**SUPPORTING INFORMATION**

**Integrated metabolic modeling, culturing and transcriptomics identify variations in *V. cholerae* essential genome**

Alyaa M. Abdel-Haleem1,2, Vaishnavi Ravikumar3, Boyang Ji4, Katsuhiko Mineta1, Xin Gao1,Jens Nielsen3,4, Takashi Gojobori\*1,2, and Ivan Mijakovic\*§3,4

1 King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Centre (CBRC), Thuwal, 23955-6900, Saudi Arabia

2 King Abdullah University of Science and Technology (KAUST), Biological and Environmental Sciences and Engineering (BESE) division, Thuwal, 23955-6900, Saudi Arabia

3 Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby, Denmark

4 Systems and Synthetic Biology, Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden

\* Corresponding authors: [takashi.gojobori@kaust.edu.sa](mailto:takashi.gojobori@kaust.edu.sa) & [ivan.mijakovic@chalmers.se](mailto:ivan.mijakovic@chalmers.se)

§ Lead contact: [ivan.mijakovic@chalmers.se](mailto:ivan.mijakovic@chalmers.se)

**Table of Contents**

[Bacterial strains 2](#_Toc12888200)

[Growth Curve and CFU Measurements 2](#_Toc12888201)

[DNA extraction sequencing 2](#_Toc12888202)

[RNA extraction and sequencing 2](#_Toc12888203)

[Genome sequencing, assembly and annotation 2](#_Toc12888204)

[Assessment of the metabolic component of *V. cholerae* essential genes 4](#_Toc12888205)

[Reconstruction of *i*AM-Vc960 network 5](#_Toc12888206)

[Refinement of *i*AM-Vc960 6](#_Toc12888207)

[Validation of *i*AM-Vc960 singe gene essentiality predictions 7](#_Toc12888208)

[Comparison to previous *V. cholerae* genome-scale metabolic model 7](#_Toc12888209)

[Refinement of *i*ETEC-1333 8](#_Toc12888210)

[*In silico* simulations of the growth of *i*ETEC1333 and *i*AM-Vc960 in a shared environment, *i*Vc-ETEC-2993 8](#_Toc12888211)

[Quality control for the co-culture model *i*Vc-ETEC-2993 9](#_Toc12888212)

[Prediction of all growth-supporting carbon, nitrogen, phosphorus, and sulfur sources 9](#_Toc12888213)

[Dual RNAseq data analysis 12](#_Toc12888214)

[Single gene essentiality predictions in mono- *vs.* co-culture models 15](#_Toc12888215)

[Variation in essential genome *vs.* species abundance 15](#_Toc12888216)

[Predicted variation in essentiality between mono- and co-cultures 16](#_Toc12888217)

[Bacteriocins result in difference in response between ETEC strains to co-growth with *V. cholerae* 17](#_Toc12888218)

[*In silico* simulations, data and software Availability 17](#_Toc12888219)

[Spent media experiments 18](#_Toc12888220)

## Bacterial strains

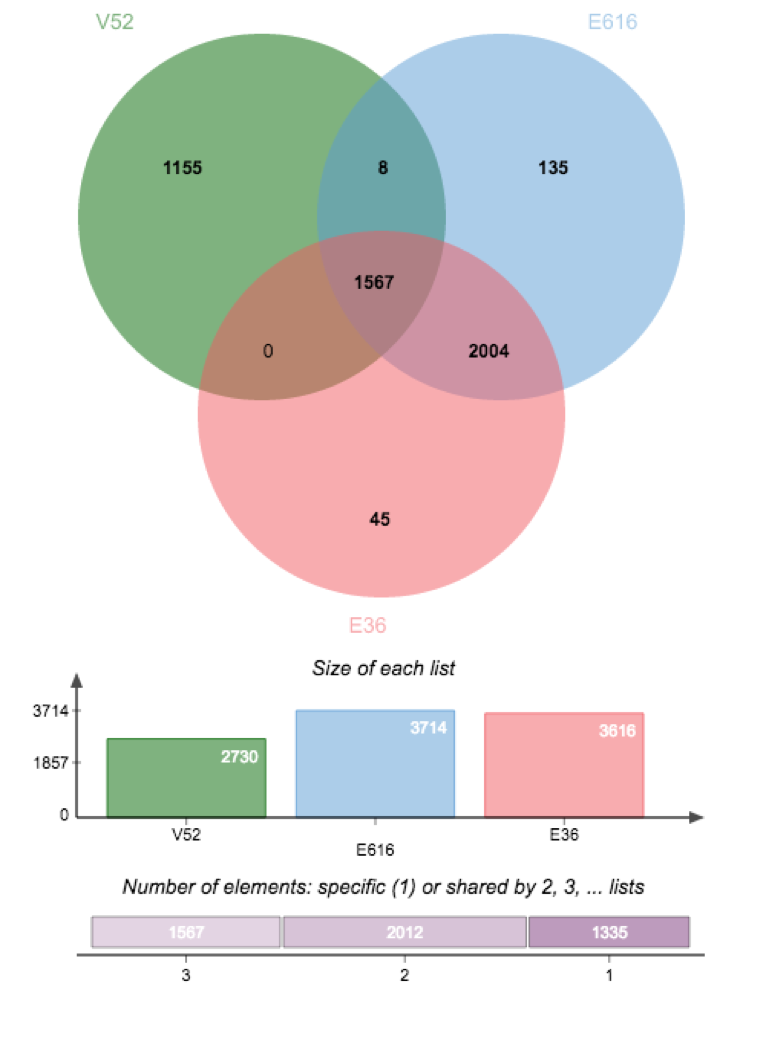
Bacterial strains were grown in M9 (Sigma Aldrich) minimal medium supplemented with 0.5 % glucose, 1 mM magnesium sulphate and 0.1 mM calcium chloride, unless otherwise specified. *V. cholerae* V52 (O37 serogroup) and the Enterotoxigenic *Escherichia coli* strains (ETEC E616 and ETEC E36) were a kind gift from Prof. Sun Nyunt Wai, Umeå University, Sweden.

## Genome sequencing, assembly and annotation

Genome sequences were assembled using SPAdes1 for *V. cholerae* V52 and SPades and plasmidSPAdes2 for ETEC E616 and ETEC E36. PATRIC3 eggNOG mapper4 were used for genome annotation. Table S13 shows the details for the genome assembly and genome annotation for each of the three bacterial strains used in this study. The genome sequence for *V. cholerae* V52 was assembled using the reference *V. cholerae* O1 biovar El Tor str. N16991 as reference, since this is the closest complete genome sequence available for *V. cholerae*. As such, our reconstruction effort here is not directed towards building a serotype-specific GEM but rather a GEM that accounts for all metabolic content of *V. cholerae* species regardless of the specific serotype. Future efforts will focus on delineating serotype differences and their functional linkage to the metabolic network like what was done for the diverse *Salmonella* serovars5.

**Table S13| Details for the genome assembly and annotation of *V. cholerae* V52, ETEC E36 and E616**.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *V. cholerae* V52 | ETEC E616 | ETEC 36 |
| PATRIC genome ID | 243277.158 | 316401.73 | 316401.72 |
| Contigs | 178 | 610 | 444 |
| Genome Length | 3941586 | 5536969 | 5403728 |
| GC Content | 47.56 | 50.37 | 50.51 |
| PATRIC CDS | 3674 | 6052 | 5793 |
| # unique specialty genes | 287 | 809 | 804 |
| # expressed genes in M9 + Glucose | 3227 | 4807 | 4615 |



**Figure S1| Number of common and unique expressed CDS in each of the bacterial pathogens sequenced in this study.** FIGfam ID as generated by PATRIC3 genome annotation service was used for generating the numbers displayed in the Venn diagram generated by jvenn6.

## 

**Figure S2| Size of V. cholerae GEM relative to GEMs of other enteric pathogens**. Proportion of metabolic genes included as GPR in GEMs of E. coli, Shigella (PMID: 24277855) and V. cholerae (this study) relative to total number of ORFs in each species. B) Comparison of iAM-Vc960 model size relative to previously published 55 E. coli and Shigella genome-scale metabolic models.

## Assessment of the metabolic component of *V. cholerae* essential genes

We had two sets of genes which we looked at to assess the proportion of metabolic genes among them. The first set is the list of specialty genes as annotated by PATRIC3 for *V. cholerae* V52. In this set, *V. cholerae* had n = 287 unique specialty genes, 102 of those (36%) were involved in metabolic processes. The majority of the metabolic specialty genes were virulence factors (n = 34), while the rest were drug targets (n = 10), human homologs (n = 12) and involved in antibiotic resistance (n = 1). The second set was generated by taking the overlap of genes that are essential for *V. cholerae* growth and survival as tested by three independent studies7-9. In this set, *V. cholerae* essential genes were classified into metabolic and non-metabolic, and a large fraction of the essential genome consisted of metabolic functions (36%) (Table S2).

## Reconstruction of *i*AM-Vc960 network

The starting model was *i*JO136610, translated by orthology to *V. cholerae* using PATRIC3 genome annotation of the *de novo* genome assembly of *V. cholerae* V52 strain,  expression evidence from the transcriptomic data of *V. cholerae* V52 monocultures (Table S9-S10) as well as Uniprot ‘[*Vibrio cholerae* serotype O1 (strain ATCC 39315 / El Tor Inaba N16961)](http://www.uniprot.org/taxonomy/243277)’. We realize that *V. cholerae* V52 and *V. cholerae* N16961 belong to different serotype groups but we make no distinction in relation to serotype groups in this study. We have used *V. cholerae* O1 El Tor N16961 genome assembly as it is the most complete and update genome annotation *for V. cholerae*. We have also sequenced and annotated the genome of *V. cholerae* V52 since it is more frequently co-isolated with ETEC in clinical settings11. Future studies will focus on assessing the linkage between metabolic capabilities and serotype groups as well as across the entire species.

For the gene-protein-reaction (GPR) building, 28 genes did not have *V. cholerae* specific gene ID. If the rule was ‘AND’ and not all orthologous were found in *V. cholerae* genome or transcriptome data generated in this study or in Uniprot, we kept the gene symbol to indicate that this specific gene has not been annotated in *V. cholerae* but it’s potentially needed for the corresponding reaction. Also, for this set of reactions, at least one of the genes was expressed according to the transcriptomic data generated in this study. For example, AACPS1 is associated to (*aas* and VC2020), there is [proteomic evidence](https://www.uniprot.org/uniprot/Q9KQH8) for VC2020 in *V. cholerae* according to Uniprot and it is also expressed in *V52* (Table S9-S10), however there is no evidence for *aas* expression and it does not have a designated VC- or VCA- formatted ID in Uniprot, so the gene symbol was used.

The objective function for *i*AM-Vc960 was derived from that of *i*JO136610 and *V. vulnificus*12. *i*AM-Vc960 was assessed for mass balance13. Metabolites charges and formulae were obtained from BiGG14 and updated in *i*AM-Vc960 to mass-balance the respective reactions.

## Refinement of *i*AM-Vc960

If the respective GPR was annotated in *V. cholerae* genome, the reaction was kept but the gene symbol instead of *V. cholerae* specific gene ID (if not assigned in the standard format of VC- or VCA- in Uniprot) was used. aquaporinZ was not annotated in *V. cholerae* but it has been reported to be present in *V. cholerae*15. Glycolate dehydrogenase (EC 1.1.99.14), subunits *glcD*/*E*/*F* are not annotated in *V. cholerae* genome. Hence, reactions GLYCTO2, GLYCTO3, GLYCTO4 were removed. A sink reaction was nevertheless added for cytosolic glycolate since no literature evidence supported the presence of glycolate dehydrogenase in *V. cholerae* genome, however *glcE* was predicted to be essential for optimal growth. Similarly, Outer membrane porin N (*ompN*) was not annotated in *V. cholerae* genome in the data generated in this study but it was associated to several transporter reactions that are needed for model consistency, hence it was retained in the model under the gene symbol (*ompN*) to mark it from *V. cholerae* annotated genes that has VC- or VCA- prefixes. Although, [*ompN*](https://www.uniprot.org/uniprot/A0A0H6LT34) is predicted to be present in *V. cholerae* at the protein level according to Uniprot but there was no experimental evidence to support its presence. Glutamate decarboxylase *panD*, associated with reaction [ASP1DC](http://bigg.ucsd.edu/models/iJO1366/genes/b0131) according to *i*JO136610, was changed to [*panP*](https://www.uniprot.org/uniprot/A0A0E4CEI8) according to gene name annotation in *V. cholerae*.

We have performed a single-reaction deletion analysis for every reaction in the single species and dual-species model. *i*AM-Vc960 has a total of 11 demand and sink reactions. 2 demand reactions (DM\_4CRSOL, DM\_MTHTHF) and 2 sink reactions (sink\_glu1sa[c], sink\_glyclt[c]) are necessary for the model to function, the remaining 7 reactions do not affect the accuracy of the growth rate or the model function.

DM\_MTHTHF: as per iJO1366; needed to allow (2R,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran to leave system. We have made it irreversible to allow leaving only and not uptake)

DM\_4CRSOL: as per iJO1366, needed to allow p-Cresol to leave the system. We have made it irreversible to allow leaving system only and not uptake.

sink\_glyclt[c]: Glycolate dehydrogenase (EC 1.1.99.14), subunits *glcD*/*E*/*F* are not annotated in the *V. cholerae* genome. Hence, reactions GLYCTO2, GLYCTO3, GLYCTO4 were removed. A sink reaction was nevertheless added for cytosolic glycolate, since no literature evidence supported the presence of a glycolate dehydrogenase-encoding gene in the *V. cholerae* genome. However, *glcE* was predicted to be essential for optimal growth.

sink\_glu1sa[c]: This sink reaction is added to supply glu1sa[c] for downstream reactions instead of being supplied by glutamyl tRNA reductase reaction in iJO1366. We removed all tRNA reactions since they are beyond the scope of direct metabolic reactions modeled in this study.

## Validation of *i*AM-Vc960 singe gene essentiality predictions

We downloaded gene essentiality data for *V. cholerae O1* str. C6706 from the Online [GEne Essentiality (OGEE)](http://ogee.medgenius.info/browse/)16,17 database. In total, 3886 genes (total number of ORFs identified in *V. cholerae*) were tested for essentiality. 458 genes were essential, 148 were essential for fitness, 3144 were non-essential and 136 were unknown. Out of the 458 essential genes, 145 were metabolic genes and were already in *i*AM-Vc960. *i*AM-Vc960 predicted 94 of those to be essential while the remaining 51 were falsely predicted by the model as non-essential. For the non-essential genes, 758 of those were already in *i*AM-Vc960. The model could predict 693 as non-essential while 65 were falsely predicted by the model as essential. The overall accuracy of the model predicted single gene essentiality was 87% (Figure 2C). This discrepancy between the model predictions and the OGEE dataset, the high confidence set that we used earlier and assuming low experimental error rate, indicates that the reconstructed *V. cholerae* reactome is incomplete and that there is further room for improvement and refinement of the *i*AM-Vc960 representing opportunities for new biological discoveries.

## Comparison to previous *V. cholerae* genome-scale metabolic model

In this study, we report on a manually-curated high quality simulable metabolic model for *V. cholerae*. A previous study has published a *V. cholerae* metabolic model that is automatically reconstructed from [path2models](https://www.ebi.ac.uk/biomodels-main/BMID000000140885). *i*AM-Vc960 exceeds the previously published *V. cholerae* model which has 658 genes, 1007 metabolites and 2084 reactions. 584 (89%) of the path2models *V. cholerae* model were already in *i*AM-Vc960. 6 metabolic genes were added to a draft version of *i*AM-Vc960. The remaining 68 genes were non-metabolic. Moreover, the previously published model has a generic objective function where all biomass coefficients are 1. We could not perform a functional comparison between *i*AM-Vc960 and the previously published *V.cholerae* model since the later was unable to produce any biomass as downloaded from the biomodels repository.

Two other automatically reconstructed and curated *V. cholerae* models were also generated as part of ModelSEED18. These models were further curated and evaluated against gene essentiality dataset8 that was already accounted for as one of the three studies which we used to compile our high confidence list of essential genes for *V. cholerae* growth. We compared the performance of *i*AM-Vc960 against the same dataset the authors have used for evaluating their models. In total, the dataset from Cameron et al.,8 was composed of 789 ORFs (metabolic and non-metabolic) potentially essential genes for *V. cholerae* growth in nutrient-rich conditions, i.e. these ORFs were not disrupted during transposon insertion library productions. 78 of those were essential in both *E. coli* and *V. cholerae* under similar growth conditions. *i*AM-Vc960 covered 18% of the potentially essential genes (that is n = 145 out of 789 ORFs). Of those, *i*AM-Vc960 predicted 73 to be essential (50%). Notably, *i*AM-Vc960 predicted all of the 75% (n = 58) of the genes that are essential in both *E. coli* and *V. cholerae*. Henry et al.,18 reported an accuracy of 75% for their *V. cholerae* model when tested against gene essentiality data. We were not able to retrieve details or parameters of their models’ gene essentiality predictions as to whether this reflects evaluation against the whole set of essential genes or only those essential in both *E. coli* and *V. cholerae* making the comparison presented here of limited use.

## Refinement of *i*ETEC-1333

Originally, *i*ETEC1333 had 2756 reactions, 1962 metabolites and 1333 genes. For consistency, we removed tRNA-related reactions, metabolites and associated genes from *i*ETEC1333. In addition, two reactions were duplicated in *i*ETEC\_1333: DADK and GTHPi. We also retained a single reaction accounting for core *E. coli* biomass ‘BIOMASS\_Ec\_iJO1366\_core\_53p95M’ and removed ‘BIOMASS\_Ec\_iJO1366\_WT\_53p95M’. Metabolite formulae for 24 metabolites were updated through BiGG to mass and charge balance the corresponding reactions.

In our hands, using *i*ETEC-1333 as downloaded from BiGG14, it could not grow on 3-Phospho-D-glycerate in aerobic conditions and couldn’t grow on adenine, ethanol or hypoxanthine in anaerobic conditions. Also, it could not grow on arginine, glycine, ethanolamine, lactate and malate in anaerobic conditions except if the uptake of the respective carbon source is increased to -1000 mmol/gDW (cf. -20 mmol/gDW which is the default used for testing). However, increasing uptake for the same carbon sources for *i*AM-Vc960 or the co-infection model did not enable their growth under similar conditions.

## *In silico* simulations of the growth of *i*ETEC1333 and *i*AM-Vc960 in a shared environment, *i*Vc-ETEC\_2293

This analysis treated a multi-species community analogously to multi-compartment metabolic models of prokaryotes as previously reported for several other microbial community models19-24. In the individual species models, multiple organelles are modeled by defining suitably labeled compartment-specific metabolites and reactions, and adding transport reactions across compartments, as dictated by the knowledge of diffusion or transporters. In a community-level model, members can be similarly treated as compartments embedded in a meta-compartment that represents the shared environment. Formally, the stoichiometric matrices of individual species can be combined with each other in a larger block matrix, to construct a community-level stoichiometric matrix. One subtle aspect of implementing FBA simulations for a microbial community based on this compartment-based stoichiometry is the identification of an appropriate objective function25. Here, the widely-employed FBA objective of biomass maximization was replaced with the maximization of a weighted sum of the biomass production fluxes for the community members26, i.e. the objective function was set to maximize the biomass function of each pathogen, simulating growth at 1:1 species composition/abundance.

## Quality control for the co-culture model *i*Vc-ETEC\_2293

Quality checks for the co-culture model showed that hydrogen uptake in absence of any carbon source was generating non-zero biomass. A deeper investigation showed that shutting the upper and lower bounds of 6 related reactions [vch\_VPAMTr, etec\_PYRt2rpp, etec\_INSt2pp\_copy2, etec\_ADNt2pp\_copy2] blocked the production of any biomass under these conditions. Similarly, no flux was allowed in VPAMTr in *i*AM-Vc960 while VALTA and ALATA\_L were made irreversible.

Testing for growth using alternate C/N/S/P sources showed that the merged model was unable to produce biomass in the combined objective function unless at least one of the individual models (prior to merging) was able to grow in the respective growth condition. In case of growth on urea as an alternative nitrogen source in anaerobic conditions, none of the individual models was able to grow in these conditions, however the merged model was able to produce biomass. On further curation, vch\_NADHNQR was found to cause the leak, so it was made irreversible in the merged model.

We tested 361 aerobic and 295 anaerobic conditions. All three models (*i*AM-Vc960, *i*ETEC1333 and *i*Vc-ETEC\_2293) were able to grow on 190 aerobic conditions (53% of the total number of tested aerobic conditions). ETEC was able to catabolize 150 nutrient sources aerobically that *V. cholerae* could not catabolize on its own under the same growth conditions. The co-culture model acquired the capability to catabolize these nutrients accordingly. In contrast, utilization of sucrose as sole carbon source aerobically was probably contributed to the co-culture model by *V. cholerae* (i.e. ETEC was not able to utilize sucrose under the same conditions). Similarly, all three models were able to grow on 143 anaerobic conditions (48%). ETEC was able to anaerobically utilize 127 nutrients that *V. cholerae* could not metabolize under the same conditions. *V. cholerae* was also able to utilize sucrose and octadecenoate C18:1 in absence of oxygen while ETEC did not have the capacity to do so.

## Prediction of all growth-supporting carbon, nitrogen, phosphorus, and sulfur sources

The possible growth-supporting carbon, nitrogen, phosphorus, and sulfur sources of V. cholerae were identified using FBA as previously detailed27. Briefly, to determine possible growth-supporting carbon sources, the lower bound of the glucose exchange reaction was constrained to zero. Then the lower bound of each carbon exchange reaction was set, one at a time, to −20 mmol gDW−1 h−1 (Table S4). Following, growth was maximized by FBA using the model-specific biomass reaction. The target substrate was considered growth supporting if the predicted growth rate was above 0.01. While identifying carbon sources, the default nitrogen, phosphorus, and sulfur sources were ammonium (nh4), inorganic phosphate (pi), and inorganic sulfate (so4), respectively. Prediction of growth-supporting sources of these other three elements was performed in the same manner as growth on carbon, with glucose as the default carbon source.

## Growth Curve and CFU Measurements

*V. cholerae* and ETEC were grown either individually (mono-cultures – V52, E616 and E36) or in combination (co-cultures – V52/E616 and V52/E36) at 37 °C at 200 rpm. Co-cultures were started with equal concentrations of each strain. The absorbance (OD600) was measured every 1h over a period of 10h for the growth curve measurements. Simultaneously, at every hour, an aliquot was taken from each culture flask, serially diluted and 5 µL were spotted (three technical replicates) on agar plates containing appropriate antibiotics (100 µg/mL of rifampicin or 15 µg/mL of tetracycline). *V. cholerae* is resistant to rifampicin and ETEC is resistant to tetracycline. V52 mono-cultures were spotted on rifampicin plates whereas ETEC E616 and E36 mono-cultures were spotted on tetracycline plates. All co-cultures were spotted on both sets of antibiotic plates. All plates were incubated for a period of 12-16h at 37 °C after which the colonies were counted and the cfu/mL was calculated based on the formula:

cfu/mL = (number of colonies/volume spotted)/dilution

To ensure biological reproducibility, experiments were performed in triplicates.

## Figure S3| CFU counting for co-cultures *vs*. mono-cultures. Investigating the impact of the co-culture on each strain’s growth by comparing single culture abundance after 10 hours of growth to the abundance of each strain in co-culture at the same time (determined using CFU counting; all strains were in transition or stationary phase)

## DNA extraction sequencing

Genomic DNA and plasmids (in case of ETEC) were extracted from bacterial cells for the purpose of whole genome sequencing. *V. cholerae* and ETEC cells (mono-cultures) were inoculated in rich LB (Sigma Aldrich) medium and grown at 37 °C at 200 rpm until stationary phase. Subsequently cells were harvested and lysed and the genomic DNA was extracted using the DNeasy® Blood & Tissue Kit (Qiagen), according to manufacturer’s instructions. Plasmid DNA from both the ETEC strains was additionally isolated using the Gene Jet Plasmid Miniprep Kit (Thermo Scientific) by following the manufacturer’s instructions.

## RNA extraction and sequencing

Sampling of cells for the purpose of RNA extraction was performed as follows: Bacterial cells (mono-cultures and co-cultures of *V. cholerae* and ETEC) were grown to mid logarithmic phase in shake flasks at 37 °C at 200 rpm. In case of the co-cultures, equal concentrations of individual mono-cultures were inoculated into the same medium from the start. Once the appropriate growth phase was reached, the cells were harvested. RNA was extracted from the harvested cells using the RNeasy®Mini Kit (Qiagen), according to manufacturer’s instructions. Experiments were carried out in triplicates. The RNA extracted was in the range of 200 – 100 ng/µL.

The RNA integrity (RNA Integrity Score ≥ 8) and quantity was determined on the Agilent 2100 Bioanalyzer (Agilent; Palo Alto, CA, USA). As ribosomal RNA comprises the vast majority of the extracted RNA population, depletion of these molecules through RiboZero-based rRNA depletion was used in efforts to increase the coverage of mRNA and to reduce rRNA reads. For this mRNA enrichment, the Illumina's RiboZeroTMProkaryotic kit was used according to manufacturer's instructions. Briefly, 2µg of total RNA samples was hybridized with prokaryotic rRNA sequence-specific 5′-biotin labeled oligonucleotide probes to selectively deplete large rRNA molecules from total RNA. Then, these rRNA-hybridized, biotinylated probes were removed from the sample with streptavidin-coated magnetic beads. The final RiboZeroTMRNA sample was subjected to thermal mRNA fragmentation using Elute, Prime, and fragment Mix from the Illumina TruSeq stranded total RNA sample preparation kit (Low-Throughput protocol). The fragmented mRNA samples were subjected to cDNA synthesis using the Illumina TruSeq stranded total RNA sample preparation kit (Low-Throughput protocol) according to manufacturer's protocol. Briefly, cDNA was synthesized from enriched and fragmented RNA using SuperScript III Reverse Transcriptase (Invitrogen) and SRA RT primer (Illumina). The cDNA was further converted into double stranded DNA using the reagents supplied in the kit, and the resulting dsDNA was used for library preparation. To this end, cDNA fragments were end-repaired and phosphorylated, followed by adenylation of 3′ends and adapter ligation. Twelve cycles of PCR amplification were then performed, and the library was finally purified with AMPure beads (Beckman Coulter) as per the manufacturer’s instructions.

The resulting libraries were quality-controlled on a Bioanalyser DNA1000 chip (Agilent). The bar-coded cDNA libraries were pooled together in equal concentrations in one pool before sequencing on Illumina HiSeq4000 using the TruSeq PE Cluster Generation Kit and TruSeq SBS Kit. Data were processed with the Illumina Pipeline Software v2.1.

## Dual RNAseq data analysis

RNAseq reads from mono-cultures were directly aligned to the genome assembly of the corresponding species. To check for reads cross mapping, we first attempted to map *V. cholerae* reads against ETEC genome assembly and vice versa. In either case, the percentage of mapped reads was < 2% (Figure S2) indicating minimal cross-mapping between the two species. Following, we constructed an artificial genome assembly of both *V. cholerae* and ETEC combined, i.e. representing the co-culture as a single entity by merging the genome assemblies of the two species. PATRIC3 was used for annotation of the merged genome assembly. *V. cholerae* and ETEC reads from the co-culture were then each separately aligned against the merged genome assembly and read counts were computed, i.e. we sequenced and annotated the genome sequences from the single and dual cultures using the same assembly and annotation pipeline to avoid differential gene calling. Although all strains used in this study (*V. cholerae* V52, ETEC E36 & E616) are clinical isolates that have been sequenced and characterized before28,29, we have generated new assemblies and annotations to subject the mono- and co-culture transcriptomes to the same processing and annotation pipelines.

For *V. cholerae* V52 genome assembly, 3674 unique PATRIC ID mapped to 2964 unique FIGfam IDs. In order to do differential expression analysis between the genome assemblies generated from the mono cultures and the co-cultures, we aggregated genes by their FIGfam IDs. Members of a FIGfam, are believed to implement the same function, they are believed to derive from a common ancestor, and they can be globally aligned30. We wanted to see if there are specific functions that will be significantly altered between the two culture conditions especially that the sequence identity between ETEC and *V. cholerae* is around 80%7. FIGfam IDs were aggregated by keeping the FIGfam ID with the maximum value of raw read counts across all replicates from both the mono- and co-cultures. Co-culture transcriptomic data clustered independently from mono-cultures by PCA (Figure S3).

*V. cholerae* V52 co-culture with ETEC E36 :

For the merged *V. cholerae* V52/ETEC E36 genome assembly, 9418 unique PATRIC ID mapped to 5437 FIGfam. The final count matrix had 2931 unique FIGfam IDs which were used for downstream analysis and differential expression analysis. Overall, 20% of *V. cholerae* quantifiable transcriptome was perturbed in response to addition of ETEC E36 to the co-culture. In particular, 431 FIGfams (15%, mapped to 458 genes) were significantly upregulated (FDR < 0.1, LogFC more than 1) in *V. cholerae* when co-cultured with ETEC E36 relative to *V. cholerae* monoculture while 137 FIGfams (5%, mapped to 138 genes) were significantly downregulated (Table S10). Differentially expressed FIGfams were significantly enriched in diverse metabolic processes, particularly, amino acid metabolism e.g. tyrosine and L-phenylalanine (P value < 0.01, odds ratio > 10) (Figure 5A, Table S10).

*V. cholerae* V52co-culture with ETEC E616:

For the merged *V. cholerae* V52 and ETEC E616 genome assembly, 9679 unique PATRIC ID mapped to 5437 FIGfam IDs. The final count matrix had 2935 unique FIGfam IDs which were used for downstream analysis and differential expression analysis. 495 FIGfams (17%) were significantly upregulated (FDR < 0.1, logFC more than 1) in *V. cholerae* when co-cultured with ETEC E616 relative to *V. cholerae* monoculture while 106 FIGfams (4%) were significantly downregulated (Table S9). Differentially expressed FIGfams were significantly enriched in diverse metabolic processes, carbohydrate metabolic processes (P value < 0.05, odds ratio = 2.630409) (Figure 5B, Table S9).

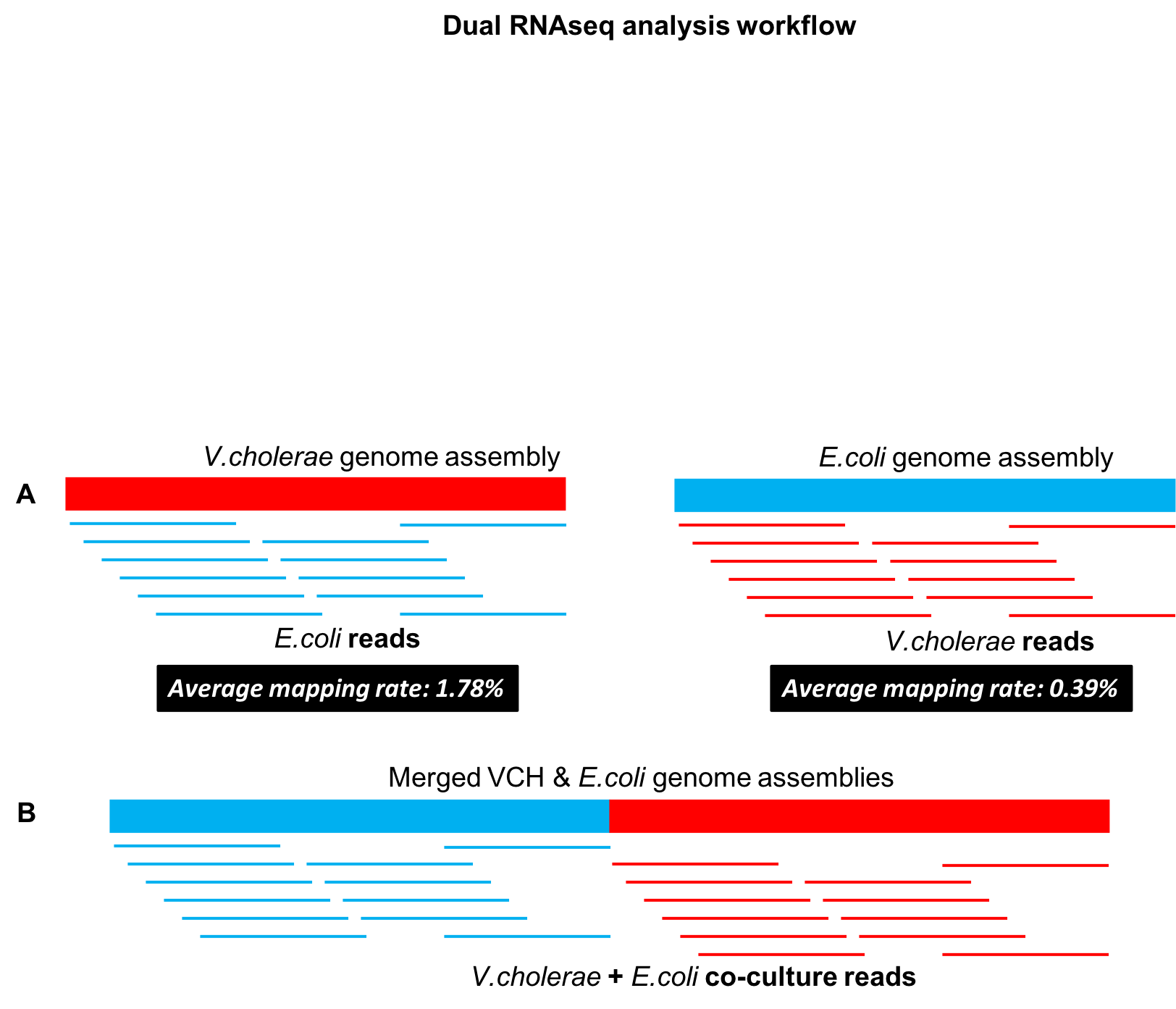
ETEC E36 co-culture with *V. cholerae* relative to its mono-culture:

Only 8% of ETEC E36 quantifiable transcriptome was perturbed and quantified in co-culture with *V. cholerae* relative to its monoculture where 3391 FIGfam IDs had non-zero total read count. Probably this is related to the significantly suppressed growth of ETEC E36 in co-culture with *V. cholerae* V52. 51 FIGfam IDs (2%) were significantly downregulated while 247 FIGfam IDs (7%) were significantly upregulated in E36 co-culture relative to its monoculture (Table S7). E36 differentially expressed genes were significantly enriched in taxis and chemotaxis in line with *V. cholerae*-mediated killing (Figure 5A).

ETEC E616 co-culture with *V. cholerae* relative to its mono-culture:

32% of ETEC E616 quantifiable transcriptome was perturbed in co-culture with *V. cholerae* relative to its monoculture where 3956 FIGfam IDs had non-zero total read count. 469 FIGfam IDs (12%) were significantly downregulated while 809 FIGfam IDs (20%) were significantly upregulated in E616 co-culture relative to its monoculture (Table S8, Figure 5B).

GOstats31 R package was used for the GO enrichment analysis and GOplot32 R package was used for visualization of GO enrichment results.



**Figure S4| Dual RNAseq analysis strategy**. A) Cross mapping rate for reads from mono-cultures against *V. cholerae* and ETEC assembled genomes. B) Merged genome assembly of the two species against which the reads from the co-cultures were mapped.

## Single gene essentiality predictions in mono- *vs.* co-culture models

Flux balance analysis33 (FBA) predicts growth phenotypes due to complete or partial metabolic gene deletions. A complete gene deletion is implemented by constraining the corresponding flux to zero. Linear programming then provide the flux distribution and maximal growth yield for the new genotype34. We have already validated *i*AM-Vc960 predicted gene essentiality against a dataset of essential genes previously identified in *V. cholerae* (Table S1, S3). We attempted to vary the species composition by varying the weights in the biomass objective function accounting for the growth of each of *V. cholera* and ETEC and we have also varied the media composition and investigate differences arising accordingly.

## Variation in essential genome *vs.* species abundance

We varied the species composition and tested for the essentiality of single genes from ETEC and *V. cholerae* for co-culture growth. 285 genes were essential for co-culture growth irrespective of the species composition while 1343 genes were non-essential. The essentiality of only 33 genes was dependent on the species composition mainly due to the 10% growth reduction cutoff (Table S11). 631 genes were essential only when *V. cholerae* was 10% while ETEC was 90%.

Difference between essentiality genes in rich vs. minimal medium (mono-cultures) -- adenylosuccinate lyase and adenylosuccinate synthetase were essential in minimal medium only. Both fructose-bisphosphate aldolase and 6-phosphofructokinase lost essentiality in rich medium (growth reduction by 12% in rich medium vs 21% in minimal medium). The difference between rich and minimal medium was more pronounced for ATP synthase subunits where growth was reduced by 46% upon deletion of any of the ATP synthase subunits in minimal media while a growth reduction of 16% was observed in rich medium.

Variation in gene essentiality in mono- vs co-cultures

glutamate racemase, D-fructose-6-phosphate amidotransferase, UDP-2,3-diacylglucosamine hydrolase, lipid transporter ATP-binding/permease protein, UDP-N-acetylglucosamine acyltransferase, UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase, UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase, UDP-N-acetylmuramoylalanyl-D-glutamate--2, 6-diaminopimelate ligase, Lipopolysaccharide export system protein LptC, NAD(P)H-dependent glycerol-3-phosphate dehydrogenase lost essentiality when ETEC is added to *V. cholerae* culture. There was no pronounced difference in gene essentiality for co-cultures in minimal vs rich media.

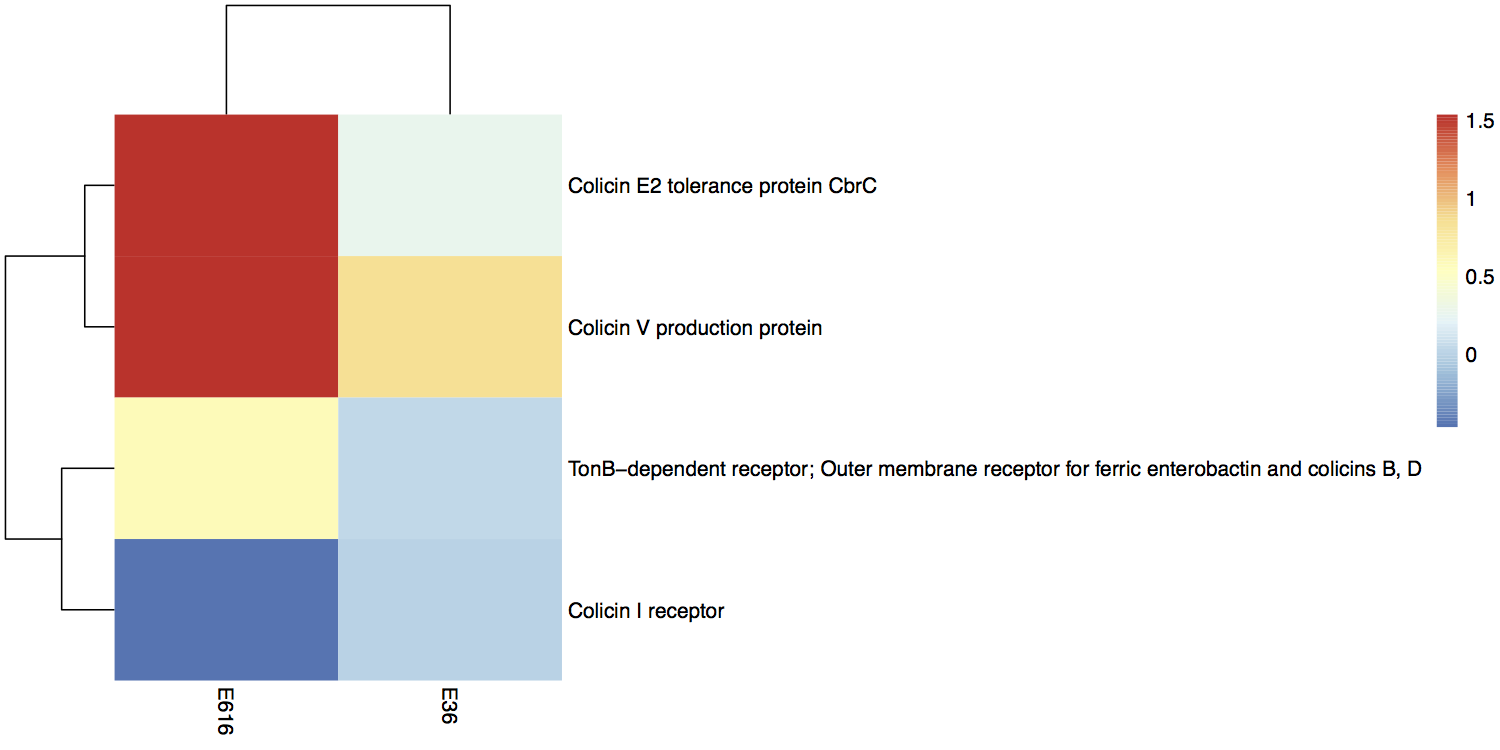
To conclude, no difference in essentiality for co-culture growth in minimal or rich medium, however the gene essentiality for mono-culture is much more affected when grown in minimal vs rich medium, i.e. the co-culture system is more resistant to gene essentiality changes and provide a context for selecting targets that would be effective across all diverse set of conditions.

### Predicted variation in essentiality between mono- and co-cultures

Interestingly, 12 *V. cholerae* genes were also predicted to totally lose their essentiality in dual infection with ETEC. Those were involved in lipid IV(A) biosynthesis (VC1877: *lpxK*, VC2396: *lpxC*), cell wall biosynthesis (VC0158: *murI*, VC2403: *murD*), *de novo purine* metabolism (VC1126: *purB*, VC2602: *purA*) and carbohydrate degradation (VC0477: *pgk*, VC0478: *fbaA*). *lpxC* performs the committed step in synthesis of Lipid A which protects bacterial cells from the damage of external agents like antibiotics. Since ETEC is not a threat for *V. cholerae*, on the contrary, as we have shown here, ETEC is boosting *V. cholerae* growth properties, it could be that *lpxC* and *lpxK* become indispensable in this setting. *lpxC* inhibitors have been shown to be effective against *V.* cholerae *in vitro* (PMID: 21167751, PMID: 23914798). We predict that this efficacy will drop if *V. cholerae* is interacting with other enteric pathogens like ETEC.

Differential gene expression analysis could potentially provide some explanation for the variation in gene essentiality between mono- and co-culture experiments. Glutamate racemase lost its essentiality in V52/ETEC co-culture. Expression data shows that it is significantly upregulated in E36 (logFC > 2, adjusted p value = 1.98E-6) and E616 (logFC > 1.5, adjusted p value = 0.06) when co-cultured with V52 relative to their mono-cultures.

## Bacteriocins result in difference in response between ETEC strains to co-growth with *V. cholerae*

****

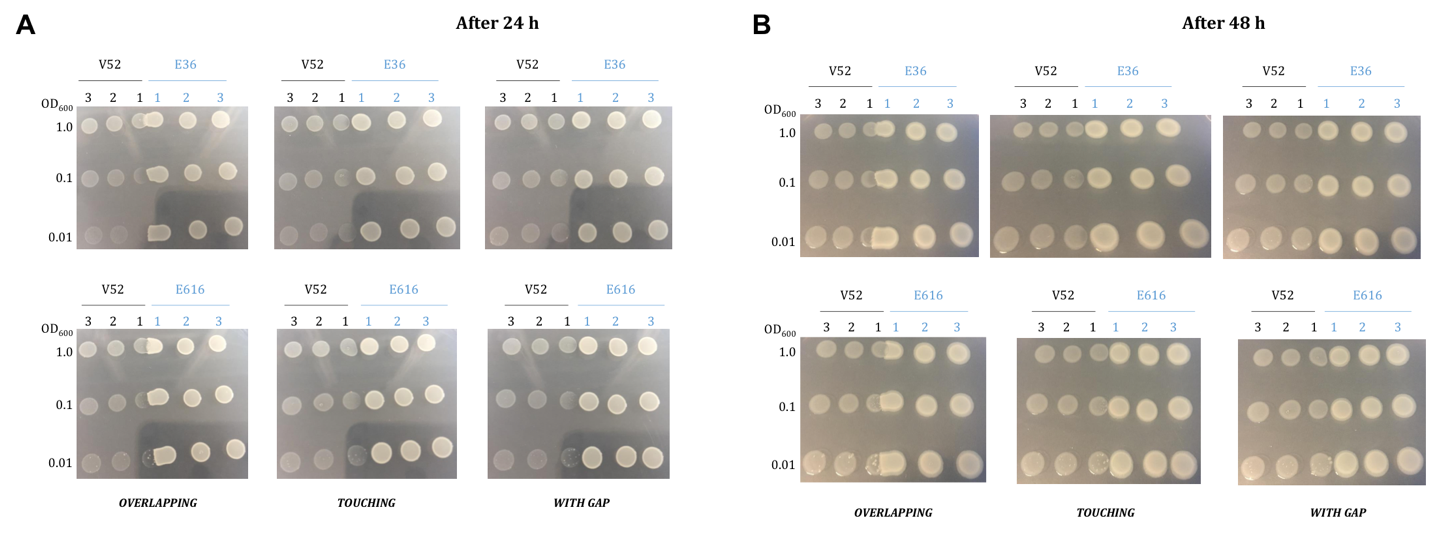
**Figure S5| Differential expression analysis of bacteriocin related genes in Enterotoxigenic E. coli**. Colicins’ production and tolerance genes are significantly up- regulated in E616 co-culture with *V. cholerae* relative to the individually grown E616. In contrast, E36, whose growth is more sensitive to co-growth with VCH failed to up-regulate bacteriocins related genes.

## *In silico* simulations, data and software Availability

All raw and processed data will be deposited to SRA archive. The *V. cholerae* *i*AM-Vc960 model will be deposited in BiGG and Biomodels repository. All code used in this project is available upon request through Rmd files and can be used for reproducibility and to display results alongside corresponding analyses.

**Testing for metabolites exchange on solid agar**

In order to assess if there is cross-talk or metabolic exchange happening at the molecular level that influences the growth of the two species under co-culture conditions, growth experiments in solid media were performed. Mutualistic growth patterns of *V. cholerae* V52 and ETEC (E36 and E616) in solid medium were observed by spotting both cultures under three different scenarios (Figure S5) – 1) “Overlapping” where the colonies merge with each other, 2) “Touching” where the colonies are just laid next to each other and 3) “Spacing” where there was a gap of 1.5 cm between the centers of the two colonies. Pictures were captured after (Figure S5A) 24h and (Figure S5B) 48h of growth at 30°C. Spotting was carried out at three different OD600 measurements to capture the effect.

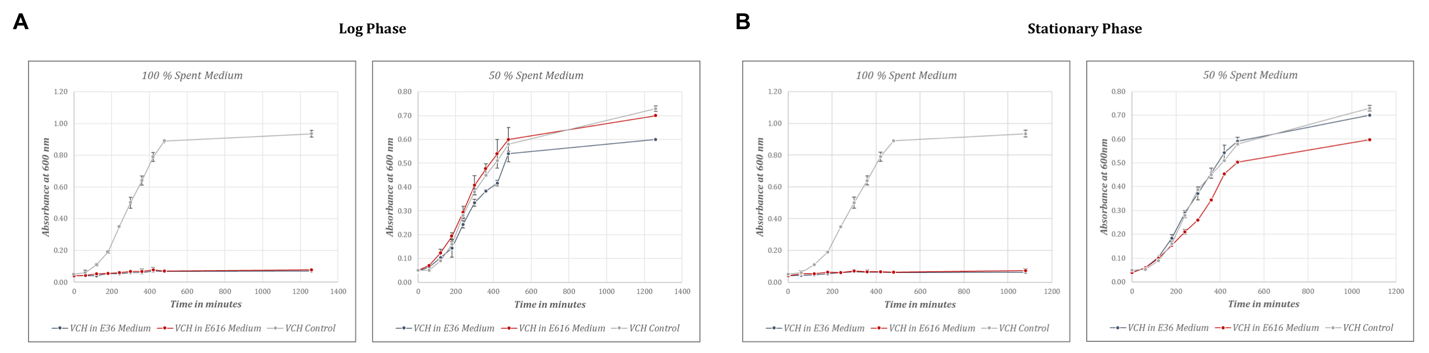


**Figure S6| Overlapping *V. cholerae* V52/ETEC colonies experiment**.Two strains of ETEC (E36 and E616) were tested in co-culture with *V. cholerae* V52. Three experimental designs were setup: ‘Overlapping’ where colonies merge with each other, ‘Touching’ where colonies are just laid next to each other and ‘With gap’ where there is a gap of 1.5 cm between the colonies. Picture was captured after (A) 24h and (B) 48h of growth at 30oC. The media used in the agar and liquid experiments were the same.

Unfortunately, using the overlapping colonies experiments, we were unable to reproduce the induction in *V. cholerae* growth rate observed when co-culturing the two species in liquid medium. An explanation for this could be that *V. cholerae* adopts distinct metabolic programs when in liquid medium (planktonic state) relative to agar (biofilm state).

## Spent media experiments

Effect of metabolites, released by ETEC, on the growth of *V. cholerae* was assessed by growing ETEC cultures to logarithmic or stationary phase in liquid medium and harvesting them. We then tested for the effect of ETEC spent medium on *V. cholerae* growth using 100% spent medium and then using ETEC spent medium mixed with water at 1:1 composition, i.e. 50% ETEC spent medium. In either cases, the “spent” medium was filtered using a 0.2 µm filter unit and *V. cholerae* cells were inoculated in the medium and grown until stationary phase. OD600 measurements were made at regular intervals. *V. cholerae* grown in fresh medium was chosen as a control for the 100% spent medium case while it was grown in frsh medium diluted with water (1:1) to control for the 50% spent medium



**Figure S7| *V. cholerae* V52/ETEC growth experiments using ETEC spent media**. Two strains of ETEC (E36 and E616) were grown to (A) log phase or (B) stationary phase, then filtered out and spent medium was used to culture *V. cholerae* V52. To control for the 100% spent medium condition, *V. cholerae* V52 was grown in completely fresh medium. To control for the 50% spent medium condition, *V. cholerae* V52 was grown in medium composed of 1:1 fresh medium to water. The media used in the agar and liquid experiments were the same.

The spent media experiments, as such, also failed to reproduce the induction in V. cholerae growth observed when both species are co-cultured and interact together, i.e. the spent medium from ETEC growth when grown in single culture failed to induce *V. cholerae* growth because the interaction between the two species is mutualistic and works only in liquid, i.e. the metabolite leading to induction of *V. cholerae* growth is not released in an ETEC mono-culture. There are several other explanations that are possible for why the spent media from ETEC is not inducing *V. cholerae* V52 growth. For instance, it could be that the sugar concentration in the spent medium is not enough. In a recent paper from Tramontano *et al.,* (PMID: 29556107), *V. cholerae* failed to grow on 4 media (out of 19 tested media); namely M9, M15A, M15B and M16. The main difference between the media that supported *V. cholerae* growth and those that did not was the type and concentration of the sugar in the medium. In the 4 media where *V. cholerae* failed to grow, there is either no glucose (M9, M16) or glucose concentration is less than 5g (0.5 g in M15B and 4g in M15A). Also, none of those four media has fructose, celllobiose, maltose or lactose compared to other media that supported maximal growth of *V. cholerae*. In light of this study, it is possible that the spent media did not have enough glucose concentration to support *V. cholerae* growth as it was already consumed by ETEC. Another potential explanation could be that a specific set of metabolites, rather than a single nutrient, is needed in order to attain the induction of *V. cholerae* growth observed when the two species are grown together. It could also be the oxygenation or pH conditions that are different between agar and liquid media.

Overall, the overlapping colonies and spent media experiments conducted as such could not conclusively prove or rule out a metabolite-mediated interaction between the two enteric pathogens.

1 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. & Pevzner, P. A. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).

2 Antipov, D., Hartwick, N., Shen, M., Raiko, M., Lapidus, A. & Pevzner, P. A. plasmidSPAdes: assembling plasmids from whole genome sequencing data. *Bioinformatics* **32**, 3380-3387, doi:10.1093/bioinformatics/btw493 (2016).

3 Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., Conrad, N., Dietrich, E. M., Disz, T., Gabbard, J. L., Gerdes, S., Henry, C. S., Kenyon, R. W., Machi, D., Mao, C., Nordberg, E. K., Olsen, G. J., Murphy-Olson, D. E., Olson, R., Overbeek, R., Parrello, B., Pusch, G. D., Shukla, M., Vonstein, V., Warren, A., Xia, F., Yoo, H. & Stevens, R. L. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res* **45**, D535-D542, doi:10.1093/nar/gkw1017 (2017).

4 Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., Rattei, T., Mende, D. R., Sunagawa, S., Kuhn, M., Jensen, L. J., von Mering, C. & Bork, P. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* **44**, D286-293, doi:10.1093/nar/gkv1248 (2016).

5 Seif, Y., Kavvas, E., Lachance, J. C., Yurkovich, J. T., Nuccio, S. P., Fang, X., Catoiu, E., Raffatellu, M., Palsson, B. O. & Monk, J. M. Genome-scale metabolic reconstructions of multiple Salmonella strains reveal serovar-specific metabolic traits. *Nat Commun* **9**, 3771, doi:10.1038/s41467-018-06112-5 (2018).

6 Bardou, P., Mariette, J., Escudie, F., Djemiel, C. & Klopp, C. jvenn: an interactive Venn diagram viewer. *BMC Bioinformatics* **15**, 293, doi:10.1186/1471-2105-15-293 (2014).

7 Chao, M. C., Pritchard, J. R., Zhang, Y. J., Rubin, E. J., Livny, J., Davis, B. M. & Waldor, M. K. High-resolution definition of the Vibrio cholerae essential gene set with hidden Markov model-based analyses of transposon-insertion sequencing data. *Nucleic Acids Res* **41**, 9033-9048, doi:10.1093/nar/gkt654 (2013).

8 Cameron, D. E., Urbach, J. M. & Mekalanos, J. J. A defined transposon mutant library and its use in identifying motility genes in Vibrio cholerae. *Proc Natl Acad Sci U S A* **105**, 8736-8741, doi:10.1073/pnas.0803281105 (2008).

9 Kamp, H. D., Patimalla-Dipali, B., Lazinski, D. W., Wallace-Gadsden, F. & Camilli, A. Gene fitness landscapes of Vibrio cholerae at important stages of its life cycle. *PLoS Pathog* **9**, e1003800, doi:10.1371/journal.ppat.1003800 (2013).

10 Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M. & Palsson, B. O. A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Mol Syst Biol* **7**, 535, doi:10.1038/msb.2011.65 (2011).

11 Chowdhury, F., Begum, Y. A., Alam, M. M., Khan, A. I., Ahmed, T., Bhuiyan, M. S., Harris, J. B., LaRocque, R. C., Faruque, A. S., Endtz, H., Ryan, E. T., Cravioto, A., Svennerholm, A. M., Calderwood, S. B. & Qadri, F. Concomitant enterotoxigenic Escherichia coli infection induces increased immune responses to Vibrio cholerae O1 antigens in patients with cholera in Bangladesh. *Infect Immun* **78**, 2117-2124, doi:10.1128/IAI.01426-09 (2010).

12 Kim, H. U., Kim, S. Y., Jeong, H., Kim, T. Y., Kim, J. J., Choy, H. E., Yi, K. Y., Rhee, J. H. & Lee, S. Y. Integrative genome-scale metabolic analysis of Vibrio vulnificus for drug targeting and discovery. *Mol Syst Biol* **7**, 460, doi:10.1038/msb.2010.115 (2011).

13 Thiele, I. & Palsson, B. O. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* **5**, 93-121, doi:10.1038/nprot.2009.203 (2010).

14 Schellenberger, J., Park, J. O., Conrad, T. M. & Palsson, B. O. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics* **11**, 213, doi:10.1186/1471-2105-11-213 (2010).

15 Flach, C. F., Qadri, F., Bhuiyan, T. R., Alam, N. H., Jennische, E., Holmgren, J. & Lonnroth, I. Differential expression of intestinal membrane transporters in cholera patients. *FEBS Lett* **581**, 3183-3188, doi:10.1016/j.febslet.2007.06.001 (2007).

16 Chen, W. H., Minguez, P., Lercher, M. J. & Bork, P. OGEE: an online gene essentiality database. *Nucleic Acids Res* **40**, D901-906, doi:10.1093/nar/gkr986 (2012).

17 Chen, W. H., Lu, G., Chen, X., Zhao, X. M. & Bork, P. OGEE v2: an update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. *Nucleic Acids Res* **45**, D940-D944, doi:10.1093/nar/gkw1013 (2017).

18 Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Linsay, B. & Stevens, R. L. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* **28**, 977-982, doi:10.1038/nbt.1672 (2010).

19 Stolyar, S., Van Dien, S., Hillesland, K. L., Pinel, N., Lie, T. J., Leigh, J. A. & Stahl, D. A. Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol* **3**, 92, doi:10.1038/msb4100131 (2007).

20 Wintermute, E. H. & Silver, P. A. Emergent cooperation in microbial metabolism. *Mol Syst Biol* **6**, 407, doi:10.1038/msb.2010.66 (2010).

21 Freilich, S., Zarecki, R., Eilam, O., Segal, E. S., Henry, C. S., Kupiec, M., Gophna, U., Sharan, R. & Ruppin, E. Competitive and cooperative metabolic interactions in bacterial communities. *Nat Commun* **2**, 589, doi:10.1038/ncomms1597 (2011).

22 Ponomarova, O. & Patil, K. R. Metabolic interactions in microbial communities: untangling the Gordian knot. *Curr Opin Microbiol* **27**, 37-44, doi:10.1016/j.mib.2015.06.014 (2015).

23 Bartell, J. A., Blazier, A. S., Yen, P., Thogersen, J. C., Jelsbak, L., Goldberg, J. B. & Papin, J. A. Reconstruction of the metabolic network of Pseudomonas aeruginosa to interrogate virulence factor synthesis. *Nat Commun* **8**, 14631, doi:10.1038/ncomms14631 (2017).

24 Ponomarova, O., Gabrielli, N., Sevin, D. C., Mulleder, M., Zirngibl, K., Bulyha, K., Andrejev, S., Kafkia, E., Typas, A., Sauer, U., Ralser, M. & Patil, K. R. Yeast Creates a Niche for Symbiotic Lactic Acid Bacteria through Nitrogen Overflow. *Cell Syst* **5**, 345-357 e346, doi:10.1016/j.cels.2017.09.002 (2017).

25 Zomorrodi, A. R. & Maranas, C. D. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput Biol* **8**, e1002363, doi:10.1371/journal.pcbi.1002363 (2012).

26 Zomorrodi, A. R. & Segre, D. Synthetic Ecology of Microbes: Mathematical Models and Applications. *J Mol Biol* **428**, 837-861, doi:10.1016/j.jmb.2015.10.019 (2016).

27 Monk, J. M., Charusanti, P., Aziz, R. K., Lerman, J. A., Premyodhin, N., Orth, J. D., Feist, A. M. & Palsson, B. O. Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci U S A* **110**, 20338-20343, doi:10.1073/pnas.1307797110 (2013).

28 von Mentzer, A., Connor, T. R., Wieler, L. H., Semmler, T., Iguchi, A., Thomson, N. R., Rasko, D. A., Joffre, E., Corander, J., Pickard, D., Wiklund, G., Svennerholm, A. M., Sjoling, A. & Dougan, G. Identification of enterotoxigenic Escherichia coli (ETEC) clades with long-term global distribution. *Nat Genet* **46**, 1321-1326, doi:10.1038/ng.3145 (2014).

29 Pukatzki, S., Ma, A. T., Sturtevant, D., Krastins, B., Sarracino, D., Nelson, W. C., Heidelberg, J. F. & Mekalanos, J. J. Identification of a conserved bacterial protein secretion system in Vibrio cholerae using the Dictyostelium host model system. *Proc Natl Acad Sci U S A* **103**, 1528-1533, doi:10.1073/pnas.0510322103 (2006).

30 Meyer, F., Overbeek, R. & Rodriguez, A. FIGfams: yet another set of protein families. *Nucleic Acids Res* **37**, 6643-6654, doi:10.1093/nar/gkp698 (2009).

31 Falcon, S. & Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257-258, doi:10.1093/bioinformatics/btl567 (2007).

32 Walter, W., Sanchez-Cabo, F. & Ricote, M. GOplot: an R package for visually combining expression data with functional analysis. *Bioinformatics* **31**, 2912-2914, doi:10.1093/bioinformatics/btv300 (2015).

33 Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nat Biotechnol* **28**, 245-248, doi:10.1038/nbt.1614 (2010).

34 Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* **99**, 15112-15117, doi:10.1073/pnas.232349399 (2002).