**Metabolic Modeling of Cystic Fibrosis Airway Communities Predicts Mechanisms of Pathogen Dominance**

Michael A. Henson,a Giulia Orazi,b Poonam Phalak,a George A. O’Tooleb

**Abstract** Cystic fibrosis (CF) is a fatal genetic disease characterized by chronic lung infections due to aberrant mucus production and the inability to clear invading pathogens. The traditional view that CF infections are caused by a single pathogen has been replaced by the realization that the CF lung usually is colonized by a complex community of bacteria, fungi, and viruses. To help unravel the complex interplay between the CF lung environment and the infecting microbial community, we developed a community metabolic model comprised of the 17 most abundant bacterial taxa, which account for >95% of reads across samples, from three published studies in which 75 sputum samples from 46 adult CF patients were analyzed by 16S rRNA gene sequencing. The community model was able to correctly predict high abundances of the “rare” pathogens Enterobacteriaceae, Burkholderia, and Achromobacter in three patients whose polymicrobial infections were dominated by these pathogens. With these three pathogens removed, the model correctly predicted that the remaining 43 patients would be dominated by *Pseudomonas* and/or *Streptococcus*. This dominance was predicted to be driven by relatively high monoculture growth rates of *Pseudomonas* and *Streptococcus* as well as their ability to efficiently consume amino acids, organic acids, and alcohols secreted by other community members. Sample-by-sample heterogeneity of community composition could be qualitatively captured through random variation of the simulated metabolic environment, suggesting that experimental studies directly linking CF lung metabolomics and 16S sequencing could provide important insights into disease progression and treatment efficacy.

**Importance** Cystic fibrosis (CF) is a genetic disease in which chronic airway infections and lung inflammation result in respiratory failure. CF airway infections are usually caused by bacterial communities that are difficult to eradicate with available antibiotics. Using species abundance data for clinically stable adult CF patients assimilated from three published studies, we developed a metabolic model of CF airway communities to better understand the interactions between bacterial species and between the bacterial community and the lung environment. Our model predicted that clinically observed CF pathogens could establish dominance over other community members across a range of lung nutrient conditions. Heterogeneity of species abundances across 75 patient samples could be predicted by assuming that sample-to-sample heterogeneity was attributable to random variations in the CF nutrient environment. Our model predictions provide new insights into the metabolic determinants of pathogen dominance in the CF lung and could facilitate the development of improved treatment strategies.

**Keywords** community metabolism, cystic fibrosis, metabolic modeling, metabolite cross-feeding

Cystic fibrosis (CF) is a genetic disease which results in excessive mucus production that reduces lung function and impedes the release of pancreatic enzymes ([1](#), [2](#)). While digestive problems are highly prevalent among CF patients ([3](#)), approximately 80

Citation: Henson MA, Orazi G, Phalak P, O’Toole GA. 2019. Metabolic modeling of cystic fibrosis airway communities predicts mechanisms of pathogen dominance. *mSystems* 4:e00026-19. https://doi.org/10.1128/mSystems.00026-19.

Editor: Nicholas Chia, Mayo Clinic

Copyright © 2019 Henson et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Michael A. Henson, mhenson@umass.edu.

Received 14 January 2019

Accepted 29 March 2019

Published 23 April 2019
to 95% of CF deaths are attributable to respiratory failure due to chronic airway infections and associated inflammation (1). The Cystic Fibrosis Foundation (CFF) estimates that approximately 70,000 CF patients are living worldwide and about 1,000 new CF cases are diagnosed in the United States each year (www.cff.org). Following Koch’s postulate (4), the traditional view of CF lung infections has been that specific airway pathogens are responsible for monomicrobial infections (5). CF bacterial pathogens that have been identified from patient sputum samples and commonly studied *in vitro* using pure culture include *Pseudomonas aeruginosa*, *Haemophilus influenzae*, *Staphylococcus aureus*, and *Burkholderia cepacia* complex, including antibiotic-resistant strains such as methicillin-resistant *S. aureus* (MRSA) and multidrug-resistant *P. aeruginosa* (MDRPA) (1), as well as less common species such as *Achromobacter xylosoxidans*, *Stenotrophomonas maltophilia*, and pathogenic *Escherichia coli* strains (6).

With the advent of culture-independent techniques such as 16S rRNA gene amplicon library sequencing, sputum and bronchoscopy samples from CF patients can be analyzed systematically with respect to the diversity and abundance of bacterial taxa present (7, 8). Numerous studies have shown that CF airway infections are rarely monomicrobial, but rather the CF lung harbors a complex community of bacteria that originate from the mouth, skin, intestine, and the environment (7–10). 16S sequencing can reliably delineate community members down to the genus level, showing that the most common genera in adult CF patient samples are *Streptococcus*, *Pseudomonas*, *Prevotella*, *Veillonella*, *Neisseria*, *Porphyromonas*, and *Catonella* (7). While the identities and relative abundances of the genera present can be determined by 16S rRNA gene sequencing, different analysis techniques are required to understand the interactions between the multiple bacterial taxa and the CF lung environment, the role of the individual microbes in shaping community composition and behavior, and the impact of community composition on the efficacy of antibiotic treatment regimens. While microbiota cooccurrence networks have provided important insights into interactions between bacterial taxa colonizing the CF lung (11, 12), these methods require species abundance data as inputs and therefore are not fully predictive.

*In silico* metabolic modeling has emerged as a powerful approach for analyzing complex microbial communities by integrating genome-scale reconstructions of single-species metabolism within mathematical descriptions of metabolically interacting communities (13, 14). Modeled species interactions typically include competition for host-derived nutrients and cross-feeding of secreted by-products such as organic acids, alcohols, and amino acids between species (15, 16). Due to challenges in developing manually curated reconstructions of poorly studied species, including those present in the CF lung, most *in silico* community models have been restricted to ~5 microbial species (17–19) and fail to adequately cover the diversity of *in vivo* communities. This limitation can be overcome in bacterial communities by using semicurated reconstructions developed through computational pipelines such as the ModelSeed (20), AGORA (21), and other methods (22). Given the availability of suitable single-strain metabolic reconstructions, a number of alternative methods have been developed for mathematical formulation and numerical solution of microbial community models (23–26). The recently developed SteadyCom method is particularly notable due to its formulation that ensures proper balancing of metabolites across the species and scalability to large communities (27). A properly formulated community model can yield information that is difficult to ascertain experimentally, including the effects of the host environment on community growth, species abundances, and cross-fed metabolite secretion and uptake rates.

In this paper, we utilized 16S rRNA gene amplicon library sequencing data from three published studies (28–30) to develop a 17-species bacterial community model for predicting species abundances in CF airway communities (Fig. 1). The 16S rRNA gene sequence data covers 75 distinct sputum samples from 46 adult CF patients and captures the heterogeneity of CF polymicrobial infections with respect to taxonomic diversity and the prevalence of pathogens, including *Pseudomonas*, *Streptococcus*, *Burkholderia*, *Achromobacter*, and *Enterobacteriaceae*. The *in silico* community model
was used to predict when each pathogen may dominate the polymicrobial infection by using the 16S rRNA gene sequence data to restrict which pathogens were present in the simulated community. By randomly varying the availability of host-derived nutrients, the model was used to simulate sample-by-sample heterogeneity of community compositions across patients and to understand how metabolite cross-feeding enhanced pathogen abundances. To our knowledge, this study represents the first attempt to metabolically model the CF airway bacterial community rather than model the individual metabolism of common CF pathogens (31–36). Furthermore, our approach of directly predicting species abundances rather than using measured abundances as model input data to constrain predictions distinguished our study from other community modeling efforts driven by 16S rRNA gene sequence data (16, 37–39).

RESULTS

Few taxonomic groups dominate the CF airway community samples. Principal-component analysis (PCA) was performed on the normalized read data of the 75 samples to evaluate sample heterogeneity. The first three principal components (PCs) captured 77.8% of the data variance, with the first PC capturing 57.3% of variance and most heavily weighting the most abundant genera Pseudomonas, Streptococcus, and Prevotella as expected (see Table S1 in the supplemental material). A considerable degree of heterogeneity was evident from a plot of the 75 samples in the coordinates defined by the first three PCs (Fig. 2A). Most striking were the outlier samples from three patients infected with Enterobacteriaceae (samples 25 to 27), Burkholderia (samples 19 to 21), or Achromobacter (samples 31 and 32) compared to the patients lacking these three organisms (i.e., the remaining 67 samples).
Because each pathogen infected only a single patient among the 46 included patients, we generated a smaller data set of 67 samples by removing these 8 samples. When PCA was performed on this reduced data set, the first three PCs explained 92.6% of the data variance (Table S2), suggesting substantially reduced heterogeneity compared to the full data set. These three PCs heavily weighted only the four taxonomic groups *Pseudomonas*, *Streptococcus*, *Prevotella*, and *Haemophilus*, with the first PC representing high *Pseudomonas* and low *Streptococcus*, the second PC component representing high *Streptococcus* and moderate *Pseudomonas*, and the third PC representing high *Haemophilus*, low *Pseudomonas*, and low *Streptococcus*. Considerable heterogeneity was evident when the 67 samples were plotted using the first two PCs accounting for 84.2% of the variance (Fig. 2B). Here the first PC represented high *Pseudomonas*, low *Streptococcus*, moderate *Prevotella*, and moderate *Haemophilus*, and the second PC represented low *Pseudomonas*, high *Streptococcus*, low *Prevotella*, and low *Haemophilus*.

Based on these results, we focused our community modeling efforts on predicting the infrequent dominance of the pathogens *Enterobacteriaceae*, *Burkholderia*, and *Achromobacter* and the heterogeneity in the abundances of *Pseudomonas*, *Streptococcus*, *Prevotella*, and *Haemophilus* across the remaining samples. *Pseudomonas*, *Streptococcus*, and *Prevotella* have been found by directly sampling the lung of CF patients via bronchoalveolar lavage (40), while *Haemophilus* is a widely accepted CF pathogen (7). The other 10 genera (Table 1) were maintained in the model to simulate competition/cooperation with the more dominant species and to determine if the relatively low abundances of these genera could be predicted.

**The community model can reproduce dominance of CF pathogens.** We simulated the growth of each species individually to compare their monoculture growth
rates with the nominal community nutrient uptake rates (Table S3). Interestingly, the three highest growth rates belonged to the rare pathogens Escherichia, Burkholderia, and Achromobacter, while the next three highest growth rates belonged to the common pathogens Pseudomonas, Streptococcus, and Staphylococcus (Fig. 3A; species numbered as in Table 1). These predictions were consistent with our modeling results for the gut microbiome (41), where opportunistic pathogens consistently had higher growth rates than commensal species. The other two species, Prevotella and Haemophilus, commonly observed in the 75 patient samples were predicted to have much

| Species no. | Species strain name | Avg reads | Sample reads >1% (%) |
|-------------|---------------------|-----------|----------------------|
| 1           | Pseudomonas aeruginosa NCGM2.51 | 0.447     | 85.3                 |
| 2           | Streptococcus sanguinis SK36    | 0.213     | 88.0                 |
| 3           | Prevotella melaninigenica ATCC 25845 | 0.098  | 74.7                 |
| 4           | Escherichia coli strain K-12 substrain MG1655 | 0.029  | 4.0                  |
| 5           | Haemophilus influenzae R2846    | 0.028     | 22.7                 |
| 6           | Burkholderia cepacia GG4        | 0.026     | 4.0                  |
| 7           | Rothia mucilaginosa DY-18       | 0.026     | 48.0                 |
| 8           | Fusobacterium nucleatum subsp. nucleatum ATCC 25586 | 0.023  | 26.7                 |
| 9           | Staphylococcus aureus subsp. aureus USA300 FPR3757 | 0.023  | 34.7                 |
| 10          | Veillonella atypica ACS-049-V-Sch6 | 0.016  | 48.0                 |
| 11          | Achromobacter xylosidans NBRC 15126 | 0.014  | 2.7                  |
| 12          | Gemella haemolysans ATCC 10379  | 0.015     | 30.7                 |
| 13          | Granulicatella adiacens ATCC 49175 | 0.012  | 36.0                 |
| 14          | Neisseria flavescens SK114      | 0.008     | 18.7                 |
| 15          | Actinomyces naeslundii strain Howell 279 | 0.009  | 21.3                 |
| 16          | Porphyromonas endodontalis ATCC 35406 | 0.006  | 20.0                 |
| 17          | Ralstonia sp. 5 7 47FAA         | 0.004     | 6.7                  |

*Shown is a list of the 17 species/strains included in the CF airway community model, the normalized fractional reads for the associated genera averaged across the 75 samples, and the percentage of samples in which the normalized reads exceeded 1%.

FIG 3 Single-species and community simulations performed with the nominal nutrient uptake rates in Table S3. (A) Single-species growth rates with the species numbered according to Table 1. (B) Comparison of predicted species abundances to the average of the normalized reads for the single patient infected with Enterobacteriaceae/Escherichia (samples 25 to 27). (C) Comparison of predicted species abundances to the average of the normalized reads for the single patient infected with Burkholderia (samples 19 to 21). (D) Comparison of predicted species abundances to the average of the normalized reads for the single patient infected with Achromobacter (samples 31 and 32). (E) Comparison of predicted species abundances to the average of the normalized reads for the 43 patients not infected with Enterobacteriaceae/Escherichia, Burkholderia, or Achromobacter (samples 1 to 18, 22 to 24, 28 to 30, and 33 to 75).
lower *in silico* growth rates. The three species representing *Fusobacterium*, *Granulicatella*, and *Porphyromonas* did not grow individually due to their inability to meet the defined ATP maintenance demand, although they could grow when strategically combined with other modeled species. For example, *Fusobacterium*, *Granulicatella*, and *Porphyromonas* were predicted to grow in coculture with *Ralstonia*, *Prevotella*, and *Actinomyces*, respectively. The species abundances predicted for a specified nutrient condition depended on both the monoculture growth rates and the ability of each species to efficiently utilize secreted metabolites to enhance its growth rate. These emergent cross-feeding relationships allowed otherwise slower-growing species to coexist with species that exhibited high monoculture growth rates.

We conducted simulations using the nominal nutrient uptake rates (Table S3) to determine if the community model could capture dominance of each rare pathogen in the absence of the other two rare pathogens. Each simulation was performed by constraining the abundances of the other two pathogens to zero, effectively producing reduced communities of 15 species. The predicted abundances from each simulation were compared to the normalized reads averaged over the patient samples which contained the associated pathogen: *Enterobacteriaceae/Escherichia* (samples 25 to 27) (Fig. 3B), *Burkholderia* (samples 19 to 21) (Fig. 3C), or *Achromobacter* (samples 31 and 32) (Fig. 3D). For each simulated case, the model correctly predicted dominance of the associated pathogen. For the *Burkholderia* - and *Achromobacter*-infected patients, the abundances of the dominant pathogen as well as less prevalent species were well predicted.

We performed simulations for the remaining 43 patients by reducing the community to 14 species by constraining the abundances of all three rare pathogens to zero. The model-predicted abundances were compared to the normalized reads averaged over the 67 samples remaining when the 8 rare pathogen-containing samples were removed (Fig. 3E). The model correctly predicted that *Pseudomonas*, *Streptococcus*, and *Prevotella* would dominate the community, although the *Prevotella* abundance was overpredicted at the expense of *Streptococcus* as well as several less abundant genera. The only other genus present in the simulated community was *Staphylococcus*, while the averaged reads showed a greater amount of diversity. Compared to the averaged data, individual samples showed less diversity, which is more consistent with model predictions as discussed below.

The community model can reproduce pathogen heterogeneity across airway samples. The CF airway communities exhibited a substantial degree of sample-to-sample heterogeneity when rare pathogens were present (Fig. 2A) or absent (Fig. 2B). We performed simulations to assess the extent to which sample-to-sample differences in taxonomic group reads could be explained by heterogeneity in the metabolic environment of the CF lung. More specifically, we randomized the community nutrient uptake rates around their nominal values (Materials and Methods; also see Table S3) to mimic heterogeneous lung environments shown to occur across CF patients (42, 43) and in longitudinal samples from a single patient (44). An objective of our future research will be to model sample-by-sample variability in individual patients as a function of disease state (e.g., clinically stable, pulmonary exacerbation, and antibiotic treatment). In this study, each simulation with a set of randomized uptake rates was termed a “simulated sample,” and we tested the hypothesis that the experimental samples could be interpreted as having been drawn from the much larger set of simulated samples we generated. Due to the relatively small number of *Enterobacteriaceae/Escherichia*, *Burkholderia*, and *Achromobacter*-containing samples, we performed only 100 randomized community simulations for each of these pathogens. In contrast, 1,000 randomized simulations were performed for communities without these three rare pathogens since the associated patient sample size was comparatively large. The single model simulation that best represented a particular patient sample was determined by the minimum least-squares error between the normalized measured reads and the predicted abundances across all simulations. For the 8 rare pathogen-containing samples, we plotted the measured reads and predicted...
abundances of the best-fit models for the five most common genera (\textit{Pseudomonas}, \textit{Streptococcus}, \textit{Prevotella}, \textit{Haemophilus}, and \textit{Staphylococcus}) and the pathogen of interest (Fig. 4; Table S4). For the remaining 67 samples, we plotted the measured reads and predicted abundances of the best-fit models for the five most common genera plus the next most abundant genus according to measured reads (Fig. 5; Table S5).

Randomized nutrient simulations were able to generate model predictions that reproduced the major features of the 3 \textit{Enterobacteriaceae}/\textit{Escherichia}-containing samples (Fig. 4A), including the high-\textit{Enterobacteriaceae}/\textit{Escherichia} reads and the presence of the other main community members (\textit{Pseudomonas}, \textit{Streptococcus}, and \textit{Prevotella}). The \textit{Streptococcus} reads were predicted relatively accurately, while \textit{Pseudomonas} reads were underpredicted and \textit{Prevotella} reads were overpredicted. As measured by the least-squares error, improved predictions were obtained for the 3 \textit{Burkholderia}-containing samples (Fig. 4B). The \textit{Burkholderia} reads were accurately reproduced, and \textit{Streptococcus} was correctly predicted to be the second most abundant genus, suggesting a synergism between these two genera. This prediction has experimental support from \textit{in vitro} experiments showing that mucin-degrading anaerobes such as streptococci promote the growth of CF pathogens such as \textit{Burkholderia cenocepacia} when mucins are provided as the sole carbon source (45). The two \textit{Achromobacter}-containing samples were well predicted in terms of \textit{Achromobacter} reads and \textit{Pseudomonas} being the other dominant genus (Fig. 4C). These predictions are consistent with an \textit{in vitro} study showing that \textit{Achromobacter} sp. enhanced the ability of multiple \textit{P. aeruginosa} strains to form biofilms (46).

Furthermore, a clinical study with 53 patients having positive cultures for \textit{A. xylosoxidans} showed that all 6 patients who were chronically infected by \textit{A. xylosoxidans} were coinfectcd with \textit{P. aeruginosa} (47). Complete comparisons of the normalized measured reads and model predicted abundances for the 8 samples with the rare pathogens are presented in...
Table S4, which shows that the model generally produced less diverse communities as measured by the richness (number of species with abundances exceeding 1%) and the equitability (the inverse Simpson metric \(48\)).

The lack of patient samples containing Enterobacteriaceae/Escherichia, Burkholderia, or Achromobacter limited our ability to analyze heterogeneity of communities with these pathogens. In contrast, the 67 samples remaining when the 8 samples containing these three pathogens were removed offered a much larger data set for heterogeneity analysis. Each of these 67 samples was matched to one of the 1,000 randomized model simulations according to the smallest least-squares error between the normalized reads of the sample and the predicted abundances of the model (Table S5). Representative results are shown for patient samples with relatively small (0.004 to 0.087, Fig. 5A), moderate (0.089 to 0.116, Fig. 5B), and large (0.123 to 0.574, Fig. 5C) error values.

Samples which were most accurately reproduced generally contained high Pseudomonas reads (84% \(\pm\) 15%) with the remainder of the community consisting of Streptococcus and Prevotella (Fig. 5A). These 22 samples were best matched by 11 distinct models, suggesting that patient samples dominated by Pseudomonas contained a higher degree of heterogeneity than the simulated samples.

The 22 samples which produced moderate prediction errors were characterized by lower and more variable Pseudomonas reads (48% \(\pm\) 28%) as well as more variable distributions of Streptococcus and Prevotella reads (Fig. 5B). The ensemble of randomized models could capture the relative amounts of these three genera but often predicted the presence of Staphylococcus not observed in the patient samples. This discrepancy could be attributable to the unmodeled ability of Pseudomonas to secrete diffusible toxins which inhibit Staphylococcus respiration and render Staphylococcus less metabolically competitive in partially aerobic environments (49) such as the CF lung. Interestingly, the model ensemble could reproduce the relatively high Ralstonia reads.

**FIG 5** Taxonomic reads for patient samples without rare pathogens compared to species abundances predicted from community models with randomized nutrient uptake rates. The genera Pseudomonas, Streptococcus, Prevotella, Haemophilus, and Staphylococcus and the next most abundant genera are shown for each case. Individual models that best fit the 67 patient samples were selected from an ensemble of 1,000 14-species models without Enterobacteriaceae/Escherichia, Burkholderia, or Achromobacter. (A) Three representative samples for which the least-squares error measures were within the smallest third of all samples. (B) Three representative samples for which the least-squares error measures were within the middle third of all samples. (C) Three representative samples for which the least-squares error measures were within the largest third of all samples. Each abundance for a patient sample is shown in the first, dark-colored bar, and each abundance predicted by the corresponding model is shown in the second, light-colored bar.
in sample 1 while also predicting no *Ralstonia* in samples 15 and 69. The 23 samples which produced the largest prediction errors were characterized by much lower *Pseudomonas* reads (13%), higher reads of *Streptococcus* and *Prevotella* (34% and 19%, respectively; e.g., samples 26 and 74 in Fig. 5C), and higher representation of less common genera. These samples also produced higher *Haemophilus* reads, primarily due to two *Haemophilus*-dominated samples (e.g., sample 39 in Fig. 5C). While the model ensemble generally was able to reproduce the observed *Streptococcus* and *Prevotella* reads in these samples, the models tended to overpredict *Pseudomonas* and *Staphylococcus* at the expense of the less common genera. In particular, the ensemble underpredicted the abundances of *Rothia*, *Fusobacterium*, and *Gemella* while the average reads of these three genera across the 23 samples summed to 16%. This discrepancy could suggest that these 23 samples were obtained from patients with less advanced CF lung disease, which correlates to higher diversity communities in vivo (30, 50).

To gain further insights into the ability of the community model to mimic sample-to-sample heterogeneity in the absence of rare pathogens, we compared read data and abundance predictions in the PC space calculated from the 67 patient samples. Each of the 1,000 model simulations was mapped into the two-dimensional space defined by the first two principal components (PCs) obtained when PCA is performed on the normalized reads of these patient samples. Predicted species abundances (red circles) from an ensemble of 1,000 models transformed into the PC space of the normalized read data. (B) Enlarged view of the lower left portion of the PCA plot in panel A. (C) Average genus reads obtained for 12 samples (samples 5, 6, 10, 39, 42, 43, 49, 57, 61, 68, 70, and 74) in panel B with elevated *Prevotella* representation compared to the average abundances predicted from the best-fit models for these samples with the species number as in Table 1.

---

**FIG 6** Principal-component analysis (PCA) of taxonomic reads for patient samples without rare pathogens and species abundances predicted from 14-species community models with randomized nutrient uptake rates. (A) Representation of the 67 patient samples (blue crosses labeled with sample number) in the two-dimensional space defined by the first two principal components (PCs) obtained when PCA is performed on the normalized reads of these patient samples. Predicted species abundances (red circles) from an ensemble of 1,000 models transformed into the PC space of the normalized read data. (B) Enlarged view of the lower left portion of the PCA plot in panel A. (C) Average genus reads obtained for 12 samples (samples 5, 6, 10, 39, 42, 43, 49, 57, 61, 68, 70, and 74) with elevated *Prevotella* representation compared to the average abundances predicted from the best-fit models for these samples with the species number as in Table 1.
The model ensemble also could reproduce variations in the second PC, which was heavily weighted by the three dominant genera and Haemophilus, for sufficiently large values of the first PC, which corresponded to relatively high Pseudomonas and low Streptococcus and Prevotella. In contrast, the model ensemble did not cover the patient samples in the lower left quadrant of the PC plot (Fig. 6B). These samples were characterized by unusual combinations of relatively high Prevotella, Haemophilus, Rothia, and/or Fusobacterium that the model could not reproduce in its present form. Of these 12 poorly modeled samples, Prevotella was highly represented in 8 samples. When the normalized reads of these 8 samples and their associated best-fit abundances were averaged, the models overpredicted Pseudomonas, Streptococcus, and Staphylococcus at the expense of the less common genera (Fig. 6C).

The community model predicts that pathogen dominance is driven by metabolite cross-feeding. To investigate putative metabolic mechanisms by which pathogens may establish dominance in the CF lung, we used model predictions to quantify rates of metabolite cross-feeding between species. For each rare pathogen (Escherichia, Burkholderia, and Achromobacter), 100 simulations performed with randomized community uptake rates were used to calculate average exchange rates of the five most significantly cross-fed metabolites between Pseudomonas, Streptococcus, and the pathogen of interest. The overall metabolite exchange rate from one species to another species was calculated by determining the minimum uptake or secretion rate for each exchanged metabolite and then summing these minimum rates over all exchanged metabolites. The overall metabolite exchange rate from one species to another species was calculated by determining the minimum uptake or secretion rate for each exchanged metabolite and then summing these minimum rates over all exchanged metabolites. The arrow thickness is proportional to the overall metabolite exchange rate between the two species. (A) Average exchange rates of the five highest cross-fed metabolites between the three most abundant species for 100 model ensemble simulations containing Escherichia. (B) Average exchange rates of the five highest cross-fed metabolites between the three most abundant species for 100 model ensemble simulations containing Burkholderia. (C) Average exchange rates of the five highest cross-fed metabolites between the three most abundant species for 100 model ensemble simulations containing Achromobacter. (D) Schematic representation of overall metabolite exchange rates for Escherichia-containing communities corresponding to panel A. Pseudomonas was omitted due to its low exchange rates compared to the other two species. (E) Schematic representation of overall metabolite exchange rates for Burkholderia-containing communities corresponding to panel B. (F) Schematic representation of overall metabolite exchange rates for Achromobacter-containing communities corresponding to panel C.
not predicted in Streptococcus monoculture even through the metabolic reconstruction supported L-lactate production (21) (www.vmh.life). While Streptococcus strains are well known to produce L-lactate as the primary product via homolactic fermentation (52, 53), we chose not to manually curate the metabolic reconstruction since in silico L-lactate synthesis was induced by the presence of other community members such as Escherichia. Pseudomonas was minimally involved in metabolite exchange due to its low average abundance (~1%) across the 100 simulations. Hence, our model suggested that organic acid cross-feeding could play a role in Enterobacteriaceae propagation in the CF lung.

More complex cross-feeding relationships were predicted for Burkholderia-containing communities that supported average Pseudomonas and Streptococcus abundances both exceeding 10%. The highest exchange rates were predicted for formate and acetate produced by Streptococcus and consumed by Burkholderia (Fig. 7B and E). The two species also exchanged amino acids, with Streptococcus providing alanine to Burkholderia and Burkholderia producing aspartate and serine for Streptococcus. Burkholderia provided the same two amino acids to Pseudomonas while receiving a small exchange of acetate in return. Pseudomonas also consumed formate secreted by Streptococcus. These model predictions suggested that acetate, formate, and alanine produced by Streptococcus via heterolactic fermentation (52) could promote Burkholderia growth in vivo. Indeed, in vitro experiments have shown that mucin-degrading anaerobes such as streptococci may promote the growth of CF pathogens such as B. cenocepacia by secreting acetate (45).

Compared to the other two pathogens, Achromobacter was predicted to be less efficient at cross-feeding, having only low uptake rates of alanine, L-lactate, and threonine secreted by the other two species. In contrast, Pseudomonas was predicted to benefit from relatively high uptake rates of formate produced by Streptococcus and succinate produced by Achromobacter. Collectively, these model predictions could help explain the enhanced ability of Burkholderia to dominate the simulated CF airway communities compared to Achromobacter (Fig. 4) despite the single-species growth rates of the two species being similar (Fig. 3).

Similar cross-feeding analyses were performed for 1,000 simulations with randomized nutrient uptake rates in 14-species communities lacking Escherichia, Burkholderia, and Achromobacter. To investigate the possibility of differential cross-feeding patterns, the simulations were split into 500 cases with the highest Pseudomonas abundances and 500 cases with the lowest Pseudomonas abundances (Fig. 8A). For each set of 500 simulations, the average exchange rates of the five most significantly cross-fed metabolites between the four most abundant species (Pseudomonas, Streptococcus, Prevotella, and Staphylococcus) were calculated. The overall metabolite exchange rate between any two species were calculated from the individual metabolite uptake and secretion rates as before.

When Pseudomonas abundances were predicted to be relatively high (average of 61%), community interactions were dominated by Pseudomonas consumption of formate, ethanol, acetate, and aspartate secreted by the other three species (Fig. 8B). Formate cross-feeding was predicted to be particularly important, which was consistent with an in vitro study showing that expression of the P. aeruginosa fdnH gene (encoding a formate dehydrogenase) was elevated in synthetic sputum medium compared to glucose minimal medium (54). Similarly, the expression of P. aeruginosa adhA (encoding an alcohol dehydrogenase) was elevated in patient-derived CF sputum compared to in vitro rich medium (55). Since P. aeruginosa strains have the capability to take up both formate and ethanol (56, 57), these in vitro studies suggest that this cross-feeding mechanism could occur in CF airway communities. Staphylococcus was the major source of exchanged formate and ethanol (Fig. 8D), a prediction consistent with studies showing that P. aeruginosa benefits from the presence of S. aureus (49, 58). Both alanine and aspartate have been shown to serve as preferred carbon sources for P. aeruginosa in a minimal medium supplemented with lyophilized CF sputum (54). However, the ensemble model did not predict exchange of L-lactate between P. aeruginosa and S.
aureus, which differs from coculture experiments that mimic the CF lung environment (49). Strong interactions between P. aeruginosa and various streptococci also have been reported (30), although the importance of metabolite cross-feeding in mediating these interactions remains incompletely understood (59). Finally, in the model Pseudomonas supplied small amounts of D-lactate for Prevotella and Staphylococcus consumption, a prediction consistent with an in vitro study showing P. aeruginosa anaerobic production of the LldA enzyme catalyzing D-lactate synthesis (60).

When Pseudomonas abundances were predicted to be relatively low (average of 32%), metabolite cross-feeding remained dominated by Pseudomonas consumption of secreted by-products and amino acids (Fig. 8C). Pseudomonas was predicted to have high consumption rates of formate produced by all three other species and L-lactate synthesized only by Streptococcus, consistent with the ability of Streptococcus salivarius (61) and P. aeruginosa (49) to synthesize and consume L-lactate, respectively. Higher exchange rates between Streptococcus and Staphylococcus were predicted when Pseudomonas abundances were relatively low (Fig. 8E). The two species cross-fed alanine and L-lactate produced by Streptococcus and aspartate and ethanol secreted by Staphylococcus. Our predicted cross-feeding relationships in Pseudomonas- and Streptococcus-dominated communities could provide insights into CF disease progression, as high abundances of Streptococcus relative to Pseudomonas have been shown to correlate with higher-diversity airway communities and improved CF clinical stability (30). Younger CF patients also are known to have more diverse airway communities (62), so such interpretations would need to be made with care.

DISCUSSION

The airways of cystic fibrosis (CF) patients are commonly infected by complex communities of interacting bacteria, fungi, and viruses which complicate disease assessment and treatment. The unique bacterial communities resident in individual
patients can be longitudinally resolved to the genus level by applying 16S rRNA gene amplicon library sequencing to sputum and bronchoscopy samples (8). While 16S rRNA gene sequencing technology provides an unprecedented capability to identify bacterial pathogens in the CF lung, other analyses are required to understand how community members interact and how these interactions impede or promote disease progression. Metabolomics represents a powerful tool to interrogate the complex metabolic environment of the CF lung (63), but the number and depth of studies published to date have been limited. Metabolic modeling is a complementary tool for probing complex microbial communities and their interactions mediated through competition for host-derived nutrients and cross-feeding of secreted metabolites (13). Community metabolic models can provide information difficult to obtain by purely experimental means, such as the combined impact of nutrient environment and metabolic interactions on community composition. Metabolic models also can predict the rates of metabolite exchange between species and identify cross-feeding relationships difficult to delineate through metabolomic analyses.

We used 16S rRNA gene sequence data from three published studies (28–30) to construct and test a metabolic model for prediction of airway community compositions in adult CF patients. The assembled data set consisted of 75 distinct samples from 46 patients who were judged to be stable or recovered from treatment in the original studies. Principal-component analysis performed on 16S read data showed considerable heterogeneity of community composition across the 75 samples, including three patients infected with Enterobacteriaceae, Burkholderia, and Achromobacter pathogens. Interestingly, each of these three patients was infected by only one of these “rare” pathogens, a characteristic we used to simplify our metabolic model simulations. The remaining 67 samples from 43 patients were largely dominated by Pseudomonas and/or Streptococcus but still exhibited substantial composition heterogeneity, which provided a sufficiently rich data set to explore sample-to-sample variability.

The community metabolic model was constructed by ranking the identified taxa according to their total reads across the 75 samples and representing each taxonomic group with a single genome-scale metabolic reconstruction obtained from the AGORA database (www.vmh.life) (21). To limit model complexity, only the 17 top-ranked taxa (16 genera and 1 combined family/genus) were included. The resulting in silico community contained the most common CF pathogens (Pseudomonas aeruginosa, Haemophilus influenzae, and Staphylococcus aureus), “rare” pathogens (Escherichia coli, Burkholderia cepacia, and Achromobacter xylosoxidans), and 11 other species commonly observed in the CF sputum samples (e.g., Prevotella melaninogenica, Rothia mucilaginosa, Fusobacterium nucleatum). The 17 modeled taxa provided substantial coverage of the read data with an average coverage of 95.6% ± 3.9% across the 75 samples. Because our in silico objective of growth rate maximization tends to produce low-diversity communities dominated by ~5 species (41), the relatively low diversity of these adult CF lung samples made them particularly well suited for analysis through metabolic modeling compared to considerably more diverse bacterial communities found elsewhere in the human body (e.g., the intestinal tract [41, 64] and chronic wounds [65]).

The community metabolic model required specification of host-derived nutrients that mimicked the CF lung environment in terms of the nutrients available, their allowed uptake rates across the community, and their allowed uptake rates by individual species. Given that the 17-species model contained 271 community uptake rates and a total of 2,378 species-specific uptake rates, a model tuning method was developed to manage the daunting complexity. A putative list of host-derived nutrients was compiled by starting with the synthetic sputum medium SCFM2 (66) and adding other nutrients either required for monoculture growth of at least one modeled species, measured in metabolomic analyses of CF sputum samples, or identified through in silico analyses. The resulting 81 nutrients were separated into 14 distinct groups (see Table S3 in the supplemental material) to facilitate tuning of nominal community uptake rates to qualitatively match average read data for the rare
pathogen samples and the *Pseudomonas/Streptococcus*-dominated samples. This tuning process proved to be the bottleneck of model development even under the simplifying assumption that the species uptake rates were not limiting. A more streamlined and experimentally driven tuning process would be facilitated by the availability of matched 16S and metabolomics data for large sets of CF sputum samples.

Despite the challenges associated with defining physiologically relevant nutrient uptake rates, the community model was able to predict species abundance in qualitative agreement with average read data for *Enterobacteriaceae*, *Burkholderia*, *Achromobacter*, and *Pseudomonas/Streptococcus*-dominated samples. The modeling effort was simplified by omitting the other two rare pathogens when simulating the 3 *Enterobacteriaceae*, 3 *Burkholderia*, and 2 *Achromobacter*-containing samples and omitting all three rare pathogens when simulating the other 67 samples, as justified through analysis of the 16S rRNA gene sequence data. The 15-species models used to simulate the rare-pathogen-containing samples were able to reproduce dominance of the associated pathogen and, to a lesser extent, the abundances of less prevalent species. However, satisfactory prediction of the 2 *Achromobacter*-containing samples required the addition of four carbon sources (arabinose, fumarate, galactonate, and xylose) which have not been measured in the CF lung to our knowledge. While there is some experimental evidence to support their inclusion, the need to add these four metabolites to elevate *in silico* *Achromobacter* growth could point to limitations of the modeled nutrients and their defined uptake rates.

The 14-species model used to simulate the rare-pathogen-free samples predicted that *Pseudomonas* and *Streptococcus* would be the dominant genera and that *Prevotella* and *Staphylococcus* also would be present in the community. These predictions provided qualitative agreement with the 16S rRNA gene sequence read data averaged across the 67 samples, although the predicted abundance of *Prevotella* was comparatively high and the predicted diversity was comparatively low. Given the uncertainty associated with identifying host-derived nutrients and translating these available nutrients into appropriate community uptake rates, we considered our predictions to provide satisfactory *in silico* recapitulation of measured community compositions across the set of four dominant CF pathogens.

A hallmark of CF lung infections is poorly understood differences in bacterial community compositions between patients and in longitudinal samples collected from a single patient (42). We performed simulations to test the hypothesis that these differences might be partially attributable to sample-to-sample variations in the nutrient environment in the CF lung. Nutrient variability was simulated by randomizing the community uptake rates around their nominal values found through manual model tuning. We performed 100 model ensemble simulations for each 15-species community containing a rare pathogen to determine if the associated patient samples could be well fit by a simulated sample. Using comparative plots of the measured reads and predicted abundances, we found that the model ensembles could satisfactorily reproduce the community compositions of the 8 rare-pathogen-containing samples. The best-fit models tended to provide good predictions of rare pathogen reads due to their relatively large values (average of 65% across the 8 samples), while the accuracy of read predictions for less prevalent species was more variable.

Due to the availability of a much larger data set of 67 patient samples, the rare-pathogen-free model consisting of 14 species afforded an opportunity to investigate sample-to-sample heterogeneity in more depth. We performed 1,000 model ensemble simulations with randomized nutrient uptake rates to find best-fit models. Patient samples with relatively high *Pseudomonas* reads tended to be well fit because the model predicted *Pseudomonas* dominance over a wide range of nutrient conditions. Less accurate but still satisfactory fits were obtained for patient samples with moderate *Pseudomonas* and relatively high *Streptococcus* reads. The model ensemble proved somewhat deficient in fitting samples with high reads of *Prevotella* or of the less common genera *Haemophilus*, *Rothia*, and *Fusobacterium*. This deficiency could be attributable to the *in silico* lung environment not containing key nutrients and/or not
specifying sufficiently high uptake rates of supplied nutrients to support high abundances of these genera.

The quality of sample fits also was correlated with the sample diversity, with the best fits having the lowest average diversity (inverse Simpson index of 0.10), moderate fits having an intermediate average diversity (inverse Simpson index of 0.18), and poor fits having the highest average diversity (inverse Simpson index of 0.23). For these three sets of samples, the best-fit models had average diversities of 0.10, 0.16, and 0.20, respectively. We believe that the lower predicted diversities were attributable to the modeling assumption that the CF lung community maximizes its collective growth rate. Using a community metabolic model of the human gut microbiota (41), we have shown that increased bacterial diversity (typically associated with health) can be achieved by simulating suboptimal growth rates under the hypothesis that disease progression correlates with a collective movement toward maximal growth. Therefore, the assumption of maximal community growth may inherently limit our ability to accurately reproduce more diverse samples and rather simulate conditions associated with disease, such as dominance of a single pathogen.

By optimizing cross-feeding of secreted metabolites, the community model was able to predict the coexistence of multiple species at the maximal community growth rate rather than just predicting a monoculture of the single species with the highest monoculture growth rate. Because the SteadyCom method (27) used to formulate and solve the community model does not allow direct incorporation of mechanisms by which one species could inhibit the growth of another species other than by nutrient competition, the predicted community growth rate always was higher than the highest individual growth rate of the coexisting species. Consequently, the formulated model was incapable of capturing more complex interactions such as *Pseudomonas* secretion of diffusible toxins that inhibit the growth of other CF pathogens (67).

Despite this limitation, the community model could be analyzed to understand the putative role of metabolite cross-feeding in shaping community composition. The model predicted that the rare pathogens *Escherichia* and *Burkholderia* were particularly efficient cross-feeders, using acetate, formate, and other secreted metabolites to establish dominance over less harmful bacteria. In contrast, the model predicted *Achromobacter* to be substantially less adept at exploiting secreted metabolites for growth enhancement. While we were able to simulate *Achromobacter* dominance through addition of four carbon sources possibly present in the CF lung, the model suggested that other nonmodeled mechanisms may be involved in promoting *Achromobacter* expansion. One possibility is that *Achromobacter* utilizes its ability to form multispecies biofilms (46, 68) to establish favorable metabolic niches for enhanced growth.

In the absence of the three rare pathogens, the model predicted that *Pseudomonas* would be the primary beneficiary of cross-fed metabolites, including acetate, alanine, and L-lactate from *Streptococcus* and aspartate, ethanol, and formate from *Staphylococcus*. Similar cross-feeding relationships have been observed in an in vitro coculture system in which *P. aeruginosa* consumed alanine and lactate secreted by *R. mucilaginosa* (69). The predicted cross-feeding behavior was an emergent property of the community model that could not be predicted from monoculture simulations and is consistent with published experimental data presented above. For example, the single-species models predicted that acetate, CO2, and formate would be the primary secreted by-products, yet the community model also cross-fed ethanol, D-lactate, L-lactate, and succinate, which were not predicted to be secreted in any monoculture simulation. We hypothesized that model ensemble simulations with relatively high and low *Pseudomonas* abundances would show differential cross-feeding patterns. While some of the specific cross-fed metabolites changed between the two cases, cross-feeding from *Streptococcus* and *Staphylococcus* to *Pseudomonas* remained the dominant feature of the simulated communities. In our assimilated data set of 75 patient samples, *Pseudomonas* reads were above 10% in 55 samples and above 50% in 35 samples. Our model predictions provide putative metabolic mechanisms that may help explain why
Pseudomonas so efficiently colonizes the adult CF lung and why Pseudomonas commonly establishes dominance over other species once colonized.

Our community metabolic model generated several predictions that could be tested experimentally with an appropriately designed in vitro community. For example, a 5-species in vitro system consisting of Pseudomonas aeruginosa, Streptococcus sanguinis, Prevotella melaninogenica, Haemophilus influenzae, and Staphylococcus aureus would provide substantial coverage of our 16S rRNA gene sequencing data, as the five genera accounted for an average of 87% of normalized reads across the 67 rare-pathogen-free samples and greater than 75% of normalized reads in 56 of these samples. Specific model predictions that could be tested in vitro include the variability of community compositions by changing nutrient levels in a synthetic CF medium and the cross-feeding of specific metabolites by genetically altering the secretion and/or uptake capabilities of these metabolites in the relevant species. The availability of such in vitro data linking the nutrient environment, cross-feeding mechanisms, and community composition would allow direct testing of a simplified 5-species model and facilitate the development of improved community models for the analysis of CF sputum samples.

MATERIALS AND METHODS

Patient data. CF airway community composition data were obtained from three published studies in which patient sputum samples were subjected to 16S rRNA gene amplicon library sequencing (28–30). The first study (28) included 30 samples from 10 clinically stable adults ranging in age from 20 to 50 years with an average age of 25 years, the second study (29) included 23 samples from 14 adults in clinically defined baseline and recovery stages ranging in age from 18 to 69 years with an average age of 34 years, and the third study (30) included 22 samples from 22 clinically stable adults ranging in age from 19 to 52 years with an average age of 28 years. Thus, in total, the assimilated data set contained 75 distinct samples from 46 patients who were clinically stable or recovered from treatment for an exacerbation event. Additional samples from these three studies corresponding to exacerbation or antibiotic treatment were not included in the modeled data set to avoid the complications of predicting these events. The top 72 taxonomic groups (typically genera) accounted for over 99.8% of total reads across the 75 samples (Fig. 1A; also see Table S6 in the supplemental material). To limit complexity, the community metabolic model described below was limited to 17 taxonomic groups that accounted for 95.6% of total reads (Fig. 1B; Table S4). Reads from the family Enterobacteriaceae and the genus Escherichia were combined and represented as a single genus. To allow direct comparison with the species abundances predicted by the model, the reads for each sample were normalized over the 17 modeled genera to sum to unity (Table S5).

Community metabolic model. For simplicity, each genus was represented by a single species commonly observed in CF airway communities (1, 6–9, 70), although we note that genera such as Streptococcus (30) can have considerable diversity with respect to species representation. As mentioned above, the combined Enterobacteriaceae/Escherichia taxonomic group was represented by the single species Escherichia coli. A genome-scale metabolic reconstruction for each species (Fig. 1C) was obtained from a large database of AGORA models (21) (www.vmh.life). Table 1 lists the representative strain used for each genus, the normalized reads fractionally associated with each genus averaged across the 75 samples (also shown in Fig. 1B), and the number of samples for which the normalized reads exceeded 1%. The community model accounted for 13,845 genes, 19,034 metabolites, and 22,412 reactions within the 17 species as well as 271 uptake and secretion reactions for the extracellular space shared by the species.

The genera Pseudomonas, Streptococcus, and Prevotella dominated most communities, in terms of both average reads for individual samples and the number of samples in which they exceeded 1%. Interestingly, Enterobacteriaceae/Escherichia, Burkholderia, and Achromobacter exceeded 0.1% in only single patients represented by 3, 3, and 2 samples, respectively. Moreover, no patients were infected by more than one of these “rare” pathogens, as the maximum reads of the other two pathogens never exceeded 0.1% in these 8 samples. Therefore, for modeling purposes the 75 samples were partitioned into 3 Enterobacteriaceae/Escherichia-containing samples with Burkholderia and Achromobacter absent, 3 Burkholderia-containing samples with Enterobacteriaceae/Escherichia and Achromobacter absent, 2 Achromobacter-containing samples with Enterobacteriaceae/Escherichia and Burkholderia absent, and 67 samples with all three rare pathogens absent.

Model tuning and simulation. The nutrient environment in the CF lung is complex and expected to vary between patients as well as between longitudinal samples for individual patients depending on disease state. While metabolomic analyses have been performed on CF sputum and bronchoscopy samples (42, 63, 70, 71), these studies were insufficient to define supplied nutrients for the metabolic model due to their limited metabolite coverage. Furthermore, we found that based on our model, the synthetic sputum medium SCFM2 used in previous in vitro CF microbiota studies (66, 72) would not support growth of any of the 17 modeled species due to the lack of ions (Co^{2+}, Cu^{2+}, Mn^{2+}, and Zn^{2+}), amino acids (asparagine and glutamine), and other metabolites (see below) essential for growth. While the medium likely would contain trace amounts of the missing ions, the requirement of these other
metabolites for growth suggests limitations for the AGORA metabolic models with respect to biosynthetic pathways leading to biomass formation. Given the semicurated nature of the AGORA models (21), such discrepancies were expected and had to be addressed by adding the missing essential metabolites to the modeled medium. A final complication was that the community model required specification of nutrient uptake rates, which were unknown even if medium component concentrations were specified due to the lack of species-dependent uptake kinetics for each nutrient. Because such uptake information is rarely available even for highly studied model organisms such as _Escherichia coli_ (73), a simplified approach was used to define nutrient uptake rates for the community model.

Supplied nutrients in the community model were defined by starting with the SCFM2 medium and adding the four ions and two amino acids listed above. We found that each species required additional metabolites in the medium to support biomass formation. These 29 additional metabolites were identified and added to the modeled medium such that all 17 species were capable of monoculture growth (see Table S3). For example, the _P. aeruginosa_ model required addition of uracil and menaquinone 7, while _in vitro_ experiments have shown that these metabolites are synthesized _de novo_ and not required in the medium (66). Next, we added four carbon sources (fructose, maltose, maltotriose, and pyruvate) and 8 other metabolites (adenosine, cytidine, glycerol, guanosine, hexadecanoate, inosine, octadecenoate, and uridine) measured in the CF lung (71) and the terminal electron acceptor O₂ to simulate aerobic respiration. Finally, we added four additional carbon sources (arabinose, fumarate, galactonate, and xylene) that increased _in silico_ Achromobacter growth such that _Achromobacter_ would be competitive with other species when it was present in the community. While these carbon sources were identified _in silico_, there is experimental evidence to support their inclusion in the simulated CF lung environment. Fumarate has been shown to be elevated in sputum samples from young CF patients (74). Arabinose and xylose are constituents of extracellular polymer substance (EPS) produced by common human pathogens the modeled genera _Pseudomonas_, _Staphylococcus_, and _Escherichia_ (75), suggesting their possible presence in the CF lung. Pathogenic _Achromobacter_ strains isolated from CF patients have been shown to grow on galactonate as a sole carbon source (76), supporting the hypothesis that _Achromobacter_ has evolved to utilize galactonate available in the CF lung.

The community uptake rates of the 86 supplied nutrients were tuned by trial and error to produce species abundances in approximate agreement with the average reads listed in Table 1, which were derived from actual patient samples. To reduce the number of adjustable rates, the nutrients were grouped together and a single uptake rate was used for each group. These 14 groups (Table S3) were defined as follows: group 1, 16 common metals and ions; group 2, 29 essential growth metabolites; group 3, 8 CF lung metabolites; group 4, 19 amino acids; group 5, the amino acids alanine and valine, which have been reported to be consumed for the uptake rate of each metabolite also are listed in Table S3. The CF lung is known to exhibit sharp O₂ gradients such that some regions are hypoxic or even anoxic (77, 78). The community model accounted for the effects of the average O₂ level through the randomized uptake rates. At the nominal oxygen uptake rate of 5 mmol/g dry weight (gDW)/h in Table S3, the 17 species had an average growth rate of 0.140 h⁻¹. At the low oxygen uptake value of 1.67 mmol/gDW/h, the 17 species had an average growth rate of 0.096 h⁻¹. Given that the maximum O₂ uptake rate of _E. coli_ has been reported as 20 mmol/gDW/h (79), the range of O₂ uptake rates in Table S4 spans from highly to moderately hypoxic lung environments.

**Community simulations.** We used the SteadyCom method (27) to perform steady-state community simulations as detailed in our previous study on the human gut microbiota (41). SteadyCom performs community flux balance analysis by computing the relative abundance of each species for maximal community growth while ensuring that all metabolites are properly balanced within each species and across the community. This simulation method is based on several simplifying assumptions, including that each sputum sample was obtained from a spatially homogeneous region of the CF lung, that all modeled species have an equal opportunity to colonize the airway, and that all propagating species have the same growth rate at steady state. Therefore, the community model was not capable of predicting sequential colonization by various species (45) or different growth rates of propagating species (80). Each species model used a non-growth-associated ATP maintenance (ATPM) value of 5 mmol/gDW/h, which is within the range reported for curated bacterial reconstructions. Cross-feeding of all 21 amino acids and 8 common metabolic by-products (acetate, CO₂, ethanol, formate, H₂, β-lactate, α-lactate, and succinate) was promoted by increasing the maximum nutrient uptake rates of these nutrients in each species model to 2.5 and 5 mmol/gDW/h, respectively. The nominal nutrient uptake rates produced a single community not directly comparable to any single patient sample (Fig. 1E), while each set of randomized uptake rates produced a unique community that was interpreted as a prediction of an individual patient sample (Fig. 1F). Outputs of each SteadyCom simulation included the community growth rate, the abundance of each species, and species-dependent uptake and secretion rates of each extracellular metabolite. The
overall difference between the normalized reads and the predicted species abundances for a single patient sample was quantified with the least-squares error measure,

$$
\Psi = \sum_{i} (p_i - \hat{p}_i)^2
$$

where $p_i$ is the normalized reads for species $i$ (Table S8), $\hat{p}_i$ is the predicted abundance of species $i$, and $n = 17$ is the number of species in the community model. Least-squares errors are a common measure of the differences between two vectors. The error measure is relative in the sense that smaller values are preferred, but the specific value that delineates “good” and “poor” model fits is problem dependent.

**Data availability.** All data used for metabolic model development and testing are provided in the supplemental material.

**SUPPLEMENTAL MATERIAL**

Supplemental material for this article may be found at https://doi.org/10.1128/mSystems.00026-19.

- **TABLE S1**, XLSX file, 0.01 MB.  
- **TABLE S2**, XLSX file, 0.01 MB.  
- **TABLE S3**, XLSX file, 0.02 MB.  
- **TABLE S4**, XLSX file, 0.01 MB.  
- **TABLE S5**, XLSX file, 0.01 MB.  
- **TABLE S6**, XLSX file, 0.1 MB.  
- **TABLE S7**, XLSX file, 0.03 MB.  
- **TABLE S8**, XLSX file, 0.1 MB.

**ACKNOWLEDGMENTS**

We acknowledge the NIH grants U01-EB019416 (M.A.H. and P.P.), R37 AI83256-06 (G.A.O.), and T32-AI007519 (G.O.) for partial support of this research. This work was supported in part by a Fellowship from the University of Massachusetts to Poonam Phalak as part of the Biotechnology Training Program (National Research Service Award T32-GM108556).

**REFERENCES**

1. Lyczak JB, Cannon CL, Pier GB. 2002. Lung infections associated with cystic fibrosis. Clin Microbiol Rev 15:194–222. [https://doi.org/10.1128/CMR.15.2.194-222.2002](https://doi.org/10.1128/CMR.15.2.194-222.2002).

2. Tang AC, Turvey SE, Alves MP, Regamey N, Tümmler B, Hartl D. 2014. Current concepts: host-pathogen interactions in cystic fibrosis airways disease. Eur Respir Rev 23:320–332. [https://doi.org/10.1183/09059180.0006113](https://doi.org/10.1183/09059180.0006113).

3. Bronstein M, Sokol R, Abman S, Chatfield B, Hammond K, Hambidge K, Stall C, Accurso F. 1992. Pancreatic insufficiency, growth, and nutrition in infants identified by newborn screening as having cystic fibrosis. J Pediatr 120:533–540. [https://doi.org/10.1016/0022-3476(85)90278-3](https://doi.org/10.1016/0022-3476(85)90278-3).

4. Orrskog S, Medin E, Tsolova S, Semenza JC. 2013. Causal inference regarding infectious aetiology of chronic conditions: a systematic review. PLoS One 8:e68861. [https://doi.org/10.1371/journal.pone.0068861](https://doi.org/10.1371/journal.pone.0068861).

5. Govan J, Nelson J. 1993. Microbiology of cystic fibrosis lung infections: themes and issues. J R Soc Med 86:11.

6. LiPuma JJ. 2010. The changing microbial epidemiology in cystic fibrosis. Clin Microbiol Rev 23:299–323. [https://doi.org/10.1128/CMR.00068-09](https://doi.org/10.1128/CMR.00068-09).

7. Filkins LM, O’Toole GA. 2015. Cystic fibrosis lung infections: polymicrobial, complex, and hard to treat. PLoS Pathog 11:e1005258. [https://doi.org/10.1371/journal.ppat.1005258](https://doi.org/10.1371/journal.ppat.1005258).

8. Lynch SV, Bruce KD. 2013. The cystic fibrosis airway microbiome. Cold Spring Harb Perspect Med 3:a009738. [https://doi.org/10.1101/cshperspect.a009738](https://doi.org/10.1101/cshperspect.a009738).

9. Van Der Gast CJ, Walker AW, Stressmann FA, Rogers GB, Scott P, Daniels TW, Carroll MP, Parkhill J, Bruce KD. 2011. Partitioning core and satellite themes and issues. J R Soc Med 86:11.

10. Lynch SV, Bruce KD. 2013. The cystic fibrosis airway microbiome. Cold Spring Harb Perspect Med 3:a009738. [https://doi.org/10.1101/cshperspect.a009738](https://doi.org/10.1101/cshperspect.a009738).

11. Layeghifard M, Li H, Wang PW, Donaldson SL, Coburn B, Clark ST, Caballero JD, Zhang Y, Tullis DE, Yau YCW, Waters V, Hwang DM, Guttman DS. 2019. Microbiome networks and change-point analysis reveal key community changes associated with cystic fibrosis pulmonary exacerbations. NPJ Biofilms Microbiomes 5:4. [https://doi.org/10.1038/s41522-018-0077-y](https://doi.org/10.1038/s41522-018-0077-y).

12. Quinn RA, Whiteson K, Lim YW, Zhao J, Conrad D, LiPuma JJ, Rohwer F, Widder S. 2016. Ecological networking of cystic fibrosis lung infections. NPJ Biofilms Microbiomes 2:4. [https://doi.org/10.1038/s41522-016-0002-1](https://doi.org/10.1038/s41522-016-0002-1).

13. Perez-Garcia O, Lear G, Singhal N. 2016. Metabolic network modeling of microbial interactions in natural and engineered environmental systems. Front Microbiol 7:673. [https://doi.org/10.3389/fmicb.2016.00673](https://doi.org/10.3389/fmicb.2016.00673).

14. Hanemaaijer M, Röling WF, Olivier BG, Khandelwal RA, Teusink B, Bruggeman FJ. 2015. Systems modeling approaches for microbial community studies: from metagenomics to inference of the community structure. Front Microbiol 6:213. [https://doi.org/10.3389/fmicb.2015.00213](https://doi.org/10.3389/fmicb.2015.00213).

15. Freilich S, Zarecki R, Eilam O, Segal ES, Henry CS, Kupiec M, Gophna U, Sharan R, Ruppin E. 2011. Competitive and cooperative metabolic interactions in bacterial communities. Nat Commun 2:589. [https://doi.org/10.1038/ncomms1597](https://doi.org/10.1038/ncomms1597).

16. Shoiea S, Karlsson F, Mardinoglu A, Nooakew I, Bordel S, Nielsen J. 2013. Understanding the interactions between bacteria in the human gut through metabolic modeling. Sci Rep 3:2532. [https://doi.org/10.1038/srep02532](https://doi.org/10.1038/srep02532).

17. Heinken A, Thiele I. 2015. Anoxic conditions promote species-specific mutualism between gut microbes in silico. Appl Environ Microbiol 81:4049–4061. [https://doi.org/10.1128/AEM.00101-15](https://doi.org/10.1128/AEM.00101-15).

18. Levy R, Borenstein E. 2013. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. Proc Natl Acad Sci U S A 110:12804–12809. [https://doi.org/10.1073/pnas.1300926110](https://doi.org/10.1073/pnas.1300926110).

19. Pinto F, Medina DA, Pérez-Correa JR, Garrido D. 2017. Modeling metabolic interactions in a consortium of the infant gut microbiome. Front Microbiol 8:2507. [https://doi.org/10.3389/fmicb.2017.02507](https://doi.org/10.3389/fmicb.2017.02507).

20. Cuevas DA, Edirisinghe J, Henry CS, Overbeek R, O’Connell TG, Edwards RA. 2016. From DNA to FBA: how to build your own genome-scale...
35. Fang K, Zhao H, Sun C, Lam CM, Chang S, Zhang K, Panda G, Godinho M, Heinken A, Thiele I. 2015. Systematic prediction of health-relevant human-microbial co-metabolism through a computational framework. Gut Microbes 6:120–130. https://doi.org/10.1080/19490976.2015.1023494.

36. Shoae S, Ghaffari P, Kovatcheva-Datchary P, Mardinoiu A, Sen P, Pujos-Guillot E, de Wouters T, Juste C, Rizkalla S, Chilloux J, Hoyles L, Nicholson JK, Dore J, Dumas ME, Clement K, Bäckhed F, Nielsen J. 2015. Quantifying diet-induced metabolic changes of the human gut microbiome. Cell Metabolism 22:320–331. https://doi.org/10.1016/j.cmet.2015.07.001.

37. Chen SJ, Simons MN, Maranas CD. 2017. SteadyCom: predicting microbial abundances while ensuring community stability. PLoS Comput Biol 13:e1005339. https://doi.org/10.1371/journal.pcbi.1005339.

38. Hampton TH, Green DM, Cutting GR, Morrison HG, Sogin ML, Gifford AH, Filkins L, Hampton T, Greenhalgh K, Jäger C, Baginska J, Wilmes P, Fleming RTM, Thiele I. 2016. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. Nat Biotechnol 35:81. https://doi.org/10.1038/nbt.3703.

39. Barnard J, Aziz RK, Fierer N, Knight R, Gilbert J, Betley J, Goyal J, Sunesen M, Greenhalgh K, Ahn C, Huang X, Schloss PD. 2016. Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity. Proc Natl Acad Sci U S A 113:E3801–E38E9. https://doi.org/10.1073/pnas.1521991113.

40. Fang K, Zhao H, Sun C, Lam CM, Chang S, Zhang K, Panda G, Godinho M, Santos VAM, Wang J. 2011. Exploring the metabolic network of the epidemic pathogen Burkholderia cenocepacia J2315 via genome-scale reconstruction. BMC Syst Biol 5:83. https://doi.org/10.1186/1752-0509-5-83.

41. Bartel JA, Yen P, Vargas J, Goldberg JB, Papin JA. 2014. Comparative metabolic systems analysis of pathogenic Burkholderia. J Bacteriol 196:210–226. https://doi.org/10.1128/JB.00997-13.

42. Noecker C, Eng A, Srinivasan S, Theriot CM, Young VB, Janssen JK, Fredrickds DN, Borenstein E. 2016. Metabolic model-based integration of microbiome taxonomic and metabolomic profiles elucidates mechanistic links between ecological and metabolic variation. mSystems 1:e00013-15. https://doi.org/10.1128/mSystems.00013-15.

43. Magnúsdóttir S, Simons M, Maranas CD. 2018. Modeling co-metabolism of the human gut microbiome.Curr Opin Biotechnol 51:90–96. https://doi.org/10.1016/j.copbio.2017.12.005.

44. Heinken A, Ravcheev DA, Baldini F, Heinend L, Fleming RM, Thiele I. 2017. Personalized modeling of the human gut microbiome reveals distinct bile acid deconjugation and biotransformation potential in healthy and IBD individuals. bioRxiv https://doi.org/10.1101/229138.

45. Zomorrodi AR, Maranas CD. 2012. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. PLoS Comput Biol 8:e1002363. https://doi.org/10.1371/journal.pcbi.1002363.

46. Heinken A, Thiele I. 2015. Systematic prediction of health-relevant human-microbial co-metabolism through a computational framework. Gut Microbes 6:120–130. https://doi.org/10.1080/19490976.2015.1023494.

47. Faria JP, Rocha M, Rocha I, Henry CS. 2018. Methods for automated genome-scale metabolic model reconstruction. Biochem Soc Trans 46:931–936. https://doi.org/10.1042/BST20172046.

48. Khandelwal RA, Oliveira BV, Röling WF, Teusink B, Bruggeman FJ. 2013. Community flux balance analysis for microbial consortia at balanced growth. PLoS One 8:e64567. https://doi.org/10.1371/journal.pone.0064567.

49. Samadpour M, Amin A, Rezaei-Tavirani M, Messori G, Aragona M, Di Nardo C, Manfredi G, Mazzoccoli G, Marasco A, Oudshoorn P, De Vos WM, Belland R, Martens-Lobenhoffer L, Reher D. 2015. Pseudomonas aeruginosa: the role of swarming motility in collective behavior of bacterial communities. mSystems 1:e000658. https://doi.org/10.1128/mSystems.000658.

50. Quinn RA, Phelan VV, Whitson KL, Garg N, Bailey BA, Lim YW, Conrad DJ, Dorrestein PC, Rohwer FL. 2016. Microbial host and xenobiotic diversity in the cystic fibrosis sputum metabolome. ISME J 10:1483. https://doi.org/10.1038/ismej.2015.207.

51. Nobakht BF, Allejadnejad R, Rezaei-Tavirani M, Taheri S, Oskouie AA. 2015. The metabolomics of airway diseases, including COPD, asthma and cystic fibrosis. Biomarkers 20:5–16. https://doi.org/10.1080/19490976.2015.1023494.

52. Henson MA, Halak R, Halak R. 2018. Suboptimal community growth mediated through metabolite crossing promotes species diversity in the gut microbiota. PLoS One 14:e0166558. https://doi.org/10.1371/journal.pone.0166558.

53. Keevil C, Marsh P, Ellwood D. 1984. Regulation of glucose metabolism in Klebsiella pneumoniae by the cyclic AMP-dependent protein kinase. J Bacteriol 157:560–567. https://doi.org/10.1128/JB.157.2.560-567.1984.

54. Palmer KL, Mashburn LM, Singh PK, Whiteley M. 2005. Cystic fibrosis airway communities. FEMS Immunol Med Microbiol 45:264–276. https://doi.org/10.1016/j.femsim.2005.06.013.
65. Wolcott RD, Hanson JD, Rees EJ, Koenig LD, Phillips CD, Wolcott RA, Cox J, Li J, Jia H, Cai X, Zhong H, Feng Q, Sunagawa S, Arumugam M, Kultima J, Scoffield JA, Wu H. 2015. Oral streptococci and nitrite-mediated interference of Pseudomonas aeruginosa. Infect Immun 83:101–107. https://doi.org/10.1128/IAI.02396-14.

66. Klepac-Ceraj V, Lemon KP, Martin TR, Allgaier M, Kembel SW, Knapp AA, Lory S, Brodie EL, Lynch SV, Bohannan BJ. 2010. Relationship between cystic fibrosis respiratory tract bacterial communities and age, genotype, antibiotics and Pseudomonas aeruginosa. Environ Microbiol 12:1293–1303. https://doi.org/10.1111/j.1462-2920.2010.02173.x.

67. O’Brien S, Fothergill JL. 2017. The role of multispecies social interactions of Pseudomonas aeruginosa in shaping bacterial community structure and environment in the cystic fibrosis airway during exacerbation. mPloS One 8:e82432. https://doi.org/10.1371/journal.pone.0082432.

68. Turner KH, Wessel AK, Palmer GC, Murray JL, Whiteley M. 2015. Essential genome of Pseudomonas aeruginosa in cystic fibrosis sputum. Proc Natl Acad Sci U S A 112:4110–4115. https://doi.org/10.1073/pnas.1419677112.

69. Meadows AL, Karnik R, Lam H, Forrestell S, Snedecor B. 2010. Application of dynamic flux balance analysis to an industrial Escherichia coli fermentation. Metab Eng 12:150–160. https://doi.org/10.1016/j.meben.2009.07.006.

70. Esther CR, Turkovic L, Rosenow T, Muhlebach MS, Boucher RC, Ranganathan S, Stick SM. 2016. Metabolomic biomarkers predictive of early structural lung disease in cystic fibrosis. Eur Respir J 48:1612–1621. https://doi.org/10.1183/13993003.00524-2016.

71. Bales PM, Renke EM, May SL, Shen Y, Nelson DC. 2013. Purification and characterization of biofilm-associated EPS exopolysaccharides from ES-KAPE organisms and other pathogens. PLoS One 8:e67950. https://doi.org/10.1371/journal.pone.0067950.

72. Turner KH, Wessel AK, Palmer GC, Murray JL, Whiteley M. 2015. Essential genome of Pseudomonas aeruginosa in cystic fibrosis sputum. Proc Natl Acad Sci U S A 112:4110–4115. https://doi.org/10.1073/pnas.1419677112.

73. Meadows AL, Karnik R, Lam H, Forrestell S, Snedecor B. 2010. Application of dynamic flux balance analysis to an industrial Escherichia coli fermentation. Metab Eng 12:150–160. https://doi.org/10.1016/j.meben.2009.07.006.

74. Esther CR, Turkovic L, Rosenow T, Muhlebach MS, Boucher RC, Ranganathan S, Stick SM. 2016. Metabolomic biomarkers predictive of early structural lung disease in cystic fibrosis. Eur Respir J 48:1612–1621. https://doi.org/10.1183/13993003.00524-2016.

75. Bales PM, Renke EM, May SL, Shen Y, Nelson DC. 2013. Purification and characterization of biofilm-associated EPS exopolysaccharides from ES-KAPE organisms and other pathogens. PLoS One 8:e67950. https://doi.org/10.1371/journal.pone.0067950.

76. Li X, Hu Y, Gong J, Zhang L, Wang G. 2013. Comparative genome characterization of Achromobacter bacteria. Appl Microbiol Biotechnol 97:6413–6425. https://doi.org/10.1007/s00253-013-5018-3.

77. Cowley ES, Kopf SH, Lalriviere A, Ziebis W, Newman DK. 2015. Pediatric cystic fibrosis sputum can be chemically dynamic, anoxic, and extremely reduced due to hydrogen sulfide formation. mBio 6:e00767-15. https://doi.org/10.1128/mBio.00767-15.

78. Worlitzsch D, Taran R, Ulrich M, Schwab U, Cekici A, Meyer KC, Birrer P, Bellon G, Berger J, Weiss T, Botzenhart K, Yankaskas JR, Randell S, Boucher RC, Döring G. 2002. Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients. J Clin Invest 109:317–325. https://doi.org/10.1172/JCI123870.

79. Carlson R, Sørensen F. 2004. Fundamental Escherichia coli biochemical pathways for biomass and energy production: identification of reactions. Biotechnol Bioeng 85:1–19. https://doi.org/10.1002/bit.10812.

80. Pienkowska K, Wielhalm L, Tümmel B. 2019. Metagenome—inferrred bacterial replication rates in cystic fibrosis sputum. J Cyst Fibros https://doi.org/10.1016/j.jcf.2019.01.003.