Shaping human gut community assembly and butyrate production by controlling the arginine dihydrolase pathway

Yiyi Liu[1,2], Yu-Yu Cheng[2], Zhichao Zhou[3], Eugenio I. Vivas[3,4], Matthew F. Warren[3]
Federico E. Rey[3], Karthik Anantharaman[3] and Ophelia S. Venturelli[1,2,3*]

[1]Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison WI 53706

[2]Department of Biochemistry, University of Wisconsin-Madison, Madison, WI 53706

[3]Department of Bacteriology, University of Wisconsin-Madison, WI 53706

[4]Gnotobiotic Animal Core Facility, University of Wisconsin-Madison, Madison, WI 53706, USA

*To whom correspondence should be addressed: venturelli@wisc.edu

ABSTRACT

The arginine dihydrolase pathway (*arc* operon) is present in a minority of diverse human gut species and enables arginine catabolism. We lack a quantitative understanding of the role of this specialized metabolic pathway in the human gut microbiome. We investigate the role of the *arc* operon in probiotic *E. coli* Nissle 1917 on community assembly and health-relevant metabolite production *in vitro* and in murine gut. The *arc* operon shapes community assembly and can enhance butyrate production at physiologically relevant environmental pH levels *in vitro*. In the presence of the *arc* operon, human gut communities display reduced variability in composition in response to variations in initial pH. Dynamic computational modeling of community assembly reveals the extent of pH-mediated inter-species interactions. Overall, we demonstrate that a specialized metabolic pathway can serve as the control knob of community assembly and beneficial metabolite production.

INTRODUCTION

The human gut microbiome substantially expands our genome's capabilities by encoding numerous metabolic pathways[1–3]. Certain metabolic pathways are frequently found in constituent members of the gut microbiota, such as the enzymes to produce acetate[4]. However, other specialized metabolic pathways such as bile acid metabolism are present in a subset of gut species[5–7]. Specialized metabolic pathways can be harbored by phylogenetically distant species and contribute uniquely to the functions performed by gut microbiota. Understanding the contributions of these unique metabolic pathways to the gut microbiome could aid in the discovery of precision strategies to manipulate their activities to our benefit.

Specialized metabolic pathways can have a major effect on the interactions between constituent members of gut microbiota. For example, *Bacteroides spp.* harbor polysaccharide utilization loci (PUL) that enable the breakdown of complex carbohydrates. These PULs can lead to competition for limiting carbohydrates as well as enhancement of the growth of constituent community members that lack the capability to directly utilize the carbohydrates[3]. However, we lack an understanding of the role of other specialized metabolic pathways beyond polysaccharide utilization loci on the interactions in the gut microbiome. For example, reduced butyrate production capability by the gut microbiome has links to several human diseases[8–10]. However, predictable modulation of butyrate production via supplementation of butyrate-producing species is challenging and cannot be fully explained by PUL-related interactions[11,12]. Therefore, further investigation is needed into other specialized metabolic pathways that could influence microbiome dynamics.

The arginine dihydrolase pathway (encoded by the *arc* operon) is a specialized metabolic pathway that converts arginine into several products including ammonia, which could impact the gut community[13]. The *arc* operon has been shown to participate in pH homeostasis in oral biofilms[14] and is involved in nitrogen cycling in soil communities[15]. Specifically in gut, this pathway is harbored by many diverse human gut species enabling utilization of arginine as an energy source under aerobic and anaerobic conditions[16–18]. The *arc* operon harbored by *Enterococcus faecium* was shown to enhance *Clostridioides difficile* toxin production through arginine depletion and cross-feeding[19]. On the other hand, *arc*-harboring *Eggerthella lenta (E.* lenta) was an influential member of the assembly and butyrate production of human gut communities extensively characterized *in vitro*[20]. In terms of mechanism, by releasing ammonia, the *arc* operon can increase environmental pH and provide a nitrogen source that could benefit the community in nitrogen-limited environments such as the mammalian gut[21]. Environmental pH modification is a major mechanism shaping the dynamics and metabolism of *in vitro* gut communities[22,23], especially the production of various short chain fatty acids (SCFA)[24,25]. Further, changes in the pH of the gastrointestinal (GI) tract are linked to specific host health states and gut ecosystem functions[26,27]. This suggests that environmental pH modification mediated by the *arc* operon could influence health-relevant gut microbiome functions.

Computational modeling of microbiomes can reveal the quantitative relationships between different species and their impact on community-level functions[28,29]. While coarse grain models like the generalized Lotka-Volterra (gLV) have been implemented for predictions of community assembly[20,30–32], each pairwise interaction is represented as a lumped constant parameter, missing mechanistic information of species interactions. Therefore, computational models that incorporate specific mechanism modalities hold promise for modeling community dynamics due to their ability to provide deeper insights into specific classes of mechanisms shaping microbial communities. Specifically for the environmental pH, a computational model that captures pH-mediated interactions provides insight into pairwise interactions in a soil microbial community[33].

By combining bottom-up design of synthetic communities, computational modeling, and synthetic biology, we investigated the role of the *arc* operon in the human-associated probiotic strain *Escherichia coli* Nissle 1917 (EC) in shaping human gut community assembly, butyrate

production, and mammalian gut colonization. Exploiting genetic control, we engineered inducible control of the *arc* operon in EC and investigated the effect of this pathway on the assembly of a defined human gut community both *in vitro* and in the murine gut. By reducing variation in external pH, the *arc* operon reduced fluctuations in community composition. Further, the presence of the *arc* operon yielded changes in butyrate production *in vitro* and transient shifts in butyrate producer abundance in the mammalian gut. To quantify the role of pH modification of the *arc* operon on community assembly, a dynamic computational model was used to forecast community assembly as a function of mono-species growth and external pH modification. The computational model revealed the quantitative contribution of pH-mediated interactions in a species-specific manner. In sum, we demonstrate that a single specialized metabolic pathway can shape community assembly, species colonization of the mammalian gut and butyrate production. A more quantitative and detailed understanding of how specialized metabolic pathways shape gut microbiota assembly and functions holds promise for engineering the gut microbiome to our benefit.

RESULTS

Arginine dihydrolase pathway increases environmental pH via arginine metabolism

We hypothesized that the *arc* operon could shape human gut community assembly based on its potential for pH modification and nitrogen release. The *arc* operon consists of four genes including arginine deiminase *arcA,* ornithine carbamoyltransferase *arcB,* carbamate kinase *arcC* and ornithine antiporter *arcD* **(Figure 1A)**. The arginine deiminase *arcA* is responsible for the conversion of arginine into citrulline, producing ammonia as a byproduct. The ornithine carbamoyltransferase *arcB* converts citrulline to ornithine and carbamyl phosphate. The carbamate kinase arcC further metabolizes carbamyl phosphate to produce more ammonia. Finally, the arginine-ornithine antiporter *arcD*, catalyzes the ATP-independent exchange between arginine and ornithine across the bacterial membrane.

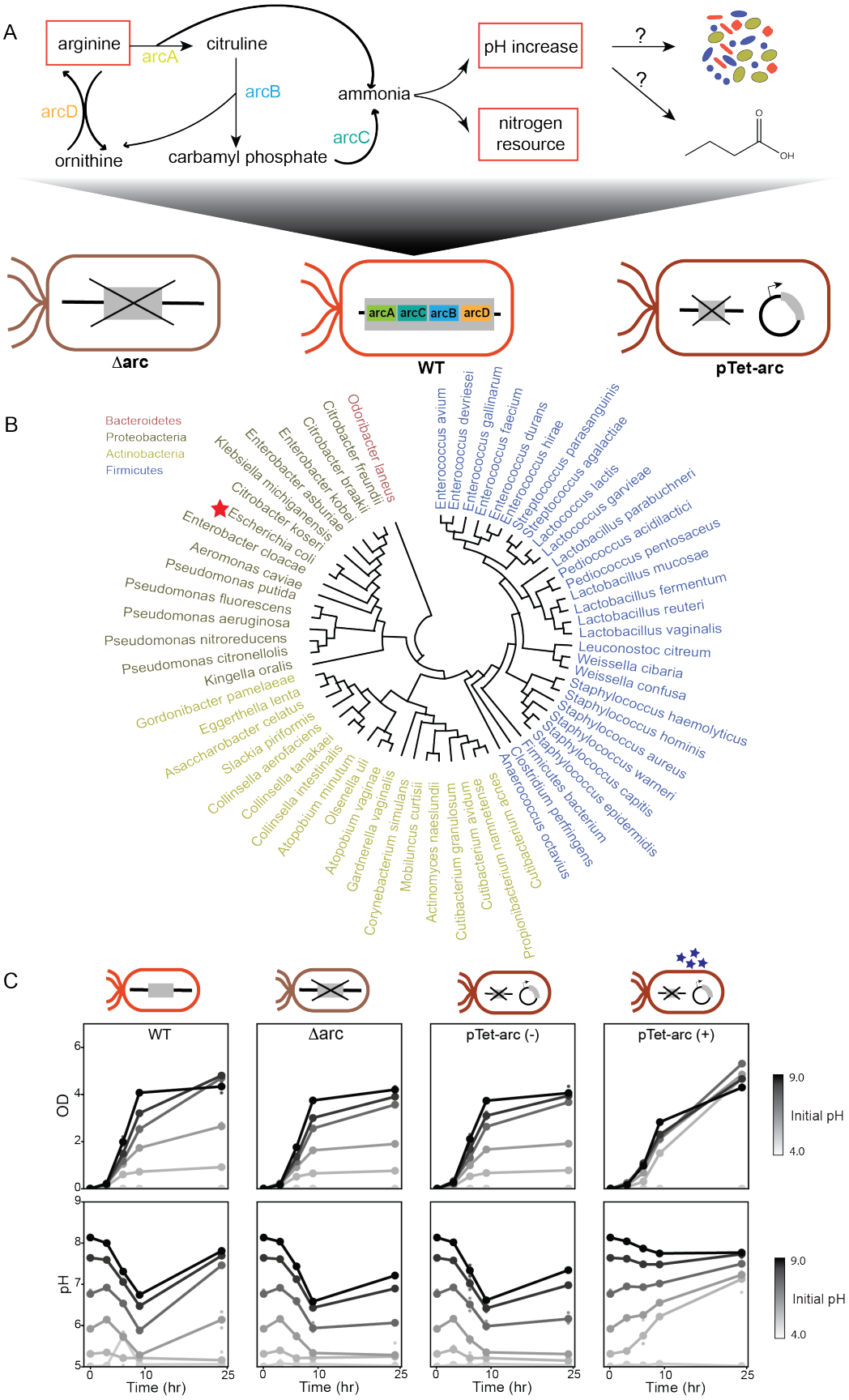
To determine the prevalence of the *arc* operon in the human gut microbiome, we performed a bioinformatics search on a metagenome-assembled genome dataset containing 154,723 microbial genomes assembled from 9,428 samples of the human microbiome[34]. The microbiome samples were derived from different body sites in individuals from diverse geographic locations, disease states, and lifestyles. Of the 533 total species identified from the annotated database, 63 species (12%) were found to harbor *arcA*, *arcB*, *arC* and *arcD* (**Figure 1B**). The *arc* operon was rarely found in Bacteroidetes, one of the most abundant phyla in human gut microbiome. Additionally, the *arc* operon was harbored by 22 pathogens associated with human gut, oral, or vaginal diseases. These results demonstrate that the *arc* operon is a specialized metabolic pathway in the human gut microbiome due to its low prevalence across species in the dataset.

The probiotic strain *Escherichia coli* Nissle 1917 (EC) harbors the *arc* operon, has extensive genetic tools, and has relevance to human health[35]. To evaluate the impact of the *arc* operon on the fitness of EC, we deleted this operon from the genome of the wildtype EC (WT) creating Δarc **(Figure 1A)**. In addition, we constructed a third strain by introducing the *arc* operon controlled by an anhydrotetracycline (aTc) inducible promoter into the *arc* deletion background **(Figure 1A)**. Using RT-qPCR, we demonstrated the expression of the *arcD* gene in the *arc* operon was substantially increased in the presence than absence of aTc in *E. coli* monocultures **(Figure S1A)**.

To quantify the effect of the *arc* operon on the growth of EC, we characterized the growth of each EC strain in media that varied in initial pH (5 to 8) and was supplemented with arginine and presence/absence of aTc (**Figure 1C**). The growth of EC was enhanced in media with higher initial pH. The Δarc and pTet-arc strain displayed similar growth and external pH trend in the absence of aTc (pTet-arc (-)), demonstrating that the pTet promoter had low leakiness. The wildtype (WT) and pTet-arc strains (pTet-arc (+)) displayed higher endpoint environmental pH than the Δarc and pTet-arc (-) conditions. Notably, the growth of pTet-arc (+) was similar across a wide range of initial pH values, whereas the other EC strains exhibited substantially larger growth variability. Specifically, pTet-arc (+) displayed growth in media with an initial pH of 5, whereas all other strains failed to grow. In addition, the environmental pH changes due to *arc* operon expression between Δarc and pTet-arc(+) at 24 hour displayed largest difference at lower initial pH (initial pH = 5) versus smallest difference at higher initial pH (initial pH=9), showing the variability of *arc* operon's ability to change environmental pH across different initial pH.

A previous study demonstrated that the *arc* operon provides an advantage for a pathogenic *E. coli* wildtype strain *in vitro* in the presence of arginine while conferring a disadvantage in the murine gut[16]. To quantitatively probe the effects of the *arc* operon on the fitness of *E. coli* in the presence and absence of *arc* operon in EC, we performed a competition experiment between the WT and Δarc strain. Specifically, we co-cultured the WT and Δarc strain

175 that each harbor a unique DNA barcode on the genome in presence of a range of arginine and
176 initial pH values. The WT strain displayed a moderate to substantial growth advantage in all *in*
177 *vitro* conditions **(Figure S1B-D)**. To determine whether this fitness advantage was observed in
178 the mammalian gut, we introduced the co-culture into germ-free mice fed a high arginine diet (5%
179 total arginine). The relative abundance of the WT was lower than the Δarc strain for most time
180 points across all four mice **(Figure S1F-G)**. Therefore, the *arc* operon in EC confers an advantage
181 *in vitro* and colonization disadvantage *in vivo*, consistent with the previous study[16]. The pH of
182 diluted fecal and cecal samples was calculated by measuring the supernatant of 0.1% (v/v) in
183 molecular grade water through a pH probe[36]. The pH of diluted fecal and cecal samples were
184 found with little variability across time, suggesting either the host dominated the regulation of pH
185 within the gut or the *arc* operon was poorly expressed *in vivo*. In sum, the expression of the *arc*
186 operon in the pTet-arc strain supported consistent growth across a wide range of initial pH values
187 and confers growth implication during mammalian colonization.
188


A
arginine
arcA
citruline
arcB
arcD
ornithine
carbamyl phosphate
arcC
ammonia
pH increase
nitrogen resource
?
?
OH
Δarc
arcA arcC arcB arcD
WT
pTet-arc
B
Bacteroidetes
Proteobacteria
Actinobacteria
Firmicutes
Odoribacter laneus
Citrobacter freundii
Citrobacter braakii
Citrobacter kobei
Enterobacter asburiae
Enterobacter aerogenes
Klebsiella michiganensis
Citrobacter koseri
Escherichia coli
Enterobacter cloacae
Aeromonas caviae
Pseudomonas putida
Pseudomonas fluorescens
Pseudomonas aeruginosa
Pseudomonas nitroreducens
Pseudomonas citronellolis
Kingella oralis
Gordonibacter pamelaeae
Eggerthella lenta
Asaccharobacter celatus
Slackia piriformis
Collinsella aerofaciens
Collinsella tanakaei
Collinsella intestinalis
Atopobium minutum
Olsenella uli
Atopobium vaginae
Gardnerella vaginalis
Corynebacterium simulans
Mobiluncus curtisii
Actinomyces naeslundii
Cutibacterium granulosum
Cutibacterium avidum
Cutibacterium namnetense
Propionibacterium sp.
Cutibacterium acnes
Anaerococcus octavius
Clostridium perfringens
Firmicutes bacterium
Staphylococcus epidermidis
Staphylococcus capitis
Staphylococcus warneri
Staphylococcus aureus
Staphylococcus hominis
Staphylococcus haemolyticus
Weissella confusa
Weissella cibaria
Leuconostoc citreum
Lactobacillus vaginalis
Lactobacillus reuteri
Lactobacillus fermentum
Lactobacillus mucosae
Pediococcus pentosaceus
Pediococcus acidilactici
Lactobacillus parabuchneri
Lactococcus garvieae
Lactococcus lactis
Streptococcus agalactiae
Streptococcus parasanguinis
Streptococcus hirae
Enterococcus durans
Enterococcus faecium
Enterococcus gallinarum
Enterococcus devriesei
Enterococcus avium
C
WT
Δarc
pTet-arc (-)
pTet-arc (+)
OD
pH
Time (hr)
Initial pH
9.0
4.0

189

Figure 1 | The *arc* operon encodes for a specialized metabolic pathway that can increase environmental pH by arginine metabolism. (A) Mechanism of arginine dihydrolase pathway and strains engineered in this study. The *arc* operon in wildtype EC (WT) encodes four color-coded genes, each responsible for various reactions in arginine metabolism. The arginine dihydrolase pathway and metabolize arginine into ammonia to increase environmental pH and provide additional nitrogen resource, which may affect community assembly and functions. EC strains with *arc* deletion (Δarc) and *arc* under inducible promoter on a plasmid (pTet-arc) were cloned for this study. (B) Phylogenetic tree of species harboring the arc operon in the human gut microbiome generated by metagenomic analysis. Species color indicates phylum. The species of interest in this study (*E. coli*) is marked by a red star. (C) Absolute abundance (calculated OD600) and supernatant pH of various EC strains under different initial media pH ranging from 5 (light shade) to 8 (dark shade) over time. Species were inoculated at an equal abundance based on OD600 measurements. Smaller sized points indicate experimental data replicates. Larger sized points and lines indicate average over all biological replicates. Different initial media pH is marked by different color of line and datapoints, with darker color corresponding to higher initial media pH.

Expression of the *arc* operon impacts *in vitro* human gut community assembly and function

Since the *arc* operon has been shown to correlate with changes in butyrate production[30] and pathogen virulence[19], we hypothesized that the *arc* operon could be used to shape community assembly and butyrate production. We designed an 8-member human gut community consisting of representatives from the major phyla *Bacteroidetes*, *Firmicutes*, and *Actinobacteria* **(Figure 2A, Supplementary Data 1)** that do not harbor the *arc* operon. This community contained two prevalent butyrate-producing species (*A. caccae* and *C. comes*), previously shown to contribute substantially to butyrate production in a 25-member human gut community[30]. In each condition, we introduced an individual EC strain. The community was cultured in a chemically defined medium that can support the growth of all species[30]. Since expression of the *arc* operon is regulated by arginine concentration[41], the media was supplemented with sufficient arginine (1%) to activate expression of the *arc* operon based on the endpoint pH, while limiting EC growth to promote community diversity **(Figure S2)**.

Since the pTet-arc (+) and Δarc conditions represent the two extremes of *arc* operon expression **(Figure S1A)**, a comparison of these conditions can provide insights into the impact of the *arc* operon on community assembly and butyrate production **(Figure 2B)**. Notably, high expression of the *arc* operon steered the community towards a very similar endpoint pH despite variations in initial media pH (**Figure 2C**). By contrast, communities lacking the *arc* operon expression exhibited strong history dependence on the initial media pH. High expression of the *arc* operon also shaped the absolute abundance and relative abundance of different species in the community across different initial pH environments **(Figure 2D, S3)**. Certain shifts in species abundance could be predicted based on monoculture growth characterization. For example, species that preferred lower initial pH in monoculture (BL, CC) were higher abundance in the absence of the *arc* operon expression. Similarly, species that preferred higher pH (e.g. AC, BC, BT and EC) were higher abundance under *arc* operon expression. Overall, the growth profiles of each species in the presence and absence of the community displayed similar trends with the exception of DF, which grew poorly in monoculture **(Figure S4A)**. Finally, the impact of the *arc* operon on the growth of the butyrate producer AC varied as a function of time, wherein AC was enhanced at earlier times and then inhibited at later times in higher initial pH conditions **(Figure S4B-E)**. This demonstrates that the presence of the *arc* operon can shape the transient shifts in species abundance as a function of time.

The presence of the *arc* operon impacted the production of butyrate by the synthetic human gut community **(Figure 2E).** At lower initial pH (pH = 5, 6), which is a pH range more physiologically relevant to the pH in human colon[42], butyrate was significantly higher for high expression of the *arc* operon (pTet-arc (+)) than in the absence of the *arc* operon (Δarc). At higher initial pH (pH=8), the reverse trend was observed where the absence of the *arc* operon yielded significantly higher butyrate. This is potentially due to butyrate producing species AC significantly enriched at lower pH with *arc* operon expression while its growth is inhibited at higher pH due to EC outgrowth **(Figure S5A)**. Notably, the relative abundance of another butyrate producing species CC was lower in pTet-arc (+) conditions **(Figure S5B)**, demonstrating the differential effect of the *arc* operon on diverse butyrate-producing bacteria. Nonetheless, the total butyrate producer abundance linearly correlated with final butyrate concentration at extreme initial pH conditions **(Figure S5C).** Interestingly, at initial pH=5.32 where we observed the greatest difference in butyrate production, we also observed greatest difference in the endpoint lactate concentration in supernatant **(Figure S6)**, where *arc* operon expression leads to lower lactate level in supernatant. This is potentially because AC, the only butyrate producer present at pH=5.32 **(Figure S5A)**, can convert lactate into butyrate[43], therefore depleting lactate in the media. While AC can also convert acetate into butyrate, such utilization is likely compensated by acetate production of other species in the community[44], therefore maintaining a relatively constant acetate

level in the supernatant. These results highlight the differential effect of the arc operon expression on productions of distinct short chain fatty acids.

To assess the effect of the *arc* operon on the sensitivity of community assembly to initial pH variation, we evaluated the pairwise Euclidean distances at the endpoint across different initial pH conditions. The variance in the Euclidean distance distribution was substantially lower in communities with high *arc* operon expression than in the absence of the *arc* operon (**Figure 2F**). This implies that the *arc* operon can stabilize the assembly of the community in response to external pH perturbations. Overall, expression of the *arc* operon in EC can reduce the variability in environmental pH during outgrowth of a human gut community, reduce the variability in community assembly in response to initial pH perturbations, and enhance butyrate production in physiologically relevant pH ranges.

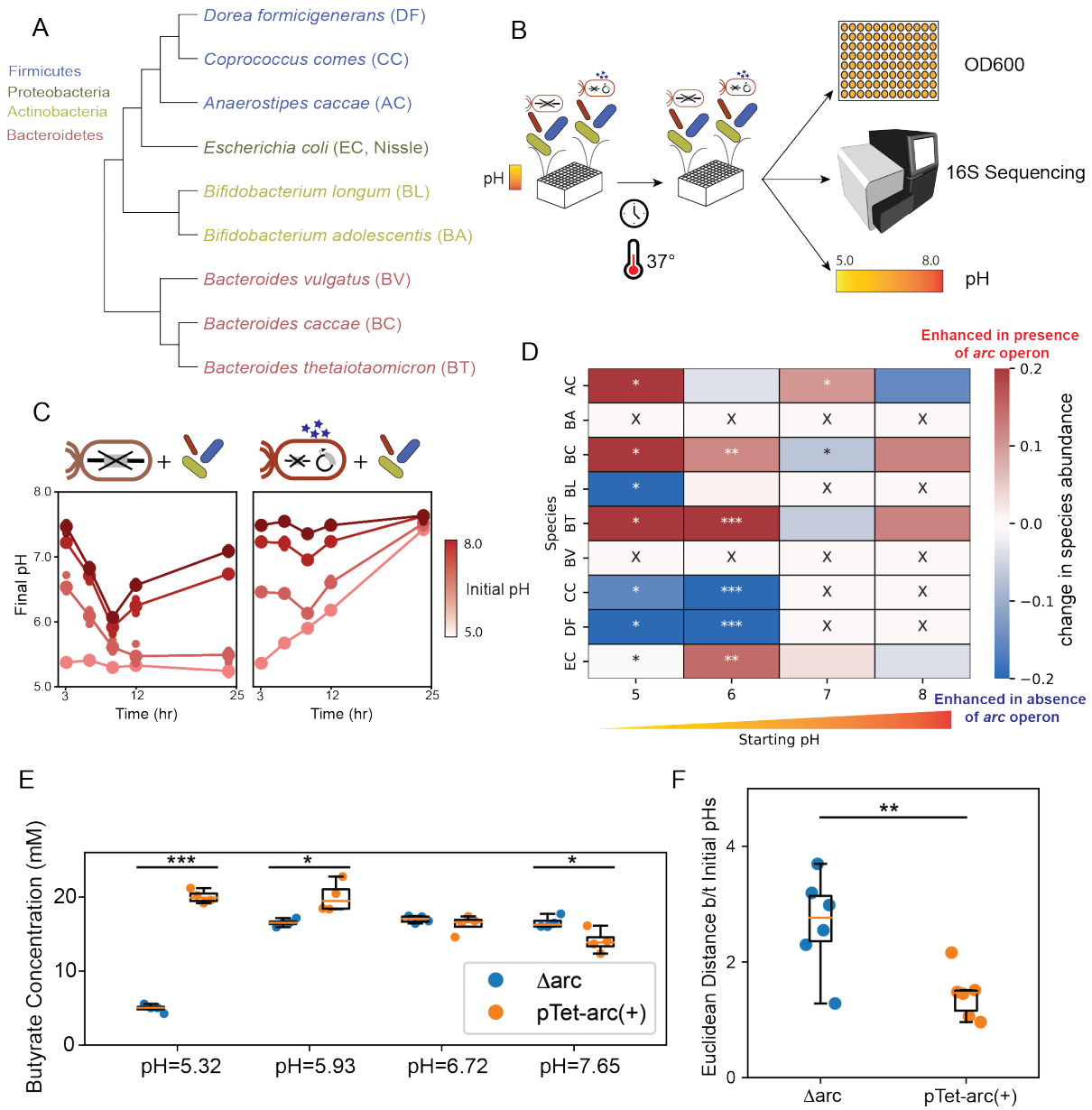


Figure 2 | Community assembly and butyrate production are impacted by the *arc* operon expression. (A) Phylogenetic tree of the synthetic human gut microbial community composed of 8 highly prevalent and diverse species. The species of interest harboring the *arc* operon (*E. coli*) is also included in the tree, with strain name noted. Species color indicates phylum. (B) Schematic of the experimental workflow of *in vitro* community coculture with various strains of EC at various starting pH. Synthetic communities are cultured in microtiter plates in anaerobic conditions and incubated at 37°C. The absolute abundance of each species is determined by measuring cell density at 600 nm (OD600). The supernatant pH is taken by PhenolRed method. The Community composition is determined using multiplexed 16S rRNA sequencing. (C) Supernatant pH of community coculture with Δarc and pTet-arc (+) strains of EC under different initial media pH ranging from 5 to 8. Species were inoculated at an equal abundance based on OD600 measurements. Datapoints indicate experimental data replicates. Lines indicate average over all biological replicates. Different initial media pH is marked by different color of line and datapoints, with darker color corresponding to higher initial media pH. (D) Heatmap of the log of the ratio

between species relative abundance (mean value of biological replicates, n = 3) in pTet-arc (+) condition and Δarc condition at various initial media pH. Species relative abundances are determined by multiplexed 16S rRNA sequencing. Red corresponds to higher relative abundance in WT conditions and blue corresponds to higher relative abundance in Δarc conditions. Low growth conditions are marked by an 'X'. Asterisks represent statistical significance: *P < 0.05, **P < 0.01, ***P < 0.001 according to an unpaired t-test. (E) Box plot of butyrate concentration (mM) at 24-hour for community+EC cocultures with either Δarc strain or pTet-arc(+) strain at various starting pH. Datapoints indicate experimental data replicates. Datapoint colors indicate strain of EC in coculture. Independent statistical t-test is performed on each pair of conditions with the same starting pH. Asterisks represent statistical significance: *P < 0.05, **P < 0.01, ***P < 0.001 according to an unpaired t-test. (F) Community structure Euclidean distances between all pairs of initial pH conditions at 24-hr for Δarc condition (blue) and pTet-arc(+) condition(orange). The Euclidean distances were calculated based on community relative abundance structure between each pair of initial pH condition for a particular strain condition. Points represent the Euclidean distance calculated for different pairs of initial pH conditions. Asterisks represent statistical significance: *P < 0.05, **P < 0.01, ***P < 0.001 according to an unpaired t-test.

A dynamic computational model reveals the contributions of external pH on community assembly

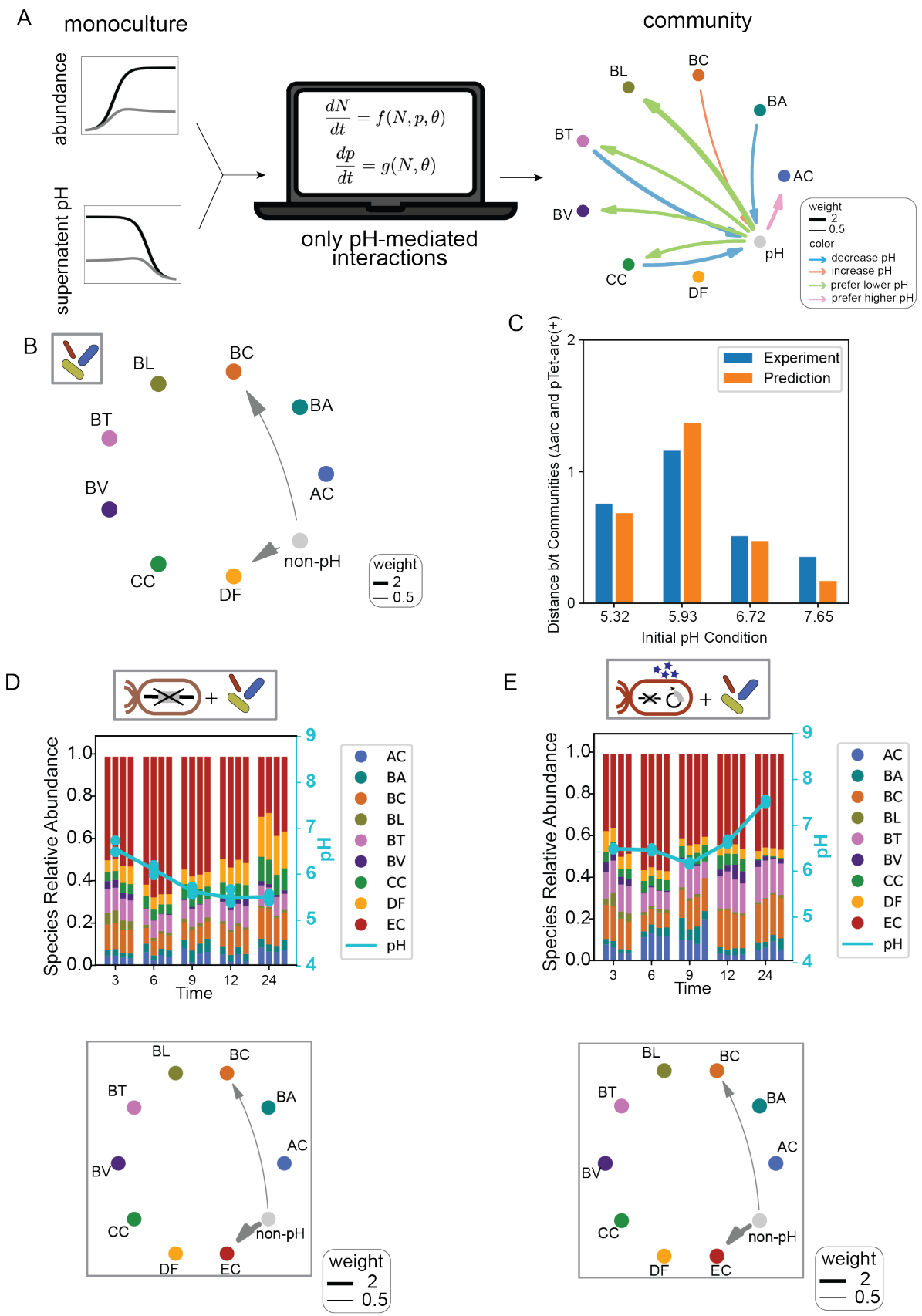
To investigate the role of environmental pH on community assembly, we used a dynamic computational model that captures pH-dependent inter-species interactions and community dynamics[33]. In the model, growth rate is maximized at preferred pH and decreases when pH deviates from this preferred value**.** This assumption is consistent with monoculture growth data as a function of variation in initial pH **(Figure S7, S8)**. For simplicity, the contribution of each species to environmental pH is assumed to be linear with species abundance. This model also captures lag phase by representing a linear rate of transition between a population that is unable to grow and a population that displays growth (Supplementary Information). Model parameters were inferred by fitting to the model to time-series measurements of single species growth in media at different initial pH **(Figure 3A, S7, S8**). The fitted model displayed a high goodness of fit to the monoculture growth across all time points (Pearson correlation coefficient = 0.99 and p-value of $6.9*10^{-23}$. **(Figure S9)**. However, the model failed to fit to the pH trend of select EC strains **(Figure S8 C-E).** This could stem from the inability of the model to capture diauxic shifts, which may lead to temporal variation in the effect of EC on environmental pH[45].

The inferred parameters indicates that some species prefer lower pH (BL, BT, BV, CC) while others prefer higher pH (AC, BA) **(Figure 3A)**. Species also display differences in their impact on environmental pH (e.g. BA, BT, and CC lower environmental pH whereas BC increases environmental pH) **(Figure 3A)**. The deviation between the model prediction and the species abundance measurements using community data provide insights into the contribution of non-pH mediated interactions (e.g. resource competition and cross-feeding) or pH-mediated interactions that deviate from monoculture behaviors (e.g. diauxic shifts in metabolism). In the pH model predictions of community-only experiments, pH decreased over time, consistent with experimental data **(Figure S10A).** The prediction of species presence or absence across different initial pH conditions was largely consistent with most experimental conditions **(Figure S10B).** In addition, most species predictions are lower than experimentally measured **(Figure S10C).** Since the computational model assumes only pH-mediated interactions among species, the discrepancy between the model prediction and experimental measurements reveals the direction and magnitude of mechanisms of interaction beyond pH. Therefore, we define a metric for non-pH mediated interactions by the sum of all predicted abundance differences across all timepoints in each initial pH condition (**Figure 3B, Figure S10C, Figure S13A**). Species that are underpredicted by the model are influenced by positive non-pH mediated interactions **(Figure 3B)**. For example, DF in monoculture failed to grow in all conditions **(Figure S8)** but displayed growth in most initial pH conditions in the community, highlighting the role of growth promoting non-pH-mediated interactions. AC and BV displayed little non-pH mediated interactions per our calculation **(Figure S13A),** suggesting the inter-species interactions for these species are predominantly pH mediated.

The model informed by monoculture data was then used to predict assembly of communities cocultures with individual EC strains to provide insights into the extent of pH-mediated interactions in the presence/absence of the *arc* operon **(Figure S11)**. Similar to the 8-member community prediction, species presence/absence patterns across conditions are largely consistent with the model prediction except initial pH condition 7.65 **(Figure S12 A, B)**. To evaluate whether the pH model can capture community assembly differences in the presence and absence of the *arc* operon, we calculated the distance between predicted and measured community abundances with EC removed, since EC was dominating the community structure. The distance between community composition in the presence and absence of the *arc* operon across pH conditions was consistent with our model predictions **(Figure 3C)**. This implies that the effects of the *arc* operon on community assembly can be largely explained by pH-mediated interactions. Further, the model accurately predicts the non-monotonic trends in the Shannon

diversity index as a function of initial environmental pH **(Figure S12C)**. These overall patterns predicted by our model are represented in network that enables visualization of the extent of non-pH mediated interactions based on the deviation between species predicted relative abundances and experimental data **(Figure 3D, E, Figure S12D-E, Figure S13B-C)**. In comparison to the interaction network of the human gut community, several species display altered non-pH-mediated interactions with the addition of EC in coculture **(Figure S13).** This implies a species-dependent change in interaction profile due to the addition of EC, regardless of *arc* operon expression. By contrast, communities with Δarc or pTet-arc(+) display similar non-pH-mediated interaction networks at each initial pH condition, whereas community structures changed significantly between coculture with either strains of EC **(Figure 3D, E, Figure S13B-C)**. This implies that the *arc* operon impacts community assembly via pH-mediated mechanisms and does not alter the types of mechanisms of inter-species interaction in the community at a given initial pH.

Across predictions for different cocultures of community only, community with Δarc, and community with pTet-arc(+), we observed lower predictive performance at higher initial pH **(Figure S9B, S12A, B)**. We speculate that this is because coculture community structure changes are not determined by *arc*-induced pH changes. We quantified the effect of the *arc* operon pH modification on community assembly by evaluating the relationship between the Euclidean distance of communities with pTet-arc(+) and Δarc and the change in pH between each pair of time points. A positive correlation would imply alterations in community composition are correlated with the *arc* operon activity of increasing environmental pH. These variables only were positively correlated for the initial pH=5.93 condition (Spearman correlation was 0.72 with p-value of 0.0004) **(Figure S12F).** These results suggest that within this pH range the *arc* operon activity had the most substantial effect on shaping community assembly through increasing pH, while in other pH ranges, the pH increases induced by *arc* operon expression cannot explain changes in community structure. In sum, a quantitative comparison of the community assembly trends predicted by the dynamic pH model fitted on monoculture data and our experimental data suggests that arc operon expression affect *in vitro* community dynamics mainly through pH-mediated interactions. Our overall computational and experimental workflow can provide quantitative insights into the contributions of pH on community dynamics and functions.


A
monoculture
abundance
supernatent pH
$\frac{dN}{dt} = f(N, p, \theta)$
$\frac{dp}{dt} = g(N, \theta)$
only pH-mediated interactions
community
BL
BC
BA
BT
AC
BV
pH
CC
DF
weight
2
0.5
color
decrease pH
increase pH
prefer lower pH
prefer higher pH
B
BL
BC
BA
BT
AC
BV
non-pH
CC
DF
weight
2
0.5
C
Experiment
Prediction
Distance b/t Communities (Δarc and pTet-arc(+))
0
1
2
5.32
5.93
6.72
7.65
Initial pH Condition
D
Species Relative Abundance
0.0
0.2
0.4
0.6
0.8
1.0
pH
4
5
6
7
8
9
3
6
9
12
24
Time
AC
BA
BC
BL
BT
BV
CC
DF
EC
pH
BL
BC
BT
BA
AC
BV
CC
non-pH
DF
EC
weight
2
0.5
E
Species Relative Abundance
0.0
0.2
0.4
0.6
0.8
1.0
pH
4
5
6
7
8
9
3
6
9
12
24
Time
AC
BA
BC
BL
BT
BV
CC
DF
EC
pH
BL
BC
BT
BA
AC
BV
CC
non-pH
DF
EC
weight
2
0.5

389

Figure 3 | A dynamic model trained on monoculture growth pinpoints pH-mediated effects of the *arc* operon on community assembly. (A) Modeling framework for predicting community assembly. The modified microbe-pH computational model represents community dynamics and environmental pH dynamics as ordinary differential equations, with interdependency between species growth and environmental pH. Model parameters were fitted on monoculture data of all gut community species and EC strains by minimizing differences between predicted and experimental species abundance and supernatant pH. The community assembly was predicted using monoculture-fitted parameters to infer species-pH interactions in a synthetic gut microbial community. The species-pH interaction network shown on the right is the predicted interactions in gut community cocultures. Arrows going from pH to species represent the species preference for higher (thicker) or lower (thinner) environmental pH. Arrows going from species to pH represent the species tendency to increase (orange) or decrease (blue) the environmental pH. Arrows going from pH to species represent species preference for higher (pink) or lower (green) environmental pH than 6.5. Arrow width represent the strength of interaction. (B) Model prediction of the effect of non-pH-mediated interactions on species in gut community cocultures at pH=5.93. The non-pH-mediated interactions were calculated by the difference between model predicted relative abundance and experimental relative abundance for each species in each coculture condition summed across all timepoints. Arrowheads represent promotion (pointed) or inhibition (blunt) of species growth by non-pH-mediated interactions. Arrow width represent the strength of interaction. (C) Community assembly Euclidean distance between community coculture with Δarc or pTet-arc(+) at various initial pH. The Euclidean distances were calculated for the absolute abundances at 24 hour of coculture for both experimental data (blue) and model prediction (orange). (D, E) Experimental coculture relative abundance and model predicted non-pH-mediated interaction network of community coculture with Δarc (D) and pTet-arc(+) (E) at initial pH=5.93, respectively. Species relative abundances are determined by multiplexed 16S rRNA sequencing. Color indicates species identity. The cyan line in relative abundance barplot indicates pH. The non-pH-mediated interactions were calculated by the difference between model predicted relative abundance and experimental relative abundance for each species in each coculture condition summed across all timepoints. Arrowheads represent promotion (pointed) or inhibition (blunt) of species growth by non-pH-mediated interactions. Arrow width represent the strength of interaction.

423 **The *in vivo* impacts of the *arc* operon on colonization and community assembly exhibit**
424 **opposite trends than *in vitro***
425
426 With the pH-mediated interactions predicted in the *in vitro* community assemblies by the
427 computational model, we wondered if similar interactions translate into mammalian models, where
428 host-microbe interaction would add another layer of complexity. Additionally, given intestinal pH
429 is important to human health and disease, we are curious to explore potential therapeutic
430 applications of the *arc* operon expression, which increases environmental pH in *in vitro* cocultures.
431 We investigated the role of the *arc* operon on community assembly *in vivo* using a germ-free
432 murine model. Gnotobiotic male mice aged 6-8 weeks were place on a customized high arginine
433 diet 3 days prior to introduction of bacteria. Different groups of mice were housed in separate
434 cages. Mice were orally gavaged with a culture containing equal species proportions. To
435 characterize the changes in community composition over time, fecal samples were collected
436 every 2-3 days for pH measurement[46,47] and 16S rRNA gene sequencing. After one week, groups
437 of mice received either the WT or Δarc strains of EC and a separate group of mice did not receive
438 an EC strain. After two weeks, mice were sacrificed and cecal contents were collected for pH
439 measurement[46,47] and 16S rRNA gene sequencing **(Figure 4A)**. Using qRT-PCR, we
440 demonstrated that the *arc* operon in mice fed the WT strain was expressed compared to the pTet-
441 arc(-) *in vitro* control **(Figure S1A)**.
442 The relative abundance of WT and Δarc strains calucated from 16S sequencing initially
443 increased at the same rate and saturated in abundance at day 4 after introduction **(Figure 4B)**.
444 The Δarc strain colonized mice at a moderately but significantly higher abundance than the WT
445 strain, consistent with its higher abundance in competition with the WT strain **(Figure S1)**. Further,
446 while EC made up 60-80% in cocultures with the community *in vitro* **(Figure S3A-D)**, EC
447 converged to approximately 30% of the cocolonization in the mammalian gut **(Figure 4B, S14A-**
448 **B)**. The colonization of certain species in the community were altered in response to the
449 presence/absence of the *arc* operon. Specifically, certain species were enriched in the presence
450 of the *arc* operon (WT condition) (CC, BC), whereas the colonization of other species was reduced
451 (AC, DF) at different timepoints **(Figure 4C)**. The effects of the *arc* operon on BC and DF were
452 transient, whereas the changes in abundance of the butyrate producers CC and AC were
453 sustained over multiple timepoints. The observed trend of susceptibility *in vivo* is in accordance
454 with *in vitro* observations of monoculture species growth in response to different initial pH
455 conditions **(Figure S4)**. These results highlight the key role of the mammalian gut in shaping
456 community dynamics that can display reversed trends from *in vitro* observations.
457 While we observed difference in select species colonization between mice colonized with
458 WT and Δarc, the fecal and cecal pH was not significantly different between groups **(Figure S14E)**,
459 indicating that the *arc* operon did not have a substantial effect on the pH of the fecal and cecal
460 samples. Further, cecal measurements of butyrate, propionate, isobutyrate, acetate and
461 isovalerate were not significantly different in the presence and absence of the *arc* operon **(Figure**
462 **S15).** These results suggest that the *arc* operon did not have a substantial impact on intestinal
463 pH and fermentation product metabolism in the mammalian gut in response to the high arginine
464 diet. Therefore, our results suggests that the *arc* operon is shown to have the potential of
465 mediating community assembly and colonization ability in murine model with varying temporal
466 dynamics but fails to have a significant effect on intestinal pH and select fermentation product
467 metabolism.
468

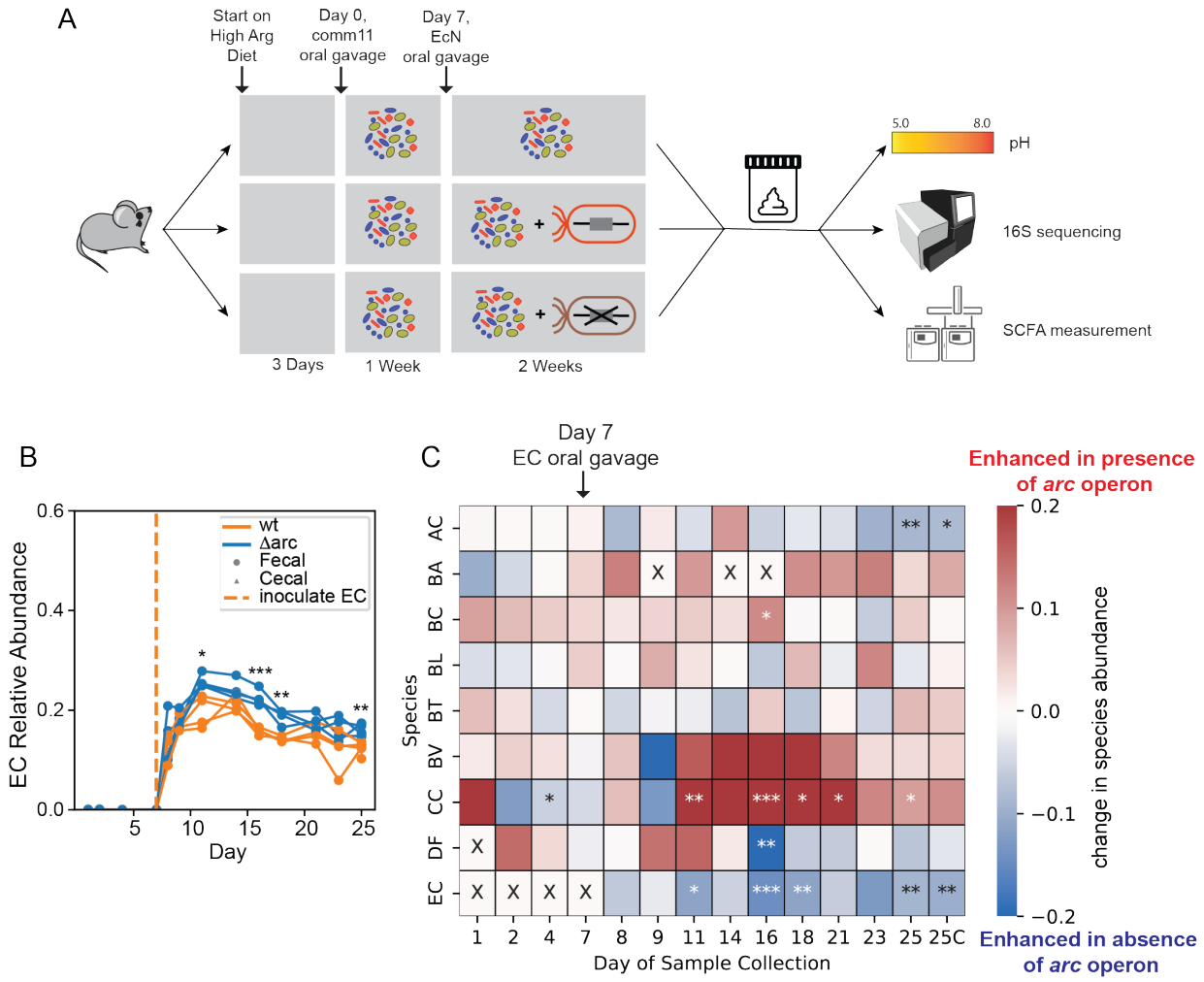


Figure 4 | The *arc* operon impacts community assembly and colonization in mammalian gut. (A) Experimental workflow of *in vivo* community coculture with various strains of EC. Mice were tagged and placed on customized high arginine diet 3 days before oral gavage of equal part community bacterial suspension. Fecal samples were collected every two to three days for determination of species relative abundances by multiplexed 16S rRNA sequencing and fecal sample pH by measurement of diluted sample with pH meter. After 1 week of colonization, WT and Δarc strains of EC were introduced by oral gavage into mice in designated groups. After 2 weeks, mice were sacrificed with cecum samples extracted for analysis of species composition, cecal sample pH, and short chain fatty acid concentrations by gas chromatography. (B) EC relative abundance from fecal and cecal sample in WT and Δarc conditions. EC relative abundances are determined by multiplexed 16S rRNA sequencing on fecal samples and cecal samples. Datapoints indicate experimental data replicates. Lines indicate average over all biological replicates. EC inoculation is marked by orange dashed line on Day 7. Asterisks represent statistical significance: *P < 0.05, **P < 0.01, ***P < 0.001 according to an unpaired t-test. (C) Heatmap of the difference of species relative abundance (mean value of biological replicates, n = 3) in WT condition and Δarc condition at various timepoints during the mouse experiment. Species relative abundances are determined by multiplexed 16S rRNA sequencing. Red corresponds to higher relative abundance in WT conditions and blue corresponds to higher relative abundance in Δarc conditions. EC inoculation is marked on Day 7. Asterisks represent statistical significance: *P < 0.05, **P < 0.01, ***P < 0.001 according to an unpaired t-test.

DISCUSSION

Our results demonstrated the possibility of tuning expression of a specialized metabolic pathway as the control knob for community dynamics and functions. Through a combination of synthetic biology and bottom-up community construction, we showed that expression of the *arc* operon can stabilize *in vitro* community structure across initial pH perturbations and increase butyrate production at physiologically relevant pH ranges. Changes in luminal pH and lower butyrate level are signatures of human gut dysbiosis[48–50], therefore *arc* operon expression has potential host health benefits through regulating community assembly and metabolite production. Further, with a computational model assuming pH-only interactions, we uncovered the extent of non-pH-mediated interactions in community assembly and showed that the *arc* operon affect *in vitro* community assembly mainly through pH-mediated interactions, providing mechanistic basis for *arc* operon as a control knob for microbiome engineering.

Robustness of microbial communities is the persistence of community structures across perturbations[51]. Specifically, designing robustness in human gut microbiome is of great interest because increased robustness is correlated with improved human health and pathogen resistance[52,53]. On the other hand, the luminal pH has been shown to significantly decrease in patients with ulcerative colitis compared to normal[48]. Therefore, retaining community structure against pH perturbations presents as a potential strategy to maintain healthy gut functions for certain disease that changes luminal pH. Our results demonstrated the effectiveness of *arc* operon in stabilizing community structure against initial pH perturbations in our *in vitro* data **(Figure 2F)**, pointing to the therapeutic potential of *arc* operon expression to maintain pH homeostasis. Future efforts could exploit control of *arc* operon expression to more precisely modulate the colonization of different species in human gut communities for precision medicine development.

Increased butyrate production by human gut microbiome is associated with protection from a wide range of diseases[49,50] and is a potential therapeutic target. With *arc* operon expression, we observed increased butyrate production at physiologically relevant pH ranges *in vitro* **(Figure 2E),** demonstrating the therapeutic potential of arc operon as a control knob for tuning beneficial metabolite production. However, co-colonization of the defined community in germ-free mice displayed no significant difference in butyrate production under the presence and absence of *arc* operon expression. These conflicting results illustrate the difficulty of scaling *in vitro* results into mammalian models, where host-microbe interactions can add another layer of complexity. Since we hypothesized significant pH-dependence in the butyrate production profile of the community, the lack of butyrate production increase could be a result of host pH homeostasis, which is supported by similar measurements of fecal and cecal pH across conditions **(Figure S13D)**. In addition, the effect of the *arc* operon on community assembly and health-relevant metabolite production could vary in response to variations in the diet or gut inflammation and dysbiosis. Future studies could explore varying carbon source and amino acid concentrations in mouse diet along with induced colitis mouse models to assess the effect of *arc* expression on butyrate production across diverse host and environmental backgrounds.

While defined microbial communities have been extensively characterized to decipher specific interactions that determine system behaviors[20], deployment of synthetic microbiomes have been hindered by a lack of understanding of generalizable mechanisms of interactions and their effects on community assembly and function[54]. The computational workflow assuming only pH-mediated interactions trained on monoculture kinetics can offer insight into extent of pH-mediated and non-pH-mediated interactions in community assembly. BT has previously been shown to utilize vitamin and xenosiderophores for fitness[55,56], which potentially explains the consistently positive non-pH-mediated interactions for BT observed in all community coculture conditions **(Figure 3B, D, E)**. The extent of different types of interactions inferred in other less-characterized species can guide mechanistic investigation of these non-model species. Additionally, our model framework

pinpoints pH as the major driver behind changes in community assembly due to *arc* operon expression **(Figure 3D, E)**, which can assist future microbiome engineering designs using *arc* operon as the control knob. Further, the deviation between model prediction and experimental data observed in all community cocultures suggest significant non-pH-mediated interactions in community assembly. These mechanisms could include resource competition, metabolite cross-feeding, and antimicrobial production. Computational workflows like the one in this work could be developed for each of these mechanisms to tease out types of interactions governing growth of each species. The understanding of important interactions for each species could reveal the key mechanisms that could be exploited to shape community functions to benefit human health.

One important non-pH-mediated effect of *arc* operon expression that we failed to capture in our computational workflow is the production of excess nitrogen resource. Nitrogen limited environments, like the mammalian gut[57], can lead to suppression of microbial activity[58]. Therefore, expression of the *arc* operon could affect *in vivo* community assembly through nitrogen supplementation. Specifically, Bacteroidales were shown to be more nitrogen limited in their growth in the large intestine than other commensals[57], corresponding with our observation of enrichment of BV under *arc* operon expression *in vivo* **(Figure 4C)**. Future studies could investigate nitrogen availability in the presence and absence of *arc* operon expression by isotope labeling ^{15}N on arginine (precursor of *arc* operon expression) in the diet. Species-dependent nitrogen response to *arc* expression could be inferred from single-cell isotopic enrichment of labelled nitrogen. Such analysis could justify discrepancies in *arc* expression effects on community assembly and function between *in vitro* and *in vivo*.

A major goal for microbiome engineering is to discover and exploit molecular control knobs to steer microbiomes to desired states. Specialized metabolic pathways, which can provide key resources for specific gut bacteria or shape global environmental parameters, could be exploited to shape human gut community functions. By integrating mechanistic computational modeling with measurements of communities constructed from the bottom-up, we revealed the extent to which a single mechanism influenced community assembly. Bottom-up assembly enables a deeper understanding of the contributions of monoculture growth and inter-species interactions in response to environmental factors. While we focused on a single metabolic pathway, higher-throughput genetic approaches could also be used to discover and rank the influence of diverse molecular pathways on target community functions[59–61]. Our strategy could be applied to the pathway hits identified via high-throughput genetic pipelines to understand their quantitative effects and develop the capability to tune these pathways to precisely control community behaviors.

MATERIALS AND METHODS

Strain maintenance and culturing

All aerobic culturing was carried out in 37°C incubator with shaking. All anaerobic culturing was carried out in an anaerobic chamber with an atmosphere of 2 ± 0.5% H2, 15 ± 1% CO2 and balance N2. All prepared media and materials for anaerobic experiments were placed in the chamber at least overnight before use to equilibrate with the chamber atmosphere. The permanent stocks of each strain used in this paper were stored in 25% glycerol at −80 °C. Batches of single-use glycerol stocks were produced for each strain by first growing a culture from the permanent stock in ABB media (Supplementary Data 2) to stationary phase, mixing the culture in an equal volume of 50% glycerol, and aliquoting 400 µL into Matrix Tubes (ThermoFisher) for storage at −80 °C. Quality control for each batch of single-use glycerol stocks included Illumina sequencing of 16S rDNA isolated from pellets of the aliquoted mixture to verify the identity of the organism. For each experiment, precultures of each species were prepared by thawing a single-use glycerol stock (SUGS) and adding 100 µL of SUGS to 5 mL fresh YBHI media (Supplementary Data 2) for stationary incubation at 37 °C for 16 hours (Supplementary Data 1). All experiments were performed in a chemically defined medium named DM38[30]. This medium was developed by a previously studied medium in lab that enabled the growth of all our species in this experiment.

Plasmid and strain construction

PCR amplifications were performed using Phusion High-Fidelity DNA polymerase (New England Biolabs) and oligonucleotides for cloning were obtained from Integrated DNA Technologies. Standard cloning methods were used to construct plasmids. Small scale electroporation was used for plasmid transformation into *E. coli*. The background strain was grown to an OD of 0.3 and then placed on ice for 10 minutes. 1mL of the cell culture was then washed with 10% glycerol two times, with the cells spun down, supernatant drawn out, and resuspended each time. After final wash, cells were resuspended in 100µL 10% glycerol and 1µg PCR product and transferred to 1mm gap electro-cuvette. Cells were electroporated using the the EC1 protocol of BioRad Micropulser Electroporator, then recovered in LB plus appropriate antibiotic for 1 hour. Transformants were selected on LB agar plates with appropriate antibiotic resistance and single colonies were stored.

The native *arc* operon (consisting of genes *arcA, arcC, arcB, arcD,* and *arcR*) was deleted from the genome of WT by replacement with a Kananycin FRT construct amplified from pKD4[62] **(Supplementary Data 4)** using primers oYL36 and oYL37 **(Supplementary Data 5)**, with Lambda red expression provided by pMP11[63] **(Supplementary Data 4)**. Transformants were selected on antibiotic selective plates and verified by colony PCR. The kanamycin cassette was then removed from the genome by transforming a FLP recombinase vector pFLP2[64]. Successful Kanamycin removal was screen for the absence of Kanamycin resistance and verified by colony PCR. Finally, the pFLP2 plasmid was cured by growth in 5% sucrose. Final colonies were screened for absence of Carbenicillin resistance. The resulting strains is Δarc.

For inducible expression of the *arc* operon, we replaced the RFP region of pBbA2a-RFP plasmid[65] **(Supplemental Data 4)** with the native *arc* operon without regulatory region (consisting of genes arcA, arcC, arcB, and arcD) from WT *E. coli* strain. The *arc* operon was amplified by oYL19 and oYL20, and the pBbA2a-RFP backbone was amplified by oYL17 and oYL18 **(Supplemental Data 5)**. Standard PCR and Gibson assembly was performed to yield pTet-arcACBD **(Supplemental Data 4)**. Finally, strain pTet-arc **(Supplemental Data 1)** was cloned by transforming plasmid pTet-arcACBD into strain Δarc.

Monoculture and community culturing dynamic growth quantification

Each species' preculture was diluted to an OD600 of 0.01 (ThermoFisher Nanodrop, 1 mL in cuvette) in DM38 and aliquoted into three or four replicates of 1.5 mL each in a 96 Deep Well (96DW) plate and covered with a semi-permeable membrane (Diversified Biotech) and incubated at 37 °C without shaking. At each time point, samples were mixed and OD600 and pH were measured by the plate reader. Cell pellet was saved at each timepoint for species abundance quantification at -80°C. For pH measurement, phenol red solution was diluted to 0.005% weight per volume in MilliQ water. Bacterial supernatant (20 µL) was added to 180 µL of phenol red solution, and absorbance was measured at 560 nm (Tecan Spark Plate Reader). A standard curve was produced by fitting the Henderson–Hasselbach equation to fresh media with a pH ranging between 3 and 11 measured using a standard electro-chemical pH probe (Mettler-Toledo). We used the following equation to map the pH values to the absorbance measurements.

$$\mathrm{pH} = a + b * \ln A_{560} \tag{1}$$

The parameters a and b were determined using a linear regression between pH and the log term for the standards in the linear range of absorbance with A representing the absorbance of each condition.

Genomic DNA extraction from cell pellets

Genomic DNA was extracted from cell pellets using a modified version of the Qiagen DNeasy Blood and Tissue Kit protocol. First, pellets in 96DW plates were removed from −80 °C and thawed in a room temperature water bath. Each pellet was resuspended in 180 µL of enzymatic lysis buffer (20 mM Tris–HCl (Invitrogen), 2 mM sodium EDTA (Sigma-Aldrich), 1.2% Triton X-100 (Sigma-Aldrich), 20 mg/mL lysozyme from chicken egg white (Sigma-Aldrich)). Plates were then covered with a foil seal and incubated at 37 °C for 30 min with orbital shaking at 600 RPM. Then, 25 µL of 20 mg mL−1 Proteinase K (VWR) and 200 µL of Buffer AL (QIAGEN) were added to each sample before mixing with a pipette. Plates were then covered by a foil seal and incubated at 56 °C for 30 min with orbital shaking at 600 RPM. Next, 200 µL of 100% ethanol (Koptec) was added to each sample before mixing and samples were transferred to a nucleic acid binding (NAB) plate (Pall) on a vacuum manifold with a 96DW collection plate. Each well in the NAB plate was then washed once with 500 µL buffer AW1 (QIAGEN) and once with 500 µL of buffer AW2 (QIAGEN). A vacuum was applied to the Pall NAB plate for an additional 10 min to remove any excess ethanol. Samples were then eluted into a clean 96DW plate from each well using 110 µL of buffer AE (QIAGEN) preheated to 56 °C. Genomic DNA samples were stored at −20 °C until further processing.

Sequencing library preparation

Genomic DNA concentrations were measured using a SYBR Green fluorescence assay and then normalized to a concentration of 1 ng µL−1 by diluting in molecular grade water using a Tecan Evo Liquid Handling Robot. First, genomic DNA samples were removed from −20 °C and thawed in a room temperature water bath. Then, 1 µL of each sample was combined with 95 µL of SYBR Green (Invitrogen) diluted by a factor of 100 in TE buffer (Integrated DNA Technologies) in a black 384-well microplate. This process was repeated with two replicates of each DNA standard with concentrations of 0, 0.5, 1, 2, 4, and 6 ng µL−1. Each sample was then measured for fluorescence with an excitation/emission of 485/535 nm using a Tecan Spark plate reader. Concentrations of each sample were calculated using the standard curve and a custom Python script was used to compute the dilution factors and write a worklist for the Tecan Evo Liquid Handling Robot to normalize each sample to 1 ng µL−1 in molecular grade water. Samples with DNA concentration <1 ng µL−1 were not diluted. Diluted genomic DNA samples were stored at −20 °C until further processing.

679
680 Amplicon libraries were generated from diluted genomic DNA samples by PCR
681 amplification of the V3–V4 of the 16S rRNA gene using custom dual-indexed primers
682 (Supplementary Data 3) for multiplexed next-generation amplicon sequencing on Illumina
683 platforms (Method adapted from Venturelli et al. Mol. Syst. Biol., 2018). Primers were arrayed in
684 skirted 96-well PCR plates (VWR) using an acoustic liquid handling robot (Labcyte Echo 550)
685 such that each well received a different combination of one forward and one reverse primer (0.1 µL
686 of each). After liquid evaporated, dry primers were stored at −20 °C. Primers were resuspended
687 in 15 µL PCR master mix (0.2 µL Phusion High Fidelity DNA Polymerase (Thermo Scientific),
688 0.4 µL 10 mM dNTP solution (New England Biolabs), 4 µL 5× phusion HF buffer (Thermo
689 Scientific), 4 µL 5 M Betaine (Sigma-Aldrich), 6.4 µL Water) and 5 µL of normalized genomic DNA
690 to give a final concentration of 0.05 µM of each primer. Primer plates were sealed with Microplate
691 B seals (Bio-Rad) and PCR was performed using a Bio-Rad C1000 Thermal Cycler with the
692 following program: initial denaturation at 98 °C (30 s); 25 cycles of denaturation at 98 °C (10 s),
693 annealing at 60 °C (30 s), extension at 72 °C (60 s); and final extension at 72 °C (10 min). 2 µL of
694 PCR products from each well were pooled and purified using the DNA Clean & Concentrator
695 (Zymo) and eluted in water. The resulting libraries were sequenced on an Illumina MiSeq using a
696 MiSeq Reagent Kit v3 (600-cycle) to generate 2 × 300 paired-end reads.
697
698 **Bioinformatic analysis for quantification of species abundance**
699 Sequencing data were demultiplexed using Basespace Sequencing Hub's FastQ Generation
700 program. Custom python scripts were used for further data processing (method adapted from
701 Venturelli et al. Mol. Syst. Biol., 2018)16. Paired end reads were merged using PEAR (v0.9.10)83
702 after which reads without forward and reverse annealing regions were filtered out. A reference
703 database of the V3–V5 16S rRNA gene sequences was created using consensus sequences
704 from next-generation sequencing data or Sanger sequencing data of monospecies cultures.
705 Sequences were mapped to the reference database using the mothur (v1.40.5)84 command
706 classify.seqs (Wang method with a bootstrap cutoff value of 60). Relative abundance was
707 calculated as the read count mapped to each species divided by the total number of reads for
708 each condition. Absolute abundance of each species was calculated by multiplying the relative
709 abundance by the OD600 measurement for each sample. Samples were excluded from further
710 analysis if >1% of the reads were assigned to a species not expected to be in the community
711 (indicating contamination) or if they had <1000 total reads and OD600 > 0.1 (indicating that there
712 were insufficient reads for analysis and this was not due to lack of community growth).
713
714 **HPLC quantification of organic acids from *in vitro* samples**
715 Supernatant samples were thawed in a room temperature water bath before addition of 2 µL of
716 H_2SO_4 to precipitate any components that might be incompatible with the running buffer. The
717 samples were then centrifuged at 3500 × rpm for 10 min and then 150 µL of each sample was
718 filtered through a 0.2 µm filter using a vacuum manifold before transferring 70 µL of each sample
719 to an HPLC vial. HPLC analysis was performed using a Shimadzu HPLC system equipped with
720 a SPD-20AV UV detector (210 nm). Compounds were separated on a 250 × 4.6 mm Rezex©
721 ROA-Organic acid LC column (Phenomenex Torrance, CA) run with a flow rate of 0.2 mL min^{-1}
722 and at a column temperature of 50 °C. The samples were held at 4 °C prior to injection. Separation
723 was isocratic with a mobile phase of HPLC grade water acidified with 0.015 N H_2SO_4 (415 µL L−1).
724 At least two standard sets were run along with each sample set. Standards were 100, 20, and
725 4 mM concentrations of butyrate, succinate, lactate, and acetate, respectively. The injection
726 volume for both sample and standard was 25 µL. The resultant data was analyzed using
727 Shimadzu LabSolutions software package with peaks manually reintegrated if necessary.
728
729 **Gnotobiotic mouse experiments**

730 All germ-free mouse experiments were performed following protocols approved by the University
731 of Wisconsin-Madison Animal Care and Use Committee. Two diets were used in this experiment:
732 regular diet with lower arginine concentration (Chow diet, Purina, LabDiet 5021) and high arginine
733 diet (Envigo, TD.210715). All strains were grown at 37 ºC anaerobically in YBHI media (Acumedia,
734 Bacto, and Sigma-Aldrich) to stationary phase. All strains for a oral gavage were mixed in equal
735 proportions based on OD600 and transferred on ice prior to oral gavage. We used 8-week old
736 C57BL/6 gnotobiotic male mice (wild-type) fed the specific diets 3 days prior to oral gavage
737 (Figure 2E, 3E). The mice were housed in biocontainment cages (Allentown Inc.) for the duration
738 of the experiment. Mice were maintained on the same experimental diets with autoclaved water.
739 Groups of mice (4 mice) fed a given condition of consortium were co-housed in a single cage.
740 Fecal samples were collected every 2-3 days after oral gavage for NGS sequencing and pH
741 measurement. At the end of the experiment, mice were euthanized, and the cecal contents were
742 collected for NGS sequencing and pH measurement.
743
744 **Genomic DNA extraction from fecal and cecal samples**
745 The DNA extraction for fecal and cecal samples was performed as described previously with
746 some modifications[66]. Fecal samples (~50 mg) were transferred into solvent-resistant screw-cap
747 tubes (Sarstedt Inc) with 500 µL 0.1 mm zirconia/silica beads (BioSpec Products) and one 3.2
748 mm stainless steel bead (BioSpec Products). The samples were resuspended in 500 µL of Buffer
749 A (200 mM NaCl (DOT Scientific), 20 mM EDTA (Sigma) and 200 mM Tris·HCl pH 8.0 (Research
750 Products International)), 210 µL 20% SDS (Alfa Aesar) and 500 µL phenol/chloroform/isoamyl
751 alcohol (Invitrogen). Cells were lysed by mechanical disruption with a bead-beater (BioSpec
752 Products) for 3 min twice to prevent overheating. Next, cells were centrifuged for 5 min at 8,000
753 x g at 4°C, and the supernatant was transferred to a Eppendof tube. We added 60 µL 3M sodium
754 acetate (Sigma) and 600 µL isopropanol (LabChem) to the supernatant and incubated on ice for
755 1 hr. Next, samples were centrifuged for 20 min at 18,000 x g at 4°C. The harvested DNA pellets
756 were washed once with 500 µL of 100% ethanol (Koptec). The remaining trace ethanol was
757 removed by air drying the samples. Finally, the DNA pellets were then resuspended into 200 µL
758 of AE buffer (Qiagen). The crude DNA extracts were purified by a Zymo DNA Clean &
759 Concentrator™-5 kit (Zymo Research) prior to PCR amplification and then NGS sequencing was
760 performed.
761
762 **GC quantification of organic acids from cecal samples**
763 Organic acids were quantified using headspace gas chromatography (HS-GC). Frozen cecal
764 samples (15 ~ 40 mg) were prepared for by adding the weighed samples to chilled 20 mL glass
765 vials (Restek, Bellefonte, PA) with 2.0 g of sodium hydrogen sulfate for acidification, distilled water
766 (300 µL – mg of cecal content), and 1.0 mL of 500 µM 2-butanol as an internal standard. Glass
767 vials were immediately sealed with aluminum crimp caps with rubber septa (Restek, Bellefonte,
768 PA), vigorously mixed by hand to mix contents, and left at room temperature overnight. Prepared
769 samples were loaded on a HS-20 headspace sampler (Shimadzu, Columbia, OH) connected to
770 a Shimadzu GC-2010 Plus GC with a flame ionization detector. The column used was an SH-
771 Stabilwax (30 m, 0.25 mm internal diameter, 0.10 µM film thickness). The GC protocol used is
772 described in Hutchison et al. (2023)[67]. Standards for acetate, propionate, isobutyrate, butyrate,
773 isovalerate, and valerate were combined and serially diluted to generate a standard curve.
774 Valerate was not detected in the cecal samples. Analytes measured had their areas under the
775 curve calculated with Shimadzu Lab Solution (Ver. 5.92) and concentrations were normalized by
776 sample mass and converted to µmol/g using a standard curve.
777
778 **Gene expression measurements of the *arc* operon in germ-free mice**

The RNA extraction of the contents of each cecum was performed as described previously with some modifications[68]. Specifically, cecal samples (100~150 mg) were transferred into solvent-resistant screw-cap tubes (Sarstedt Inc) with 500 µL of acid washed beads (212-300 µm, Sigma). The samples were resuspended in 500 µL of Acid-Phenol:Chloroform (with IAA, 125:24:1, pH 4.5, Thermo Fisher Scientific), 500 µL of Buffer B (200mM NaCl (DOT Scientific), 20 mM EDTA (Sigma), pH 8.0), 210 µL 20% SDS (Alfa Aesar). Cells were lysed by mechanical disruption with a bead-beater (BioSpec Products) for 2 min twice to prevent overheating. Next, cells were centrifuged for 10 min at 8,000 x g at 4°C, and the supernatant was transferred into an Eppendof tube. We added 60 µL 3M sodium acetate (pH 5.5, Sigma) and 600 µL isopropanol (LabChem) to the supernatant and incubated at −80°C for 6 min. Next, samples were centrifuged for 15 min at 18,200 x g at 4°C. The harvested RNA pellets were washed once with 900 µL of 100% ethanol (Koptec). Finally, the RNA pellets were resuspended into 100 µL of RNase-free water. The crude RNA extracts were purified by a RNeasy Mini kit (Qiagen) with a DNase I (Qiagen) treatment step to eliminate DNA contamination in the sample. To remove any remaining DNA, the RNA products were treated with Baseline-ZERO DNase I (Epicentre) and then purified again using the RNeasy Mini kit.

We performed cDNA synthesis with 0.5-1 µg of total purified RNA using the iScript Select cDNA Synthesis Kit (Bio-Rad Laboratories). We performed quantitative reverse transcription PCR (qRT-PCR) on the Bio-Rad CFX connect Real-Time PCR instrument with SYBR™ Green PCR Master Mix (Thermo Fisher Scientific) using primers qPCR_arc_fwd and qPCR_arc_rev for arc expression measurement. We also included qPCR controls measuring the 16S rRNA level (primers: qPCR_EC16S_fwd and qPCR_EC16S_rev) and cysG gene expression level (primers: qPCR_ECcysG_fwd and qPCR_ECcysG_rev). We computed the fold changes of the target genes by normalizing to the two reference genes 16S rRNA gene and cysG[69] using the geometric mean[70]. We used three technical replicates per gene for qRT-PCR.

Bioinformatic analysis for *arc*-harboring species identification

The previous human microbiome metagenome-assembled genome (MAG) dataset containing 154,723 MAGs from 9,428 human gut microbiomes from diverse geographic location, body sites, diseases, and lifestyles was used as the reference database[71]. These genomes extend to > 95% the median mappability of gut microbiomes. All MAGs were annotated by Prodigal v2.6.3[72]. Only genomes with >= 95% completeness and < 10% contamination were used in the downstream analysis (45,502 MAGs). The KOfam HMMs[73] (KOfam 27-Aug-2021 release) for the four *arc* operon genes (K01478, arcA; K00611, arcB; K00926, arc; K03758, arcD) were used as the reference to search against all the MAGs by using hmmsearch implemented in HMMER v3.1b2[74]. The hmmsearch hits and genome taxonomy were summarized accordingly.

pH model and training

The pH model shown was adapted from the growth model with pH mediation developed by the Gore lab[33] **(equation 1-3)**.

$$\frac{\mathrm{d}D_i}{\mathrm{d}t} = -k_{dead,i}D_i \tag{2}$$

$$\frac{\mathrm{d}N_i}{\mathrm{d}t} = N_i\left(\mu_i - \frac{N_i}{K_i}\right)\left(e^{-\frac{(p-p_{pref,i})^2}{\sigma_i^2}} - \delta_i\right) + k_{dead,i}D_i \tag{3}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \sum_{i=species} \gamma_i N_i \tag{4}$$

Bacterial species density N_i for species i follows a logistic growth defined by maximum growth rate μ_i and carrying capacity K_i. A Gaussian pH modification is applied to the growth rate which is defined around preferred pH $p_{pref,i}$ with standard deviation σ_i and death rate δ_i. The species growth rate reaches maximum at a preferred pH and decreases outside a growth-promoting pH range. The environmental pH p is affected by the summation of each species by a pH-change-factor γ_i, where the factor can take values of both positive and negative depending on the species being an acidifier or basifier. To capture the initial lag phase of batch culture, D_i was introduced as the species density unavailable for replication. The unavailable density decreases with an exponential curve defined by $k_{dead,i}$, granting cells to replicate at the same rate in N_i.

Monoculture timeseries growth data and supernatant pH measurements at various initial pH were fit with the pH model for each species of interest **(equations 1-3)**. Parameters were estimated by stimulating growth curve and pH changes using Python ODE solver scipy.integrate.solve_ivp and minimizing the sum of squared error between model fitting and experimental data of growth curve and pH by Python optimization package scipy.optimize.least_squares.

The pH model fitted on monoculture growth was used to predict community dynamics. Parameters optimized for monoculture of each species of interest were used to simulate community growth curve and pH changes using Python ODE solver scipy.integrate.solve_ivp. Differences between model prediction and experimental result were evaluated by mean squared error of all species relative abundances. Shannon diversity index was calculated following equation 4 for both experimental data and model prediction at a given initial pH condition, where N_i represents bacterial density for species i and N_{tot} represents total bacterial density for all species.

$$Shannon\ Diversity\ Index = -\sum_{i=1}^{species} \frac{N_i}{N_{tot}} \ln \frac{N_i}{N_{tot}} \tag{5}$$

Data Availability

Additional data supporting the findings described in this work are available from the corresponding author under reasonable request. The processed data for all community experiments and simulation results will be available through Zenodo and Github prior to publication.

Code Availability

Scripts for model training and simulations will be available through Github prior to publication.

Acknowledgements

We would like to thank all current and past members of the Venturelli lab for their helpful discussions. We would like to thank Ryan Clark, Susan Hromada, and Bryce Connors for their advice related to anaerobic culturing. We would like to thank Jun Feng for their advice on RNA extraction and fecal sample processing techniques. Research was sponsored by the National Institutes of Health under Grant Number R35GM124774, R01EB030340 and Army Research Office W911NF-19-1-0269.

Author contributions

Y.L, Y.Y.C., and O.S.V conceived the research. Y.L., Y.Y.C., M.F.W. and E.I.V. carried out experiments. F.E.R. assisted in the design of the *in vivo* experiments. Y.L., Z.C.Z., K.A. and O.S.V. performed bioinformatics and data analysis. O.S.V. secured funding. Y.L. and O.S.V. wrote the manuscript and all authors provided feedback on the manuscript.

Ethics declarations

The authors declare no competing interests.

881 **REFERENCES**
882

883 1. Schirmer, M. *et al.* Linking the Human Gut Microbiome to Inflammatory Cytokine
884 Production Capacity. *Cell* **167**, 1125-1136.e8 (2016).
885 2. Jones, B. v., Begley, M., Hill, C., Gahan, C. G. M. & Marchesi, J. R. Functional and
886 comparative metagenomic analysis of bile salt hydrolase activity in the human gut
887 microbiome. *Proc Natl Acad Sci USA* **105**, 13580–13585 (2008).
888 3. Feng, J. *et al.* Polysaccharide utilization loci in Bacteroides determine population fitness
889 and community-level interactions. *Cell Host Microbe* **30**, 200-215.e12 (2022).
890 4. Wolfe, A. J. Glycolysis for Microbiome Generation. *Microbiol Spectr* **3**, (2015).
891 5. Lucas, L. N. *et al.* Dominant Bacterial Phyla from the Human Gut Show Widespread
892 Ability To Transform and Conjugate Bile Acids. *mSystems* **6**, (2021).
893 6. Rowland, I. *et al.* Gut microbiota functions: metabolism of nutrients and other food
894 components. *European Journal of Nutrition 2017 57:1* **57**, 1–24 (2017).
895 7. Visconti, A. *et al.* Interplay between the human gut microbiome and host metabolism.
896 *Nature Communications 2019 10:1* **10**, 1–10 (2019).
897 8. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of
898 colonic regulatory T cells. *Nature 2013 504:7480* **504**, 446–450 (2013).
899 9. Rivera-Chávez, F. *et al.* Depletion of Butyrate-Producing Clostridia from the Gut
900 Microbiota Drives an Aerobic Luminal Expansion of Salmonella. *Cell Host Microbe* **19**,
901 443–454 (2016).
902 10. Fung, K. Y. C., Cosgrove, L., Lockett, T., Head, R. & Topping, D. L. A review of the
903 potential mechanisms for the lowering of colorectal oncogenesis by butyrate. *British*
904 *Journal of Nutrition* **108**, 820–831 (2012).
905 11. Gilijamse, P. W. *et al.* Treatment with Anaerobutyricum soehngenii: a pilot study of safety
906 and dose–response effects on glucose metabolism in human subjects with metabolic
907 syndrome. *npj Biofilms and Microbiomes 2020 6:1* **6**, 1–10 (2020).
908 12. Geirnaert, A. *et al.* Butyrate-producing bacteria supplemented in vitro to Crohn's disease
909 patient microbiota increased butyrate production and enhanced intestinal epithelial barrier
910 integrity. *Sci Rep* **7**, (2017).
911 13. Zhang, H. *et al.* The cyanobacterial ornithine–ammonia cycle involves an arginine
912 dihydrolase. *Nature Chemical Biology 2018 14:6* **14**, 575–581 (2018).
913 14. Nascimento, M. M. *et al.* The effect of arginine on oral biofilm communities. *Mol Oral*
914 *Microbiol* **29**, 45–54 (2014).
915 15. Enowashu, E., Poll, C., Lamersdorf, N. & Kandeler, E. Microbial biomass and enzyme
916 activities under reduced nitrogen deposition in a spruce forest soil. *Applied Soil Ecology*
917 **43**, 11–21 (2009).
918 16. Billard-Pomares, T. *et al.* The arginine deiminase operon is responsible for a fitness
919 trade-off in extended-spectrum-β-lactamase-producing strains of Escherichia coli.
920 *Antimicrob Agents Chemother* **63**, (2019).
921 17. Marquis, R. E., Bender, G. R., Murray, D. R. & Wong, A. Arginine deiminase system and
922 bacterial adaptation to acid environments. *Appl Environ Microbiol* **53**, 198–200 (1987).
923 18. Wauven, C. v., Pierard, A., Kley-Raymann, M. & Haas, D. Pseudomonas aeruginosa
924 mutants affected in anaerobic growth on arginine: evidence for a four-gene cluster
925 encoding the arginine deiminase pathway. *J Bacteriol* **160**, 918–934 (1984).
926 19. Smith, A. B. *et al.* Enterococci enhance Clostridioides difficile pathogenesis. *Nature 2022*
927 *611:7937* **611**, 780–786 (2022).
928 20. Venturelli, O. S. *et al.* Deciphering microbial interactions in synthetic human gut
929 microbiome communities. *Mol Syst Biol* **14**, e8157 (2018).
930 21. Reese, A. T. *et al.* Microbial nitrogen limitation in the mammalian large intestine. *Nature*
931 *Microbiology 2018 3:12* **3**, 1441–1450 (2018).

932 22. Jenni Firrman, LinShu Liu, Karley Mahalak, Ceylan Tanes, Kyle Bittinger, Vincent Tu,
933 Jamshed Bobokalonov, Lisa Mattei, Huanjia Zhang, P. V. den A. The impact of
934 environmental pH on the gut microbiota community structure and short chain fatty acid
935 production. *FEMS Microbiol Ecol* 1–11 (2016)
936 23. Wang, S. P. *et al.* Pivotal Roles for pH, Lactate, and Lactate-Utilizing Bacteria in the
937 Stability of a Human Colonic Microbial Ecosystem. *mSystems* **5**, (2020).
938 24. Belenguer, A. *et al.* Impact of pH on lactate formation and utilization by human fecal
939 microbial communities. *Appl Environ Microbiol* **73**, 6526–6533 (2007).
940 25. Ilhan, Z. E., Marcus, A. K., Kang, D.-W., Rittmann, B. E. & Krajmalnik-Brown, R. pH-
941 Mediated Microbial and Metabolic Interactions in Fecal Enrichment Cultures. *mSphere* **2**,
942 (2017).
943 26. Nugent, S. G., Kumar, D., Rampton, D. S. & Evans, D. F. Intestinal luminal pH in
944 inflammatory bowel disease: possible determinants and implications for therapy with
945 aminosalicylates and other drugs. *Gut* **48**, 571–577 (2001).
946 27. Sofi, M. H. *et al.* PH of drinking water influences the composition of gut microbiome and
947 type 1 diabetes incidence. *Diabetes* **63**, 632–644 (2014).
948 28. Leggieri, P. A. *et al.* Integrating Systems and Synthetic Biology to Understand and
949 Engineer Microbiomes. *Annu Rev Biomed Eng* **23**, 169–201 (2021).
950 29. Qian, Y., Lan, F. & Venturelli, O. S. Towards a deeper understanding of microbial
951 communities: integrating experimental data with dynamic models. *Curr Opin Microbiol* **62**,
952 84–92 (2021).
953 30. Clark, R. L. *et al.* Design of synthetic human gut microbiome assembly and butyrate
954 production. *Nature Communications 2021 12:1* **12**, 1–16 (2021).
955 31. Hromada, S. *et al.* Negative interactions determine Clostridioides difficile growth in
956 synthetic human gut communities. *Mol Syst Biol* **17**, e10355 (2021).
957 32. Ostrem Loss, E., Thompson, J., Cheung, P. L. K., Qian, Y. & Venturelli, O. S.
958 Carbohydrate complexity limits microbial growth and reduces the sensitivity of human gut
959 communities to perturbations. *Nature Ecology & Evolution 2023 7:1* **7**, 127–142 (2023).
960 33. Ratzke, C. & Gore, J. Modifying and reacting to the environmental pH can drive bacterial
961 interactions. *PLoS Biol* **16**, e2004248 (2018).
962 34. Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over
963 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*
964 **176**, 649-662.e20 (2019).
965 35. Altenhoefer, A. *et al.* The probiotic Escherichia coli strain Nissle 1917 interferes with
966 invasion of human intestinal epithelial cells by different enteroinvasive bacterial
967 pathogens. *FEMS Immunol Med Microbiol* **40**, 223–229 (2004).
968 36. Ma, G. *et al.* In vivo fermentation of a Pleurotus eryngii polysaccharide and its effects on
969 fecal microbiota composition and immune response. *Food Funct* **8**, 1810–1821 (2017).
970 37. Shepherd, E. S., Deloache, W. C., Pruss, K. M., Whitaker, W. R. & Sonnenburg, J. L. An
971 exclusive metabolic niche enables strain engraftment in the gut microbiota. *Nature* **557**,
972 434–438 (2018).
973 38. Sheth, R. U., Cabral, V., Chen, S. P., Wang, H. H. & Wang, H. H. Manipulating Bacterial
974 Communities by in situ Microbiome Engineering. *Trends in Genetics* **32**, (2016).
975 39. Venturelli, O. S. *et al.* Deciphering microbial interactions in synthetic human gut
976 microbiome communities. *Mol Syst Biol* **14**, (2018).
977 40. Presser, K. A., Ratkowsky, D. A. & Ross, T. Modelling the growth rate of Escherichia coli
978 as a function of pH and lactic acid concentration. *Appl Environ Microbiol* **63**, 2355–2360
979 (1997).
980 41. Makhlin, J. *et al.* Staphylococcus aureus ArcR controls expression of the arginine
981 deiminase operon. *J Bacteriol* **189**, 5976–5986 (2007).

42. Fallingborg, J. Intraluminal pH of the human gastrointestinal tract. *Dan Med Bull* **46**, 183–196 (1999).
43. Flint, H. J., Duncan, S. H., Scott, K. P. & Louis, P. Links between diet, gut microbiota composition and gut metabolism. *Proceedings of the Nutrition Society* **74**, 13–22 (2015).
44. Baranwal, M. *et al.* Recurrent neural networks enable design of multifunctional synthetic human gut microbiome dynamics. *Elife* **11**, (2022).
45. Enjalbert, B., Cocaign-Bousquet, M., Portais, J. C. & Letisse, F. Acetate exposure determines the diauxic behavior of Escherichia coli during the glucose-acetate transition. *J Bacteriol* **197**, 3173–3181 (2015).
46. Hu, J. L., Nie, S. P., Min, F. F. & Xie, M. Y. Polysaccharide from seeds of plantago asiatica L. increases short-chain fatty acid production and fecal moisture along with lowering pH in mouse colon. *J Agric Food Chem* **60**, 11525–11532 (2012).
47. Zhao, R., Cheng, N., Nakata, P. A., Zhao, L. & Hu, Q. Consumption of polysaccharides from Auricularia auricular modulates the intestinal microbiota in mice. *Food Research International* **123**, 383–392 (2019).
48. Nugent, S. G., Kumar, D., Rampton, D. S. & Evans, D. F. Intestinal luminal pH in inflammatory bowel disease: possible determinants and implications for therapy with aminosalicylates and other drugs. *Gut* **48**, 571–577 (2001).
49. Rivera-Chávez, F. *et al.* Depletion of Butyrate-Producing Clostridia from the Gut Microbiota Drives an Aerobic Luminal Expansion of Salmonella. *Cell Host Microbe* **19**, 443–454 (2016).
50. Furusawa, Y. *et al.* Commensal microbe-derived butyrate induces the differentiation of colonic regulatory T cells. *Nature 2013 504:7480* **504**, 446–450 (2013).
51. Lee, K. K., Park, Y. & Kuehn, S. Robustness of microbiome function. *Curr Opin Syst Biol* **36**, 100479 (2023).
52. Raman, A. S. *et al.* A sparse covarying unit that describes healthy and impaired human gut microbiota development. *Science* **365**, (2019).
53. Cheng, A. G. *et al.* Design, construction, and in vivo augmentation of a complex gut microbiome. *Cell* **185**, 3617-3636.e19 (2022).
54. Lawson, C. E. *et al.* Common principles and best practices for engineering microbiomes. *Nature Reviews Microbiology* **17**, 725–741 (2019)
55. Wexler, A. G. *et al.* Human gut bacteroides capture vitamin B12 via cell surface-exposed lipoproteins. *Elife* **7**, (2018).
56. Zhu, W. *et al.* Xenosiderophore Utilization Promotes Bacteroides thetaiotaomicron Resilience during Colitis Article Xenosiderophore Utilization Promotes Bacteroides thetaiotaomicron Resilience during Colitis. *Cell Host Microbe* **27**, 376-388.e8 (2020).
57. Reese, A. T. *et al.* Microbial nitrogen limitation in the mammalian large intestine. *Nature Microbiology* **3**, 1441–1450 (2018).
58. Hu, S., Chapin, F. S., Firestone, M. K., Field, C. B. & Chiariello, N. R. Nitrogen limitation of microbial decomposition in a grassland under elevated CO2. *Nature 2001 409:6817* **409**, 188–191 (2001).
59. Liu, H. *et al.* Functional genetics of human gut commensal Bacteroides thetaiotaomicron reveals metabolic requirements for growth across environments. *Cell Rep* **34**, 108789 (2021).
60. Huang, Y. Y. *et al.* Functional screens of barcoded expression libraries uncover new gene functions in carbon utilization among gut Bacteroidales. *bioRxiv* 2022.10.10.511384 (2022)
61. Lan, F. *et al.* Massively parallel single-cell sequencing of genetic loci in diverse microbial populations. *bioRxiv* 2022.11.21.517444 (2022)

1031 62. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in
1032 Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of*
1033 *Sciences* **97**, 6640–6645 (2000).
1034 63. Mehrer, C. R., Incha, M. R., Politz, M. C. & Pfleger, B. F. Anaerobic production of
1035 medium-chain fatty alcohols via a β-reduction pathway. *Metab Eng* **48**, 63–71 (2018).
1036 64. Hoang, T. T., Karkhoff-Schweizer, R. R., Kutchma, A. J. & Schweizer, H. P. A broad-host-
1037 range Flp-FRT recombination system for site-specific excision of chromosomally-located
1038 DNA sequences: application for isolation of unmarked Pseudomonas aeruginosa
1039 mutants. *Gene* **212**, 77–86 (1998).
1040 65. Lee, T. S. *et al.* BglBrick vectors and datasheets: A synthetic biology platform for gene
1041 expression. *Journal of Biological Engineering 2011 5:1* **5**, 1–14 (2011).
1042 66. Goodman, A. L. *et al.* Extensive personal human gut microbiota culture collections
1043 characterized and manipulated in gnotobiotic mice. *Proc Natl Acad Sci USA* **108**, 6252–
1044 6257 (2011).
1045 67. Hutchison, E. R. *et al.* Dissecting the impact of dietary fiber type on atherosclerosis in
1046 mice colonized with different gut microbial communities. *npj Biofilms and Microbiomes* **9**,
1047 1–12 (2023).
1048 68. Rey, F. E. *et al.* Dissecting the in Vivo Metabolic Potential of Two Human Gut Acetogens.
1049 *J Biol Chem* **285**, 22082 (2010).
1050 69. Zhou, K. *et al.* Novel reference genes for quantifying transcriptional responses of
1051 Escherichia coli to protein overexpression by quantitative PCR. *BMC Mol Biol* **12**, 1–9
1052 (2011).
1053 70. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by
1054 geometric averaging of multiple internal control genes. *Genome Biol* **3**, 1–12 (2002).
1055 71. Pasolli, E. *et al.* Extensive Unexplored Human Microbiome Diversity Revealed by Over
1056 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. *Cell*
1057 **176**, 649-662.e20 (2019).
1058 72. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site
1059 identification. *BMC Bioinformatics* **11**, 119 (2010).
1060 73. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and
1061 adaptive score threshold. *Bioinformatics* **36**, 2251 (2020).
1062 74. Finn, R. D. *et al.* HMMER web server: 2015 update. *Nucleic Acids Res* **43**, W30 (2015).
1063