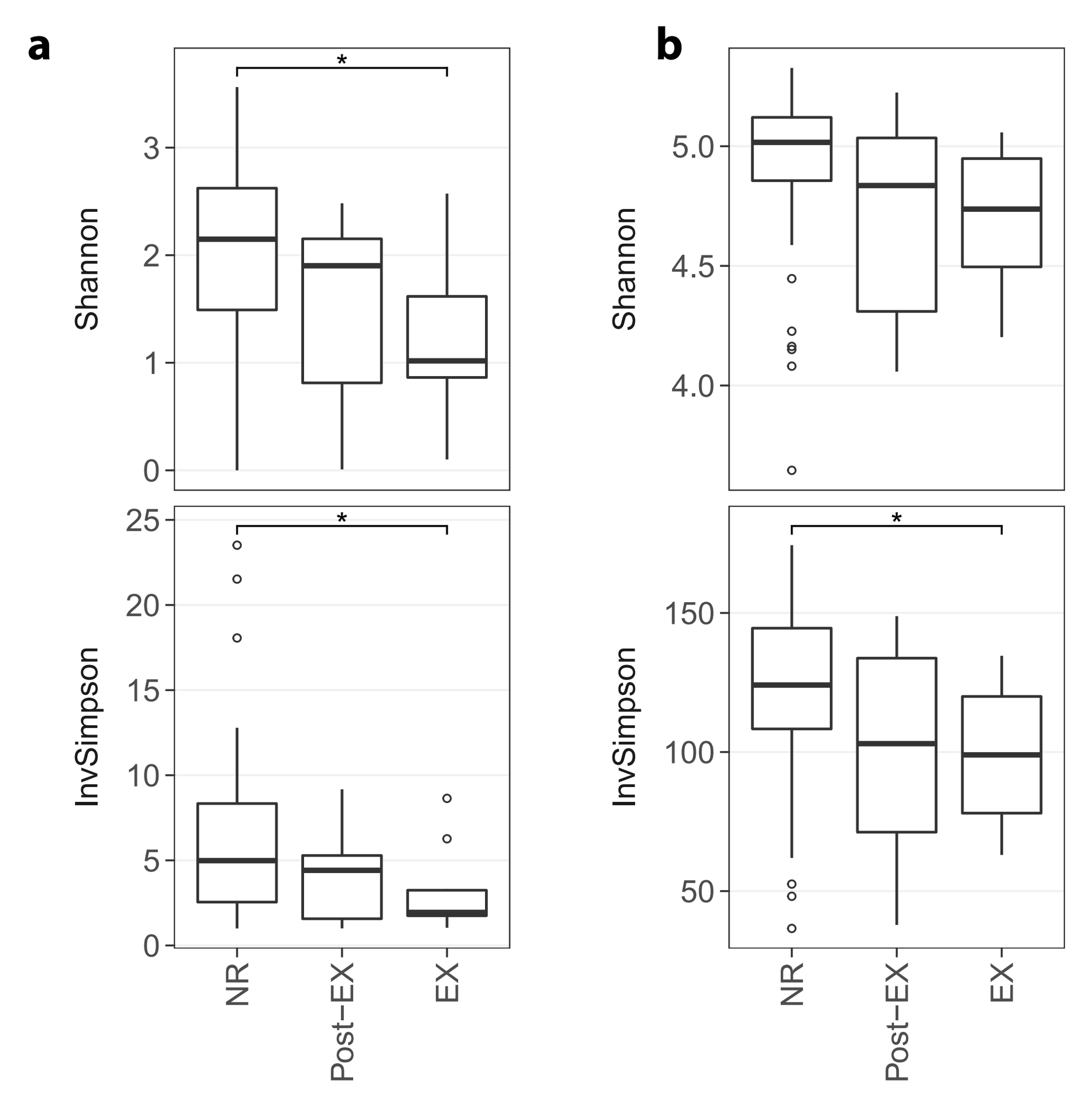
Nature **464,** 59–65 (2010).
13. Bacci, G. *et al.* A different microbiome gene repertoire in the airways of cystic fibrosis patients with severe lung disease. *Int. J. Mol. Sci.* **18,** 1654 (2017).
14. Franzosa, E. A. *et al.* Relating the metatranscriptome and metagenome of the human gut. *Proc. Natl. Acad. Sci.* **111,** E2329–E2338 (2014).
15. Lim, Y. W. *et al.* Metagenomics and metatranscriptomics: Windows on CF-associated viral and microbial communities. *J. Cyst. Fibros.* **12,** 154–164 (2013).
16. Segata, N. On the Road to Strain-Resolved Comparative Metagenomics. *mSystems* **3,** e00190-17 (2018).
17. Beghini, F. *et al.* Large-scale comparative metagenomics of Blastocystis, a common member of the human gut microbiome. *ISME J.* **11,** 2848–2863 (2017).
18. Zolfo, M., Tett, A., Jousson, O., Donati, C. & Segata, N. MetaMLST: Multi-locus strain-level bacterial typing from metagenomic samples. *Nucleic Acids Res.* **45,** e7 (2017).
19. Zhao, J. *et al.* Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc. Natl. Acad. Sci.* **109,** 5809–5814 (2012).
20. Delhaes, L. *et al.* The airway microbiota in cystic fibrosis: A complex fungal and bacterial community-implications for therapeutic management. *PLoS One* **7,** e36313 (2012).
21. Willner, D. *et al.* Comparison of DNA extraction methods for microbial community profiling with an application to pediatric bronchoalveolar lavage samples. *PLoS One* **7,** e34605 (2012).
22. Lim, Y. W. *et al.* Clinical insights from metagenomic analysis of sputum samples from patients with cystic fibrosis. *J. Clin. Microbiol.* **52,** 425–437 (2014).
23. Willner, D. *et al.* Case studies of the spatial heterogeneity of DNA viruses in the cystic fibrosis lung. *Am. J. Respir. Cell Mol. Biol.* **46,** 127–31 (2012).
24. Whiteson, K. L. *et al.* Breath gas metabolites and bacterial metagenomes from cystic fibrosis airways indicate active pH neutral 2,3-butanedione fermentation. *ISME J.* **8,** 1247–58 (2014).
25. Surette, M. G. The cystic fibrosis lung microbiome. *Ann. Am. Thorac. Soc.* **11,** 61–65 (2014).
26. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. *Nat. Biotechnol.* **35,** 833–844 (2017).
27. Truong, D. T., Tett, A., Pasolli, E., Huttenhower, C. & Segata, N. Microbial strain-level population structure and genetic diversity from metagenomes. 626–638 (2017). doi:10.1101/gr.216242.116.Freely
28. Scholz, M. *et al.* Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nat. Methods* **13,** 435–438 (2016).
29. Farrell, P. M. *et al.* Diagnosis of Cystic Fibrosis: Consensus Guidelines from the Cystic Fibrosis Foundation. *J. Pediatr.* **181,** S4–S15 (2017).
30. Bacci, G. *et al.* Pyrosequencing Unveils Cystic Fibrosis Lung Microbiome Differences Associated with a Severe Lung Function Decline. *PLoS One* **11,** e0156807 (2016).
31. Kerem, E. *et al.* Standards of care for patients with cystic fibrosis: a European consensus. *J. Cyst. Fibros.* **4,** 7–26 (2005).
32. Smyth, A. R. *et al.* European cystic fibrosis society standards of care: Best practice guidelines. *J. Cyst. Fibros.* **13,** S30–S53 (2014).
33. Taylor-Robinson, D. *et al.* Understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: a longitudinal study. *Thorax* **67,** 860–866 (2012).
34. Paganin, P. *et al.* Changes in Cystic Fibrosis Airway Microbial Community Associated with a Severe Decline in Lung Function. *PLoS One* **10,** e0124348 (2015).
35. Farrell, P. M. *et al.* Guidelines for diagnosis of cystic fibrosis in newborns through older adults: Cystic Fibrosis Foundation consensus report. *J Pediatr* **153,** S4–S14 (2008).

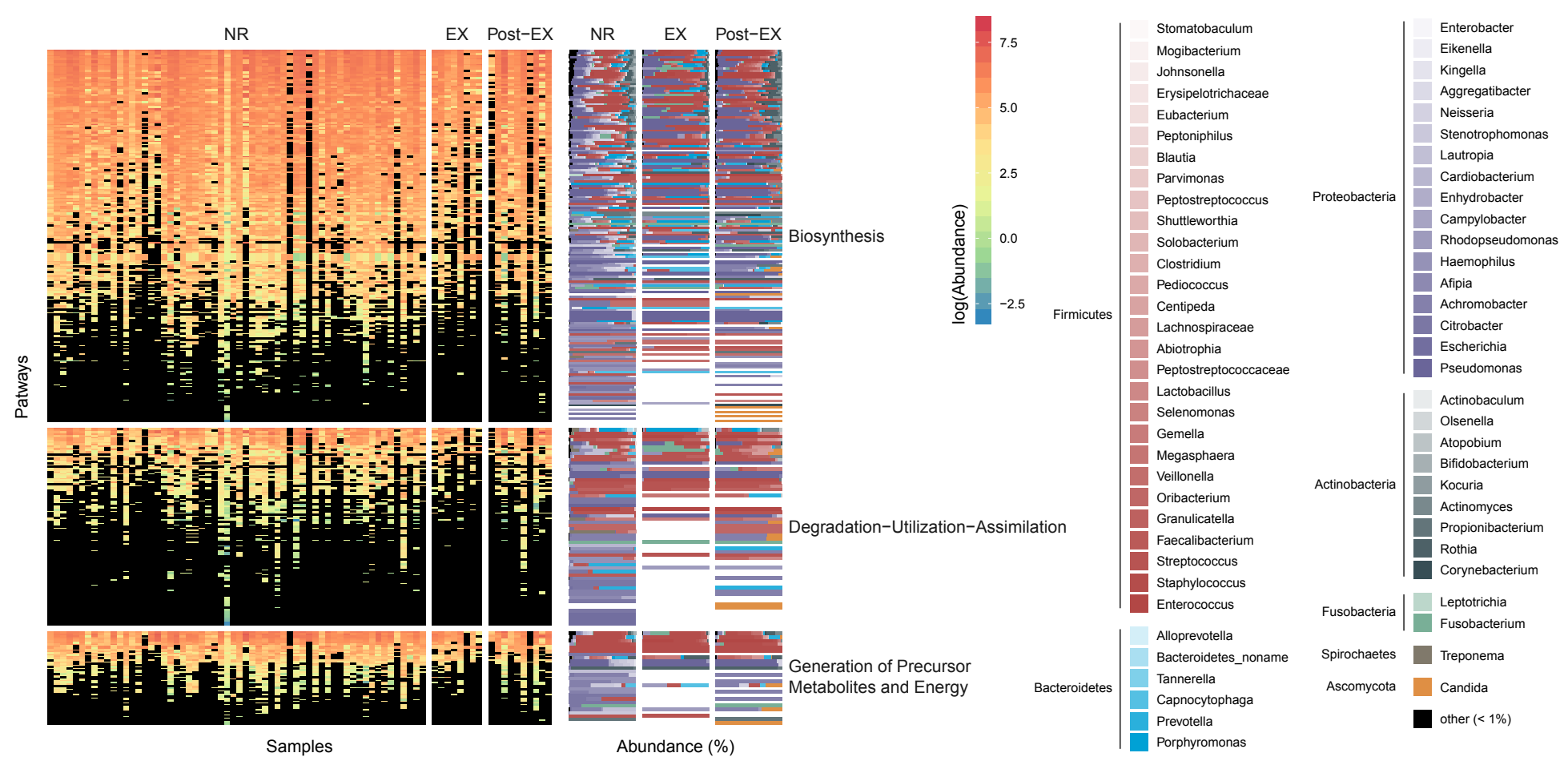
36. Bacci, G., Bazzicalupo, M., Benedetti, A. & Mengoni, A. StreamingTrim 1.0: A Java software for dynamic trimming of 16S rRNA sequence data from metagenetic studies. *Mol. Ecol. Resour.* **14,** 426–434 (2014).
37. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–9 (2012).
38. Abubucker, S. *et al.* Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.* **8,** e1002358 (2012).
39. Truong, D. T. *et al.* MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat. Methods* **12,** 902–903 (2015).
40. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. MetaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **27,** 824–834 (2017).
41. Hyatt, D. *et al.* Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11,** 119 (2010).
42. Huerta-Cepas, J. *et al.* EGGNOG 4.5: A hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* **44,** D286–D293 (2016).
43. Pearson, W. R. in *Current protocols in bioinformatics* **42,** 3.1.1-3.1.8 (2013).
44. Jia, B. *et al.* CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45,** D566–D573 (2017).
45. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/. *R Found. Stat. Comput. Vienna, Austria.* (2012).
46. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47 (2015).
47. Gentleman, R. C. *et al.* voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **5,** R80 (2004).
48. Faust, K., Lahti, L., Gonze, D., de Vos, W. M. & Raes, J. Metagenomics meets time series analysis: unraveling microbial community dynamics. *Curr. Opin. Microbiol.* **25,** 56–66 (2015).
49. Shade, A. *et al.* Fundamentals of microbial community resistance and resilience. *Front. Microbiol.* **3,** 417 (2012).
50. Lozupone, C. A., Stombaugh, J. I., Gordon, J. I., Jansson, J. K. & Knight, R. Diversity, stability and resilience of the human gut microbiota. *Nature* **489,** 220–30 (2012).
51. Doolittle, W. F. & Booth, A. It's the song, not the singer: an exploration of holobiosis and evolutionary theory. *Biol. Philos.* **32,** 5–24 (2016).
52. Eng, A. & Borenstein, E. Taxa-function robustness in microbial communities. *Microbiome* **6,** 45 (2018).
53. Louca, S. *et al.* High taxonomic variability despite stable functional structure across microbial communities. *Nat. Ecol. Evol.* **1,** 0015 (2016).
54. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
55. Thaiss, C. A. *et al.* Persistent microbiome alterations modulate the rate of post-dieting weight regain. *Nature* **540,** 1–27 (2016).


a)
NR
EX
Post−EX
Abundance (%)
10^{2}
10^{0}
10^{-2}
10^{-4}
Species
Firmicutes
Proteobacteria
Bacteroidetes
Actinobacteria
Fusobacteria
Spirochaetes
Ascomycota
Sample
b)
NR
EX
Post−EX
Staphylococcus aureus
Streptococcus parasanguinis
Veillonella unclassified
Enterococcus faecalis
Gemella sanguinis
Granulicatella unclassified
Veillonella parvula
Gemella haemolysans
Granulicatella adiacens
Streptococcus mitis oralis pneumoniae
Firmicutes
Pseudomonas aeruginosa
Pseudomonas unclassified
Haemophilus parainfluenzae
Neisseria unclassified
Neisseria flavescens
Haemophilus influenzae
Lautropia mirabilis
Achromobacter xylosoxidans
Rhodopseudomonas palustris
Stenotrophomonas maltophilia
Proteobacteria
Porphyromonas sp oral taxon 279
Prevotella melaninogenica
Prevotella histicola
Prevotella pleuritidis
Prevotella nanceiensis
Capnocytophaga unclassified
Capnocytophaga gingivalis
Capnocytophaga sp oral taxon 329
Prevotella sp C561
Prevotella pallens
Bacteroidetes
Rothia mucilaginosa
Rothia dentocariosa
Actinomyces graevenitzii
Rothia aeria
Bifidobacterium longum
Atopobium parvulum
Actinomyces odontolyticus
Actinomyces sp ICM39
Propionibacterium acnes
Rothia unclassified
Actinobacteria
Fusobacterium nucleatum
Fusobacterium periodonticum
Leptotrichia unclassified
Leptotrichia sp oral taxon 215
Leptotrichia goodfellowii
Leptotrichia wadei
Leptotrichia shahii
Fusobacteria
Treponema vincentii
Treponema maltophilum
Treponema socranskii
Treponema medium
Spirochaetes
Candida albicans
Candida unclassified
Ascomycota
10^{-2} 10^{0} 10^{2} 10^{-2} 10^{0} 10^{2} 10^{-2} 10^{0} 10^{2}
Abundance (%)


a)
b)
PCoA 2 (17.2%)
PCoA 1 (26.5%)
PCoA 2 (24.9%)
PCoA 1 (48.5%)
Genotype
het (F508)
hom (F508)
other
Exacerbation
EX
NR
Post−EX


a
0.001
b
0.002
c
0.01
d
0.002
B01
B02
B03
B06
G10
G24
G28
G30
G31
G34
M05
M19
M21
M22
M23
M24
M25
M26
M28
M29
M31
M33
M34
reference


a
b
Shannon
InvSimpson
*
0
1
2
3
4.0
4.5
5.0
0
5
10
15
20
25
50
100
150
NR
Post−EX
EX


NR
EX
Post−EX
NR
EX
Post−EX
Patways
Biosynthesis
Degradation−Utilization−Assimilation
Generation of Precursor
Metabolites and Energy
Samples
Abundance (%)
log(Abundance)
7.5
5.0
2.5
0.0
−2.5
Firmicutes
Stomatobaculum
Mogibacterium
Johnsonella
Erysipelotrichaceae
Eubacterium
Peptoniphilus
Blautia
Parvimonas
Peptostreptococcus
Shuttleworthia
Solobacterium
Clostridium
Pediococcus
Centipeda
Lachnospiraceae
Abiotrophia
Peptostreptococcaceae
Lactobacillus
Selenomonas
Gemella
Megasphaera
Veillonella
Oribacterium
Granulicatella
Faecalibacterium
Streptococcus
Staphylococcus
Enterococcus
Bacteroidetes
Alloprevotella
Bacteroidetes_noname
Tannerella
Capnocytophaga
Prevotella
Porphyromonas
Proteobacteria
Enterobacter
Eikenella
Kingella
Aggregatibacter
Neisseria
Stenotrophomonas
Lautropia
Cardiobacterium
Enhydrobacter
Campylobacter
Rhodopseudomonas
Haemophilus
Afipia
Achromobacter
Citrobacter
Escherichia
Pseudomonas
Actinobacteria
Actinobaculum
Olsenella
Atopobium
Bifidobacterium
Kocuria
Actinomyces
Propionibacterium
Rothia
Corynebacterium
Fusobacteria
Leptotrichia
Fusobacterium
Spirochaetes
Treponema
Ascomycota
Candida
other (< 1%)

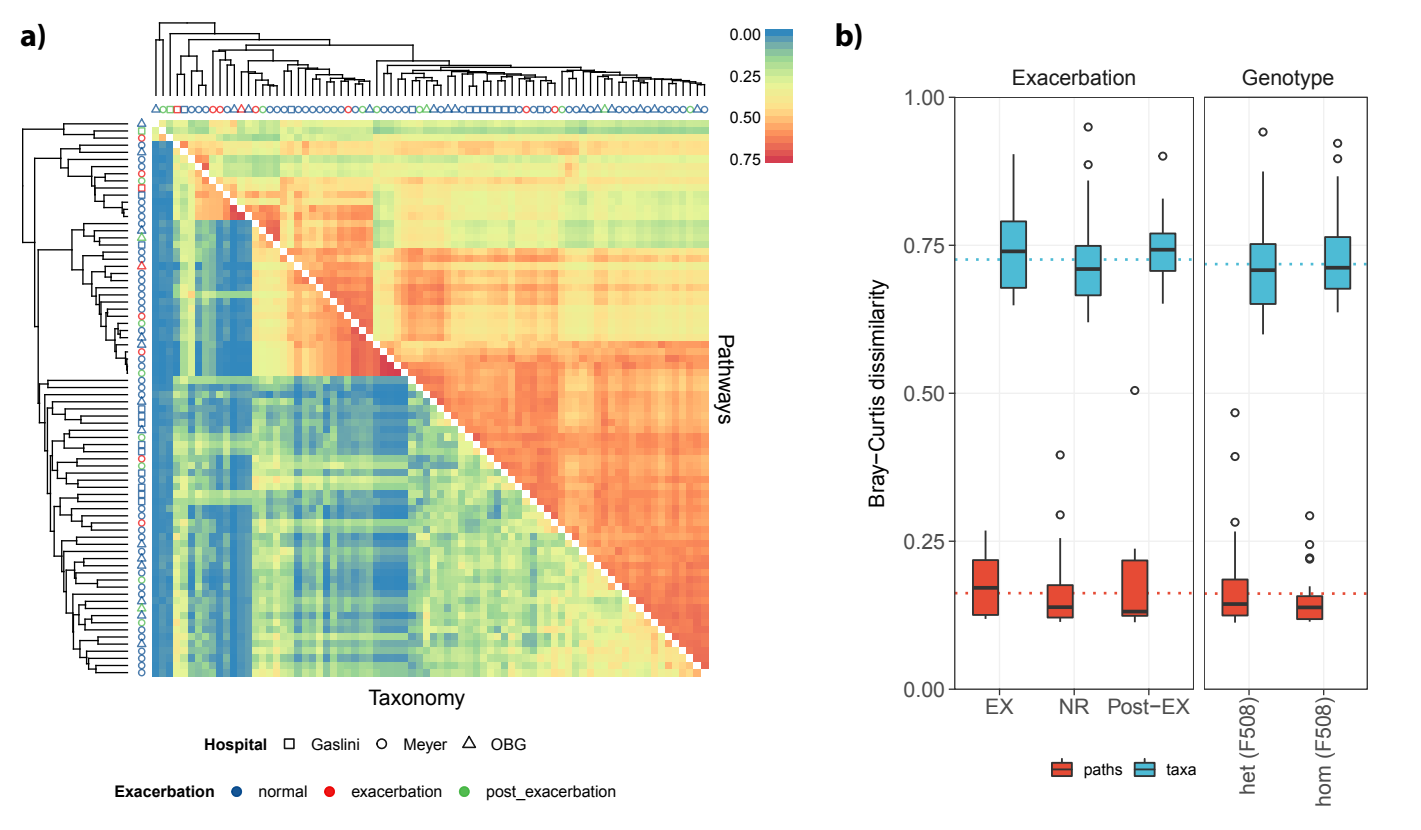

a)
0.00
0.25
0.50
0.75
Pathways
Taxonomy
Hospital
Gaslini
Meyer
OBG
Exacerbation
normal
exacerbation
post_exacerbation
b)
Exacerbation
Genotype
1.00
0.75
0.50
0.25
0.00
Bray−Curtis dissimilarity
EX
NR
Post−EX
het (F508)
hom (F508)
paths
taxa

Table 1: Characteristics of patients enrolled in the study. ID, study id; Hospital, hospital in which patient has been enrolled; Genotype, CFTR genotype; Status, clinical status (S, stable; SD, severe decliner); Gender, gender; Age, enrollment's age; n, number of samples collected; Exacerbation, yes if an exacerbation event has occurred during the study (no otherwise); FEV_1, mean value of forced expiratory volume in 1 second plus/minus the standard error on the mean.

ID	Hospital	Genotype	Status	Gender	Age	n	Exacerbation	FEV_1
B01	OBG	F508/2183AA>G	S	M	27	5	yes	37.0 ± 1.70
B02	OBG	F508/N1303K	SD	F	26	3	no	54.7 ± 3.48
B03	OBG	F508/4016insT	S	F	30	4	no	55.0 ± 1.08
B06	OBG	F508/F508	SD	F	21	4	no	60.2 ± 3.42
G10	Gaslini	F508/F508	S	M	51	4	no	54.0 ± 3.08
G24	Gaslini	F508/F508	S	F	49	3	yes	NA ± NA
G28	Gaslini	F508/F508	NA	F	38	2	no	42.5 ± 1.50
G30	Gaslini	F508/F508	S	F	50	1	no	54
G31	Gaslini	G1244E/G42X	SD	F	53	2	no	41.5 ± 1.50
G34	Gaslini	F508/F508	S	F	39	1	no	47
M05	Meyer	F508/F508	SD	M	32	4	no	34.8 ± 0.85
M19	Meyer	F508/F508	S	M	24	4	no	44.0 ± 2.04
M21	Meyer	F508/N1303K	SD	M	27	4	yes	51.5 ± 4.35
M22	Meyer	F508/2789+5G->A	S	F	29	5	yes	50.4 ± 1.03
M23	Meyer	F508/G542X	S	F	30	4	yes	37.0 ± 1.47
M24	Meyer	F508/F508	S	M	32	4	no	35.2 ± 0.85
M25	Meyer	F508/296+1G->T	SD	F	41	4	no	42.5 ± 2.02
M26	Meyer	F508/3849+10	SD	F	49	5	yes	39.6 ± 1.94
M28	Meyer	F508/N1303K	S	M	23	4	no	39.0 ± 1.08
M29	Meyer	F508/G542X	S	F	12	4	no	43.5 ± 3.75
M31	Meyer	F508/F508	SD	F	11	3	yes	32.7 ± 4.41
M33	Meyer	F508/G85E	SD	F	13	5	yes	35.4 ± 5.78
Total:22	Gaslini:6 Meyer:12 OBG:4	het(F508):47 hom(F508):29 other:2	S:12 SD:9	F:15 M:7	32.1 ± 2.73	79	no:14 yes:8	43.5 ± 1.09

Table 2: Permutational multivariate analysis of variance (PERMANOVA) on both taxonomic distribution and metabolic pathways. The analysis based on taxonomic distribution was reported in the upper part of the table whereas the analysis based on metabolic pathways was reported at the bottom. Df, degrees of freedom; SumOfSqs, sum of squares; R^2, r-squared statistic (reported as proportion); F, F-statistic; Pr(>F), p-value associated to the F-statistic. Significant effects, namely those reporting a p-value lower than 0.05, were reported in bold.

	Df	SumOfSqs	R^2	F	Pr(>F)
Taxonomy					
Exacerbation	**2**	**0.68**	**0.03**	**1.91**	**0.0300**
Genotype	**1**	**0.77**	**0.03**	**4.30**	**0.0020**
Sample	**18**	**11.97**	**0.52**	**3.74**	**0.0010**
FEV_1 value	1	0.27	0.01	1.53	0.1349
Days	1	0.28	0.01	1.58	0.1229
Exacerbation:Genotype	1	0.11	0.01	0.64	0.7642
Residual	49	8.72	0.38		
Pathway					
Exacerbation	**2**	**0.20**	**0.04**	**2.37**	**0.0220**
Genotype	**1**	**0.14**	**0.03**	**3.42**	**0.0080**
Sample	**18**	**2.43**	**0.48**	**3.20**	**0.0010**
FEV_1 value	1	0.09	0.02	2.14	0.0989
Days	1	0.05	0.01	1.26	0.2458
Exacerbation:Genotype	1	0.08	0.02	1.96	0.1169
Residual	49	2.07	0.41	NA	NA