Culture-enriched metagenomic sequencing enables in-depth profiling of the cystic fibrosis lung microbiota

Fiona J Whelan[1], Barbara Waddell[2], Saad A Syed[3], Shahrokh Shekarriz[3], Harvey R Rabin[2,4], Michael D Parkins[2,4], & Michael G Surette[2,3,5*]

November 13, 2019

[1]School of Life Sciences, University of Nottingham, Nottingham, United Kingdom
[2]Department of Microbiology, Immunology and Infectious Diseases, The University of Calgary, Calgary, Canada
[3]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Canada
[4]Department of Medicine, The University of Calgary, Calgary, Canada
[5]Department of Medicine, McMaster University, Hamilton, Canada

Contact information:
Michael G. Surette,
Department of Biochemistry and Biomedical Sciences,
Department of Medicine, Farncombe Family Digestive Health Research Institute,
McMaster University,
1280 Main St W, HSC 3N-9,
Hamilton, ON,
L8S 4K1,
Canada

Amplicon sequencing (e.g. of the 16S rRNA gene) identifies the presence and relative abundance of microbial community members; however, metagenomic sequencing is needed to identify the genetic content and functional potential of a community. Metagenomics is challenging in samples dominated by host DNA such as those from the skin, tissue, and respiratory tract. Within, we combine advances in amplicon and metagenomic sequencing with culture-enriched molecular profiling to study the human microbiota. Using the cystic fibrosis lung as an example, we culture an average of 82.13% of the OTUs representing 99.3% of the relative abundance identified in direct sequencing of sputum samples; importantly, culture-enrichment identified 63.3% more OTUs than direct sequencing. We developed the PLate Coverage Algorithm (PLCA) to determine a representative subset of culture plates on which to conduct culture-enriched metagenomics, resulting in the recovery of greater taxonomic diversity – including of low abundance taxa – with better metagenome-assembled genomes, longer contigs, and better functional annotations when compared to culture-independent methods. The PLCA is also applied as a proof-of-principle to a previously published gut microbiota dataset. Culture-enriched molecular profiling can be used to better understand the role of the human microbiota in health and disease.

The field of microbiology began with the visualization of microbes [1], and continued once we learned to control their growth. The advent of next-generation sequencing revolutionized microbiology, by allowing microbial genomics and community analysis without the requirement of culture. These technologies have expanded our understanding of the human microbiome, identifying the effect that environment, diet, and host genetics has on these complex communities [2, 3, 4, 5]. Culture-independent studies have made crucial contributions to microbiome research; however, without bacterial culture, we lose the advantages of classical microbiology, resulting in a lack of mechanistic studies and the alternative focus on microbial "dysbiosis" and microbiome diversity [6, 7].

While it is commonly cited that the majority of the human microbiota is unculturable, numerous studies conclusively counter this. In 1974, Finegold *et al.* cultured ∼300 species from 40 fecal specimens using both aerobic and anaerobic culture [8]. More recently, Goodman *et al.* cultured almost half of the human gut microbiota, recovered 316 Operational Taxonomic Units (OTUs) by culture of the 631 OTUs identified by culture-independent techniques [9]. Lagier *et al.* cultured 340 species of bacteria from 3 stool samples [10]. Further, recent studies recovered 88% of family-level OTUs [11] and 95% of all OTUs identified in fecal specimens [12]; importantly, both of these studies also identified more OTUs via culture-dependent than culture-independent methods. Other human-associated communities have been profiled with culture, including urine [13], skin [14], oral [15], and cystic fibrosis communities [16].

Marker gene profiling (e.g. 16S rRNA gene sequencing) provides a simple and rapid method to assess the taxonomic composition of a community. While metagenomics can assess functional capacity in addition to taxonomy, the usefulness of these data is dependent on how well short-read sequences can be assembled into contigs. The quality of assembly can be impacted by the complexity of the community, the sequencing technology, and/or the proportions of host DNA contamination [17, 18, 19]. Culture-enrichment is uniquely positioned to improve metagenomic assembly by allowing the proliferation of microbes (thus abating host DNA contamination) on media that "biologically bin" samples, thus decreasing their complexity. Coupling culture-enrichment with computational approaches allows us to separate promiscuous and fastidious microbes to better resolve these communities. Here, we use sputum from cystic fibrosis as an example of a complex microbial community where host DNA can represent a large proportion of the community (>99%) [20, 21, 22, 23]. We present a culture-enriched metagenomic strategy which overcomes these limitations and improves metagenomic results, providing a more comprehensive profile of the microbial community.

In this study, we merge culture-dependent and -independent techniques to better understand microbial communities. Culture-enriched 16S rRNA gene sequencing established that 82.13% of all OTUs - representing 99.3% of the relative abundance - in the cystic fibrosis lung microbiota are culturable. Further, culture-enrichment increased OTU recovery when compared to culture-independent sequencing. We introduce the PLate Coverage Algorithm (PLCA) which uses 16S rRNA gene sequencing to optimize culture-enriched metagenomics, and show that culture-enriched metagenomics improves the recovery of metagenomic-assembled genomes and produces more thorough functional annotations than direct metagenomic approaches. We identify the advantages of culture-enriched metagenomics: increased taxonomic and functional information due to a decrease in contaminating host DNA and the ability to computationally and biologically bin microbial species.

Results

In this study, we devised a strategy for culture-enriched metagenomic profiling (**Fig 1**). This strategy is based on previous culture-enrichment for amplicon-based sequencing conducted on the cystic fibrosis lung microbiome which was able to culture 43 of 48 bacterial families identified in sputum [16]. Upon collection, samples were immediately plated onto the 13 different media types (**Extended Data Fig 1**) under aerobic and anaerobic conditions. 16S rRNA gene sequencing was performed on the sputum sample (*direct sequencing*) as well as on the collective organisms grown in each media/environment pairing (*culture-enriched sequencing*) for a total of 26 culture-enriched samples per sputum (**Fig 1**). The OTU diversity in the direct sequencing and distribution in

the culture-enriched sequencing was used in conjunction with the PLate Coverage Algorithm (PLCA, details below) to determine a representative subset of culture-enriched plates which adequately represent the sample. Shotgun metagenomic sequencing was performed on the original sample and the culture-enriched subset.

The majority of the cystic fibrosis lung microbiota is culturable

We first identified the culturable fraction of the cystic fibrosis lung microbial community in 20 sputum samples from 10 patients. We defined an OTU to be culturable if it contained $\geq$ 10 reads and was recovered from $\geq$ 1 culture-enriched plate at a relative abundance of $\geq$ 0.01%. Across the dataset, an average of 82.13% (range: 64.6-100%) of OTUs identified by direct sequencing were culturable; this culturable fraction represents an average of 99.3% (range: 97.8-100%) of the relative abundance in the associated direct sequencing results (**Fig 2a**).When the genus-level OTU taxonomic assignments were compared to a list of species previously identified via culture-enrichment [24], and to previous culture-enrichment of the cystic fibrosis lung [16] and gut [12], we identified 18 genera cultured in this study which had not previously been identified via large-scale culture-enrichment (**Supplementary Table 1**). Of the OTUs which were never cultured across the dataset, we observe an over-representation of the Spirochaetes (7 of 7 OTUs identified in culture-independent sequencing) and the SR1 (1 of 1), and many members of the Tenericutes (7 of 22), and TM7 (2 of 3) phyla (**Fig 2b, blue ring**). Together, these results indicate that most OTUs in the cystic fibrosis lung microbiota are culturable and that those OTUs which are not cultured are taxonomically restricted, and historically challenging to culture groups. Importantly, culture-independent methods do not distinguish DNA from viable versus non-viable organisms; therefore, some bacteria not recovered may not be viable in these samples. These results still hold if a more stringent definition of culturable is used (**Extended Data Fig 2a-b**).

Culture-enrichment increases OTU recovery

Culture-enriched 16S rRNA gene sequencing consistently recovered more OTUs then direct sequencing (**Fig 2b, Extended Data Fig 2c-e**). For example, in the direct sequencing of Sample 1, 49 OTUs were recovered, 42 of which were also identified by culture-enrichment (**Fig 2a**); in addition to these 42 OTUs, an additional 124 OTUs were identified by culture-enrichment (**Extended Data Fig 2c**). This enrichment in OTU recovery did not correlate with variability in α-diversity of the original sample (**Extended Data Fig 2e**). We hypothesized that the ability to enrich may be due to the recovery of low abundance taxa. To test this hypothesis, we re-sequenced a sample to a depth 24x deeper than the original direct sequencing (972,834 vs. 41,199 reads) and rarefied at

decreasing depths (range: 500,000-1,000 reads). We observed that the number of OTUs recovered only by culture decreases as the sequencing depth increases (**Extended Data Fig 3**). These cultured OTUs were typical members of the cystic fibrosis lung microbiota including *Streptococcus sp., Prevotella sp.,* and *Veillonella sp*, indicating that culture allows for the enrichment of taxa present at low abundance in the original sample.

Culture-enrichment's increase in OTU recovery is dependent on media type and oxygen availability.

The variety of media, and environmental conditions used is important in capturing the diversity of microbial communities. The use of both anaerobic and aerobic conditions encourage the recovery of different taxa as evident in the taxonomic distribution and β-diversity relationships of the 16S rRNA gene sequencing results of direct and culture-enriched sequencing (**Fig 3a-b**). For example, *Veillonella sp.*, and *Prevotella sp.* were recovered exclusively under anaerobic conditions; conversely, *Rothia sp.* and *Pseudomonas sp.* were obtained at greater abundances in aerobic culture (**Fig 3a**). The α-diversity of each culture condition (media + aerobic/anaerobic environment) varied with each sample (**Supplementary Fig 2**), and no single condition consistently best recapitulated the originating sputum sample (**Extended Data Fig 4**). Hierarchical clustering of the taxa recovered from each culture condition further indicates the importance of using both selective and non-selective culture conditions (**Fig 3c, Extended Data Figs 5-6**). While some organisms, such as *Streptococcus sp.*, will grow under many conditions, others, such as *Neisseria sp., Rothia sp.,* and *Stenotrophomonas sp.* were only recovered from a subset of culture conditions. Further, there were also OTU-dependent differences in growth patterns within some genera (e.g. *Prevotella* OTUs, **Fig 3d; Extended Data Fig 6**). Across the dataset, anaerobic culture recovered almost half of all cultured OTUs; further, we nearly doubled the number of recovered OTUs by expanding our culture conditions beyond the 4 typically employed in a clinical laboratory (**Fig 3e**).

Another advantage of culture-enrichment is that it allows for the post-hoc recovery of organisms of interest from frozen bacterial stocks (see *Methods*). As an example, *Stenotrophomonas sp.* were isolated from two bacterial stocks in which it accounted for relatively low relative abundances (1.3 and 1.5%). To facilitate recovery, the stocks were replated on a media type specific to *Stenotrophomonas sp.* (**Fig 3f**). Taxonomic assignment of isolated colonies as *Stenotrophomonas maltophilia* were confirmed via full-length 16S rRNA gene sequencing (**Supplementary Table 2**).

The PLCA informs culture-enriched metagenomic sequencing

The conditions used for culture-enrichment are necessarily broad; the cystic fibrosis lung microbiota, like other human-associated communities, can be comprised of a wide range of organisms, from common pathogens (e.g. *Pseudomonas, Staphylococcus*, and *Haemophilus* [25]) to anaerobes (e.g. *Prevotella, Fusobacterium*, and *Veillonella* [26, 27, 28]) and emerging pathogens (*Stenotrophomonas maltophilia*, and *Achromobacter* [29]). While the lung could harbor any of these organisms, an individual's microbiota is a unique subset of these possibilities. This means that while the variety of culture conditions used is necessary to capture the diversity across individuals, not every plate is needed to enrich the microbiota of a given sample. This is also true of other human-associated communities; the gut, for example, harbours many organisms which are prevalent across the population, whereas some species can be quite specific to the individual [30]. It is not known *a priori* which subset of plates would best recapitulate a given community; however, we can use 16S rRNA gene sequencing to identify the taxonomic distribution across cultured plates, and to choose the subset of plates which best recapitulates the community on which to conduct culture-enriched metagenomic sequencing. As such, we implemented the PLate Coverage Algorithm (PLCA) which determines the minimum number of culture-enriched plates necessary to capture the taxonomic diversity of a sample with culture-enriched metagenomics. The PLCA is not specific to a particular set of culture conditions, or particular microbiome, but instead can be used with any community which contains a culturable majority.

There are two versions of the PLCA. The *denovo PLCA* recapitulates the culture-enriched community independent of the direct sequencing, whereas the *adjusted PLCA* focuses on the OTUs recovered from the direct sequencing results (**Supplementary Fig 3**). That is to say, the user can decide between recovery of all cultured organisms (*denovo PLCA*), or of preferential recovery of the abundant organisms from the original community (*adjusted PLCA*). The use of either version is dependent on whether the user is interested in the composition of the original sample - for example, when answering clinically relevant research questions - or in questions concerning sample biodiversity. Using these algorithms at different thresholds (see *Methods*) across the dataset highlights the uniqueness of these communities: no two samples have the same optimal plate set, and every culture condition is necessary for at least one sample as indicated by each condition in the *denovo* and *adjusted* outcomes being necessary for the culture-enriched metagenomic sequencing of at least one sample in the dataset (**Fig 4ab, Extended Data Fig 7**).

Culture-enriched metagenomics provides improved taxonomic and functional resolution

To test the PLCA, we used both versions of the algorithm on a representative sample. The PLCA indicated culture-enriched metagenomics on 5 (denovo PLCA) and 3 (adjusted PLCA) culture plates: Beef (aerobic) and AIA, KVLB, MAC, and McKay (anaerobic) for denovo PLCA; TSY and McKay (aerobic) and CHOC (anaerobic) for the adjusted PLCA. Moderate taxonomic concordance was observed between the 16S rRNA gene and metagenomic sequencing, with some differences in composition and relative abundance (**Extended Data Fig 8**). Following metagenomic co-assembly, binning, and taxonomic assignment (see below), we compared the observed versus expected results of each algorithm. The denovo PLCA recovered 10 metagenomic bins with matching taxonomic assignments to the 10 OTUs above the PLCA threshold in addition to a further 6 bins matching OTUs below the threshold (**Fig 4c, Supplementary Table 3**). Similarly, the adjusted PLCA recovered 10 of 11 expected OTUs and matched an additional 14 OTUs below the adjusted PLCA threshold (**Fig 4c, Supplementary Table 4**). When the adjusted PLCA was applied to the other 19 samples in this dataset, culture-enriched metagenomic sequencing of a collective 91 (of a possible 520) plates resulted in the recovery of 83.4% of targeted species with an additional 413 metagenomic bins recovered from other species (**Extended Data Fig 9a**). We also applied the PLCA to previously published culture-enrichment data from the gut microbiota [12], establishing that this algorithm is not specific to the lung microbiome or a specific set of culture conditions (**Extended Data Fig 9b**).

For each of the denovo and adjusted PLCA plate sets and the direct sequencing, we co-assembled shotgun metagenomic reads into contigs which were binned and defined as metagenome-assembled genomes (MAGs) or non-MAG bins based on sequence composition (see *Methods*). Culture-enriched metagenomic sequencing resulted in 7 MAGs and 9 non-MAG bins (denovo PLCA) and 12 MAGs and 12 non-MAG bins (adjusted PLCA); in contrast, 1 MAG and 1 non-MAG bin were recovered from direct metagenomic sequencing (**Fig 5**). When the direct sequencing metagenomic reads were mapped onto the culture-enriched metagenomic bins, significant coverage was only observed in 1 bin in both the denovo and adjusted PLCA results (**Fig 5, blue dots**). Both direct sequencing metagenomic bins were taxonomically assigned as *Pseudomonas sp.*; in contrast, the culture-enriched bins were taxonomically diverse, spanning 14 genera (**Fig 5**).

The increased taxonomic diversity obtained via culture-enrichment directly translates into increased functional information about this microbial community (**Fig 6**). Consistently, across Clusters of Orthologous Groups (COG) functional categories and predictions of virulence genes, phage, antibiotic resistance, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs), and secondary metabolites, culture-enriched metagenomic sequencing provides

a greater diversity and number of functional identifications. For example, genes contributing to the GacS/GacA two-component system were identified in *Pseudomonas sp.* bins from both culture-enrichment datasets but not in the direct sequencing of the sputum sample. Previous research has shown that strains of *P. aeruginosa* lacking this system are less able to colonize in mouse models [31], indicating their importance for *P. aeruginosa* virulence. The detection of these genes in the culture-enriched data suggests that this system is present in the cystic fibrosis lung microbiota but is not identified in the direct metagenomic sequencing due to poor sequencing depth/assembly. Among the *Pseudomonas sp.* results, there were 10, 16, and 17 "Perfect" hits against the Comprehensive Antibiotic Resistance Database (CARD) in the direct, denovo PLCA, and adjusted PLCA datasets, respectively; among these, 2 beta-lactamases (OXA-50, PDC-1), and 2 repressors (nalD, nfxB) were found only in the culture-enriched sequencing. These *Pseudomonas*-specific results indicate that even when bins overlap in taxonomic assignment between techniques, that culture enriches for functional annotation; in addition to this, the functional characterisations of all other bins (15 and 23 in the denovo and adjustedPLCA, respectively) would not have been possible from direct sequencing alone.

Previous research has established the presence of heterogeneous populations of single species in cystic fibrosis airways (e.g. in *P. aeruginosa* [32], *Burkholderia cepacia complex* [33], and *Stenotrophomonas maltophilia* [34]) as well as in the human microbiome as a whole (e.g. *Bifidobacterium longum* populations in the infant gut microbiome [35], various strains following fecal microbiota transplantation [36]). As such, we calculated the genetic variability within each metagenomic bin by identifying haplotypes within open reading frames (ORFs) ([37]; **Fig 6d**). In some metagenomic bins, we identify a consistent and small number of ORF haplotypes, suggesting that the bin represents a single genomic population (i.e. one strain). However, in most bins, we see great haplotype diversity indicative of heterogeneous populations (i.e. multiple strains). As expected, the number of haplotypes per gene within these bins is diverse, indicating the known spectrum of evolutionary pressures within bacterial genomes [38]. On average, more gene haplotypes were identified in the *Pseudomonas sp.* culture-enriched bins then in the direct sequencing; however, the direct sequencing identified a greater number of prevalent haplotype gene outliers. There was no correlation between bin completeness or MAG status and mean haplotype frequency.

Discussion

The decrease in cost and increase in massively parallelized sequencing technology has revolutionized the way that we study the human microbiota, leading to an increased understanding of these communities and how they relate to health and disease. The gut has arguably been the most well-studied human microbiome, with numerous studies linking it to various diseases (e.g. IBD [39], and

IBS [40]) and conditions (e.g. obesity [41] and pregnancy [42]). This community lends itself well to these investigations; its composition can be approximated via fecal matter – readily available without intervention – and consists of a dense microbial community with little host DNA contamination. However, many other important human-associated communities have high-levels of non-microbial DNA, including communities of the skin [45], tissue biopsies [19], and oral microbiome samples [46]. The respiratory microbiome is a low biomass community, often contaminated with DNA from endothelial cells or from the DNA associated with an acute immune response in diseases such as asthma and cystic fibrosis [47]. Because of the nature of such samples, metagenomic sequencing must be paired with the *in silico* removal of most sequencing reads due to host contamination, meaning that only the most abundant members of the community can be assembled into MAGs. We demonstrate that culture-enriched metagenomics - in conjunction with traditional, culture-independent sequencing - can improve the resolution of these communities.

Within, we show that the cystic fibrosis lung microbiota is culturable, and that culture-enrichment increases OTU diversity (**Fig 2**). This follows directly from Sibley *et al.* who, using T-RFLP and 454 sequencing, were also able to identify a culturable majority within the cystic fibrosis lung microbiota [16]. Although culture-enrichment consistently recapitulated $> 97\%$ of the direct sequencing results, a few organisms were not cultured. Many, including members of the Spirochaetes, SR1, Tenericutes, and TM7, consist of organisms which are difficult to culture [48, 49, 50], but were also at low abundance in these samples. In contrast, the organisms recovered only by culture are common members of the cystic fibrosis lung microbiota [25], including *Rothia, Prevotella, Veillonella, Fusobacterium*, and *Streptococcus* species. Selective media allow for the proliferation of low abundance organisms which can be below the level of detection of standard sequencing approaches. It is not uncommon, for example, for the cystic fibrosis lung microbiota to reach a density of at least 10^8 CFUs/mL [51, 52, 53]. If amplicon sequencing produces 50,000 reads per sample, an organism identified by a single read would equate to $\leq 0.002\%$ relative abundance or $2\text{x}10^4$ CFUs/mL. This is not to dismiss culture-independent approaches; culture-enrichment benefits from being combined with direct sequencing in order to maintain the relative abundance ratios of the original community and to recover important uncultured organisms.

In order to most effectively combine culture-independent and -dependent approaches, we designed the PLate Coverage Algorithm (PLCA) which determines the cultured metagenomic sequencing necessary to recapitulate the original microbial community as determined by amplicon sequencing. We show that the PLCA recovers targeted OTUs as well as a substantial number of additional OTUs (**Fig 4**). Importantly, PLCA-assisted culture-enriched metagenomics vastly improved the taxonomic and functional outputs of sequencing. The inability of traditional metagenomic sequencing to distinguish microbe from host can result in the need for incredible sequencing depths of samples with high

host contamination. Combining culture with culture-independent sequencing is one way of mitigating host contamination due to culture's ability to enrich for viable microbes.

The PLCA is not specific to the cystic fibrosis microbiome and can be used on any microbial community in which most of its membership is culturable. The culture conditions chosen will impact the ability to culture the sample and the performance of the PLCA. Ideally, a combination of selective, non-selective, and enrichment media should be used. Further, the PLCA is also not specific to any one 16S rRNA gene processing pipeline. Within, we have used a 16S rRNA gene sequencing pipeline which has been previously validated against other available approaches [54]; however, 16S rRNA gene sequencing processing is a moving target, with new (or improved) methodologies constantly being published, and compared/validated against already existing methods. Because the PLCA is agnostic to how the data is processed, it can be applied to datasets processed with any method. As the field continues to progress, the PLCA will only benefit from the improvements in taxonomic resolution available from these technological improvements.

The combination of culture-enrichment with direct sequencing enhances the observed taxonomic diversity and provides greater insight into human-associated microbial communities. We have shown that culture-enriched metagenomics provides deeper resolution of these communities. With this data, we can better predict, for example, mechanisms of antimicrobial resistance, virulence factors, and - in general - gain a better understanding of each organism's gene repertoire. Further, having these organisms in culture means that we can carry out *in vitro*, mechanistic studies to better understand these communities in the context of human health and disease. Culture-enriched metagenomics, as exemplified here for cystic fibrosis sputum, provides an approach for the study of microbiome samples which are comprised mostly of human rather than microbial DNA. This method can be applied to any community where the majority of its members are culturable.

Methods

Sputum collection and culture-enrichment

Upon receiving informed consent, sputum samples were collected from December 4th 2013 to October 6th 2014 from willing participants visiting the Calgary Adult Cystic Fibrosis Clinic (ethical approval granted by the Calgary Health Region Ethics Board, REB-24123). Two samples were collected from each patient (with the noted exceptions): one at the onset of pulmonary exacerbation (as defined by Fuchs *et al.* [55]) and a second during a follow up appointment (1 week-4 months) following the resolution of symptoms and antibiotic discon-

tinuation. In one case, a patient wasn't able to produce a followup sputum sample; in another, a patient experienced two exacerbations before a follow up appointment so 3 samples were collected.

Samples were transported to an anaerobic environment within 1 minute of expectoration and plated within 4 hours of production. Samples were homogenized by passage through a 18 gauge needle and 1 mL syringe. Once homogeneous, 300μL was set aside for direct sequencing. The remainder was used for culture enrichment. The thirteen solid agar media used in this study included (**Extended Data Fig 1**): Actinomycetes isolation agar (AIA; BD), brain heart infusion agar (BHI; BD), cooked meat broth with 1.5% agar (Beef; Fluka), Columbia agar base with 5% sheep's blood (CBA; BD), GC powder (BD) with 5% hemoglobin, and 0.5% IsoVitaleX (CHOC; BD), Columbia CNA agar with 5% sheep's blood (CNA; BD), fastidious anaerobe agar (FAA; Acumedia), tryptic soy agar with 0.1μg/mL kanamycin, 7.5μg/mL vancomyin, 10μg/mL Vitamin K, 0.05ng/mL hemin, and 5% laked blood (KVLB), MacConkey agar (MAC; BD), mannitol salt agar (MSA; BD), McKay media [56], phenylethyl alcohol agar with 5% sheep's blood (PEA; BD), and tryptic soy agar with 1.5% yeast extract (TSY; BD). To Beef, BHI, and TSY, the following additional additives were included: 10μg/mL colisitin sulfate, 0.5mg/mL L-Cysteine, 1.0ng/mL Vitamin K, and 10ng/mL hemin. These media were chosen based on previous successful isolation of bacterial species from cystic fibrosis sputum by Sibley *et al.* [16], and include non-selective, selecitive, and enrichment media types (as defined in [57]; **Extended Data Fig 1**). Comparisons of this culturing method to that employed by a typical microbiology lab in **Figure 3** included the following media types: aerobic CBA, MAC, MSA, and anaerobic CHOC.

Culture enrichment was performed by placing 100μL of sputum diluted in BHI with 0.05% L-Cysteine to 10^{-3} and 10^{-5} on to each of the above media. Two sets of plating were performed, one which was incubated aerobically (5% CO_2, 37°C) and another anaerobically (5% CO_2, 5% H_2, 90% NO_2, 37°C), resulting in 52 plates per sample.

After 3-5 days (aerobic) and 5-7 days (anaerobic) of growth, plates were imaged and growth acquired by adding 2mL of BHI broth to each plate and lifting colonies. 1mL of this broth was frozen directly for DNA extraction while the remaining 1mL was frozen in skim milk (final concentration 10%) for any potential growth or re-isolation experiments. For the first few culture-enrichment sample sets, plates with no visible growth were processed like any other plate (see below); however, we consistently were unable to obtain visible PCR products on a 2% agarose gel from those plates which did not have visible growth. Thus, any plate which resulted in no visible bacterial colonies was discarded and omitted from all downstream processing.

In order to demonstrate the reproducibility of sputum sample collection and culture-enrichment methods, we carried out an additional experiment where 2

sputum samples from each of 3 patients were collected in the clinic before and after physiotherapy (3x2 biological replicates). The consistency of these biological replicates indicate the similarity of sputum communities when collected in quick succession. These samples were then plated on 6 media in triplicate (6x3 technical replicates across 6 sputum samples, n=108). The results demonstrate the consistency in replicate sputum samples and in culture enrichment (**Extended Data Fig 10**).

DNA isolation and Illumina sequencing

Genomic DNA was isolated from culture-enriched plates and sputum as previously described [58] with the exception of use of lifted colonies/homogenized sputum as input instead of Copan Swabs as performed in [16]. Dilutions resulting from the same culture conditions were combined into one genomic DNA isolation for a maximum of 26 culture-enriched samples per sputum sample. The variable 3 region of the 16S rRNA gene was amplified using universal primers as adapted from [58, 59]. The PCR reaction consisted of 5pmol of each primer, 1ng template DNA, 200μM dNTPs, 1.5mM $MgCl_2$, and 1 U Taq polymerase. The PCR protocol used is as follows: 95°C for 5 minutes, followed by 30 cycles of 95°C for 30 seconds, 50°C for 30 seconds, and 72°C for 30 seconds, with a final 72°C for 7 minutes. Presence of a PCR product was verified by electrophoresis (2% agarose gel). PCR products were sequenced using the Illumina MiSeq platform using 2x250 paired-end reads.

DNA from select culture-enriched samples and the sputum sample were sonicated to 300bp and library preparations were made using the NEBNext DNA Library Prep Master Mix Set for Illumina (New England Biolabs) and sequenced using the Illumina HiSeq platform with 2x250 paired-end reads.

All sequencing results are publically available (BioProject ID: PRJNA503799).

16S rRNA sequence processing and analysis

16S rRNA paired-end reads were processed using sl1p [54]. Briefly, reads were trimmed of any remaining primers using cutadapt [60] and discarded using sickle based on a quality threshold of 30 (`https://github.com/najoshi/sickle`). Paired-end reads were assembled using PANDAseq [61]. OTUs were picked using AbundantOTU+ with a 97% clustering threshold [62] and chimeras removed using USEARCH [63] as implemented in QIIME [64]. The Ribosomal Database Project classifier [65] was used to assign taxonomy against the 4th February 2011 release of the Greengenes database [66], and a phylogeny was created by pruning the Greengenes phylogeny to those taxa present in the dataset. OTU tables were created with QIIME [64]. Any OTU which was not assigned a bacterial taxonomy or which only had one instance across the full dataset (singleton) was culled. Any sample with < 1000 reads was discarded (**Supplementary Table 5**). The result of this culling process, in combination with only sequencing

plates with visual growth, resulted in a total of 531 samples (20 sputum samples and 511 plates). The mean sequence depth across this dataset was 68,160 reads per sample (range 2,032-159,381), with a mean number of OTUs of 94.1 (range 10-311).

Taxonomic summaries over multiple samples were performed by calculating the maximum relative abundance across samples, and normalizing to 100%. Principal Coordinate Analysis (PCoA) plots were calculated using phyloseq [67] and ggplot2 [68] in R after proportional normalization [69]. An OTU was considered present in the direct or culture-enriched sequencing if it had a relative abundance of $> 0.01\%$ (all exceptions noted). Phylogenies were decorated with GraPhlAn [70]. Heatmaps were generated with pheatmap [71]. In the sequencing depth experiments shown in **Extended Data Fig 3**, rarefaction was performed at varying depths using QIIME's alpha rarefaction function.

Recovery of isolates from frozen culture-enriched stocks

Improved isolation of *Stenotrophomonas maltophilia* from frozen skim milk stocks of select plates was performed using a selective medium as described in [72]. Isolates were Sanger sequenced using the 8F (5'-AGAGTTTGATCCTGGCTCAG-3') and 926R (5'-CCGTCAATTCCTTTRAGTTT-3') primers to the 16S rRNA gene, resulting in a 900nt product. The identity of the isolates were confirmed by comparisons to the Human Oral Microbiome Database (HOMD) and to NCBI's 16S ribosomal RNA sequences (Bacteria and Archaea) Database.

The PLate Coverage Algorithm (PLCA)

Taking 16S rRNA gene sequencing results as input, the PLCA calculates, for each sample, the optimal subset of cultured plates which should be included in culture-enriched metagenomics in order to recapitulate the microbial community. The PLCA (**Supplementary Fig 3a-b**) first identifies any OTU above the user-supplied relative abundance threshold that was only cultured on a single plate, and that plate is added to the "plate set" for culture-enriched sequencing. Next, for all OTUs not already identified in the plate set, the plate with the most OTUs present above the threshold is added to the plate set. This continues until all OTUs are accounted for in the plate set and this list is output to the user. The PLCA incorporates a user-adjustable relative abundance threshold to determine which cultured OTUs the algorithm should include in the resulting plate set. In the adjusted PLCA, a second threshold determines the cutoff of OTU inclusion from the direct sequencing. Altering these thresholds results in varying plate and OTU recovery (including OTUs below the threshold which are included by consequence of being present on a plate which is part of the optimal plate set; **Fig 4a, Extended Data Fig 7**). The plate set for the adjusted PLCA is not always a direct subset of the denovo PLCA because when only the organisms present in the sputum are considered, there is often a better combination of plates that minimizes the number of total plates needed for sequencing.

The PLCA is freely available from `http://github.com/fwhelan/PLCA`.

Metagenomic sequence processing and analysis

Resultant Illumina paired-end reads from the 20 sputum samples and their associated adjustedPLCA plate sets (and one additional denovoPLCA plate set for comparison) were processed first by using cutadapt to trim Illumina adapters and primers [60]. Sickle, with a quality threshold of 30, was used to remove low-quality sequences (`https://github.com/najoshi/sickle`). The direct sequencing was decontaminated of host-associated reads using DeconSeq [73]. Metagenomic assembly of the direct sequencing reads was conducted using Megahit [74]. A co-assembly of the culture-enriched reads (as determined by the denovo PLCA and/or adjusted PLCA) was also conducted using Megahit [74]. The results of all assemblies were separately binned using MaxBin-2.2.1 [75]; MaxBin was chosen based on its performance in the CAMI Challenge [76]. Quality was assessed with checkM [77], and taxonomic assignments for each bin determined using KrakenUniq (formally KrakenHLL) [78] and supplementary scripts (`https://github.com/shekas3/BinTaxaAssigner`). Assembly statistics, including mean/maximum contig length, and N50 values are provided in **Supplementary Table 6**. Metagenome-assembled genomes (MAGs) were defined as metagenomic bins containing $\geq 70\%$ completeness and $< 10\%$ contamination as previously described [79]; non-MAG bins are any bin which does not meet the criteria of a MAG. Only contigs >1000bp were binned. Bowtie2 was used to map direct sequencing reads onto culture-enriched metagenomic bins. Even though extensive decontamination and quality control measures were performed, the direct sequencing suffered from considerable host DNA contamination, resulting in reads within these bins mapping with high stringency to the human genome. This resulted in bins from the direct sequencing which were much larger than the size of the closest reference genome (**Supplementary Fig 4**). In contrast, culture-enriched bins with high "contamination" indicated the inability of the binning algorithm to differentiate between closely-related species. For example, b6 in denovo and b9 in adjusted had bin lengths almost double the closest reference and contained taxonomic signatures of two *Streptococcus sp.* (**Supplementary Fig 4**).

To compare the taxonomic composition of 16S rRNA gene and metagenomic sequencing, we used the Kraken2 (version 2.0.7) classifier [80] to classify metagenomic reads from each culture-enriched sample against the 2011 Greengenes database [66].

To show that the PLCA was not specific to the cystic fibrosis lung microbiota, the algorithm was applied to a previously published culture-enrichment study of the gut microbiota [12]. The denovo and adjusted PLCA were applied, with default thresholds, to sample IBS3, and the taxonomic recovery from culture-enriched metagenomic sequencing was predicted based on the 16S

rRNA gene sequencing profiles of the resulting plate set.

Functional annotations of each metagenomic bin were performed with a variety of software. Virulence gene counts were determined by use of blastn (E-value cutoff of 10^{-8}) [81] against PATRIC's virulence factor library [82] and counting the number of hits to unique virulence genes per bin. COG functional category counts were determined using eggnog-mapper with default parameters [83, 84]. Phage counts were determined by Phaster [85] and predictions of antibiotic resistance genes were conducted with CARD (2.0.2) in conjunction with the RGI (4.1.0) [86]. Secondary metabolites were predicted with PRISM 3 [87]. The presence of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) were determined using MinCED [88]. The direct sequencing was plagued with host DNA contamination resulting in excessive bin lengths, and, possibly, abundant type I errors in the identified functionality of the community (**Fig 6, Supplementary Fig 4, Supplementary Table 7**). All bar charts were made using R's ggplot2 [68] and heatmaps were visualized using R's pheatmap [71]. Haplotype diversity of the open reading frames of each bin were calculated using Hansel and Gretel [37].

Data availability

All sequencing results are publically available (BioProject ID: PRJNA503799). The PLCA algorithm is available from `https://github.com/fwhelan/PLCA`.

Code availability

All code developed by the authors is available under a GNU license at `http://github.com/fwhelan/PLCA` and `https://github.com/shekas3/BinTaxaAssigner`.

References

[1] Van Leeuwenhoek, A. Microscopical observations about animals in the scurf of the teeth. *Philos Trans R Soc Lond B Biol Sci* **14**, 568–574 (1683).

[2] Turnbaugh, P. J. *et al.* The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. *Science Translational Medicine* **1**, 6ra14–6ra14 (2009). URL `http://www.ncbi.nlm.nih.gov/pubmed/20368178http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2894525http://stm.sciencemag.org/cgi/doi/10.1126/scitranslmed.3000322`.

[3] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–14 (2012). URL `http://dx.doi.org/10.1038/nature11234`.

[4] Spor, A., Koren, O. & Ley, R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Reviews Microbiology* **9**, 279–290 (2011). URL http://www.nature.com/doifinder/10.1038/nrmicro2540.

[5] Rothschild, D. *et al.* Environment dominates over host genetics in shaping human gut microbiota. *Nature* (2018). URL http://www.nature.com/doifinder/10.1038/nature25973.

[6] Olesen, S. W. & Alm, E. J. Dysbiosis is not an answer. *Nature Microbiology* **1**, 16228 (2016). URL http://www.nature.com/articles/nmicrobiol2016228.

[7] Shade, A. Diversity is the question, not the answer. *The ISME Journal* **11**, 1–6 (2017). URL http://www.nature.com/doifinder/10.1038/ismej.2016.118.

[8] Finegold, S. M., Attebery, H. R. & Sutter, V. L. Effect of diet on human fecal flora: comparison of Japanese and American diets. *The American Journal of Clinical Nutrition* **27**, 1456–69 (1974). URL http://www.ncbi.nlm.nih.gov/pubmed/4432829.

[9] Goodman, A. L. *et al.* Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proceedings of the National Academy of Sciences* **108**, 6252–7 (2011). URL http://www.pnas.org/content/108/15/6252.short.

[10] Lagier, J.-C. *et al.* Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* **18**, 1185–93 (2012). URL http://www.ncbi.nlm.nih.gov/pubmed/23033984.

[11] Rettedal, E. A., Gumpert, H. & Sommer, M. O. A. Cultivation-based multiplex phenotyping of human gut microbiota allows targeted recovery of previously uncultured bacteria. *Nature Communications* **5**, 4714 (2014). URL http://www.nature.com/ncomms/2014/140828/ncomms5714/abs/ncomms5714.html.

[12] Lau, J. T. *et al.* Capturing the diversity of the human gut microbiota through culture-enriched molecular profiling. *Genome Medicine* **8**, 72 (2016). URL http://www.ncbi.nlm.nih.gov/pubmed/27363992.

[13] Hilt, E. E. *et al.* Urine is not sterile: use of enhanced urine culture techniques to detect resident bacterial flora in the adult female bladder. *Journal of Clinical Microbiology* **52**, 871–6 (2014). URL http://jcm.asm.org/content/52/3/871.short.

[14] Myles, I. A. *et al.* A method for culturing Gram-negative skin microbiota. *BMC Microbiology* **16**, 60 (2016). URL http://www.ncbi.nlm.nih.gov/pubmed/27052736.

[15] Thompson, H., Rybalka, A., Moazzez, R., Dewhirst, F. E. & Wade, W. G. In-vitro culture of previously uncultured oral bacterial phylotypes. *Applied and Environmental Microbiology* **81**, 8307–8314 (2015). URL http://www.ncbi.nlm.nih.gov/pubmed/26407883.

[16] Sibley, C. D. *et al.* Culture enriched molecular profiling of the cystic fibrosis airway microbiome. *PLOS ONE* **6**, e22702 (2011). URL http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0022702.

[17] Oh, J. *et al.* Biogeography and individuality shape function in the human skin metagenome. *Nature* **514**, 59–64 (2014). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4185404{\&}tool=pmcentrez{\&}rendertype=abstract.

[18] Wang, W.-L. *et al.* Application of metagenomics in the human gut microbiome. *World J Gastroenterol World J Gastroenterol January* **21**, 803–814 (2015). URL http://www.wjgnet.com/esps/http://www.wjgnet.com/esps/helpdesk.aspxhttp://www.wjgnet.com/1007-9327/full/v21/i3/803.htmhttp://dx.doi.org/10.3748/wjg.v21.i3.803.

[19] Zhang, C. *et al.* Identification of low abundance microbiome in clinical samples using whole genome sequencing. *Genome Biology* **16** (2015). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4661937/pdf/13059{_}2015{_}Article{_}821.pdf.

[20] Lim, Y. W. *et al.* Metagenomics and metatranscriptomics: Windows on CF-associated viral and microbial communities. *Journal of Cystic Fibrosis* **12**, 154–164 (2013). URL https://www.sciencedirect.com/science/article/pii/S1569199312001403.

[21] Huang, Y. J. & LiPuma, J. J. The Microbiome in Cystic Fibrosis. *Clinics in Chest Medicine* **37**, 59–67 (2015). URL http://www.ncbi.nlm.nih.gov/pubmed/26857768.

[22] Zhao, J. *et al.* Decade-long bacterial community dynamics in cystic fibrosis airways. *Proceedings of the National Academy of Sciences* **109**, 5809–14 (2012). URL http://www.ncbi.nlm.nih.gov/pubmed/22451929http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3326496.

[23] Whelan, F. J. *et al.* Longitudinal sampling of the lung microbiota in individuals with cystic fibrosis. *PLOS ONE* **12**, e0172811 (2017). URL http://dx.plos.org/10.1371/journal.pone.0172811.

[24] Lagier, J.-C. *et al.* Culture of previously uncultured members of the human gut microbiota by culturomics. *Nature Microbiology* **1**, 16203 (2016). URL http://www.nature.com/articles/nmicrobiol2016203.

[25] Surette, M. G. The cystic fibrosis lung microbiome. *Annals of the American Thoracic Society* **11 Suppl 1**, S61–5 (2014). URL http://www.ncbi.nlm.nih.gov/pubmed/24437409.

[26] Field, T. R., Sibley, C. D., Parkins, M. D., Rabin, H. R. & Surette, M. G. The genus Prevotella in cystic fibrosis airways. *Anaerobe* **16**, 337–344 (2010). URL https://www.sciencedirect.com/science/article/pii/S1075996410000600.

[27] van der Gast, C. J. *et al.* Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *The ISME Journal* **5**, 780–791 (2011). URL http://www.nature.com/articles/ismej2010175.

[28] Tunney, M. M. *et al.* Detection of anaerobic bacteria in high numbers in sputum from patients with cystic fibrosis. *American Journal of Respiratory and Critical Care Medicine* **177**, 995–1001 (2008). URL http://www.atsjournals.org/doi/abs/10.1164/rccm.200708-1151OC{\#}.VV0tvlVViko.

[29] Parkins, M. D. & Floto, R. A. Emerging bacterial pathogens and changing concepts of bacterial pathogenesis in cystic fibrosis. *Journal of Cystic Fibrosiss* (2015). URL http://www.ncbi.nlm.nih.gov/pubmed/25881770.

[30] Pop, M. *et al.* Individual-specific changes in the human gut microbiota after challenge with enterotoxigenic Escherichia coli and subsequent ciprofloxacin treatment. *BMC Genomics* **17**, 440 (2016). URL http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-2777-0.

[31] Coleman, F. T. *et al.* Hypersusceptibility of cystic fibrosis mice to chronic Pseudomonas aeruginosa oropharyngeal colonization and lung infection. *Proceedings of the National Academy of Sciences* **100**, 1949–1954 (2003). URL http://www.ncbi.nlm.nih.gov/pubmed/12578988http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC149939http://www.pnas.org/cgi/doi/10.1073/pnas.0437901100.

[32] Jorth, P. *et al.* Regional Isolation Drives Bacterial Diversification within Cystic Fibrosis Lungs. *Cell Host & Microbe* (2015). URL http://www.sciencedirect.com/science/article/pii/S193131281500298X.

[33] Lieberman, T. D. *et al.* Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nature genetics* **46**, 82–7 (2014). URL http://www.ncbi.nlm.nih.gov/pubmed/24316980http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3979468.

[34] Pompilio, A. *et al.* Stenotrophomonas maltophilia Phenotypic and Genotypic Diversity during a 10-year Colonization in the Lungs of a Cystic Fibrosis Patient. *Frontiers in Microbiology* **7**, 1551 (2016). URL http://journal.frontiersin.org/Article/10.3389/fmicb.2016.01551/abstract.

[35] Ferretti, P. *et al.* Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut

Microbiome. *Cell Host & Microbe* **24**, 133–145.e5 (2018). URL http://www.ncbi.nlm.nih.gov/pubmed/30001516https://linkinghub.elsevier.com/retrieve/pii/S1931312818303172.

[36] Li, S. S. *et al.* Durable coexistence of donor and recipient strains after fecal microbiota transplantation. *Science* **352**, 586–589 (2016). URL http://science.sciencemag.org/content/352/6285/586.abstract.

[37] Nicholls, S. M. *et al.* Probabilistic recovery of cryptic haplotypes from metagenomic data URL https://www.biorxiv.org/content/biorxiv/early/2017/03/17/117838.full.pdf.

[38] Creevey, C. J., Doerks, T., Fitzpatrick, D. A., Raes, J. & Bork, P. Universally Distributed Single-Copy Genes Indicate a Constant Rate of Horizontal Transfer. *PLoS ONE* **6**, e22099 (2011). URL http://dx.plos.org/10.1371/journal.pone.0022099.

[39] Frank, D. N. *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proceedings of the National Academy of Sciences* **104**, 13780–13785 (2007). URL http://www.pnas.org/cgi/doi/10.1073/pnas.0706625104.

[40] Collins, S. M. A role for the gut microbiota in IBS. *Nature Reviews Gastroenterology & Hepatology* **11**, 497–505 (2014). URL http://www.nature.com/doifinder/10.1038/nrgastro.2014.40.

[41] Ley, R. E., Turnbaugh, P. J., Klein, S. & Gordon, J. I. Microbial ecology: Human gut microbes associated with obesity. *Nature* **444**, 1022–1023 (2006). URL http://www.nature.com/doifinder/10.1038/4441022a.

[42] Gohir, W., Whelan, F. J., Surette, M. G., Moore, C. & Jonathan, D. Pregnancy-related changes in the maternal gut microbiota are dependent upon the mother's periconceptional diet. *Gut microbes* **6**, 310–320 (2015).

[43] Zeevi, D. *et al.* Personalized Nutrition by Prediction of Glycemic Responses. *Cell* **163**, 1079–1095 (2015). 1611.06654.

[44] Naseribafrouei, A. *et al.* Correlation between the human fecal microbiota and depression. *Neurogastroenterol Motil* **26**, 1155–1162 (2014). URL http://www.ncbi.nlm.nih.gov/pubmed/24888394http://doi.wiley.com/10.1111/nmo.12378.

[45] Grice, E. A. & Segre, J. A. The skin microbiome (2011). URL www.nature.com/reviews/micro.

[46] Wang, J. *et al.* Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. *Scientific Reports* **3** (2013). URL www.theseed.org. arXiv:1011.1669v3.

[47] Dickson, R. P., Erb-Downward, J. R., Martinez, F. J. & Huffnagle, G. B. The Microbiome and the Respiratory Tract. *Annual review of physiology* (2015). URL http://www.ncbi.nlm.nih.gov/pubmed/26527186.

[48] Chi, B., Chauhan, S. & Kuramitsu, H. Development of a system for expressing heterologous genes in the oral spirochete Treponema denticola and its use in expression of the Treponema pallidum flaA gene. *Infection and Immunity* **67**, 3653–6 (1999). URL http://www.ncbi.nlm.nih.gov/pubmed/10377154http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC116559.

[49] Camanocha, A. & Dewhirst, F. E. Host-associated bacterial taxa from Chlorobi, Chloroflexi, GN02, Synergistetes, SR1, TM7, and WPS-2 Phyla/candidate divisions View Crossmark data Host-associated bacterial taxa from Chlorobi, Chloroflexi, GN02, Synergistetes, SR1, TM7, and WPS-2 Phyla/can. *Journal of Oral Microbiology* **6** (2546). URL http://www.tandfonline.com/action/journalInformation?journalCode=zjom20.

[50] Marcy, Y. *et al.* Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proceedings of the National Academy of Sciences* **104**, 11889–94 (2007). URL http://www.pnas.org/content/104/29/11889.short.

[51] Meyer, K. C., Sharma, A., Rosenthal, N. S., Peterson, K. & Brennan, L. Regional Variability of Lung Inflammation in Cystic Fibrosis. *American Journal of Respiratory and Critical Care Medicine* **156**, 1536–1540 (1997). URL http://www.ncbi.nlm.nih.gov/pubmed/9372672http://www.atsjournals.org/doi/abs/10.1164/ajrccm.156.5.9701098.

[52] Stressmann, F. A. *et al.* Does bacterial density in cystic fibrosis sputum increase prior to pulmonary exacerbation? *Journal of Cystic Fibrosis* **10**, 357–365 (2011). URL http://linkinghub.elsevier.com/retrieve/pii/S1569199311000749.

[53] Quigley, E. M. M., Eamonn, D. & Quigley, M. M. Gut Bacteria in Health and Disease. Tech. Rep. (2013). URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3983973/pdf/GH-09-560.pdf.

[54] Whelan, F. J. & Surette, M. G. A comprehensive evaluation of the sl1p pipeline for 16S rRNA gene sequencing analysis. *Microbiome* **5**, 100 (2017). URL http://microbiomejournal.biomedcentral.com/articles/10.1186/s40168-017-0314-2.

[55] Fuchs, H. J. *et al.* Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group. *The New England Journal of Medicine* **331**, 637–42 (1994). URL http://www.ncbi.nlm.nih.gov/pubmed/7503821.

[56] Sibley, C. D. *et al.* McKay agar enables routine quantification of the 'Streptococcus milleri' group in cystic fibrosis patients. *Journal of Medical Microbiology* **59**, 534–40 (2010). URL http://www.ncbi.nlm.nih.gov/pubmed/20093379.

[57] Whelan, F. J., Rossi, L., Stearns, J. C. & Surette, M. G. Culture and Molecular Profiling of the Respiratory Tract Microbiota. 49–61 (2018). URL https://doi.org/10.1007/978-1-4939-8728-3{_}4,http://link.springer.com/10.1007/978-1-4939-8728-3{_}4.

[58] Whelan, F. J. *et al.* The loss of topography in the microbial communities of the upper respiratory tract in the elderly. *Annals of the American Thoracic Society* **11**, 513–21 (2014). URL http://www.ncbi.nlm.nih.gov/pubmed/24601676.

[59] Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G. & Neufeld, J. D. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Applied and Environmental Microbiology* **77**, 3846–52 (2011). URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3127616{\&}tool=pmcentrez{\&}rendertype=abstract.

[60] Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011). URL http://journal.embnet.org/index.php/embnetjournal/article/view/200.

[61] Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31 (2012). URL http://www.biomedcentral.com/1471-2105/13/31.

[62] Ye, Y. Identification and Quantification of Abundant Species from Pyrosequences of 16S rRNA by Consensus Alignment. *Proceedings (IEEE Int Conf Bioinformatics Biomed)* 153–157 (2011). URL http://www.ncbi.nlm.nih.gov/pubmed/22102981http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3217275.

[63] Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–1 (2010). URL http://bioinformatics.oxfordjournals.org/content/26/19/2460.

[64] Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–6 (2010). URL http://dx.doi.org/10.1038/nmeth.f.303.

[65] Wang, Q., Garrity, G. M., Tiedje, J. M. & Cole, J. R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* **73**, 5261–7 (2007). URL http://www.ncbi.nlm.nih.gov/pubmed/17586664http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1950982.

[66] DeSantis, T. Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology* **72**, 5069–5072 (2006).

[67] McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLOS ONE* **8**, e61217 (2013). URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3632530{\&}tool=pmcentrez{\&}rendertype=abstract`.

[68] Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, New York, 2009). URL `http://had.co.nz/ggplot2/book`.

[69] McMurdie, P. J. & Holmes, S. Waste not, want not: why rarefying microbiome data is inadmissible. *PLOS Computational Biology* **10**, e1003531 (2014). URL `http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003531`.

[70] Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C. & Segata, N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**, e1029 (2015). URL `http://www.ncbi.nlm.nih.gov/pubmed/26157614`.

[71] Kolde, R. Pheatmap: pretty heatmaps (2012).

[72] Denton, M., Hall, M., Todd, N., Kerr, K. & Littlewood, J. Improved isolation of Stenotrophomonas maltophilia from the sputa of patients with cystic fibrosis using a selective medium. *Clinical Microbiology and Infection* **6**, 395–396 (2000). URL `http://www.sciencedirect.com/science/article/pii/S1198743X14649867`.

[73] Schmieder, R. & Edwards, R. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLOS ONE* **6**, e17288 (2011). URL `http://www.ncbi.nlm.nih.gov/pubmed/21408061http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3052304`.

[74] Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* **31**, 1674–1676 (2015). URL `http://www.ncbi.nlm.nih.gov/pubmed/25609793https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btv033`.

[75] Wu, Y.-W., Tang, Y.-H., Tringe, S. G., Simmons, B. A. & Singer, S. W. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**, 26 (2014). URL `http://www.microbiomejournal.com/content/2/1/26`.

[76] Sczyrba, A. *et al.* Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software. *Nature Publishing Group* (2017). URL https://www.nature.com/nmeth/journal/vaop/ncurrent/pdf/nmeth.4458.pdf.

[77] Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* **25**, 1043–1055 (2015). URL http://genome.cshlp.org/content/25/7/1043.long.

[78] Breitwieser, F. P. & Salzberg, S. L. KrakenHLL: Confident and fast metagenomics classification using unique k-mer counts. *bioRxiv* 262956 (2018). URL https://www.biorxiv.org/content/biorxiv/early/2018/02/09/262956.full.pdfhttps://www.biorxiv.org/content/early/2018/02/09/262956.

[79] Lee, S. T. M. *et al.* Tracking microbial colonization in fecal microbiota transplantation experiments via genome-resolved metagenomics URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5418705/pdf/40168{_}2017{_}Article{_}270.pdf.

[80] Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome biology* **15**, R46 (2014). URL http://genomebiology.com/2014/15/3/R46.

[81] Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990). URL http://linkinghub.elsevier.com/retrieve/pii/S0022283605803602.

[82] Mao, C. *et al.* Curation, integration and visualization of bacterial virulence factors in PATRIC. *Bioinformatics (Oxford, England)* **31**, 252–8 (2015). URL http://www.ncbi.nlm.nih.gov/pubmed/25273106http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4287947.

[83] Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution* **34**, 2115–2122 (2017). URL http://academic.oup.com/mbe/article/34/8/2115/3782716/Fast-GenomeWide-Functional-Annotation-through.

[84] Huerta-Cepas, J. *et al.* eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Research* **44**, D286–D293 (2016). URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1248.

[85] Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research* **44**, W16–W21 (2016). URL http://www.ncbi.nlm.nih.gov/pubmed/27141966http:

//www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4987931https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw387.

[86] Jia, B. *et al.* CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Research* **45**, D566–D573 (2017). URL http://www.ncbi.nlm.nih.gov/pubmed/27789705http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5210516https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1004.

[87] Skinnider, M. A. *et al.* Genomes to natural products PRediction Informatics for Secondary Metabolomes (PRISM). *Nucleic Acids Research* **43**, gkv1012 (2015). URL https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1012.

[88] Bland, C. *et al.* CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* **8**, 209 (2007). URL http://www.ncbi.nlm.nih.gov/pubmed/17577412http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1924867http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-209.

Acknowledgements

This research was funded in part by a Canadian Institutes of Health Research (CIHR) Doctoral Scholarship, a Cystic Fibrosis Canada (CFC) Studentship, and a Marie Skłodowska-Curie Individual Fellowship (GA #: 793818) awarded to FJW, and grants from CIHR, CFC, and a Tier 1 Canada Research Chair (CRC) to MGS. The authors would like to thank the patients and health care professionals at the Calgary Adult Cystic Fibrosis Clinic for their participation and assistance with this study. We acknowledge critical intellectual conversations with J.T. Lau, and J.C. Szamosi. We wish to acknowledge that this research was conducted on traditional territory shared by the Haudenosaunee confederacy and the Anishinaabe nations as well as the peoples of the Treaty 7 region in Southern Alberta.

Author Contributions

FJW is the primary author of this prepared manuscript. HRR, and MDP collected patient information and enrolled willing participants for this study. BW collected, processed, and cultured all sputum samples in addition to all biological and technical controls. FJW and SAS isolated DNA from culture/sputum material, and ran PCR reactions to amplify the 16S rRNA gene variable 3 region. SAS performed the enrichment of *Stenotrophomonas*. FJW prepared DNA for metagenomic sequencing. FJW processed and analysed all 16S rRNA

gene, and metagenomic sequencing results. SS provided code for the taxonomic assignment of metagenomic bins. FJW, MDP, and MGS conceptualized the experimental outline. FJW conducted all data analyses and wrote this manuscript. All authors edited and approved the manuscript.

Competing Interests

The authors declare no competing interests.

Corresponding author

Correspondence to Michael G. Surette.

Figure Legends & Tables

Figure 1: **The culture-enriched metagenomic sequencing work flow.** Sputum samples collected from cystic fibrosis patients were plated onto 13 selective and non-selective media and incubated either aerobically or anaerobically. 16S rRNA gene sequencing was conducted on the sputum sample (*direct sequencing*) as well as on each media type (*culture-enriched sequencing*).

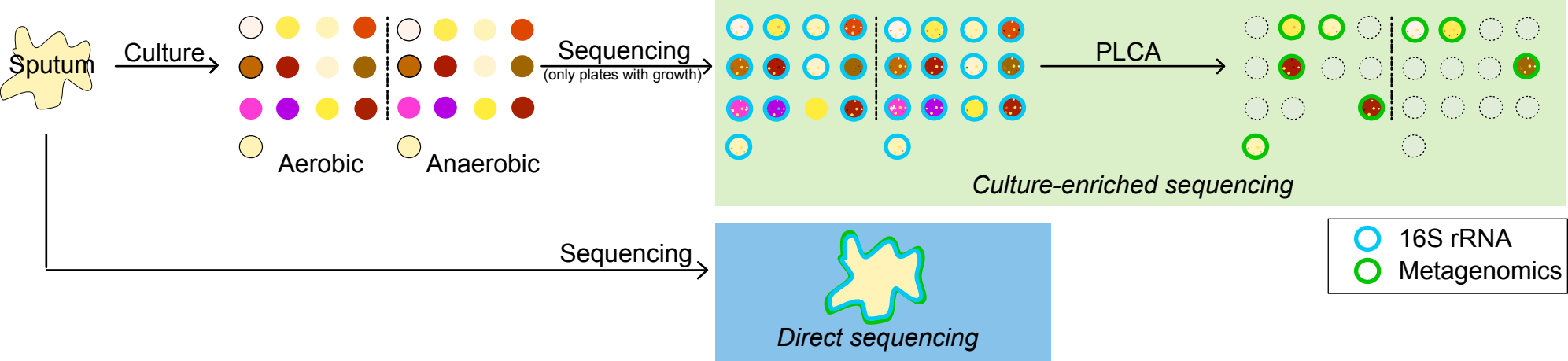
Figure 2: **The majority of the cystic fibrosis lung microbiota is culturable. a.** The majority of OTUs identified in the direct 16S rRNA gene sequencing were also recovered by culture-enriched sequencing. **b.** In fact, 63.3% of OTUs across the dataset were identified only by culture-enriched sequencing (**green ring**). In contrast, 5.7% of OTUs were not cultured including many Tenericutes (^), and TM7 (*****), and all Spirochaetes (**&**), and SR1 (+). Similar results were obtained with a more stringent relative abundance cutoff (**Extended Data Fig 1a-b**).

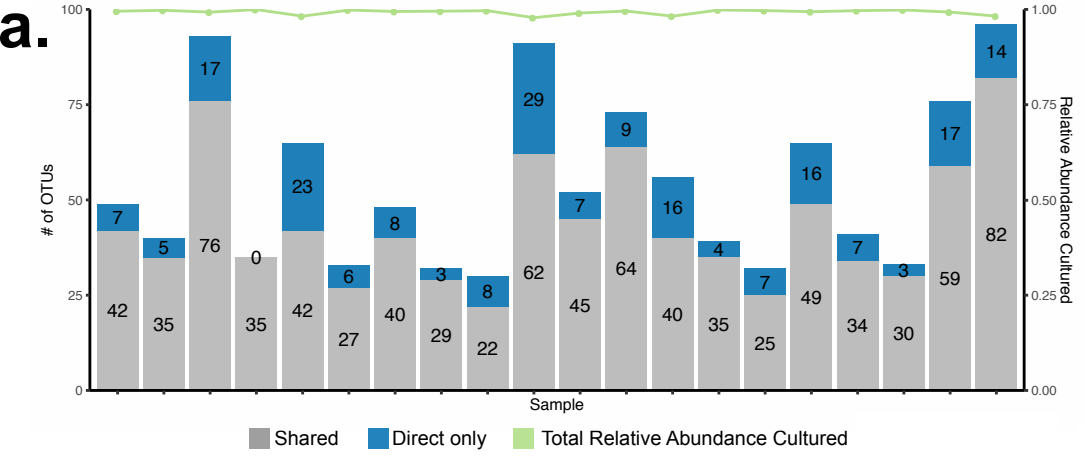
Figure 3: **Taxonomic diversity captured across culture-enrichment conditions.** The use of both anaerobic and aerobic conditions allows greater organism recovery as shown by a pair of direct and culture-enriched sequencing of a representative sample (taxonomic summaries (**a**) and Unweighted Unifrac β-diversity metric (n=26, **b**)). In **a**, culture-enriched plates, indicated with circles as in **Figure 1**, are displayed alphabetically (**Extended Data Fig 1**). In **b**, the direct sequencing clusters with some aerobic samples due to the abundance of Pseudomonas sp., and not due to a lack of bacterial growth (**Supplementary Fig 1**). **c.** A heatmap showing the maximum observed relative abundance (range 0-1) of each genus across culturing conditions. Aerobic (Aer) and anaerobic (Ana) culture condition acronyms and recipes are provided in the *Methods*. Genus-level labelling is available in **Extended Data Fig 4**. **d.** Within a genus, different OTUs can also have different culture preferences (also see **Extended Data Fig 5**). **e.** The number of OTUs obtained from culture-enrichment is compared to the number obtained if only aerobic culturing was used, or if culture was restricted to that of a standard clinical microbiology laboratory (CBA.Aer, MAC.Aer, MSA.Aer, CHOC.Ana). **f.** Cultured organisms can be recovered from frozen bacterial stocks. Here, *Stenotrophomonas sp.* were isolated from stocks with a relative abundance of 1.3% on CNA.Aer and 1.5% on TSY.Aer, respectively. Plates with low abundance of *Stenotrophomonas sp.* were purposefully chosen to indicate the power of this approach.

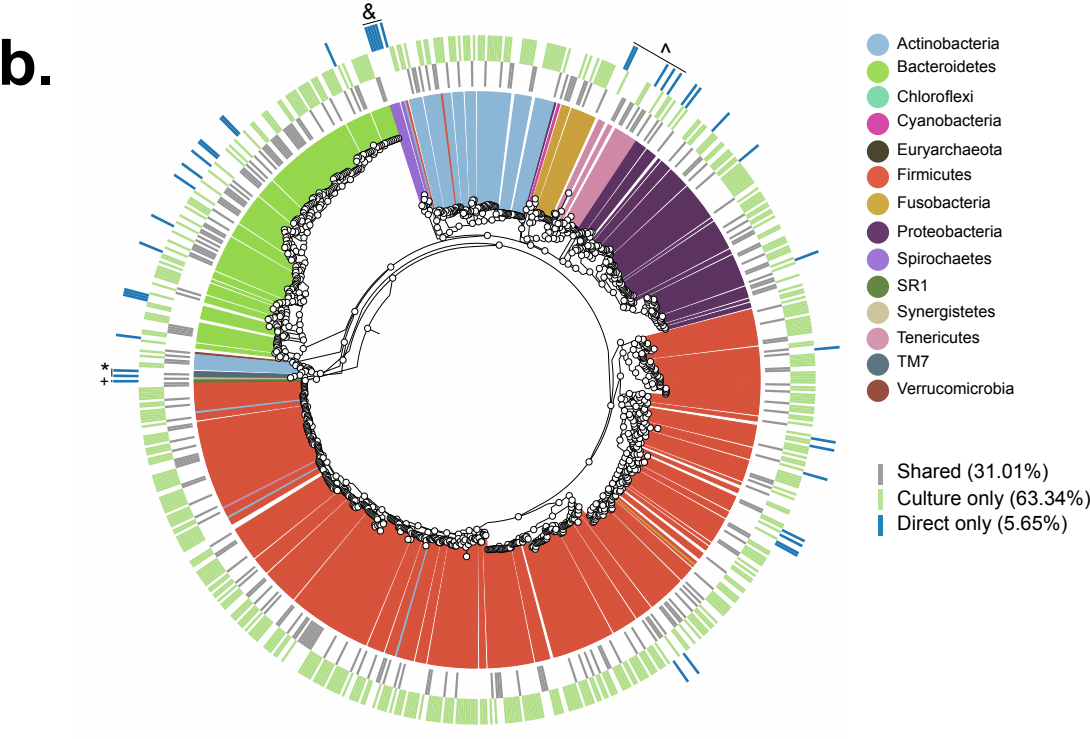
Figure 4: **The PLCA determines an optimal plate set for culture-enriched metagenomics. a.** The number of plates and OTUs necessary to capture the taxonomic diversity above varying thresholds of a representative sample using the denovo and adjusted PLCAs. The resultant OTUs are divided into those obtained above the threshold and those captured by consequence of being present on plates within the optimal plate set. A similar output for all other samples is available (**Extended Data Fig 6**). **b.** The plate subsets for the denovo and adjusted PLCAs for each sample in the dataset. Each culture condition is represented with a gray dot which is coloured in the samples (S) in which it is part of the PLCAs optimal plate set. **c.** The number of identified OTUs obtained with the denovo (**blue**) and adjusted PLCA (**orange**) when applied to Sample 1 with the thresholds (**dotted lines**) displayed in **b**. Because the aim of the denovo PLCA is to recover the most abundant cultured organisms, the OTUs identified by culture-enrichment are displayed; in contrast, the adjusted PLCA aims to recover abundant OTUs from the original sample and thus the OTUs identified in the direct sequencing are shown.

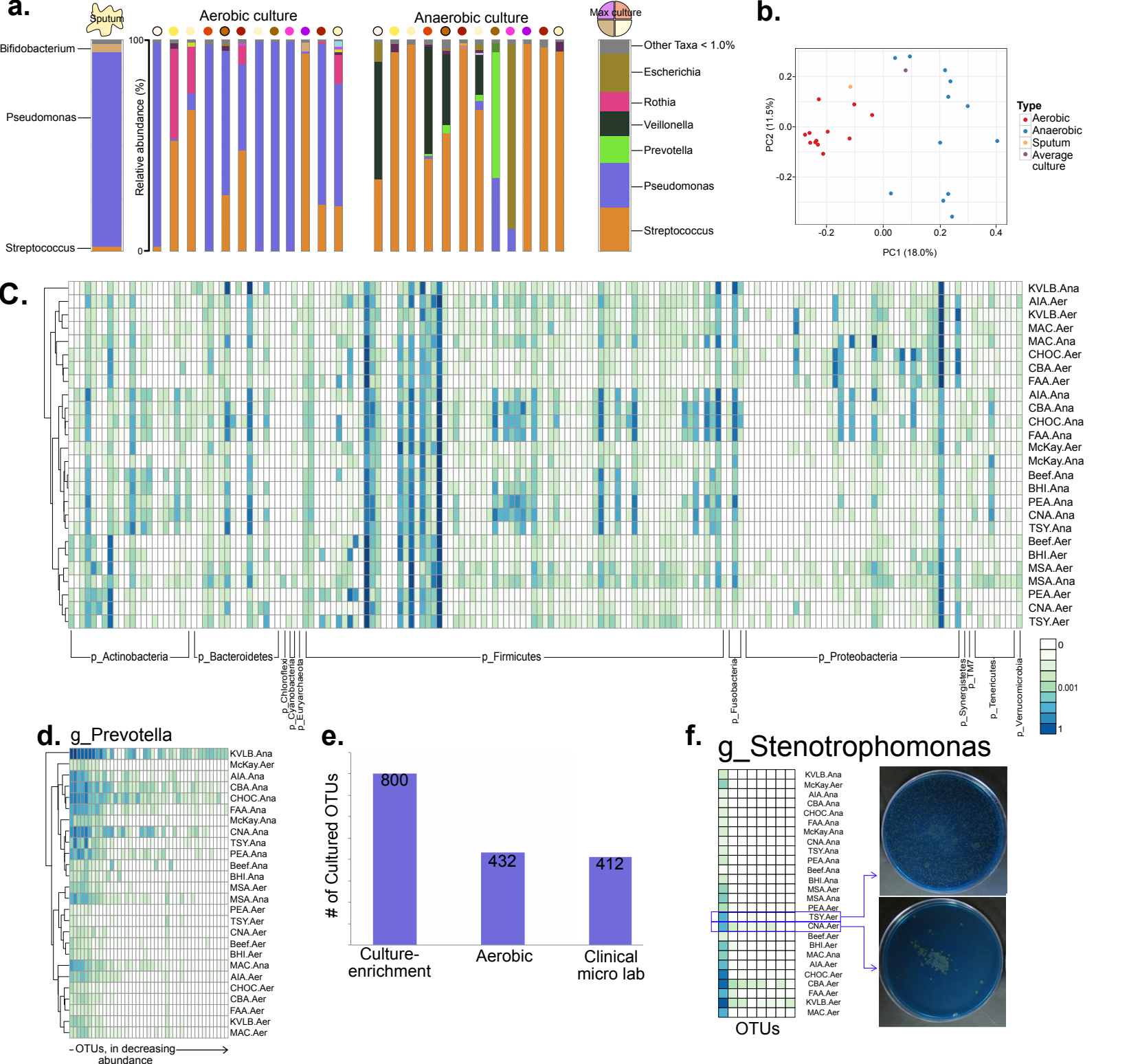
Figure 5: **MAG and non-MAG bins resulting from culture-enriched and direct metagenomic sequencing of the first sample in the dataset**. Three sets of metagenomic sequencing were performed on a representative sample in order to compare approaches. The statistics for the bins resulting from direct (**blue**) and culture-enriched (**green**) metagenomic sequencing. Culture-enriched metagenomics resulted in more metagenomic bins and more MAGs (**hashed overlay**) compared to direct sequencing. The coverage and contamination values (in parentheses) are displayed above each bin. Blue dots indicate the percent coverage of direct sequencing reads to each bin.

Figure 6: **The taxonomic and functional diversity of direct and culture-enriched metagenomic sequencing. a-c.** Bins are ordered by completeness. From left to right, estimations of virulence gene counts, the prevalence of proteins within COG functional categories, phage predictions, antibiotic resistance genes, counts of CRISPR genes, and secondary metabolite predictions were estimated for direct (**a**) and culture-enriched (**b-c**) metagenomic sequencing. Predictions are split into rows based on metagenomic bin assignments (bold assignments indicate MAGs). **d.** Haplotype diversity of the open reading frames within each bin. Asterisks identify MAGs and bins are boxed if they correspond to *Pseudomonas sp.*. Sample numbers are present on the x-axis. Boxplots display the 1st and 3rd quartiles, a horizontal line to indicate the median, and whiskers extending to 1.5 times the interquartile range. 2o Metab. = Secondary Metabolites; NRP = Nonribosomal peptide; PK = Polyketide; NRP/PK = Nonribosomal peptide/polyketide; UTCT = Unknown thiotemplated cluster type.

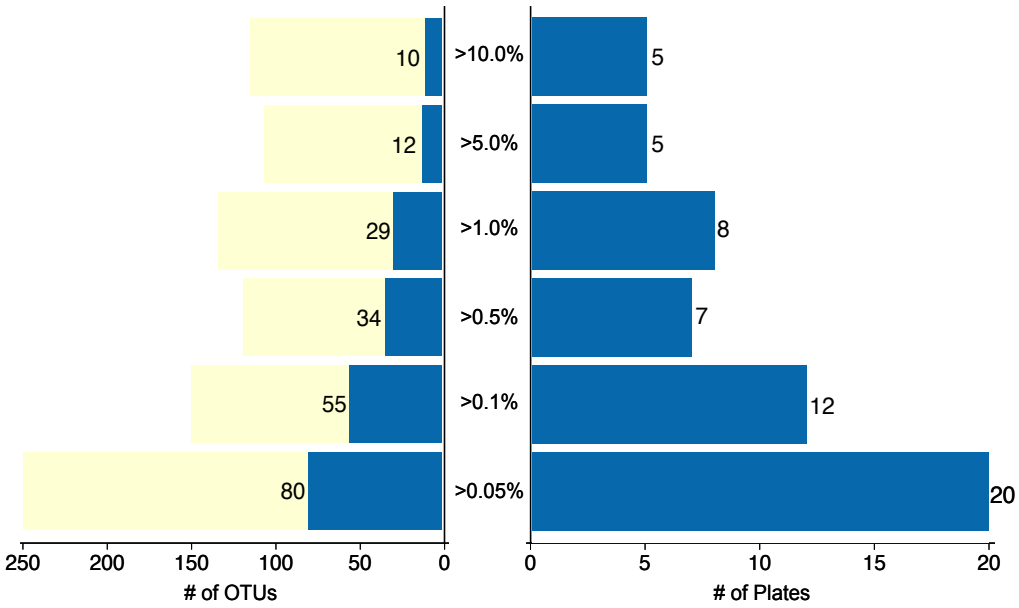

Sputum
Culture
Aerobic
Anaerobic
Sequencing
(only plates with growth)
PLCA
Culture-enriched sequencing
Sequencing
Direct sequencing
16S rRNA
Metagenomics

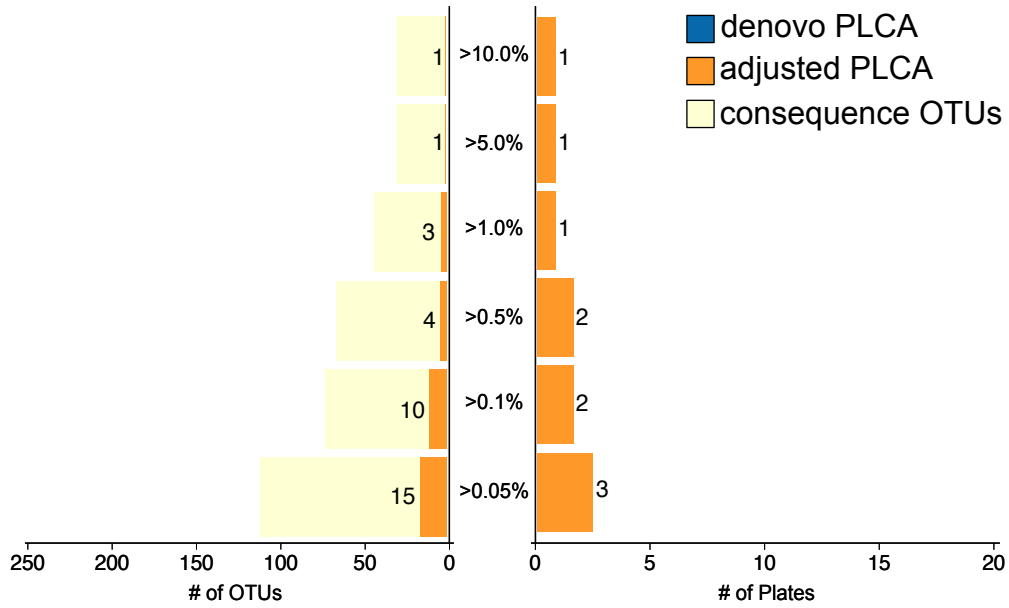

a.
# of OTUs
Relative Abundance Cultured
Sample
Shared
Direct only
Total Relative Abundance Cultured


b.
Actinobacteria
Bacteroidetes
Chloroflexi
Cyanobacteria
Euryarchaeota
Firmicutes
Fusobacteria
Proteobacteria
Spirochaetes
SR1
Synergistetes
Tenericutes
TM7
Verrucomicrobia
Shared (31.01%)
Culture only (63.34%)
Direct only (5.65%)

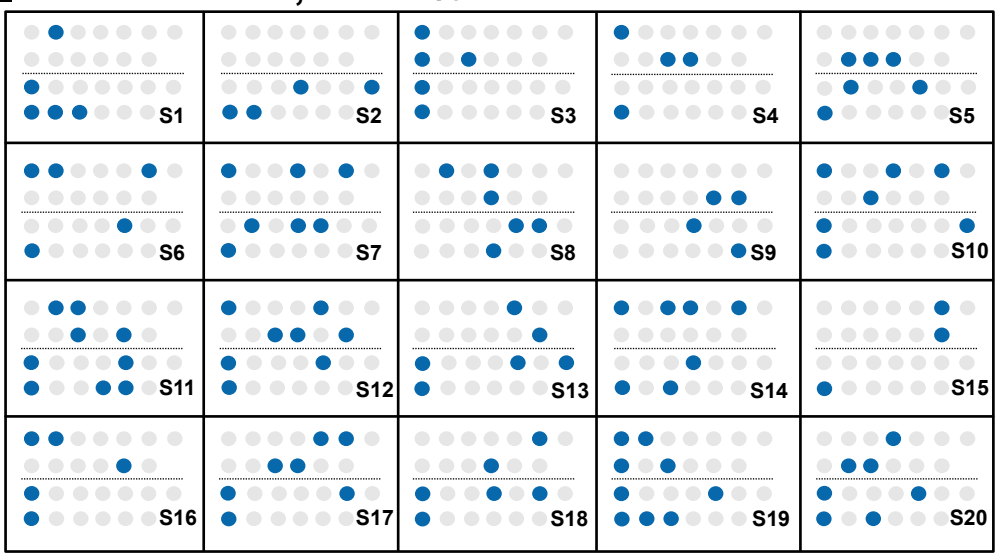

a.
Sputum
Aerobic culture
Anaerobic culture
Max culture
Bifidobacterium
Pseudomonas
Streptococcus
Relative abundance (%)
100
0
Other Taxa < 1.0%
Escherichia
Rothia
Veillonella
Prevotella
Pseudomonas
Streptococcus
b.
PC2 (11.5%)
PC1 (18.0%)
0.2
0.0
-0.2
-0.2
0.00
0.2
0.4
Type
Aerobic
Anaerobic
Sputum
Average culture
c.
KVLB.Ana
AIA.Aer
KVLB.Aer
MAC.Aer
MAC.Ana
CHOC.Aer
CBA.Aer
FAA.Aer
AIA.Ana
CBA.Ana
CHOC.Ana
FAA.Ana
McKay.Aer
McKay.Ana
Beef.Ana
BHI.Ana
PEA.Ana
CNA.Ana
TSY.Ana
Beef.Aer
BHI.Aer
MSA.Aer
MSA.Ana
PEA.Aer
CNA.Aer
TSY.Aer
p_Actinobacteria
p_Bacteroidetes
p_Chloroflexi
p_Cyanobacteria
p_Euryarchaeota
p_Firmicutes
p_Fusobacteria
p_Proteobacteria
p_Synergistetes
p_TM7
p_Tenericutes
p_Verrucomicrobia
0
0.001
1
d.
g_Prevotella
KVLB.Ana
McKay.Aer
AIA.Ana
CBA.Ana
CHOC.Ana
FAA.Ana
McKay.Ana
CNA.Ana
TSY.Ana
PEA.Ana
Beef.Ana
BHI.Ana
MSA.Aer
MSA.Ana
PEA.Aer
TSY.Aer
CNA.Aer
Beef.Aer
BHI.Aer
MAC.Ana
AIA.Aer
CHOC.Aer
CBA.Aer
FAA.Aer
KVLB.Aer
MAC.Aer
OTUs, in decreasing abundance
e.
# of Cultured OTUs
800
432
412
Culture-enrichment
Aerobic
Clinical micro lab
f.
g_Stenotrophomonas
KVLB.Ana
McKay.Aer
AIA.Ana
CBA.Ana
CHOC.Ana
FAA.Ana
McKay.Ana
CNA.Ana
TSY.Ana
PEA.Ana
Beef.Ana
BHI.Ana
MSA.Aer
MSA.Ana
PEA.Aer
TSY.Aer
CNA.Aer
Beef.Aer
BHI.Aer
MAC.Ana
AIA.Aer
CHOC.Aer
CBA.Aer
FAA.Aer
KVLB.Aer
MAC.Aer
OTUs

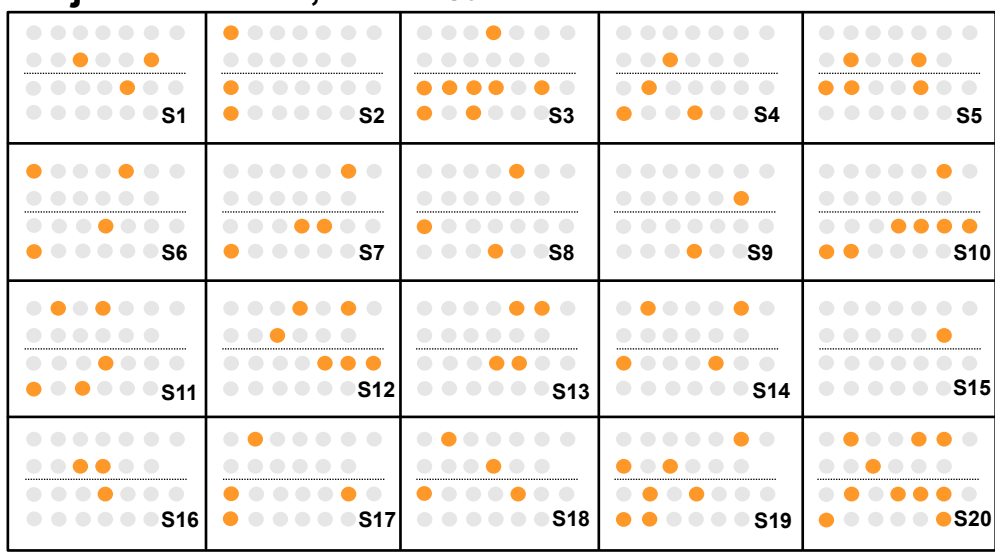
a.


10
12
29
34
55
80
>10.0%
>5.0%
>1.0%
>0.5%
>0.1%
>0.05%
5
5
8
7
12
20
250 200 150 100 50 0
# of OTUs
0 5 10 15 20
# of Plates

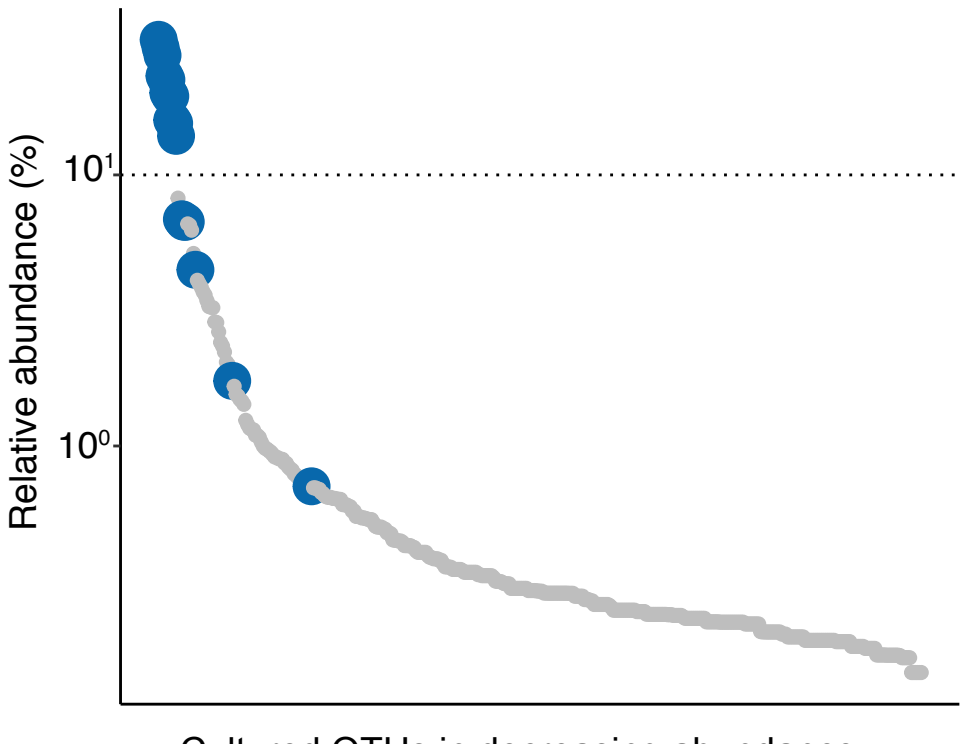

denovo PLCA
adjusted PLCA
consequence OTUs
1
1
3
4
10
15
>10.0%
>5.0%
>1.0%
>0.5%
>0.1%
>0.05%
1
1
1
2
2
3
250 200 150 100 50 0
# of OTUs
0 5 10 15 20
# of Plates

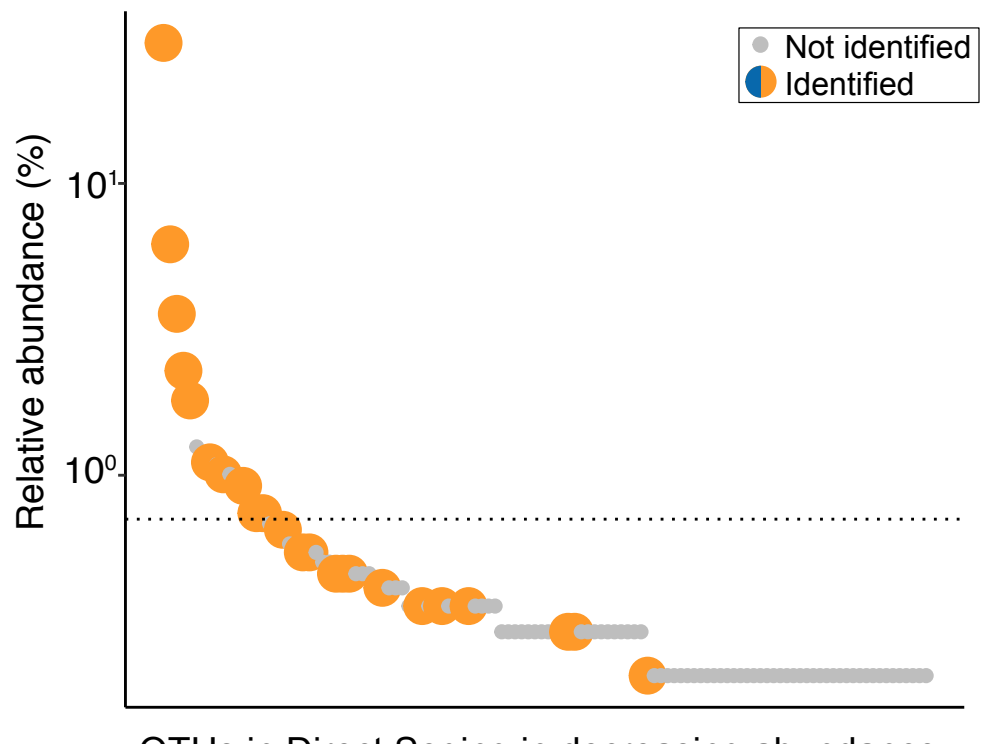
b.

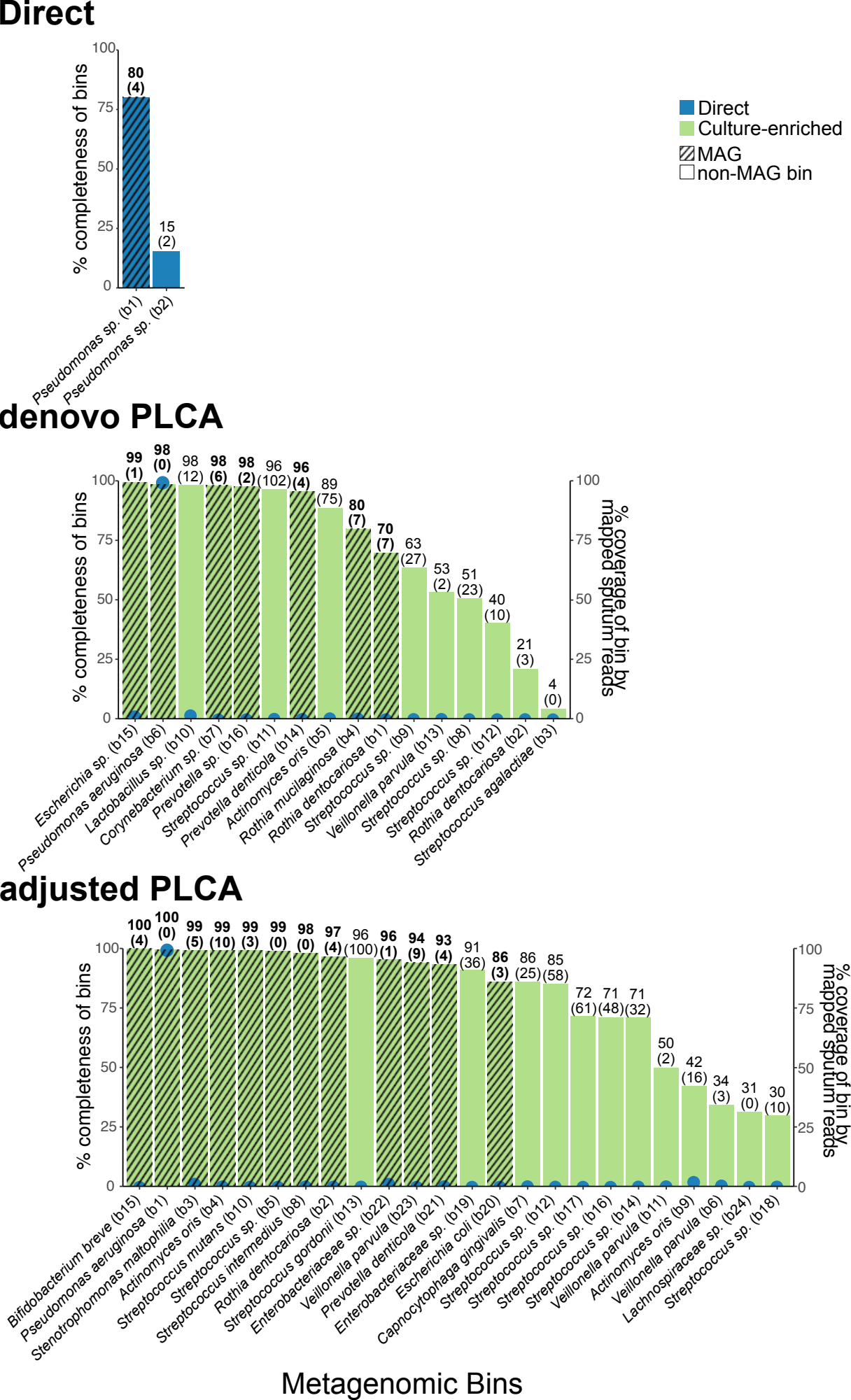

denovo PLCA, > 10.0%
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
S13
S14
S15
S16
S17
S18
S19
S20

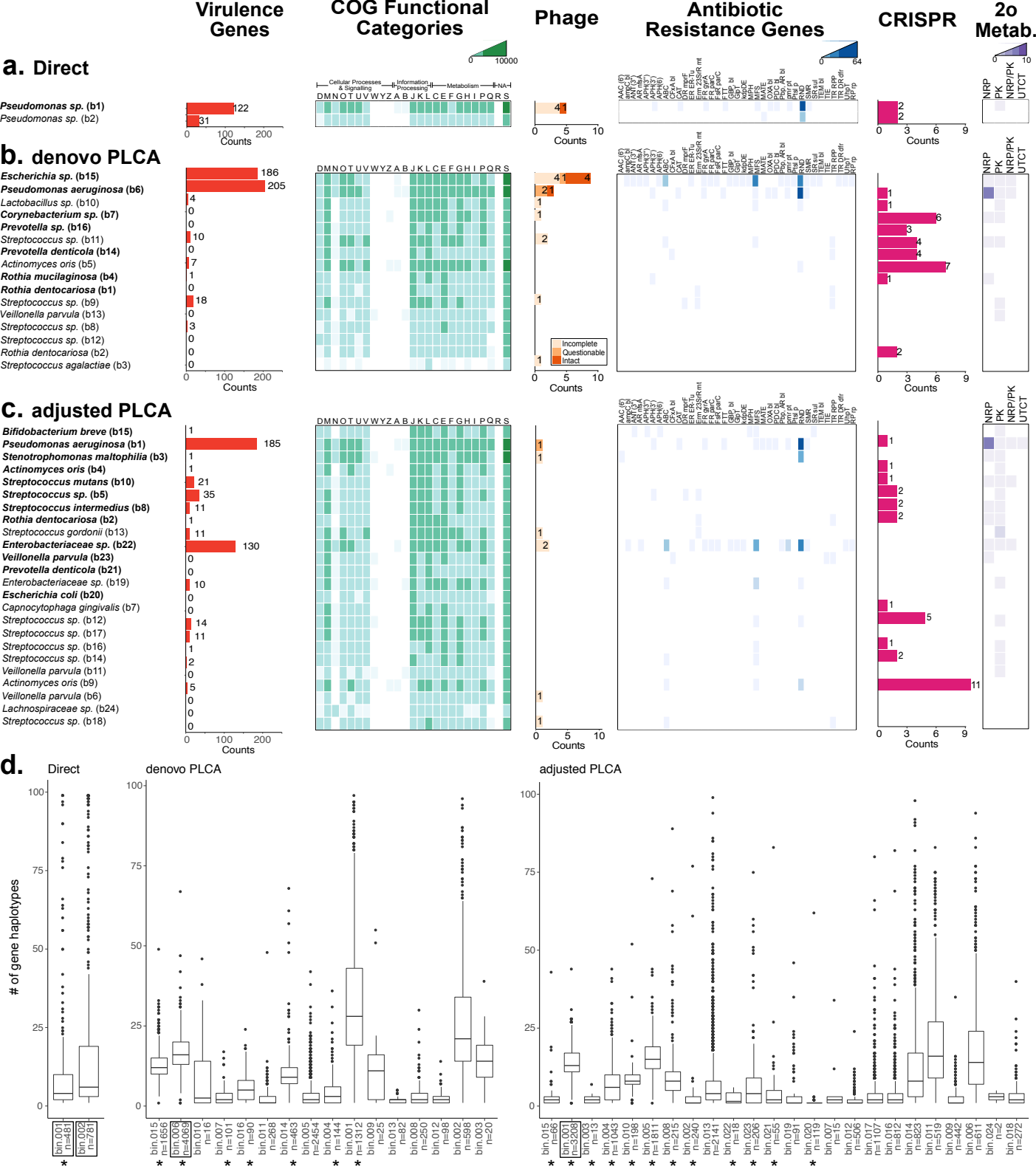

adjusted PLCA, > 0.05%
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
S13
S14
S15
S16
S17
S18
S19
S20

c.


Relative abundance (%)
10^1
10^0
Cultured OTUs in decreasing abundance


Not identified
Identified
Relative abundance (%)
10^1
10^0
OTUs in Direct Seqing in decreasing abundance


Direct
% completeness of bins
100
75
50
25
0
80
(4)
15
(2)
Pseudomonas sp. (b1)
Pseudomonas sp. (b2)
Direct
Culture-enriched
MAG
non-MAG bin
denovo PLCA
% completeness of bins
% coverage of bin by mapped sputum reads
99 (1)
98 (0)
98 (12)
98 (6)
98 (2)
96 (102)
96 (4)
89 (75)
80 (7)
70 (7)
63 (27)
53 (2)
51 (23)
40 (10)
21 (3)
4 (0)
Escherichia sp. (b15)
Pseudomonas aeruginosa (b6)
Lactobacillus sp. (b10)
Corynebacterium sp. (b7)
Prevotella sp. (b16)
Streptococcus sp. (b11)
Prevotella denticola (b14)
Actinomyces oris (b5)
Rothia mucilaginosa (b4)
Rothia dentocariosa (b1)
Streptococcus sp. (b9)
Veillonella parvula (b13)
Streptococcus sp. (b8)
Streptococcus sp. (b12)
Rothia dentocariosa (b2)
Streptococcus agalactiae (b3)
adjusted PLCA
% completeness of bins
% coverage of bin by mapped sputum reads
100 (4)
100 (0)
99 (5)
99 (10)
99 (3)
99 (0)
98 (0)
97 (4)
96 (100)
96 (1)
94 (9)
93 (4)
91 (36)
86 (3)
86 (25)
85 (58)
72 (61)
71 (48)
71 (32)
50 (2)
42 (16)
34 (3)
31 (0)
30 (10)
Bifidobacterium breve (b15)
Pseudomonas aeruginosa (b1)
Stenotrophomonas maltophilia (b3)
Actinomyces oris (b4)
Streptococcus mutans (b10)
Streptococcus sp. (b5)
Streptococcus intermedius (b8)
Rothia dentocariosa (b2)
Streptococcus gordonii (b13)
Enterobacteriaceae sp. (b22)
Veillonella parvula (b23)
Prevotella denticola (b21)
Enterobacteriaceae sp. (b19)
Escherichia coli (b20)
Capnocytophaga gingivalis (b7)
Streptococcus sp. (b12)
Streptococcus sp. (b17)
Streptococcus sp. (b16)
Streptococcus sp. (b14)
Veillonella parvula (b11)
Actinomyces oris (b9)
Veillonella parvula (b6)
Lachnospiraceae sp. (b24)
Streptococcus sp. (b18)
Metagenomic Bins


Virulence Genes
COG Functional Categories
Phage
Antibiotic Resistance Genes
CRISPR
2o Metab.
a. Direct
Pseudomonas sp. (b1)
Pseudomonas sp. (b2)
b. denovo PLCA
Escherichia sp. (b15)
Pseudomonas aeruginosa (b6)
Lactobacillus sp. (b10)
Corynebacterium sp. (b7)
Prevotella sp. (b16)
Streptococcus sp. (b11)
Prevotella denticola (b14)
Actinomyces oris (b5)
Rothia mucilaginosa (b4)
Rothia dentocariosa (b1)
Streptococcus sp. (b9)
Veillonella parvula (b13)
Streptococcus sp. (b8)
Streptococcus sp. (b12)
Rothia dentocariosa (b2)
Streptococcus agalactiae (b3)
c. adjusted PLCA
Bifidobacterium breve (b15)
Pseudomonas aeruginosa (b1)
Stenotrophomonas maltophilia (b3)
Actinomyces oris (b4)
Streptococcus mutans (b10)
Streptococcus sp. (b5)
Streptococcus intermedius (b8)
Rothia dentocariosa (b2)
Streptococcus gordonii (b13)
Enterobacteriaceae sp. (b22)
Veillonella parvula (b23)
Prevotella denticola (b21)
Enterobacteriaceae sp. (b19)
Escherichia coli (b20)
Capnocytophaga gingivalis (b7)
Streptococcus sp. (b12)
Streptococcus sp. (b17)
Streptococcus sp. (b16)
Streptococcus sp. (b14)
Veillonella parvula (b11)
Actinomyces oris (b9)
Veillonella parvula (b6)
Lachnospiraceae sp. (b24)
Streptococcus sp. (b18)
Counts
Cellular Processes & Signalling
Information Processing
Metabolism
NA
Incomplete
Questionable
Intact
NRP
PK
NRP/PK
UTCT
d.
Direct
denovo PLCA
adjusted PLCA
# of gene haplotypes