

Essential genes for *Haemophilus parainfluenzae* survival and biofilm growth

Thais H. de Palma,¹ Chris Powers,¹ Morgan J. McPartland,¹ Jessica Mark Welch,² Matthew Ramsey¹

AUTHOR AFFILIATIONS See affiliation list on p. 11.

ABSTRACT *Haemophilus parainfluenzae* (*Hp*) is a Gram-negative, highly prevalent, and abundant commensal in the human oral cavity, and an infrequent extraoral opportunistic pathogen. *Hp* occupies multiple niches in the oral cavity, including the supragingival plaque biofilm. Little is known about how *Hp* interacts with its neighbors in healthy biofilms nor its mechanisms of pathogenesis as an opportunistic pathogen. To address this, we identified the essential genome and conditionally essential genes in *in vitro* biofilms aerobically and anaerobically. Using transposon insertion sequencing (TnSeq) with a highly saturated *mariner* transposon library in two strains, the ATCC33392 type-strain (*Hp* 392) and oral isolate EL1 (*Hp* EL1), we show that the essential genomes of *Hp* 392 and *Hp* EL1 are composed of 395 (20%) and 384 (19%) genes, respectively. The core essential genome, consisting of 341 (17%) essential genes conserved between both strains, was composed of genes associated with genetic information processing, carbohydrate, protein, and energy metabolism. We also identified conditionally essential genes for aerobic and anaerobic biofilm growth, which were associated with carbohydrate and energy metabolism in both strains. RNAseq analysis determined that most genes upregulated during anaerobic growth are not essential for *Hp* 392 anaerobic survival. The completion of this library and analysis under these conditions gives us a foundational insight into the basic biology of *H. parainfluenzae* in differing oxygen conditions, similar to its *in vivo* habitat. This library presents a valuable tool for investigation into conditionally essential genes for an organism that lives in close contact with many microbial species in the human oral habitat.

IMPORTANCE *Haemophilus parainfluenzae* is a highly abundant human commensal microbe, present in most healthy individuals where it colonizes the mouth. *H. parainfluenzae* correlates with good oral health and may play a role in preservation of healthy host status. Also, *H. parainfluenzae* can cause opportunistic infections outside of the oral cavity. To date, little is known about how *H. parainfluenzae* colonizes the human host, despite being such a frequent and abundant part of our human microbiome. Here, we demonstrate the creation and use of a powerful tool, a TnSeq library, used to identify genes necessary for both the outright growth of this organism and also genes conditionally essential for growth in varying oxygen status which it can encounter in the human host. This tool and these data serve as a foundation for further study of this relatively unknown organism that may play a role in preserving human health.

KEYWORDS biofilms, facultative anaerobes, oral microbiology, *Haemophilus*, TnSeq, commensal, dental plaque

The oral cavity harbors a diverse microbial community with more than 700 species of bacteria capable of colonizing the mouth (1–3). Oral microbes are exposed to mechanical and physical disturbances such as oral hygiene practices, mastication, salivary flow, and others (4). Thus, microbes must attach to host substrates and/or each

Editor Babak Momeni, Boston College, Chestnut Hill, Massachusetts, USA

Address correspondence to Matthew Ramsey, mramsey@uri.edu.

The authors declare no conflict of interest.

See the funding table on p. 11.

Received 13 May 2024

Accepted 21 July 2024

Published 21 August 2024

Copyright © 2024 de Palma et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

other forming an attached community known as a biofilm to persist in the mouth (5, 6). The biofilm formed on the tooth surface above the gum line is known as supragingival plaque (SUPP). Among the factors that affect the composition and organization of many biofilms, oxygen stands out as a major factor (7). While the mouth is largely aerobic, oxygen availability in SUPP varies. The periphery of the biofilm is rich in oxygen, while its interior is mostly anaerobic and rich in CO₂ (8). This oxygen gradient within SUPP allows for aerobes, obligate anaerobes, and facultative anaerobes to coexist (9). One facultative anaerobe commonly found in SUPP is *Haemophilus parainfluenzae* (*Hp*), a Gram-negative bacterium. *Hp* is also found in high abundance in other niches of the oral cavity, such as the tongue dorsum, keratinized gingiva, and saliva (4, 10–12), demonstrating that *Hp* is a generalist in the mouth (4).

Hp is commonly isolated from the mouth of healthy individuals (13) and is not typically associated with oral disease. However, elsewhere in the body, *Hp* can act as an opportunistic pathogen. *Hp* is a member of the HACEK group (*Haemophilus*, *Aggregatibacter*, *Cardiobacterium hominis*, *Eikenella corrodens*, and *Kingella*), which is made up of bacterial species and genera that commonly colonize the oropharynx and can cause infective endocarditis (14). Besides endocarditis, *Hp* has also been infrequently associated with other diseases including meningitis (15, 16), septicemia (17), pleural effusion (18), urethritis (19), prosthetic joint infection (20), and respiratory infections such as pneumonia (17, 21), pharyngitis and epiglottitis (17), and chronic obstructive pulmonary disease (22).

To fully understand what is required for *Hp* to exist as a key member of the healthy oral microbiome and an infrequent extraoral opportunistic pathogen, it is useful to know the minimal set of genes that *Hp* needs to survive. A current method to determine essential genes is to use a high-throughput method known as transposon sequencing (TnSeq) (23). This technique has been used to assess bacteria fitness under different conditions (24–26) and has also been applied to identify the essential genome of other bacterial species that colonize the oral cavity and the upper respiratory tract such as *Streptococcus mutans* (27), *Aggregatibacter actinomycetemcomitans* (25), *Porphyromonas gingivalis* (24), *Treponema denticola* (28), and *Haemophilus influenzae* (29). In addition to TnSeq, transcriptomes can shed light on the adaptation of an organism to its environment and allow for hypotheses generation for mechanisms that may be essential for fitness in that environment. For example, transcriptome measurement via RNAseq has been used to identify the influence of oxygen in gene expression in other microorganisms (30) and has proven to be a very sensitive and comprehensive method for characterizing bacterial transcriptomes (31).

Here, we describe the essential genes necessary for *Hp* survival and the conditionally essential genes required for *Hp* to survive in a biofilm aerobically and anaerobically. Given the pronounced genomic variation among *Hp* (32, 33), we characterized essential genes in two different *Hp* strain backgrounds. Comparison of essential genes in both strains allowed us to identify the genes that were conserved and essential both in *Hp* 392 and *Hp* EL1 -named here the “absolute essential genome of *Hp*” and differentiate them from strain-specific essential genes. Identifying the essential genome of *Hp* and the conditionally essential genes for biofilm growth aerobically and anaerobically provide a basic framework of likely genes necessary for *in vivo* colonization and present a valuable tool for the study of *Hp* in both oral niches as a commensal and extraoral niches as an opportunistic pathogen.

RESULTS

Transposon library development and sequencing

We created saturating *mariner* transposon libraries via methods we have used previously (25), conjugating a delivery vector into two *Hp* strains, the ATCC3392 type-strain and the oral isolate EL1 (34), generating ~200,000 and ~100,000 individual mutants respectively. Similar to previous methods (25, 26, 35), we generated and sequenced Illumina libraries to map *mariner* insertion sites in each library. Libraries were generated in triplicate for

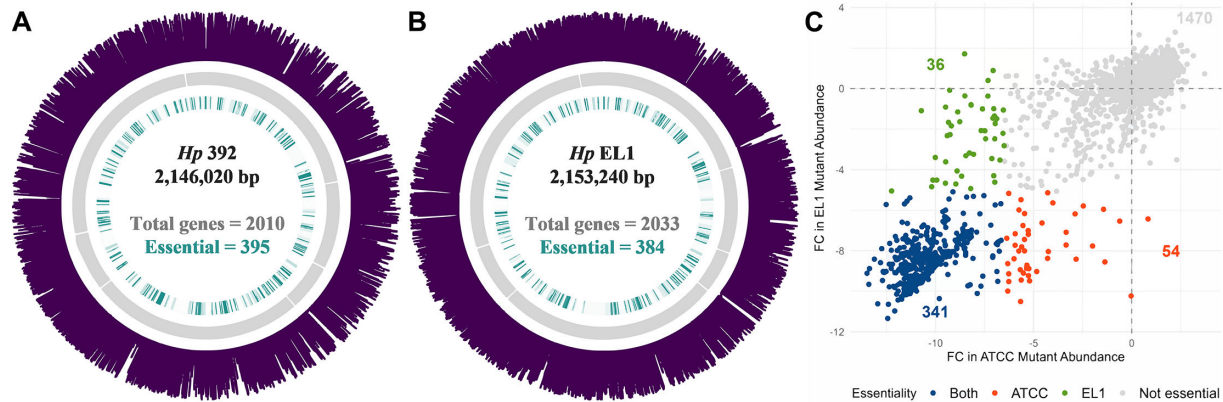


FIG 1 The essential genome of *Hp* 392 (A) and *Hp* EL1 (B). The outer (purple) ring indicates transposon insertions, next inward (gray) represents open reading frames, and the inner ring (green) indicates essential genes. Figure (C) shows the log₂ fold change (FC) in mutant abundance for *Hp* 392 (x-axis) versus EL1 (y-axis). Each point corresponds to an ortholog and genes <200 bp were excluded from the analysis.

each experiment. We identified 195,115 unique insertions present in at least one of three replicates in *Hp* 392 and 144,946 in *Hp* EL1, resulting in a coverage of approximately one insertion at every 11 and 15 bp, respectively, indicating that the mutant pools were well saturating and randomly distributed (Fig. 1A and B).

The essential genome for *H. parainfluenzae*

Analysis of insertion adjacent sequences determined which genes were not represented in the mutant library, that is, genes that render *Hp* unfit for growth and thus are absolutely essential for *in vitro* aerobic conditions on nutrient-rich medium used for library selection. We then used a Monte Carlo-based approach as previously described (25, 35, 36) to verify essential genes. Essential genes are expected to have zero, or significantly fewer, transposon insertions compared to non-essential genes, as mutations here render these strains unable to grow and are thus absent from the outgrown library. We observed that 395 (20%) and 384 (19%) genes are essential in *Hp* 392 and *Hp* EL1, respectively, and that most essential genes had zero transposon insertions (Fig. S1). We then grouped these genes into their respective KEGG categories (37–39) and determined that most of the essential genes were, as expected, involved in key functions including genetic information processing and common energy metabolism pathways (Fig. 2). A total of 102 and 109 genes that were <200 bp in length were excluded from the analysis for *Hp* 392 and *Hp* EL1, respectively, due to low insertion coverage, leading to unacceptably high variability for the determination of gene essentiality.

Core essential genes for *H. parainfluenzae*

Comparison of the type-strain *Hp* 392 genome to the commensal oral isolate EL1 revealed that 1,835 genes are shared between both strains, representing 91% and 90% of their respective genomes (Table S7). Of these 1,835 shared genes, we found that 341 were absolutely essential for survival *in vitro* (Fig. 1C). We then compared these 341 core essential genes via Anvi'o (40) to 16 *Hp* closed genomes available in the NCBI database (Fig. 3) and found that all 341 genes were conserved between all 16 genomes. These data indicate that the core essential genes detected in only two strains under laboratory conditions are also well conserved in natural populations. This is an unexpected observation, given the genomic plasticity of *Hp* (32, 33).

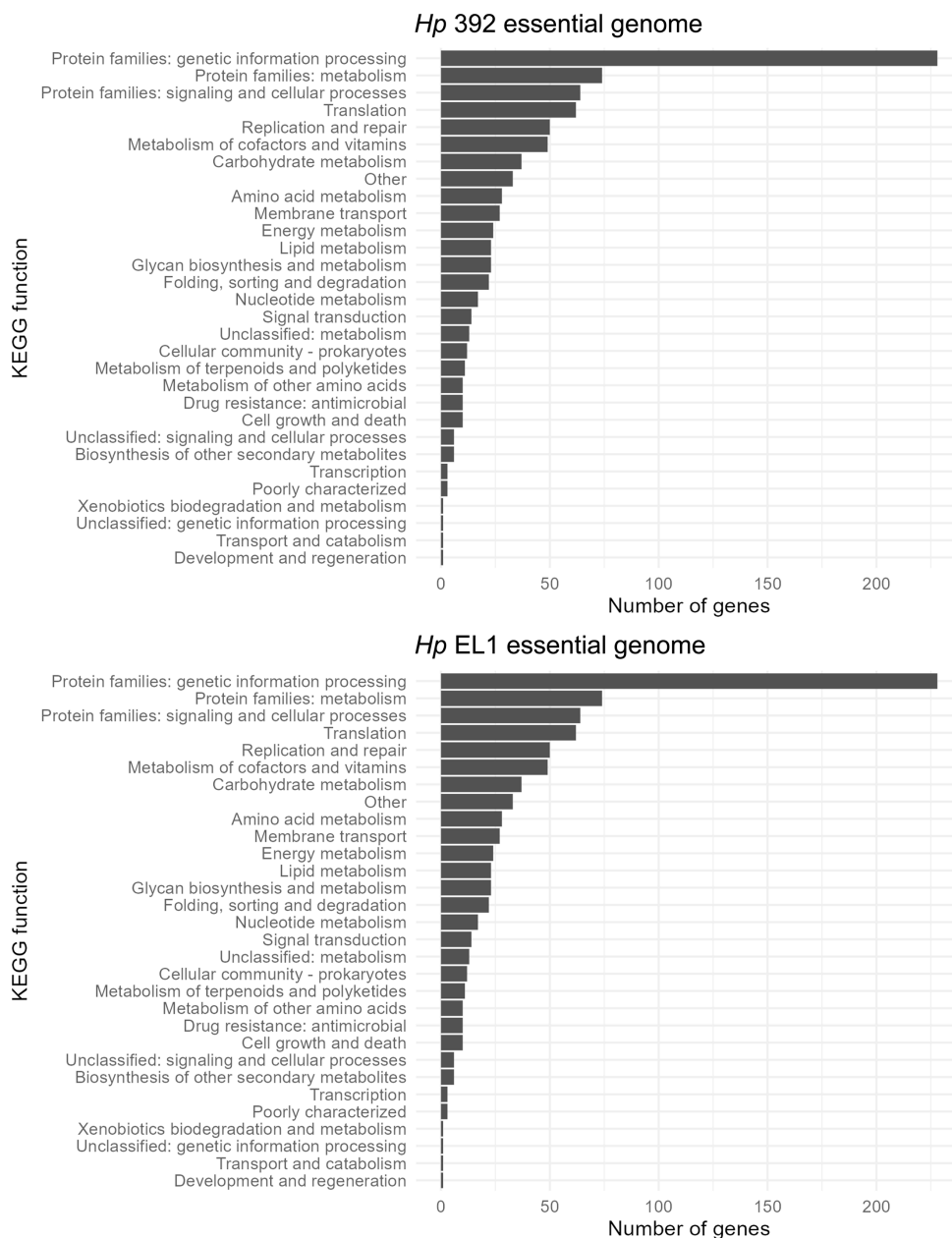


FIG 2 KEGG group assignments for absolute essential genes in *Hp*. The essential genome of *Hp* 392 (A) and *Hp* EL1 (B) grouped into KEGG categories. Genes <200 bp were not included in the analysis. Most essential genes are involved in central cell functional processes including genetic information processing and common energy metabolism pathways.

H. parainfluenzae conditionally essential genes for *in vitro* aerobic and anaerobic biofilm growth

To identify genes conditionally essential *in vitro* for either aerobic or anaerobic biofilm growth, *Hp* was grown using a colony biofilm model (41) on a permeable membrane on solid medium as we have done previously (42) and incubated either in the presence or absence of oxygen in 5% CO₂ for 24 h. Colony biofilms were scraped from the membranes, and DNA extraction and sequencing were performed to assess which genes were “conditionally” essential for growth in aerobic and anaerobic biofilms for both strains (Table 1). We next classified the conditionally essential genes based on their predicted KEGG functions (Fig. 4), observing that they are largely involved with carbohydrate and

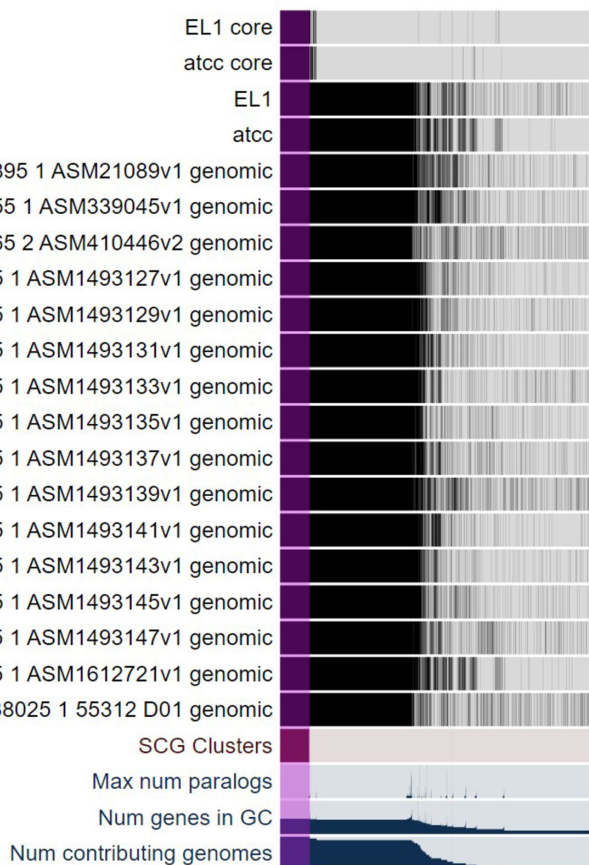


FIG 3 The core essential genome of *Hp*. Comparison of 18 genomes of *Hp* and the core essential genome of *Hp* ATCC 392 and *Hp* EL1 (top two rows) with the full genomes and the 16 available *Hp* genomes. Purple indicates the core essential genome, that is, the TnSeq identified absolute essential genes that are conserved in all species.

energy metabolism as anticipated. A full list of all genes and their predicted KEGG functions is presented in Tables S3 and S4.

The *Hp* 392 anaerobic transcriptome

To determine which genes are differentially regulated in aerobic versus anaerobic biofilm conditions, we assayed *Hp* 392 grown aerobically and anaerobically in a colony biofilm assay via RNAseq. This analysis revealed 167 genes that were significantly differentially expressed (either up- or downregulated ≥ 2 -fold) between the two conditions (Table S5A and B). Ninety-eight differentially expressed genes (DEGs) were upregulated in anaerobic biofilm conditions and 69 downregulated (Table S5A and B, respectively). Genes upregulated were largely involved in genetic information processing (26%) such as: 16S rRNA (guanine1516-N2)-methyltransferase, sigma factor RpoE regulatory protein *rseC*, formamidopyrimidine-DNA glycosylase, transcription termination encoding gene *nusA*, and translation initiation factor 2; metabolism (11%) such as: dihydroneopterin

TABLE 1 Number of conditionally essential genes for *Hp* 392 and *Hp* EL1 aerobically and anaerobically^a

Strain	Aerobic biofilm	Anaerobic biofilm	Aerobic and anaerobic biofilm
<i>Hp</i> 392	73	55	50
<i>Hp</i> EL1	119	121	108
Both	36	28	27

^aPreviously essential genes and those <200 bp are not shown.

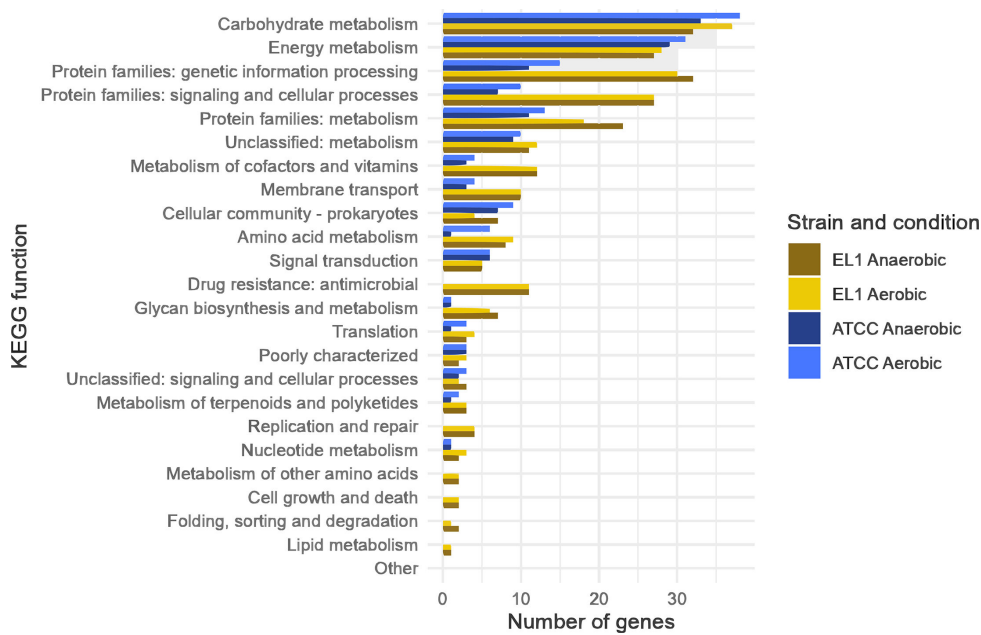


FIG 4 Genes essential for aerobic and anaerobic growth in a colony biofilm model for *Hp* 392 and *Hp* EL1. Conditionally essential genes for aerobic and anaerobic growth in *in vitro* biofilms in *Hp* 392 (blue), and *Hp* EL1 (yellow) grouped into higher KEGG functions.

aldolase, monofunctional biosynthetic peptidoglycan transglycosylase, D-alanine--D-alanine ligase, and acyl-CoA thioesterase II; and signaling and cellular processes (10%) for instance: *ampG* permease, lipid A export permease/ATP-binding gene *msbA*, and phospholipid ABC transporter ATP-binding encoding gene *miaF*. The downregulated genes were involved in signaling and cellular processes (38%) such as: TonB-dependent receptor, galactokinase; and L-lactate permease; genetic information processing (19%) including: ribosome-associated inhibitor A, galactose operon repressor, chaperone protein-encoding gene *clpB*, *dnaK*; and membrane transport (13%) including multiple sugar, amino acids, and metal transporters. Overall transcriptome responses were largely unremarkable, and we saw broad similarity between both environments, likely due to the transiently anaerobic nature of even aerobic biofilms.

Comparison of *in vitro* anaerobic biofilm transcriptome to conditionally essential genes for anaerobic biofilm growth in *Hp* 392

To identify whether genes that are induced under anaerobic biofilm growth conditions are also conditionally essential for anaerobic biofilm growth, we compared the conditionally essential genes (TnSeq data) to differentially expressed genes (RNAseq data) (Fig. 5). We observed that out of 98 genes upregulated in anaerobic conditions, only three (Murein DD-endopeptidase MepM, LPS-core synthesis glycosyltransferase PM0509, and Bacterial ribosome SSU maturation protein RimP, potentially involved in acid stress) (43) were also considered essential for anaerobic biofilm growth. Moreover, one downregulated gene (phosphoenolpyruvate carboxykinase (ATP)) was also essential. Taken together, these results indicate that there is almost no overlap between gene essentiality and differential expression, reminiscent of previous findings in *Pseudomonas aeruginosa* (26).

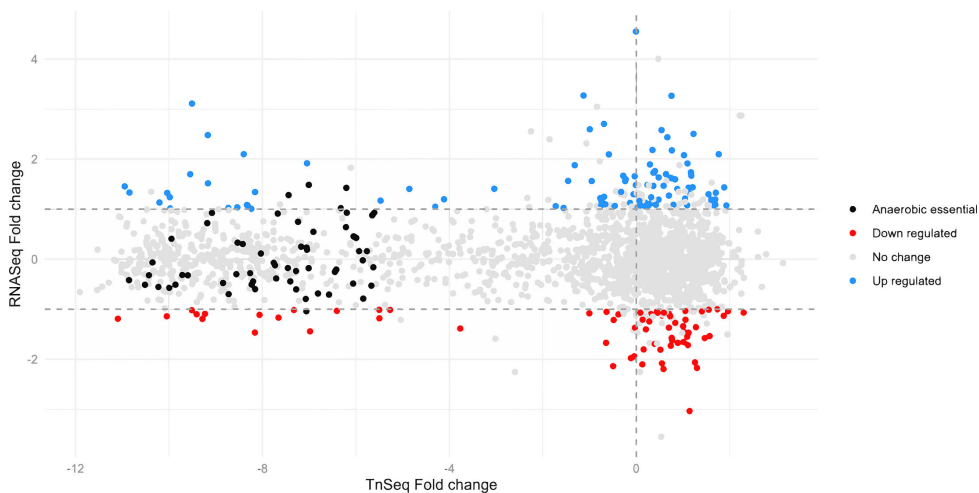


FIG 5 The majority of *Hp* 392 genes conditionally essential for anaerobic biofilm growth are not differentially regulated between aerobic and anaerobic conditions. Ninety-eight genes, indicated in blue, were significantly upregulated ($P \leq 0.05$, fold change ≥ 2) in anaerobic compared to aerobic conditions, while 67 genes, indicated in red, were significantly downregulated ($P \leq 0.05$, fold change ≤ -2). Genes uniquely essential for colony growth under anaerobic conditions are indicated in black with only three total being significantly differentially expressed.

MATERIALS AND METHODS

Strains and growth conditions

Strains and plasmids used are listed in Table S1. *Hp* ATCC 33392 and *Hp* EL1 (34) were used to build *mariner* mutant libraries. *Hp* strains were grown in brain heart infusion (BHI) supplemented with 15 $\mu\text{g}/\text{mL}$ of nicotinamide adenine dinucleotide (NAD), 15 $\mu\text{g}/\text{mL}$ of hemin, and 5% yeast extract (BHIYE-HP). For agar plates, the medium was supplemented with 1.6% agar. Cultures were incubated at 37°C, in 5% CO₂ overnight unless otherwise specified. *Escherichia coli* strains were grown in Luria-Bertani (LB) broth and agar supplemented with 0.3 mM of diaminopimelate (DAP) (for the MFD*pir* conjugation host) and incubated at 37°C overnight.

H. parainfluenzae transposon library construction

Mariner mutant pools were developed as previously described (25) with some modifications. Individual colonies of *Hp* 392 and *Hp* EL1 were inoculated into 5 mL of BHIYE-HP and incubated overnight at 37°C, in 5% CO₂. Single colonies of *E. coli* MFD*pir*, harboring the *mariner* transposon delivery plasmid pMR361-K, were inoculated into 5 mL of LB and incubated at 37°C, shaken at 200 rpm. Next, five independent conjugations for *Hp* 392 and four conjugations for EL1 were made. For that, samples were washed with phosphate-buffered saline, and OD₆₀₀ was adjusted to 1 for *Hp* strains and 0.1 *E. coli* MFD*pir*. *Hp* cultures were then heat shocked at 46°C for 6 min. Next, 10 μL of *Hp* and 10 μL of *E. coli* MFD*pir* were combined and plated on BHIYE-HP supplemented with 0.3 mM of DAP and incubated at 37°C, 5% CO₂ overnight. Colonies were scraped from the plate and resuspended in 2 mL of BHIYE diluted in 50% glycerol. Cells were then plated on the counter-selection plate containing BHIYE-HP supplemented with 0.1 mM isopropyl- β -D-thiogalactopyranoside (IPTG), 40 mg/mL kanamycin, and 30 mg/mL nalidixic acid (to select against *E. coli* donor strains). Counterselection plates were incubated at 37°C, 5% CO₂ for 48 h. Colonies were then scraped and resuspended in 2 mL of BHIYE plus 50% glycerol. The total number of cells in each conjugation was calculated via CFU counting. Approximately 40,000 cells were grown in eight plates for each of the five conjugations of *Hp* 392, and 25,000 in five plates for each of the four conjugations of *Hp* EL1, totaling approximately 200,000 colonies for *Hp* 392, and more than 100,000 for EL1. Plates were incubated for 48 h, and colonies were scraped, homogenized, and aliquoted

into cryogenic storage medium with 25% glycerol. Mutant pool stocks were stored at -80°C .

Identification of essential genes

Approximately 2.5×10^6 cells of *Hp* 392 and *Hp* EL1 from the mutant pools' stocks were inoculated into 3 mL of BHIYE-HP broth and incubated for 16 h at 37°C , 5% CO_2 . Cultures were centrifuged for 1 min at $14,000 \times g$, and cells were collected for DNA extraction using the Epicentre kit following the manufacturer's instructions. Samples were lysed using a Mini-Beadbeater (Biospec) in 2 mL vials preloaded with Lysing Matrix B (MP Biomedicals). Library growth was performed in biological triplicates using three different aliquot tubes to start respective cultures.

The core genome of *H. parainfluenzae*

Anvi'o (40, 44) was used to identify the core genome and the core essential genome of *Hp*. To do so, the genomes of *Hp* 392 and *Hp* EL1, and the 341 essential genes shared by both strains were aligned to 16 closed *Hp* genomes available at the NCBI database (Table S5). We used scripts available on GitHub (<https://github.com/mramsey01/TnSeq-analysis-1>). Briefly, modified FASTA files were loaded into the Anvi'o default pangenome pipeline for alignment and comparison.

TnSeq Illumina library preparation

To locate *mariner* insertion sites in the genome, Illumina libraries were prepared largely as described previously (26, 35). DNA was extracted and sheared to 400 bp average length (peak power of 175.0 and 200 cycles/Burst) using a Covaris S220 Focused-ultrasonicator. Small and large fragments (<100 or >700 bp) of DNA were removed using KAPA beads according to the manufacturer's instructions. Next, two rounds of PCR were performed using primers (Table S2) targeting the transposon and adding the Illumina adapters for sequencing similar to previous methods (26, 35). Libraries were submitted to the Rhode Island Genomics and Sequencing Center (RIGSC) for quantification (qPCR) and quality control (Bioanalyzer). Libraries were then sent to SeqCenter and sequenced on an Illumina NextSeq 550 1-by-75 single-end run at the Microbial Genome Sequencing Center (MIGS) (Pittsburgh, PA).

H. parainfluenzae conditionally essential genes for *in vitro* aerobic and anaerobic biofilm growth

Hp 392 and *Hp* EL1 mutant pools were grown in 3 mL of BHIYE-HP broth and incubated for 16 hours as described above. Next, the OD_{600} was adjusted to 0.1 and 10 μL of culture was pipetted onto sterile polycarbonate membranes (25 mm diameter, 0.2 μm pore size, MilliporeSigma) on BHIYE-HP plates using a colony biofilm model (41, 45) to compare to growth conditions we performed previously (42). Three polycarbonate membranes were placed on each plate, and five 10 μL spots of culture were pipetted onto each membrane for a total of 15 spots per plate. Two plates were prepared. One plate was incubated at 37°C , 5% CO_2 (for genes essential for colony growth in aerobic conditions), and the other was incubated in a Coy anaerobic chamber with an atmosphere of 90% N_2 , 5% CO_2 , and 5% H_2 (for genes essential for colony growth in anaerobic conditions). After 24 h of incubation, membranes were transferred to an Eppendorf tube and washed with sterile BHIYE medium. Cells were centrifuged (1 min at $14,000 \times g$) and the pellets were collected for DNA extraction.

RNAseq analysis

RNAseq reads obtained from a previous study (42) from *H. parainfluenzae* ATCC 33392 grown using the colony biofilm model were aligned, mapped, and differentially expressed genes were determined as previously described (42), with the exception that

we re-sequenced the genome of *Hp* 392 and used this now-closed genome and its RAST annotation (46), generated under default settings, as the reference, to allow direct comparison between the RNAseq and TnSeq data sets.

TnSeq data analysis

We utilized a previously established pipeline (25) with some modifications. All scripts utilized are available on GitHub (<https://github.com/mramsey01/TnSeq-analysis-1>). Prior to genome alignment, the initial sequence data were trimmed and filtered with Trimmomatic (47) using a sliding window of four nucleotides and removing sequences when -phred33 quality score <20, and when sequences were shorter than 30 nucleotides. These quality-filtered sequences were then aligned to their respective genomes using Bowtie2 (48) under default settings. Next, *mariner* insertion sites and their associated read counts were calculated (Table S7). Local smoothing (LOESS) was used as previously described by reference (35) to correct for potential multifork chromosomal replication bias of insertion frequency. The Monte Carlo method was then applied as described previously (25, 36) to generate a pseudo data set with the expected number of insertions per TA site. This method generates 100 pseudo data sets through sampling with replacement from the observed locations of the insertions and their respective read counts. Pseudo data sets were then compared to the observed distribution of read counts to determine where the number of observed read counts differs from the expected number of reads in the pseudo data sets. DESeq (49, 50) was used to compare the observed versus expected (pseudo) data sets and calculate differential mutant abundance and significance. Then, following the protocols established by reference (35), a clustering algorithm was used to fit a bimodal curve on the distribution of calculated differential mutant abundances to cluster genes into two groups. Genes with substantially fewer reads were clustered under the “reduced” group, while genes for which the number of reads did not differ substantively from the expected number of reads were clustered under the “unchanged” group. The clustering algorithm also estimates an uncertainty measure for the assignment to each group. A gene was considered essential when three conditions were met: (i) the observed number of insertions was significantly (P value ≤ 0.01) different from the expected number of insertions in the pseudo data sets; (ii) the observed number of insertions was substantially smaller than the expected number of reads so that the gene was clustered under the “reduced” group; and (iii) the uncertainty of the placement of that gene in the reduced group was smaller than 0.01 (on a 0 to 1 scale).

DISCUSSION

TnSeq is a tool widely used to identify genes essential for survival (51, 52), many of which fall in the realm of “housekeeping” genes. This data alone has utility as a tool to identify potential drug targets (53). Further, TnSeq can be utilized in mutant strain backgrounds to find synthetically lethal genes involved in different synthesis or degradation pathways (54). TnSeq is also often used to compare different growth environments, for example, *in vitro* versus *in vivo*, finding conditionally essential genes, which are useful to generate hypothetical mechanisms for further testing and potential therapeutic targets (26).

Here, we use TnSeq to describe the essential genome and the conditionally essential genes necessary for *in vitro* biofilm growth under aerobic and anaerobic conditions in two strains of *Hp* (*Hp* 392 and *Hp* EL1). We chose to assess essential genomes using growth on a nutrient-rich medium so that accessory synthesis pathways would not be required and, thus, not deemed essential, as they would have been using a minimal, defined medium. This strategy allows us to identify only the most essential elements necessary for survival and allows for greater resolution of the completed library to interrogate other environments to find conditionally essential genes, making this a useful tool for future investigations. As expected, we observed that the majority of essential *Hp* genes are associated with central metabolic pathways and cell machinery, similar to results published in other studies (25, 36). We also identified a small subset

of essential genes that were strain-specific (Table S7). Studies have demonstrated great genetic variation among *Hp* isolates. Thus, strain variation was an expected outcome and a motivating reason for performing these experiments (32, 33) with more than one strain. *Hp* 392 and *Hp* EL1 were isolated at very different times, geographic sites, and from distinct body sites. *Hp* 392 was isolated from a septic finger infection in London, England, around 1949, whereas *Hp* EL1 was isolated by our lab from the supragingival plaque of a healthy adult in Rhode Island in 2018. As the type-strain, *Hp* 392 was procured from the ATCC collection, and it is unknown how often it has been passaged prior to deposition and if the strain has been “domesticated” to an *in vitro* environment. The diverging histories of our two *Hp* strains ensure a greater likelihood that our data is more representative of *Hp* as a species. To further investigate the possibility of strain variation, we compared the genomes of our two strains here to 16 other *Hp* genomes available from NCBI (Fig. 3). We found that all 341 essential genes shared between our strains were also present in each of the other genomes. This finding is similar to findings in *A. actinomycetemcomitans* (25) and indicates that the essential genes we identified are indeed part of the core of all the examined *Hp* genomes.

We would expect biofilm growth to require gene functions in addition to the absolute essential genes initially identified by TnSeq. We, therefore, used our libraries to identify conditionally essential genes for biofilm growth in aerobic and anaerobic conditions using a colony biofilm model (41) grown on the same nutrient-rich medium that was used for the selection of the initial library. Most of the genes identified as essential for biofilm growth were essential regardless of oxygen availability. It is well-documented that in dense biofilms, oxygen gradients are formed through bacterial respiration (7, 55). While finding mutants that were conditionally essential for both oxygen conditions was expected, they were even more similar than we initially anticipated.

Some of the genes conditionally essential anaerobically for both *Hp* strains are the Na⁺-translocating NADH: quinone oxidoreductase (NQR) subunits A-F(*nqr*ABCDEF). NQR is a redox-driven sodium pump, well characterized in *Vibrio cholerae*, and operates in the respiratory chain, oxidizing cytoplasmic NADH and reducing it to ubiquinone (56, 57). Also anaerobically essential, were the four genes organized in operon that code for the fumarate reductase enzyme. It is unknown if this is acting as part of the reductive TCA cycle or if fumarate is serving as a terminal electron acceptor in this context (58) and requires further investigation. Given the broad similarity between conditions, and the diverse nature of mutants conditionally essential for growth in this context, we did not further study individual mutant phenotypes here and opted instead to report on and make available the mutant library as a whole.

We were curious if the conditionally essential genes for biofilm growth would also be upregulated in biofilm growth. To assess this, we used our earlier transcriptome data (42) of colony biofilms under aerobic and anaerobic conditions for *Hp* 392 and compared genes differentially expressed in anaerobic biofilm growth to conditionally essential genes for anaerobic biofilm growth (Fig. 5). Interestingly, only three genes essential for anaerobic biofilm growth were significantly upregulated during anaerobiosis. Of those, two genes (Murein DD-endopeptidase and LPS-core synthesis glycosyltransferase) products are involved in cell metabolism, and one (Bacterial ribosome SSU maturation protein RimP) is involved in genetic information processing, but has also shown to increase sensitivity to acid stress in another species (43). The low overlap between gene essentiality and differential expression indicates that these conditionally essential genes are transcribed under aerobic conditions, where they are not essential, as well as anaerobic conditions, where they are essential. These data are similar to previous findings in *P. aeruginosa* (26), indicating that mutant fitness and elevated gene expression are not well correlated. Likewise, other researchers have found that for *Listeria monocytogenes* (59), genes significantly upregulated in anaerobiosis were not essential for anaerobic growth. It is also possible that basal level expression of many of these genes is sufficient for fitness in the short timeframes of our experimental model, but that regulation may be a factor under longer durations, where minor differences in efficiency

are more relevant. Based on these results, we can argue that using transcriptomes to predict gene essentiality is not very effective. Increasing transcription of some genes may, however, contribute to how *Hp* adapts to changes in oxygen levels, allowing *Hp* to thrive in different portions of the biofilm.

In summary, we present the essential genomes for the type-strain and oral isolate of *H. parainfluenzae*, revealing that this essential genome is highly conserved between *Hp* strains. We also identified genes necessary for aerobic and anaerobic biofilm growth *in vitro*, showing that many genes involved in anaerobiosis are also essential for aerobic biofilm growth. Between these data and transcriptomes of aerobic versus anaerobic growth, we determined that *Hp* grows in a highly heterogeneous environment for oxygen availability. These data may serve as a starting point for understanding *Hp* growth in oral biofilms and how it interacts with other members of polymicrobial oral biofilms in diverse niches within the oral cavity it inhabits.

ACKNOWLEDGMENTS

The authors would like to thank members of the M. Ramsey lab for valuable help with all phases of this manuscript. We thank Dr. Janet Atoyan in the URI Genomics Core Facility for assistance with sequencing library synthesis and quality control work.

This work was funded by the NIDCR/NIH (R01DE027958 – M.R.) and the USDA National Institute of Food and Agriculture, Hatch Formula project accession number 1017848 (M.R.).

T.P. and M.R. designed all experiments. T.P. performed all biological experiments. T.P. and C.P. analyzed and interpreted the data. T.P., M.M., and J.M.W. wrote the manuscript. M.R., M.M., and J.M.W. edited the manuscript.

AUTHOR AFFILIATIONS

¹Department of Cell and Molecular Biology, University of Rhode Island, Kingston, Rhode Island, USA

²Department of Microbiology, ADA Forsyth Institute, Cambridge, Massachusetts, USA

AUTHOR ORCIDs

Thais H. de Palma  <http://orcid.org/0000-0002-7948-3660>

Chris Powers  <http://orcid.org/0000-0001-5770-156X>

Morgan J. McPartland  <http://orcid.org/0009-0004-9751-0943>

Jessica Mark Welch  <http://orcid.org/0000-0003-0696-2334>

Matthew Ramsey  <http://orcid.org/0000-0001-7158-5120>

FUNDING

Funder	Grant(s)	Author(s)
HHS NIH National Institute of Dental and Craniofacial Research (NIDCR)	R01DE027958	Jessica Mark Welch Matthew Ramsey
U.S. Department of Agriculture (USDA)	1017848	Matthew Ramsey

AUTHOR CONTRIBUTIONS

Thais H. de Palma, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing | Chris Powers, Investigation, Software, Visualization, Writing – original draft, Writing – review and editing | Morgan J. McPartland, Validation, Writing – review and editing | Jessica Mark Welch, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Supervision, Visualization, Writing – original draft, Writing – review and editing | Matthew Ramsey, Conceptualization, Funding acquisition, Supervision, Writing – review and editing

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental material (mSystems00674-24-s0001.docx). Supplemental figures, tables, and legends.

Table S7 (mSystems00674-24-s0002.xlsx). Essential genes for *Hp* 392 and EL1.

REFERENCES

- Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. 2005. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* 43:5721–5732. <https://doi.org/10.1128/JCM.43.11.5721-5732.2005>
- Chen T, Yu W-H, Izard J, Baranova OV, Lakshmanan A, Dewhirst FE. 2010. The human oral microbiome database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database (Oxford)* 2010:baq013. <https://doi.org/10.1093/database/baq013>
- Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner ACR, Yu W-H, Lakshmanan A, Wade WG. 2010. The human oral microbiome. *J Bacteriol* 192:5002–5017. <https://doi.org/10.1128/JB.00542-10>
- Mark Welch JL, Dewhirst FE, Borisy GG. 2019. Biogeography of the oral microbiome: the site-specialist hypothesis. *Annu Rev Microbiol* 73:335–358. <https://doi.org/10.1146/annurev-micro-090817-062503>
- Kolenbrander PE, Palmer RJ Jr, Rickard AH, Jakubovics NS, Chalmers NI, Diaz PI. 2006. Bacterial interactions and successions during plaque development. *Periodontol* 2000 42:47–79. <https://doi.org/10.1111/j.1600-0757.2006.00187.x>
- Marsh PD. 2005. Dental plaque: biological significance of a biofilm and community life-style. *J Clin Periodontol* 32 Suppl 6:7–15. <https://doi.org/10.1111/j.1600-051X.2005.00790.x>
- Ahn S-J, Burne RA. 2007. Effects of oxygen on biofilm formation and the AtIA autolysin of *Streptococcus mutans*. *J Bacteriol* 189:6293–6302. <https://doi.org/10.1128/JB.00546-07>
- Wessel AK, Arshad TA, Fitzpatrick M, Connell JL, Bonnez RT, Shear JB, Whiteley M. 2014. Oxygen limitation within a bacterial aggregate. *mBio* 5:e00992. <https://doi.org/10.1128/mBio.00992-14>
- Diaz PI, Chalmers NI, Rickard AH, Kong C, Milburn CL, Palmer RJ, Kolenbrander PE. 2006. Molecular characterization of subject-specific oral microflora during initial colonization of enamel. *Appl Environ Microbiol* 72:2837–2848. <https://doi.org/10.1128/AEM.72.4.2837-2848.2006>
- Kilian M, Schiott CR. 1975. Haemophili and related bacteria in the human oral cavity. *Arch Oral Biol* 20:791–796. [https://doi.org/10.1016/0003-9969\(75\)90055-2](https://doi.org/10.1016/0003-9969(75)90055-2)
- Mark Welch JL, Rossetti BJ, Rieken CW, Dewhirst FE, Borisy GG. 2016. Biogeography of a human oral microbiome at the micron scale. *Proc Natl Acad Sci U S A* 113:E791–E800. <https://doi.org/10.1073/pnas.1522149113>
- Utter DR, Borisy GG, Eren AM, Cavanaugh CM, Mark Welch JL. 2020. Metapangenomics of the oral microbiome provides insights into habitat adaptation and cultivar diversity. *Genome Biol* 21:293. <https://doi.org/10.1186/s13059-020-02200-2>
- Caselli E, Fabbri C, D'Accolti M, Soffritti I, Bassi C, Mazzacane S, Franchi M. 2020. Defining the oral microbiome by whole-genome sequencing and resistome analysis: the complexity of the healthy picture. *BMC Microbiol* 20:120. <https://doi.org/10.1186/s12866-020-01801-y>
- Sen Yew H, Chambers ST, Roberts SA, Holland DJ, Julian KA, Raymond NJ, Beardsley J, Read KM, Murdoch DR. 2014. Association between HACEK bacteremia and endocarditis. *J Med Microbiol* 63:892–895. <https://doi.org/10.1099/jmm.0.070060-0>
- Bachman DS. 1975. *Haemophilus meningitis*: comparison of *H. influenzae* and *H. parainfluenzae*. *Pediatrics* 55:526–530. <https://doi.org/10.1542/peds.55.4.526>
- Black CT, Kupferschmid JP, West KW, Grosfeld JL. 1988. *Haemophilus parainfluenzae* infections in children, with the report of a unique case. *Clin Infect Dis* 10:342–346. <https://doi.org/10.1093/clinids/10.2.342>
- Oill PA, Chow AW, Guze LB. 1979. Adult bacteremic *Haemophilus parainfluenzae* infections. Seven reports of cases and a review of the literature. *Arch Intern Med* 139:985–988. <https://doi.org/10.1001/archinte.139.9.985>
- Cremades R, Galiana A, Rodriguez JC, Santos A, Lopez P, Ruiz M, Garcia-Pachon E, Royo G. 2011. Identification of bacterial DNA in noninfectious pleural fluid with a highly sensitive PCR method. *Respiration* 82:130–135. <https://doi.org/10.1159/000322003>
- Sturm AW. 1986. *Haemophilus influenzae* and *Haemophilus parainfluenzae* in nongonococcal urethritis. *J Infect Dis* 153:165–167. <https://doi.org/10.1093/infdis/153.1.165>
- Bailey C, Duckett S, Davies S, Townsend R, Stockley I. 2011. *Haemophilus parainfluenzae* prosthetic joint infection. The importance of accurate microbiological diagnosis and options for management. *J Infect* 63:474–476. <https://doi.org/10.1016/j.jinf.2011.08.009>
- Pillai A, Mitchell JL, Hill SL, Stockley RA. 2000. A case of *Haemophilus parainfluenzae* pneumonia. *Thorax* 55:623–624. <https://doi.org/10.1136/thorax.55.7.623>
- Smith CB, Golden CA, Kanner RE, Renzetti AD. 1976. *Haemophilus influenzae* and *Haemophilus parainfluenzae* in chronic obstructive pulmonary disease. *Lancet* 1:1253–1255. [https://doi.org/10.1016/s0140-6736\(76\)91733-5](https://doi.org/10.1016/s0140-6736(76)91733-5)
- van Opijnen T, Bodi KL, Camilli A. 2009. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* 6:767–772. <https://doi.org/10.1038/nmeth.1377>
- Klein BA, Duncan MJ, Hu LT. 2015. Defining essential genes and identifying virulence factors of *Porphyromonas gingivalis* by massively parallel sequencing of transposon libraries (Tn-seq), p 25–43. In Lu LJ (ed), *Gene essentiality, methods in molecular biology*. Springer New York, New York, NY. https://doi.org/10.1007/978-1-4939-2398-4_3
- Narayanan AM, Ramsey MM, Stacy A, Whiteley M. 2017. Defining genetic fitness determinants and creating genomic resources for an oral pathogen. *Appl Environ Microbiol* 83:e00797-17. <https://doi.org/10.1128/AEM.00797-17>
- Turner KH, Everett J, Trivedi U, Rumbaugh KP, Whiteley M. 2014. Requirements for *Pseudomonas aeruginosa* acute burn and chronic surgical wound infection. *PLoS Genet* 10:e1004518. <https://doi.org/10.1371/journal.pgen.1004518>
- Walker AR, Shields RC. 2022. Identification and analysis of essential genes in *Streptococcus mutans* with transposon sequencing, p 237–258. In Zhang R (ed), *Essential genes and genomes, methods in molecular biology*. Springer US, New York, NY. https://doi.org/10.1007/978-1-0716-1720-5_13
- Yang Y, Stewart PE, Shi X, Li C. 2008. Development of a transposon mutagenesis system in the oral spirochete *Treponema denticola*. *Appl Environ Microbiol* 74:6461–6464. <https://doi.org/10.1128/AEM.01424-08>
- Gawronski JD, Wong SMS, Giannoukos G, Ward DV, Akerley BJ. 2009. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proc Natl Acad Sci U S A* 106:16422–16427. <https://doi.org/10.1073/pnas.0906627106>
- Vergara-Irigaray M, Fookes MC, Thomson NR, Tang CM. 2014. RNA-seq analysis of the influence of anaerobiosis and FNR on *Shigella flexneri*. *BMC Genomics* 15:438. <https://doi.org/10.1186/1471-2164-15-438>
- Croucher NJ, Thomson NR. 2010. Studying bacterial transcriptomes using RNA-seq. *Curr Opin Microbiol* 13:619–624. <https://doi.org/10.1016/j.mib.2010.09.009>

32. Kerr GRD, Forbes KJ, Williams A, Pennington TH. 1993. An analysis of the diversity of *Haemophilus parainfluenzae* in the adult human respiratory tract by genomic DNA fingerprinting. *Epidemiol Infect* 111:89–98. <https://doi.org/10.1017/S0950268800056715>
33. Watts SC, Judd LM, Carzino R, Ranganathan S, Holt KE. 2021. Genomic diversity and antimicrobial resistance of *Haemophilus* colonizing the airways of young children with cystic fibrosis. *mSystems* 6:e0017821. <https://doi.org/10.1128/mSystems.00178-21>
34. de Palma TH, Loomis E, Perera D, Ramsey M. 2023. Whole-genome sequence of *Haemophilus parainfluenzae* EL1 isolated from healthy supragingival plaque. *Microbiol Resour Announc* 12:e0047623. <https://doi.org/10.1128/MRA.00476-23>
35. 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>
36. Lewin GR, Stacy A, Michie KL, Lamont RJ, Whiteley M. 2019. Large-scale identification of pathogen essential genes during coinfection with sympatric and allopatric microbes. *Proc Natl Acad Sci U S A* 116:19685–19694. <https://doi.org/10.1073/pnas.1907619116>
37. Kanehisa M. 2019. Toward understanding the origin and evolution of cellular organisms. *Protein Sci*. 28:1947–1951. <https://doi.org/10.1002/pro.3715>
38. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. 2023. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res*. 51:D587–D592. <https://doi.org/10.1093/nar/gkac963>
39. Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 28:27–30. <https://doi.org/10.1093/nar/28.1.27>
40. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, et al. 2021. Community-led, integrated, reproducible multi-omics with anvio. *Nat Microbiol* 6:3–6. <https://doi.org/10.1038/s41564-020-00834-3>
41. Anderl JN, Franklin MJ, Stewart PS. 2000. Role of antibiotic penetration limitation in *Klebsiella pneumoniae* biofilm resistance to ampicillin and ciprofloxacin. *Antimicrob Agents Chemother* 44:1818–1824. <https://doi.org/10.1128/AAC.44.7.1818-1824.2000>
42. Perera D, McLean A, Morillo-López V, Cloutier-Leblanc K, Almeida E, Cabana K, Mark Welch J, Ramsey M. 2022. Mechanisms underlying interactions between two abundant oral commensal bacteria. *ISME J* 16:948–957. <https://doi.org/10.1038/s41396-021-01141-3>
43. Poonam null, Yennamalli RM, Bisht GS, Shrivastava R. 2019. Ribosomal maturation factor (RimP) is essential for survival of nontuberculous mycobacteria *Mycobacterium fortuitum* under *in vitro* acidic stress conditions. *3 Biotech* 9:127. <https://doi.org/10.1007/s13205-019-1659-y>
44. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>
45. Eigentler L, Davidson FA, Stanley-Wall NR. 2022. Mechanisms driving spatial distribution of residents in colony biofilms: an interdisciplinary perspective. *Open Biol* 12:220194. <https://doi.org/10.1098/rsob.220194>
46. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, et al. 2008. The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>
47. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
48. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
49. Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* 11:R106. <https://doi.org/10.1186/gb-2010-11-10-r106>
50. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15:550. <https://doi.org/10.1186/s13059-014-0550-8>
51. Barquist L, Boinett CJ, Cain AK. 2013. Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol* 10:1161–1169. <https://doi.org/10.4161/rna.24765>
52. Cain AK, Barquist L, Goodman AL, Paulsen IT, Parkhill J, van Opijnen T. 2020. A decade of advances in transposon-insertion sequencing. *Nat Rev Genet* 21:526–540. <https://doi.org/10.1038/s41576-020-0244-x>
53. Mobegi FM, van Hijum SAFT, Burghout P, Bootsma HJ, de Vries SPW, van der Gaast-de Jongh CE, Simonetti E, Langereis JD, Hermans PWM, de Jonge MI, Zomer A. 2014. From microbial gene essentiality to novel antimicrobial drug targets. *BMC Genomics* 15:958. <https://doi.org/10.1186/1471-2164-15-958>
54. Greene NG, Fumeaux C, Bernhardt TG. 2018. Conserved mechanism of cell-wall synthase regulation revealed by the identification of a new PBP activator in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 115:3150–3155. <https://doi.org/10.1073/pnas.1717925115>
55. Klementiev AD, Jin Z, Whiteley M. 2020. Micron scale spatial measurement of the O₂ gradient surrounding a bacterial biofilm in real time. *mBio* 11:10. <https://doi.org/10.1128/mBio.02536-20>
56. Agarwal S, Bernt M, Toulouse C, Kurz H, Pfannstiel J, D'Alvise P, Hasselmann M, Block AM, Häse CC, Fritz G, Steuber J. 2020. Impact of Na⁺-translocating NADH:quinone oxidoreductase on iron uptake and *nqrM* expression in *Vibrio cholerae*. *J Bacteriol* 202:e00681-19. <https://doi.org/10.1128/JB.00681-19>
57. Steuber J, Vohl G, Casutt MS, Vorburger T, Diederichs K, Fritz G. 2014. Structure of the *V. cholerae* Na⁺-pumping NADH:quinone oxidoreductase. *Nature* 516:62–67. <https://doi.org/10.1038/nature14003>
58. Kröger A. 1978. Fumarate as terminal acceptor of phosphorylative electron transport. *Biochim Biophys Acta* 505:129–145. [https://doi.org/10.1016/0304-4173\(78\)90010-1](https://doi.org/10.1016/0304-4173(78)90010-1)
59. Müller-Herbst S, Wüstner S, Mühlig A, Eder D, M. Fuchs T, Held C, Ehrenreich A, Scherer S. 2014. Identification of genes essential for anaerobic growth of *Listeria monocytogenes*. *Microbiology* 160:752–765. <https://doi.org/10.1099/mic.0.075242-0>