SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
MICROBIOME
RESPIRATORY TRACT DISEASES

Received
20 January 2014

Accepted
24 February 2014

Published
11 March 2014

Correspondence and requests for materials should be addressed to J.J.L. (jlipuma@umich.edu)

Modeling the Impact of Antibiotic Exposure on Human Microbiota

Jiangchao Zhao[1], Susan Murray[2] & John J. LiPuma[1,3]

[1]Department of Pediatrics and Communicable Diseases, University of Michigan Medical School, Ann Arbor, Michigan, USA, [2]Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA, [3]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan, USA.

Human-associated microbial communities play important roles in health and disease. Antibiotic administration is arguably one of the most important modifiable determinants of the composition of the human microbiota. However, quantitatively modeling antibiotic use to account for its impact on microbial community dynamics presents a challenge. We used antibiotic therapy of chronic lung infection in persons with cystic fibrosis as a model system to assess the influence of key variables of therapy on measures of microbial community perturbation. We constructed multivariate linear mixed models with bacterial community diversity as the outcome measure and various scales of antibiotic weighting as predictors, while controlling for other variables. Antibiotic weighting consisted of three components: (i) dosing duration; (ii) timing of administration relative to sample collection; and (iii) antibiotic type and route of administration. Antibiotic weighting based on total dose and proximity to the time of sampling was most predictive of bacterial community change. Using this model to control for antibiotic use enabled the identification of other significant independent predictors of microbial community diversity such as dominant taxon, disease stage, and gender. Quantitative modeling of antibiotic use is critical in understanding the relationships between human microbiota and disease treatment and progression.

The advent of next generation sequencing has enabled culture-independent profiling of complex human-associated bacterial communities (microbiota) in unprecedented detail. Application of this approach has dramatically expanded our understanding of the diversity of microbiota associated with various human body habitats[1], different human populations[2], and periods of health and illness[3–7]. Metagenomic, metabolomic, and metatranscriptomic analyses have now begun to investigate the functional attributes of these microbial communities to better understand the role human microbiota play in health maintenance, predisposition to and pathogenesis of disease, and the response to therapy[8–13]. Such studies view the human body as an ecosystem, with human health being dependent in part upon the services provided by the host-associated microbiota. In this regard, the application of ecological theory to study human microbiota is gaining increasing attention[14]. Microbial community assembly theory, in particular, is finding application in efforts to understand the processes that shape diversity in local assemblages during periods of ill health, treatment and recovery.

A key element of community assembly theory is an appreciation of the effects of community disturbance on diversity and the recovery of communities after perturbation. With respect to the human ecosystem, antibiotic therapy serves as a paradigm for disturbance of host-associated communities; in fact, antibiotic administration may be considered the most important and common form of disturbance of the human microbiota[14–19]. In studies of human gut microbiota, for example, the effect of antibiotic administration on diversity is far greater than the routine temporal variability in community composition[16,19,20].

Despite the prominent role that antibiotic therapy plays in effecting changes in microbial community composition and reassembly of local communities, robust quantitative models of antibiotic use are lacking. This limits incorporating measures of antibiotic use in studies to assess the relationships between antibiotic-driven community perturbation, movement, and reassembly. As important, this presents a challenge to studies wherein antibiotic use must be controlled as one of a number of variables that impact community diversity and disease progression. This is especially relevant in studies of chronic infectious diseases characterized by recurrent, intensive antibiotic administration.

We used treatment of persons with cystic fibrosis (CF), a condition characterized by persistent bacterial infection of the airways managed with chronic maintenance antibiotic therapy as well as intensive episodic antibiotic treatment, to develop a quantitative model of antibiotic use. We tested the effect of antibiotic administration on airway bacterial community diversity (our outcome measure of interest) by considering the duration

Table 1 | Characteristics of patients in the training and validation data sets

Variables	Training Set (n = 6)	Validation Set (n = 60)
Samples per patient, Mean (Range)	20 (12, 30)	6 (2, 23)
Gender, Count (%)		
Male	6 (100)	32 (53)
Female	0 (0)	28 (47)
Disease Severity[1], Count (%)		
Mild	3 (50)	13 (22)
Moderate	3 (50)	18 (30)
Severe	0 (0)	29 (48)
CFTR Genotype, Count (%)		
δ F508 homozygous	4 (67)	24 (39)
δ F508 heterozygous	1 (17)	29 (49)
Others	1 (17)	7 (12)

[1]Patients were assigned to one of three disease severity categories[22].

Table 2 | Characteristics of samples in the training and validation data sets

Variables	Training Set (n = 116)	Validation Set (n = 362)
Age in years[1], Count (%)		
<17	0 (0)	41 (11)
17–26	77 (66)	165 (46)
27–37	39 (34)	91 (25)
>37	0 (0)	65 (18)
Disease Stage[2], Count (%)		
Early	44 (38)	65 (18)
Intermediate	31 (27)	176 (49)
Late	41 (35)	121 (33)
Dominant OTUs[3], Count (%)		
Pseudomonas	95 (82)	205 (57)
Burkholderia	0 (0)	43 (12)
Streptococcus	8 (7)	52 (14)
Others	13 (11)	62 (17)

[1]Age of patient when sample was obtained.
[2]Specimens were assigned to one of three disease stage categories, defined by per cent predicted forced expiratory volume in one second (%FEV_1) values at the time of sample collection: early (%$FEV_1 > 70$), intermediate ($70 \geq$ %$FEV_1 \geq 40$), or advanced (%$FEV_1 < 40$).
[3]OTUs: operational taxonomic units; The dominant OTU was defined as the most abundant OTU detected in the sample.

and timing of administration relative to the day of sampling, as well as antibiotic class and route of administration. A training data set, composed of 116 sputum samples that had been extensively characterized with respect to microbial community profiles, antibiotic exposure, and metrics of patient health, was used to develop a test model[21]. This model was then validated with a larger data set consisting of 362 similarly characterized respiratory samples. We show how this model may be adapted to other studies of the relationship between microbial community dynamics, antibiotic use and disease progression.

Results

Characterization of patients and sputum samples. The patients and sputum samples included in the training and validation data sets were characterized with respect to two types of variables (Table 1 and 2). Fixed, *patient-specific* variables included gender, CFTR genotype, and disease severity (or aggressiveness) phenotype[21,22]. Time-dependent, *sample-specific* variables included patient age, lung function, and disease stage[21] at the time of sample collection. Lung function was measured as percent predicted forced expiratory volume in one sec (%FEV_1). Disease stage was defined as early when serial %FEV_1 measures were >70; intermediate when %FEV_1 measures were between 70 and 40; and advanced when %FEV_1 measures were <40. The dominant operational taxonomic unit (OTU) detected in each sample by deep-sequencing was also included as a sample-specific variable. This was defined as the most abundant OTU detected in the sample. Bacterial community diversity of each sputum sample was measured by calculating the inverse Simpson index, which takes into account both the number of OTUs (richness) present in the sample and their relative abundance (evenness).

The *training set* samples (n = 116) were from six men previously described by us in detail[21]. All six patients had a mild or moderate disease severity phenotype. Samples were collected when these patients were between 18 and 30 years of age. The samples were roughly evenly distributed among periods when these patients were in early, intermediate, or late stages of lung disease. The dominant OTU in most (82%) samples represented the genus *Pseudomonas*. The patients (n = 60) and sputum samples (n = 362) in the *validation set* were more heterogeneous. Men represented 53% of patients and there was a greater distribution of patients with mild, moderate and severe disease severity phenotypes. The validation set samples were also more diverse with respect to the dominant OTU detected. Of note, none of the patients in either the training set or the validation set were smokers.

Antibiotic weighting score development. The antibiotic exposure associated with each sample was measured by assessing the antibiotic administration to the source patient during a 30-day window prior to sample collection. The duration of exposure (no. of days receiving the antibiotic), the timing of administration relative to the day of sampling (e.g., 20 days vs 2 days prior to sampling), and the antibiotic class and route of administration were determined for each sample.

These variables were used to develop antibiotic weighting components that yielded scores used as covariates in models predicting bacterial community diversity.

Weight component A. (*wcA*; Equation 1 in Materials and Methods, and Fig. S1) accounts for the duration of antibiotic use during this 30 day window by assessing the number of days an antibiotic was ($wcA = 1$) or was not ($wcA = 0$) administered (Fig. 1A). *wcA* for each antibiotic was determined from observed sample level data without subjective assessment and was constructed similarly for all samples in both the training and validation data sets.

Weight component B. (*wcB*; Equation 2 in Materials and Methods) accounts for the proximity of antibiotic use relative to the sampling day. Four weighting schemes, described by the formulas in Equation 2, were assessed, including (i) an equal weight for each day irrespective of proximity to the sampling date, (ii) a linear increase in weights with increasing proximity to the sampling date, and either (iii) a concave or (iv) a convex increase in weights with increasing proximity to the sampling date (Fig. 1B and Table S1). A score for each antibiotic administered during the 30 days prior to the date of each sputum sample was calculated as a product of *wcA* and *wcB* (Equation 3 in Materials and Methods). The sum of the scores for all antibiotics administered in association with a sputum sample was calculated to provide the *total antibiotic exposure* for each sample (Equation 4 in Materials and Methods).

Using the total antibiotic exposure for each sputum sample, the training dataset was analyzed to determine which of the four *wcB* weighting schemes best predicted the inverse Simpson index, which had been previously calculated for each sample, based on the Akaike Information Criterion (AIC), after adjusting for age and %FEV_1 at the sampling time. A comparison of AICs indicated that the convex increasing weighting scheme provided the best prediction for the

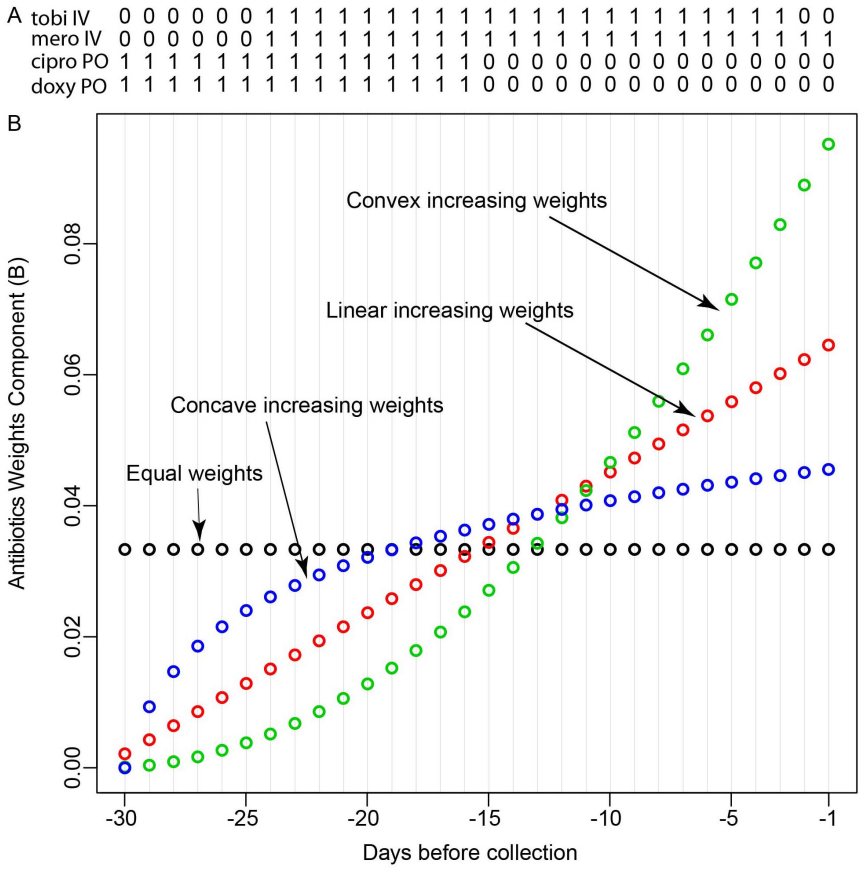

Figure 1 | Antibiotic weight components (A) (*wcA*) and B (*wcB*). Panel (A) depicts daily *wcA* values for patient P2 during the 30 days prior to collection of sample 27. This patient received four antibiotics during this time (tobramycin-IV, meropenem-IV, ciprofloxacin-PO, and doxycycline-PO). A value of 1 indicates antibiotic administration on that day, while 0 indicates no antibiotic administration. Panel (B) depicts *wcB* profiles during the 30 days prior to sampling. These profiles indicate equal weighting (black) as well as linear (red), concave (blue), and convex (green) increasing weights as days approach the sampling time. The data points for each profile (circles) were drawn based on values calculated by Equation 2 in the text and each value was listed in Table S1.

inverse Simpson index (Table S2). When the larger and independently sampled validation sample set was similarly analyzed, using the same four *wcB* weighting schemes, the convex increasing weighing scheme again provided the best prediction of the inverse Simpson index. Analyses on 500 bootstrap samples from the validation set demonstrated the stability of this *wcB* weighting scheme. The *wcB* convex increasing weights were ranked as the best fit in 75% of the bootstrap samples, outperforming the other choices, and were therefore used in the remainder of the study whenever *wcB* was considered.

Weight component C. (*wcC*; Equation 5 in Materials and Methods) accounts for the effects of antibiotic type and route of administration on predicting the inverse Simpson index. Since only 16 of the 37 antibiotic types observed in the validation set were used in the training set, we based our initial evaluation of this weight component on analyses of the combined training and validation sets. First, the ability of each antibiotic associated with a sputum sample (i.e., administered within 30 days of sample collection) to predict the inverse Simpson index was assessed, based on the *wcA* and *wcB* weighting components and adjusting for age and $\%FEV_1$ at sampling. No significant interactions between multiple antibiotics were detected. Next, the AICs for the 37 antibiotic types were modeled and ranked from lowest (best prediction model) to highest (worst prediction model). Antibiotic type coefficients from the 37 AIC models were also ranked based on largest to smallest impact on inverse Simpson index (Table S3).

The AIC and coefficient ranks were summed and sorted from lowest to highest and grouped into terciles. *wcC* values of 0.5, 0.33 or 0.17 were assigned to each combination of antibiotic type and route of administration for the best, intermediate, and worst predictors of the inverse Simpson index, respectively (Equation 5 in Materials and Methods and Table S3). Antibiotics administered by the IV route were more likely to be in the top tercile (i.e., *wcC* value of 0.5) than were antibiotics administered orally or by inhalation (50% of IV administered antibiotics had *wcC* of 0.5 compared to 16% of oral/inhaled antibiotics; Fisher's exact test $p = 0.04$).

The *combined antibiotic weighting score* (Equation 6 in Materials and Methods) for each sputum sample was calculated by multiplying *wcA*, *wcB*, and *wcC* for each of the 30 days prior to sputum collection and then summing these scores across the 30 days. Inclusion of *wcC* improved the prediction of community diversity (inverse Simpson index) in each of the training, validation and combined sets as opposed to using only *wcA* and *wcB* alone. A permutation test indicated that inclusion of *wcC* in the combined antibiotic weighting score gave a significantly lower AIC value (better model fit; $p < 0.001$) than would have occurred under 5000 random permutations of the *wcC* values across the 37 antibiotic types. The distribution of the combined antibiotic load score across validation samples is shown in Figure S2.

Predictors of community diversity in CF. To illustrate the utility of the antibiotic weighting scoring schemes, we included the combined antibiotic weighting score as a covariate in a multivariate model

predicting community diversities in our collection of CF sputum samples. This allowed us to explore associations between bacterial community diversity (inverse Simpson index) and the patient- and sample-specific variables associated with these samples. Since disease stage was defined based on lung function (%FEV_1), these correlated variables could not both be included in the model. We therefore included disease stage as a covariate in the model since it has a stronger predictive ability (AIC 1572.14) than lung function (AIC 1579.56).

Table 3 shows results from the multivariate linear mixed model predicting the inverse Simpson index in the validation set. An increase of one unit in the antibiotic weighting score calculated over the month prior to sampling is associated with a 1.25 point decrease in the inverse Simpson index on average (95% CI, 0.44, 2.05 point decrease, $p = 0.002$) after adjusting for dominant OTU, disease severity, gender, CFTR genotype, disease stage, and patient age. We observed significant differences in community diversity with respect to dominant OTU (composite $P < 0.001$). Samples dominated by *Pseudomonas* or *Burkholderia* showed the least diversity, while samples dominated by *Streptococcus* showed the greatest diversity (approximately 3 inverse Simpson index points higher than *Pseudomonas* and *Burkholderia*, on average) after adjusting for age, gender, disease severity, disease stage, CFTR genotype, and antibiotic weighting in the previous month. On average, men had an approximately 1 point higher inverse Simpson index than women after adjusting for other factors ($p = 0.004$). We observed decreasing diversity with advancing disease stage after adjusting for other factors; on average, early disease stage had a 1.15 point higher inverse Simpson index than late disease stage ($p = 0.02$). We did not observe significant associations between community diversity and disease severity or CFTR genotype after adjusting for age, gender, dominant OTU, disease stage, and antibiotic use in the previous month. In addition, we did not observe a significant decrease in diversity with increasing age after adjusting for other factors.

Figure 2 shows the community diversity of a "standardized" CF patient predicted by specific variables after adjusting for other variables. For example, after controlling for other variables, the estimated community diversity is 5.7 if the community is dominated by *Streptococcus* and 2.6 if dominated by *Burkholderia*.

Discussion

It is not unexpected that antibiotic administration perturbs bacterial communities in human hosts, thereby confounding efforts to better understand the microbial community signatures of disease progression, particularly in chronic diseases where prolonged and repeated antibiotic use is required. Quantifying the impact of antibiotic therapy on complex bacterial communities is challenging, involving a myriad of variables such as dosing duration and timing relative to sample collection, and antibiotic type and route of administration. In this study, we propose a model to address this challenge by developing antibiotic weighting scores that account for these variables. Using these scores, a combined antibiotic score can be calculated and included as a covariate in multivariate models that assess other factors that may be important predictors of community structure. We chose community diversity – as measured by the inverse Simpson index – as the outcome parameter in this study. However, the principles and approach we describe could be applied to account for antibiotic use in analyses assessing predictors of any number of other microbiota community outcome measures.

Weight component A takes into account the duration of antibiotic administration. In our model we limited this only to antibiotic administration in the 30 days prior to sample collection. This is an admittedly arbitrary interval that could be modified in the model as appropriate for other studies. Further, a defined interval such as this does not account for cumulative antibiotic load that may have accrued over very long intervals. Such use, involving potential 'legacy effects' of prior antibiotic courses, could have a significant impact on shaping human-associated microbiota. Although a comprehensive accounting of long-term prior antibiotic use is often limited by the unavailability of reliable medical records, the increasing use of detailed electronic medical records may enable analyses of this sort in future studies.

In accounting for the timing of antibiotic administration relative to the sample date, (weight component B), antibiotic weighting based on a convex increasing weighting scheme provided a better prediction of community diversity than did the other schemes evaluated. This expected result likely reflects community resilience after antibiotic perturbation; a longer interval between antibiotic administration and sampling provides greater opportunity for the recovery of

Table 3 | Multivariate linear mixed model including antibiotic use as a covariate

		95% Confidence Interval			
Parameters	Coefficient	Lower Bound	Upper Bound	Wald P-Value	Composite P-Value
Intercept	5.73	3.98	7.48	<0.001	<0.001
Dominant OTUs					<0.001
Pseudomonas	−1.59	−2.29	−0.89	<0.001	
Burkholderia	−1.83	−3.11	−0.56	0.005	
Streptococcus	1.24	0.40	2.08	0.004	
Others	0.00				
Disease Severity					0.256
Mild	0.64	−0.80	2.08	0.38	
Moderate	−0.41	−1.40	0.59	0.42	
Severe	0.00				
Gender					0.004
Male	1.16	0.39	1.93	0.004	
Female	0.00				
CFTR (ΔF 508)					0.119
Homozygous	−1.11	−2.33	0.11	0.07	
Heterozygous	−1.21	−2.39	−0.04	0.04	
Other	0.00				
Disease Stage					0.067
Early	1.15	0.18	2.11	0.02	
Intermediate	0.46	−0.25	1.16	0.21	
Late	0.00				
Antibiotic Usage	−1.25	−2.05	−0.44	0.002	0.002
Age	−0.04	−0.09	0.01	0.117	0.117

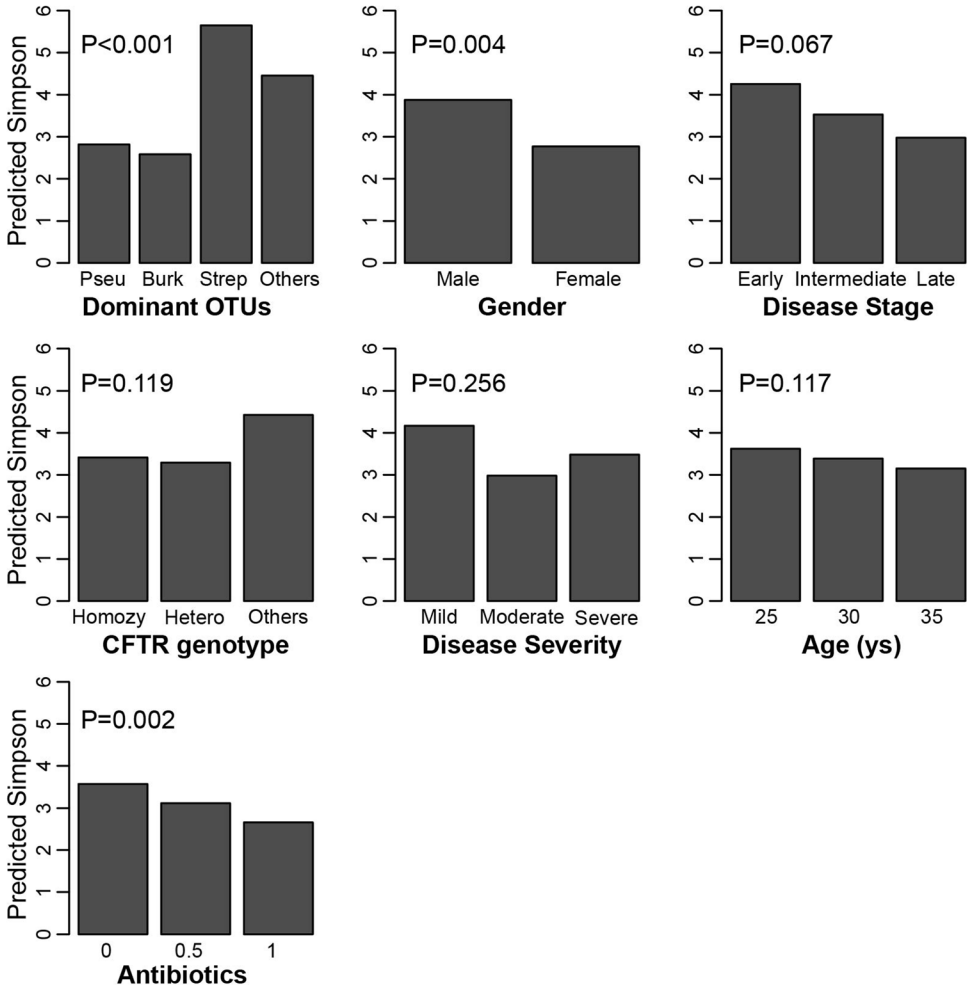

Figure 2 | Estimated community diversity by each predictor. Pseu: *Pseudomonas*, Burk: *Burkholderia*, Strep: *Streptococcus*. The predicted values by each predictor were calculated by controlling for other predictors based on the "standardized" CF patient profile: 57%, 12%, 14%, and 17% chance of being dominated by *Pseudomonas*, *Burkholderia*, *Streptococcus* or other bacteria, respectively; 18%, 49%, and 33% chance of being in early, intermediate or late disease stage, respectively; 22%, 30%, and 48% chance of having a mild, moderate or severe disease phenotype; 64% chance of being male; 44%, 44% and 12% chance of being delta F508 homozygous, delta F508 heterozygous or another CFTR genotype, respectively; average age = 28.13 years and an antibiotic load = 0.11.

members of the community affected by the antibiotic. Thus, an antibiotic administered closer to the time of sampling would be expected to have a greater impact on the microbial community than would that same antibiotic administered at a longer interval from the sampling time.

Specific antibiotic type is clearly a major component of any model designed to account for variables impacting the effect of antibiotic administration on human microbiota. We reasoned that route of antibiotic administration would also impact antibiotic effect and included this variable in our model as well. In developing this weighting component, we assessed the bacterial communities in 478 sputum samples from 66 persons with CF. Each combination of antibiotic type and route of administration used to treat these individuals was provided a weight based on its ability to predict community diversity, our outcome measure of interest. As such, this weighting was a function of the resistance of these communities to the perturbation caused by the combinations of antibiotic type and administration route included in our study. In studies of other human-associated microbial communities, other combinations of antibiotic type and route of administration would be expected to have different impacts. Assigning weights to these combinations therefore requires an analysis of a suitably sized data set relevant to the microbiome being investigated.

We observed that, in general, antibiotic administration through an IV route had a significantly higher probability of effecting a greater change in community diversities, after controlling for subject age and $\%FEV_1$ at the time of sampling, than did other routes of administration. This may be due, in part, to the broad spectrum antimicrobial activities of agents such as cefepime and meropenem, which are administered exclusively via an IV route. A more complete analysis would need to take into account the antimicrobial susceptibilities of all species detected with deep-sequencing, including those that are not (or cannot be) routinely cultured *in vitro*.

To demonstrate the utility of modeling antibiotic use in studies correlating the human microbiome and disease, we analyzed sets of CF sputum samples for which we had determined bacterial community composition by deep sequencing. We included the antibiotic weighting score assigned to each sample as a covariate, allowing us to adjust for antibiotic use in a multivariate linear mixed model assessing other potential predictors of bacterial community diversity. In this analysis, we observed that community diversity was significantly associated with the taxonomic affiliation of the dominant OTU. *Streptococcus* dominated communities, for example, had significantly higher diversities than communities dominated by either *Pseudomonas* or *Burkholderia*, consistent with a recent study showing greater community diversity and higher relative abundance of

Streptococcus in airway communities among outpatients with CF compared to inpatients[23]. Our analysis also identified gender as a significant predictor of CF airway community diversity, with females having lower diversities than males. The adjustment for antibiotic use, as well as age, disease stage and the other variables included in the model suggests additional unidentified factors that reduce airway community diversity in women compared to men[24]. Disease severity in CF describes the rate of lung function decline relative to patient age[22]. In our analysis, after adjusting for antibiotic use and other variables, disease aggressiveness was not found to be significantly associated with community diversity, suggesting either that a larger sample size is needed to detect such an association or that other factors, most likely host-associated variables, play critical roles in CF lung disease progression. An important observation in our analysis was that antibiotic use was found to be an independent predictor of decreased CF airway bacterial community diversity. The correlation between decreasing airway community diversity and lung disease progression in CF has been noted in several recent studies[21,25–27]; however, the causal relationship between decreasing diversity and decreasing lung function has been the subject of controversy[28]. The results of our study support our previous observation suggesting that decreasing airway community diversity is driven primarily by antibiotic therapy[21].

In summary, we have described an approach to account for antibiotic exposure in studies examining the relationships between human microbiota and disease. More specifically, we propose a scheme, based on weighing variables associated with antibiotic use, to develop an antibiotic score that, in turn, can be included as a covariate in models exploring correlations between bacterial communities and human disease progression, particularly in conditions associated with repeated antibiotic therapy. We applied this scheme in a multivariate analysis of potential predictors of bacterial airway community diversity in a large cohort of persons with CF. Our findings sharpen previous observations of associations between decreased airway community diversity, lung disease progression, and variables such as gender and dominant community OTU. We show that antibiotic therapy is an independent predictor of decreased airway community diversity in CF. We expect that the specific weighting schemes we developed for our dataset will need to be modified to best suit studies of other disease conditions. Nevertheless, we propose that the approach we describe will have broad applicability to such studies.

Methods

Patients and clinical samples. A total of 478 sputum samples, collected from 66 adults with CF receiving care at the University of Michigan Health System, were included in this study. Sample collection and medical record review were approved by the University of Michigan Institutional Review Board, which waived the requirement for informed consent. Sputum specimens were collected during the course of routine medical care and stored at −80°C in 0.5 mL aliquots as described previously[21]. A minority proportion of the samples was relegated for exploratory analysis. This *training data set* consisted of 116 sputum samples collected from six male CF patients during periods of 8 to 9 years as described previously[21]. The majority of sputum samples was used for validation analyses; this *validation data set* consisted of 362 sputum samples from 60 patients not included in the training data set.

Sputum DNA extraction and pyrosequencing. DNA extraction and pyrosequencing were performed as described previously[21]. Briefly, sputum aliquots were thawed on ice and homogenized with 0.5 mL Sputolysin® (EMD Chemicals, San Diego, CA) before DNA was purified by an automated nucleic acid purification platform (MagNA Pure Compact System, Roche, Indianapolis, IN) according to the manufacturer's protocol. The 16S rDNA V3–V5 region was amplified with bar-coded primers and sequenced by the Human Genome Sequencing Center at Baylor College of Medicine using protocols developed for the Human Microbiome Project (http://www.hmpdacc.org/tools_protocols/tools_protocols.php) as described[17].

DNA sequence processing and analysis. Raw sequences were analyzed using mothur v1.24[29]. Reads were denoised by the PyroNoise component of the AmpliconNoise suite of programs[30]. Sequences containing homopolymers greater than 8 bp, 1 mismatch in the barcode or 2 in the primer, one or more ambiguous bases were removed. Remaining sequences that were at least 200 bp but less than 590 bp in length were further curated to remove chimeric sequences using UCHIME[31] and to further reduce sequencing noise by a preclustering methodology[32] before being assigned to operational taxonomic units (OTUs) using an average neighbor algorithm with a 0.03 dissimilarity cutoff. The total number of reads for each community was first normalized to 568, the smallest number of reads among the 478 samples by random sampling, to control for differences in sequencing depth before calculation of community diversity.

Statistical methods. The inverse Simpson index[33] was calculated and used as a measure of bacterial community diversity and was modeled via linear mixed models (IBM SPSS Statistics package version 20) so that associations between predictors and community measures would take into account autocorrelation of repeated samples from the same individual. The training data set was used to study the effects of the duration of antibiotic administration and the proximity of administration to the time of sample collection. Both the training and validation data sets were used to assess the effects of antibiotic type and route of administration on community measures. The Akaike Information Criterion (AIC) was used to compare the predictive value of antibiotic weighting definitions applied to the training and validation data in developing the final antibiotic weighting scheme; since sample size affects AIC values, these comparisons were only made within models of the same cohort. Using the validation data set, mixed model analyses studied multivariate associations between the final antibiotic weighting scheme and community measures, accounting for other clinical factors of interest.

Antibiotic score calculations. Antibiotic weight component A (*wcA*) for antibiotic *j*, $j = 1,\ldots,37$, is defined by:

$$wcA_{ij} = \begin{cases} 1 (\textit{if antibiotic } j \textit{ was used on day } i) \\ 0 (\textit{if antibiotic } j \textit{ was not used on day } i) \end{cases} \quad (1)$$

for the $i = 1, \ldots,30$ days approaching the sampling time.

Antibiotic weight component B (*wcB*) contenders shown in Fig. 1B are defined by:

$$wcB_i = \begin{cases} \frac{1}{30} & \text{Equal weight} \\ \frac{i}{\sum_{i=1}^{30} i} & \text{Linear increasing weight} \\ \frac{\log_{10}(i)}{\sum_{i=1}^{30} \log_{10}(i)} & \text{Concave increasing weight} \\ \frac{i^2}{\sum_{i=1}^{30} i^2} & \text{Convex increasing weight} \end{cases} \quad (2)$$

for the $i = 1, \ldots,30$ days approaching the sampling time.

The score for antibiotic $j = 1,\ldots,37$, based on *wcA* and *wcB* becomes:

$$\text{Antibiotic } j \text{ score} = \sum_{i=1}^{30} wcA_{ij} \times wcB_i \quad (3)$$

The total antibiotic exposure calculated from treating all antibiotics as equally weighted is:

$$\sum_{j=1}^{37} \sum_{i=1}^{30} wcA_{ij} \times wcB_i \quad (4)$$

Weight component C (*wcC*) was assigned for antibiotic type $j = 1,\ldots,37$, via:

$$wcC_j = \begin{cases} 0.05 & \text{(if best diversity reduction tercile)} \\ 0.33 & \text{(if middle diversity reduction tercile)} \\ 0.17 & \text{(if worst diversity reduction tercile)} \end{cases} \quad (5)$$

The final antibiotic exposure calculated from *wcA*, *wcB* and *wcC* is:

$$\text{WT} = \sum_{j=1}^{37} \sum_{i=1}^{30} wcC_j \times wcA_{ij} \times wcB_{ij} \quad (6)$$

1. Costello, E. K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694–1697 (2009).
2. Yatsunenko, T. *et al.* Human gut microbiome viewed across age and geography. *Nature* **486**, 222–227 (2012).
3. Kostic, A. D. *et al.* Genomic analysis identifies association of Fusobacterium with colorectal carcinoma. *Genome Res* **22**, 292–298 (2012).
4. Karlsson, F. H. *et al.* Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat Commun* **3**, (2012).
5. Yoshimoto, S. *et al.* Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. *Nature* **499**, 97–101 (2013).
6. Arthur, J. C. *et al.* Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**, 120–123 (2012).
7. Atarashi, K. *et al.* Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* **500**, 232–236 (2013).
8. Koren, O. *et al.* Host Remodeling of the Gut Microbiome and Metabolic Changes during Pregnancy. *Cell* **150**, 470–480 (2012).
9. Jansson, J. *et al.* Metabolomics reveals metabolic biomarkers of Crohn's disease. *PLoS One* **4**, e6386 (2009).

10. Holmes, E., Li, J. V., Athanasiou, T., Ashrafian, H. & Nicholson, J. K. Understanding the role of gut microbiome-host metabolic signal disruption in health and disease. *Trends Microbiol* **19**, 349–359 (2011).
11. Gajer, P. *et al.* Temporal dynamics of the human vaginal microbiota. *Sci Transl Med* **4**, 132ra152 (2012).
12. Westermann, A. J., Gorski, S. A. & Vogel, J. Dual RNA-seq of pathogen and host. *Nat Rev Microbiol* **10**, 618–630 (2012).
13. Sharma, C. M. *et al.* The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464**, 250–255 (2010).
14. Costello, E. K., Stagaman, K., Dethlefsen, L., Bohannan, B. J. & Relman, D. A. The application of ecological theory toward an understanding of the human microbiome. *Science* **336**, 1255–1262 (2012).
15. Lemon, K. P., Armitage, G. C., Relman, D. A. & Fischbach, M. A. Microbiota-targeted therapies: an ecological perspective. *Sci Transl Med* **4**, 137rv135 (2012).
16. Dethlefsen, L. & Relman, D. A. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. *Proc Natl Acad Sci U S A* **108 Suppl 1**, 4554–4561 (2011).
17. Jernberg, C., Lofmark, S., Edlund, C. & Jansson, J. K. Long-term impacts of antibiotic exposure on the human intestinal microbiota. *Microbiology* **156**, 3216–3223 (2010).
18. Jakobsson, H. E. *et al.* Short-term antibiotic treatment has differing long-term impacts on the human throat and gut microbiome. *PLoS One* **5**, e9836 (2010).
19. Jernberg, C., Lofmark, S., Edlund, C. & Jansson, J. K. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *ISME J* **1**, 56–66 (2007).
20. Dethlefsen, L., Huse, S., Sogin, M. L. & Relman, D. A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biol* **6**, e280 (2008).
21. Zhao, J. *et al.* Decade-long bacterial community dynamics in cystic fibrosis airways. *Proc Natl Acad Sci U S A* **109**, 5809–5814 (2012).
22. Konstan, M. W., Wagener, J. S. & VanDevanter, D. R. Characterizing aggressiveness and predicting future progression of CF lung disease. *J Cyst Fibros* **8 Suppl 1**, S15–19 (2009).
23. Filkins, L. M. *et al.* Prevalence of streptococci and increased polymicrobial diversity associated with cystic fibrosis patient stability. *J Bacteriol* **194**, 4709–4717 (2012).
24. Nick, J. A. *et al.* Effects of Gender and Age at Diagnosis on Disease Progression in Long-term Survivors of Cystic Fibrosis. *Am J Resp Crit Care* **182**, 614–626 (2010).
25. Cox, M. J. *et al.* Airway microbiota and pathogen abundance in age-stratified cystic fibrosis patients. *PLoS One* **5**, e11044 (2010).
26. Klepac-Ceraj, V. *et al.* Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa. *Environmental Microbiology* **12**, 1293–1303 (2010).
27. van der Gast, C. J. *et al.* Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J* **5**, 780–791 (2011).
28. VanDevanter, D. R. & LiPuma, J. J. Microbial diversity in the cystic fibrosis airways: where is thy sting? *Future Microbiol* **7**, 801–803 (2012).
29. Schloss, P. D. *et al.* Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**, 7537–7541 (2009).
30. Quince, C., Lanzen, A., Davenport, R. J. & Turnbaugh, P. J. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**, 38 (2011).
31. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200 (2011).
32. Huse, S. M., Welch, D. M., Morrison, H. G. & Sogin, M. L. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ Microbiol* **12**, 1889–1898 (2010).
33. Magurran, A. E. (ed.) *Ecological Diversity and Its Measurement* (Princeton University Press, Princeton, NJ, 1988).

Acknowledgments

This work was funded by the National Institutes of Health CTSA grant UL1RR024986 and NHLBI grant 1RC1HL100809-01. J.J.L. also was supported by the Cystic Fibrosis Foundation. Additional support was provided by the Charles Woodson Pediatric Research Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare no potential conflict of interest. We thank Bridget Foster for technical support.

Author contributions

J.Z., S.M. and J.J.L. designed the project. J.Z. and S.M. performed the data analyses. J.Z., S.M. and J.J.L. wrote the paper.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Zhao, J.C., Murray, S. & LiPuma, J.J. Modeling the Impact of Antibiotic Exposure on Human Microbiota. *Sci. Rep.* **4**, 4345; DOI:10.1038/srep04345 (2014).