

# Functional genomics of the rapidly replicating bacterium *Vibrio natriegens* by CRISPRi

Henry H. Lee<sup>1,4\*</sup>, Nili Ostrov<sup>1,4</sup>, Brandon G. Wong<sup>2</sup>, Michaela A. Gold<sup>1</sup>, Ahmad S. Khalil<sup>2,3</sup> and George M. Church<sup>1,3\*</sup>

**The fast-growing Gram-negative bacterium *Vibrio natriegens* is an attractive microbial system for molecular biology and biotechnology due to its remarkably short generation time<sup>1,2</sup> and metabolic prowess<sup>3,4</sup>. However, efforts to uncover and utilize the mechanisms underlying its rapid growth are hampered by the scarcity of functional genomic data. Here, we develop a pooled genome-wide clustered regularly interspaced short palindromic repeats (CRISPR) interference (CRISPRi) screen to identify a minimal set of genes required for rapid wild-type growth. Targeting 4,565 (99.7%) of predicted protein-coding genes, our screen uncovered core genes comprising putative essential and growth-supporting genes that are enriched for respiratory pathways. We found that 96% of core genes were located on the larger chromosome 1, with growth-neutral duplicates of core genes located primarily on chromosome 2. Our screen also refines metabolic pathway annotations by distinguishing functional biosynthetic enzymes from those predicted on the basis of comparative genomics. Taken together, this work provides a broadly applicable platform for high-throughput functional genomics to accelerate biological studies and engineering of *V. natriegens*.**

We began by investigating laboratory culturing conditions that support rapid and consistent *Vibrio natriegens* growth. We opted for lysogeny broth (LB) supplemented with NaCl to a final concentration of 3% (w/v) (LB3), which approximates ocean salinity and provides consistent growth (Supplementary Fig. 1 and Supplementary Table 1). At 37 °C, we measured the generation time of *V. natriegens* to be 15.2 min in bulk culture and 14.8 min using single-cell, time-lapse microscopy (Fig. 1a,b, Supplementary Video 1 and Supplementary Figs. 1 and 2). A longer generation time (30.8 min) was observed in minimal glucose medium supplemented with 2% (w/v) NaCl (Supplementary Fig. 1). These generation times are comparable with those found in other studies<sup>3–5</sup>.

We then produced two closed circular chromosomes of 3.24 Mb (chromosome 1; chr1) and 1.92 Mb (chromosome 2; chr2) by de novo sequence assembly, followed by annotation using RAST<sup>6</sup> (RefSeq NZ\_CP009977–8, Fig. 1c, Supplementary Tables 2 and 3 and Methods)<sup>7</sup>. We identified 4,578 open reading frames (ORFs), of which approximately 63% were on chr1 and approximately 37% were on chr2. Consistent with the broad metabolic capacity described for *Vibrio* spp.<sup>8</sup>, nearly half of the annotated ORFs are involved in carbohydrate, RNA or protein metabolism (Fig. 1d). The *V. natriegens* genome carries 11 ribosomal RNA operons and 129 transfer RNA genes, more than are found in the genomes of *Vibrio cholerae* N16961 (8 and 98, respectively)<sup>9</sup> or *Escherichia coli* MG1655 (7 and

87, respectively)<sup>10</sup>. We further identified 36,599 putative methylated adenine residues at GATC motifs. Finally, we mapped the origin and terminus of each chromosome by quantifying sequencing coverage of actively dividing cells<sup>11</sup>, and found a single bidirectional origin on each chromosome (Fig. 1e), which displayed high sequence similarity to those in other *Vibrio* spp. (Supplementary Fig. 3)<sup>9</sup>.

We next sought to develop genetic tools to enable high-throughput assessment of gene functionality. Whereas transposon mutagenesis is commonly used to generate single-gene-mutant libraries<sup>12</sup>, low insertion efficiency prohibited scalable saturation mutagenesis in *V. natriegens* (Supplementary Table 8 and Supplementary Discussion). Instead, we used the clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein 9 (Cas9) system, which has been demonstrated in well-studied model organisms<sup>13–16</sup> but has not been scaled for functional genomics in emerging microbial systems.

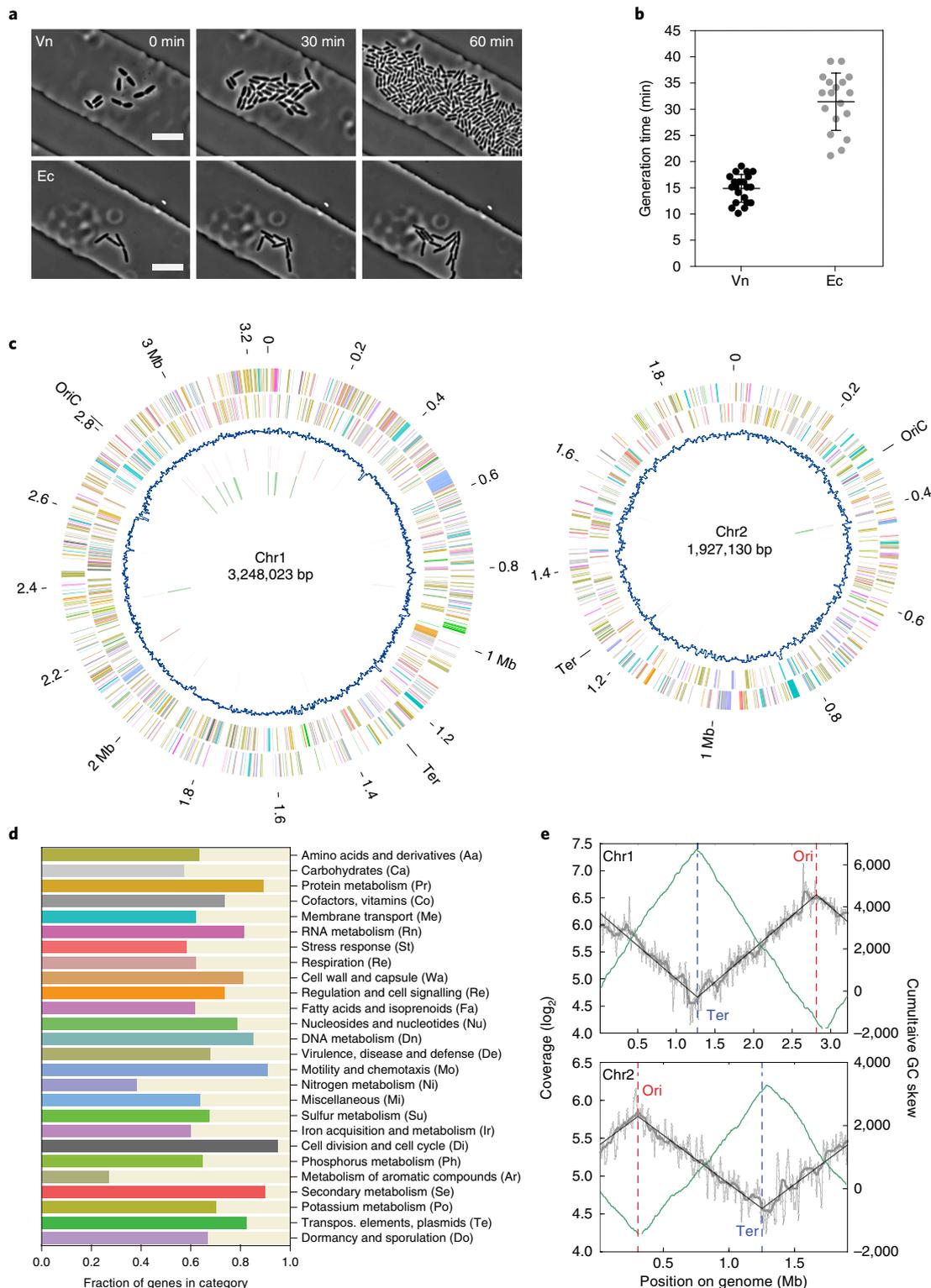
To test CRISPR–Cas9 functionality, we created a reporter *V. natriegens* strain expressing green fluorescent protein (GFP), and tested the *Streptococcus pyogenes* Cas9 nuclease with or without GFP-targeting guide RNA (gRNA) (plasmid pRSF-paraB-spCas9-gRNA, Supplementary Figs. 5–7 and Supplementary Table 4). No transformants were detected following electroporation of Cas9 nuclease with a GFP-targeting gRNA, even without induction of Cas9 expression (Supplementary Fig. 9). This result indicates significant toxicity due to double-stranded breaks and undetectable levels of non-homologous end joining (NHEJ). We thus concluded that the CRISPR–Cas9 system alone cannot be used for efficient generation of mutants by NHEJ.

Transcriptional inhibition of gene expression by CRISPR interference (CRISPRi) relies on the nuclease-deficient variant dCas9 to block RNA polymerase activity<sup>15</sup>. On the basis of design rules elucidated in other model bacteria<sup>14,15</sup>, we designed guides targeting the template and non-template strand proximal to the GFP transcriptional start site and measured fluorescence in the presence of dCas9 (plasmid pRSF-paraB-dCas9-gRNA, Supplementary Table 4). Consistent with previous studies, we observed greater inhibition when targeting the non-template strand (more than 13-fold, Fig. 2a). As with Cas9 nuclease, induction of dCas9 was not required for activity and led to a longer lag phase indicative of marginal toxicity (Supplementary Fig. 9). We thus opted to use the presence and absence of the dCas9 plasmid in lieu of induction for subsequent experiments.

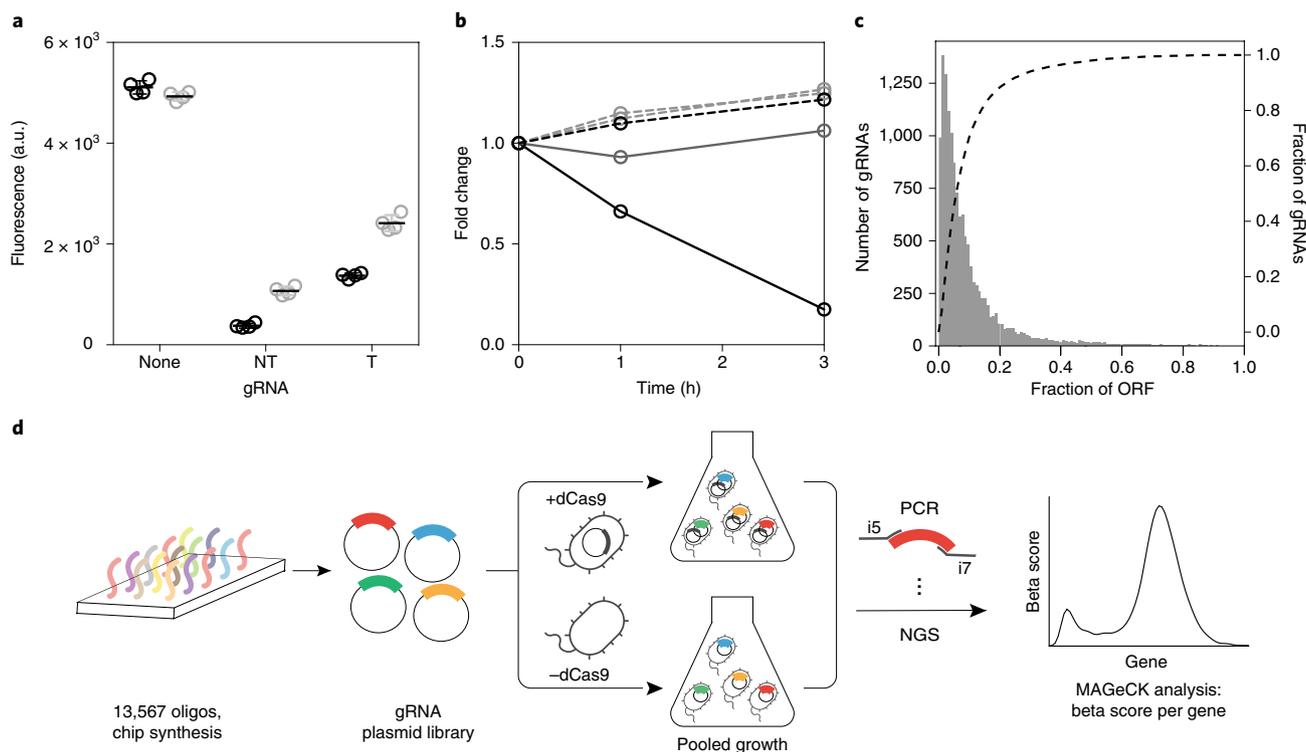
For high-throughput screening of single-gene inhibition by CRISPRi, we transformed a pool of gRNA plasmids such that each cell carries dCas9 and a single gRNA targeting inhibition of one

<sup>1</sup>Department of Genetics, Harvard Medical School, Boston, MA, USA. <sup>2</sup>Department of Biomedical Engineering and Biological Design Center, Boston University, Boston, MA, USA. <sup>3</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA.

<sup>4</sup>These authors contributed equally: Henry H. Lee, Nili Ostrov. \*e-mail: [hlee@genetics.med.harvard.edu](mailto:hlee@genetics.med.harvard.edu); [gchurch@genetics.med.harvard.edu](mailto:gchurch@genetics.med.harvard.edu)



**Fig. 1 | Profiling of *V. natriegens* growth and genome. a**, Visualization of *V. natriegens* (Vn) and *E. coli* (Ec) in a microfluidic chemostat (cultures grown at 37 °C in LB3 and LB media, respectively). Designated cell-trap heights were used to keep each species growing in a monolayer. Scale bar, 10  $\mu$ m;  $n = 3$  independent experiments. **b**, Single-cell growth measurement of *V. natriegens* and *E. coli* at 37 °C (in LB3 and LB media, respectively). Data are mean  $\pm$  s.d.;  $n = 18$  *E. coli* cells,  $n = 21$  *V. natriegens* cells. **c**, *V. natriegens* chromosomes. From outside inward: two outer circles represent protein-coding genes on the plus and minus strands, respectively. Colour coding by RAST annotation, as in **d**. The third circle represents GC content relative to mean GC content of the respective chromosome, using a sliding window of 3,000 base pairs (bp). tRNA and rRNA genes are shown in the fourth and fifth circles, respectively. **d**, Fraction of genes in each RAST category on chr1 (darker colours) and chr2 (light yellow). **e**, Filtered sequence coverage (black) and GC skew (green) for each chromosome, as measured for exponentially growing *V. natriegens* in LB3 at 37 °C. Positions of the origin (Ori, red) and terminus (Ter, blue) are indicated.



**Fig. 2 | CRISPRi screen in rich and minimal media.** **a**, dCas9 inhibition of chromosomally integrated GFP, using gRNA targeting the template (T) or non-template (NT) strand. Inhibition was observed with (black) or without (grey) induction of dCas9. Data are mean  $\pm$  s.d.;  $n = 4$  biological replicates. **b**, Small-scale pooled CRISPRi screen. gRNAs are targeted to a putative essential gene *lptF<sub>vn</sub>* (solid black line), a putative neutral gene *flgC<sub>vn</sub>* (dashed black line), GFP (dashed grey line) and *E. coli* gene *lptF<sub>ec</sub>* (solid grey line, target gene not present in genome). The library was grown in LB3 at 37 °C in the presence of dCas9. Fold change in gRNA abundance is shown. **c**, Genome-wide gRNA library. Number (grey bars) and percentage (dotted line) of gRNAs across ORFs are shown. **d**, Schematic overview of pooled genome-wide CRISPRi screen. NGS, next-generation sequencing; a.u., arbitrary units.

protein-coding gene. The library was grown as a competitive culture, and the abundance of each gRNA was determined by sequencing of gRNA cassettes. Since the abundance of gRNA reflects the fitness cost of reduced gene expression, we expect rapid depletion of gRNAs targeting genes that, when inhibited, impair cellular growth.

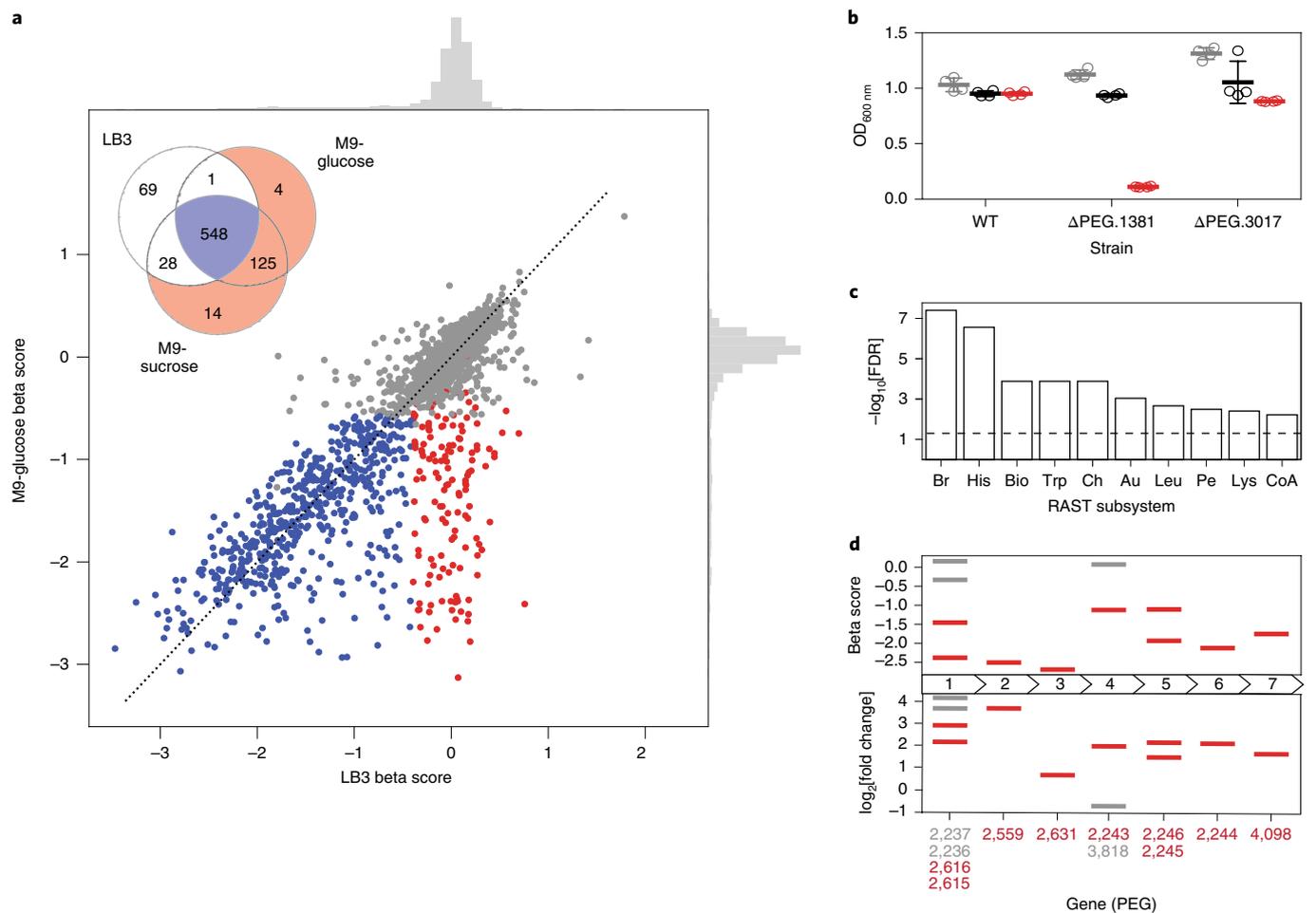
We initially used a small library of five gRNAs to test the feasibility of a pooled CRISPRi approach. gRNAs were selected for inhibition of one putative essential gene (*lptF<sub>vn</sub>*, a homologue of an essential *E. coli* gene), one putative non-essential gene (*flgC<sub>vn</sub>*) and two control genes that are not present in the genome (GFP and *E. coli* gene *lptF*). All gRNA plasmids (vector pCTX-R6K-gRNA, Supplementary Fig. 8) were co-transformed into *V. natriegens* bearing dCas9 (plasmid pdCas9 bacteria<sup>15</sup>). Plasmids were extracted and sequenced to determine the abundance of each gRNA over 3 h of competitive growth. We observed a significant reduction in abundance of the gRNA targeting the putative essential *lptF<sub>vn</sub>* gene, whereas all other guides remained relatively unchanged (Fig. 2b and Supplementary Fig. 10). These results demonstrate that gRNA abundance can be used as proxy for a gene's fitness in pooled competitive culture.

Next, we scaled up our gRNA library to cover all predicted ORFs in the *V. natriegens* genome. Transcriptional knockdown of a gene by CRISPRi can be achieved by targeting either its transcriptional start site, which blocks initiation, or its coding region, which blocks elongation<sup>15</sup>. As transcriptional start sites have not yet been mapped in this organism, we designed gRNAs targeting the non-template strand proximal to each gene's coding region. We created a library of 13,567 unique gRNAs to target 4,565 (99.7%) RAST-predicted protein-coding genes. Our library includes up to three unique gRNAs per gene, with 88% of all guides within the first 20% of the coding

sequence (Fig. 2c and Supplementary Table 5). Library screens were performed in LB3, M9-glucose or M9-sucrose medium for 8 h (Fig. 2d and Supplementary Fig. 11). To estimate fitness for each gene from its targeting gRNAs, we generated a beta score and statistical significance using MAGeCK's maximum likelihood estimation (MLE)<sup>17</sup> (Supplementary Fig. 5). A negative beta score indicates depletion of a gRNA under the tested condition.

Using sucrose as the sole carbon source, we identified 14 genes with significant negative beta scores (false discovery rate,  $FDR \leq 0.05$ ) that are unique to growth in this medium (Fig. 3a and Supplementary Table 5). These include phosphoglucomutase, several membrane transport proteins and sucrose-6-phosphate hydrolase, a key component of intracellular sucrose catabolism. As only one of two annotated sucrose-6-phosphate hydrolase genes was identified in our screen, we generated an in-frame deletion of each gene and measured their fitness under non-competitive growth conditions. Consistent with the CRISPRi results, we found that the sucrose-6-phosphate hydrolase located on chr1 (PEG.1381, beta score for sucrose,  $\beta_{suc} = -1.6907$ ), but not the one located on chr2 (PEG.3071,  $\beta_{suc} = 0.12$ ), was required for growth in M9-sucrose medium (Fig. 3b). This demonstrates the utility of our approach for identifying functionally relevant genes.

Overall, our glucose and sucrose minimal media screen yielded 143 unique genes ( $FDR \leq 0.05$ ) enriched in RAST subcategories for amino acid and vitamin biosynthesis (Fig. 3c and Supplementary Fig. 12a). When mapped onto the respective pathways, this gene set can be used to distinguish functional genes from a larger set of computationally assigned biosynthetic enzymes<sup>7,18</sup>. For example, of all predicted L-leucine biosynthetic genes, only one of two aceto-lactate synthase complexes (PEGs 2236, 2237 and 2615, 2616) and



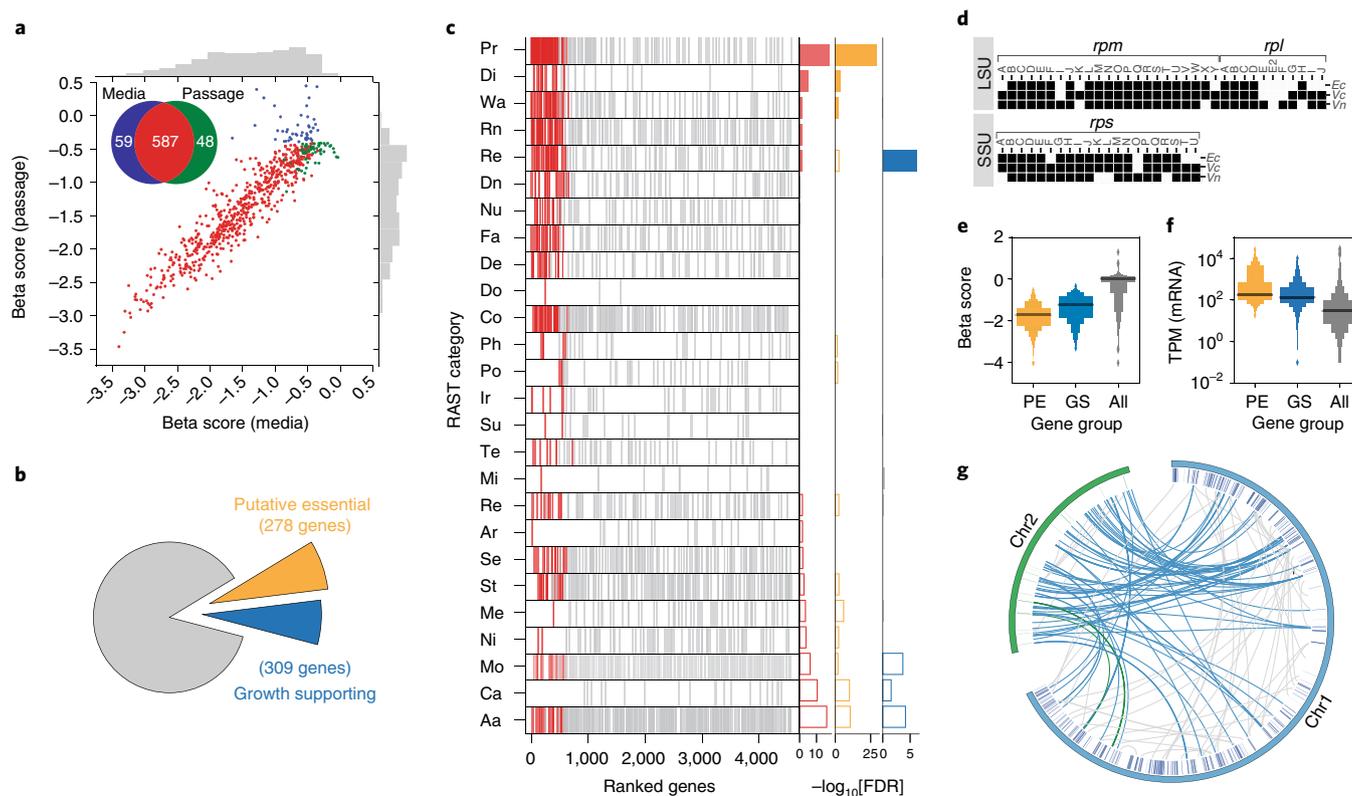
**Fig. 3 | CRISPRi in rich and minimal media. a**, Distribution of gene scores in rich and minimal media screen are shown: blue dots represent genes with significantly negative score in both rich and minimal media; red dots represent genes with significantly low score only in minimal medium; grey dots represent all other genes in the population. Histograms show beta score distributions of all targeted genes. Inset: the number of genes with significantly negative scores in each growth medium. **b**, Single-gene knockout of sucrose-6-phosphate hydrolase located on chr1 ( $\Delta$ PEG.1381) or its paralog located on chr2 ( $\Delta$ PEG.3017) in LB3 (grey), M9-glucose (black) or M9-sucrose (red) media. Data are mean  $\pm$  s.d.;  $n=4$  biological replicates. WT, wild type. **c**, Top ten RAST categories enriched in genes that are uniquely depleted in minimal medium ( $n=143$  genes). Br, branched-chain amino acid biosynthesis; His, histidine biosynthesis; Bio, biotin biosynthesis; Trp, tryptophan synthesis; Ch, chorismate; Au, auxin biosynthesis; Leu, leucine biosynthesis; Pe, peptide ABC transport system; Lys, lysine biosynthesis DAP pathway; CoA, coenzyme-A biosynthesis. Statistics are derived from one-tailed hypergeometric test for enrichment assuming independence between subsystems. FDR controlled with Benjamini-Hochberg procedure. **d**, Leucine (L-Leu) biosynthetic pathway starting from 2-pyruvate. Beta score (top) and change in expression between LB3 and M9 (fold change, bottom) are shown for each protein encoding genes (PEG number); functional genes identified by CRISPRi in minimal medium are shown in red, other computationally assigned genes are shown in grey. Leucine biosynthetic genes numbered as follows: (1) acetolactate synthase (EC 2.2.1.6); (2) ketol-acid reductoisomerase (EC 1.1.1.86); (3) dihydroxy-acid dehydratase (EC 4.2.1.9); (4) isopropylmalate synthase (EC 2.3.3.13); (5) 3-isopropylmalate dehydratase (EC 4.2.1.33); (6) 3-isopropylmalate dehydrogenase (EC 1.1.1.85); and (7) aminotransferase (EC 2.6.16/2.6.1.24).

one of two isopropylmalate synthases (PEGs 2243 and 3818) were identified as functional in our screen (Fig. 3d). Similar results were obtained for tryptophan and isoleucine biosynthetic pathways. Moreover, we show that the functional enzymes are not always the most differentially expressed (Fig. 3d and Supplementary Fig. 12b,c). Thus, whereas transcriptional enrichment corroborates the importance of amino acid biosynthesis pathways in minimal medium (Supplementary Table 7), expression level is not sufficient to infer gene functionality.

We next sought to identify a minimal set of genes necessary for rapid wild-type growth in rich medium (LB3). In our first experiment, we found 646 genes with significant negative beta score (Fig. 3a). When repeated, the same screen resulted in a slightly larger number of significant genes (735); however, further passaging of the culture

yielded only 648 significant negative-scoring genes. Of these, 587 (92.4%) overlapped between the two experiments (Fig. 4a and Supplementary Fig. 13). Given the high overlap, these data indicate that passaging reduces variability and provides more stringent results. We thus define 587 genes identified in both experiments as ‘core genes’ that are required for rapid growth in rich media (Supplementary Table 6).

We further analysed the passaging experiments for evidence of CRISPRi escape by examining 100 genes that were identified as significant in the first, but not the last passage (Supplementary Fig. 13e and Supplementary Table 5). We expected CRISPRi escape to result in increased beta score in later passages. However, the majority of these genes (73%) displayed a decrease in beta score, suggesting that rather than escaping dCas9 inhibition, they were likely to have



**Fig. 4 | Analysis of *V. natriegens* core genes.** **a**, Distribution of beta scores for all significantly depleted genes in rich medium experiments ( $FDR \leq 0.05$ ). Core genes, shown in red, display consistently negative beta scores in both experiments. Negative-scoring genes that were not included in the core set are shown in blue (media experiment) and green (passage experiment). **b**, Core genes were divided into putative essential genes (yellow) and growth-supporting genes (blue). All other genes are shown in grey. **c**, Distribution of all targeted genes across RAST categories. Single genes, sorted by beta score, are displayed as grey lines in the respective category. Core genes are shown in red. Histogram displays over-represented (filled bars) or under-represented categories (empty bars) within all core genes (red), putative essential genes (yellow) and growth-supporting genes (blue). Statistics are derived from one-tailed hypergeometric test for enrichment and depletion assuming independence between categories. FDR was controlled with the Benjamini–Hochberg procedure.  $n = 587$  core genes. RAST category names as shown in Fig. 1b. **d**, Ribosomal proteins essentiality as identified in *V. natriegens* (Vn) screen. Black square denotes essential gene. Empty square denotes non-essential gene. Essentiality shown for Vc, *E. coli* (Ec)<sup>21</sup> and *V. cholerae* (Vc)<sup>19</sup>. LSU, large subunit; SSU, small subunit. **e**, Beta score as calculated from CRISPRi screen for all genes (grey), putative essential genes (PE) and growth-supporting genes (GS) are shown as a letter-value plot with median. **f**, RNA-seq transcript counts as TPM for all genes (grey) and for the set of core genes, as measured at 37 °C in LB3 medium. PE genes shown in yellow, GS genes shown in blue. Representative counts for a single biological replicate (out of three independent biological replicates) are shown as a letter-value plot with median. **g**, Duplicated annotations of core genes. Core genes are shown as hash marks on chr1 (blue) and chr2 (green). Lines connect core genes with the respective duplicate annotation gene. Blue lines indicate core genes on chr1 with duplicates on chr2. Green lines indicate core genes on chr2 with duplicates on chr1. Gray lines indicate core genes and respective duplicates located on the same chromosome.

decreased significance due to the increase in the total number of low-scoring genes (Supplementary Fig. 13f). Of the remaining 27 genes for which beta scores increased after passaging, only seven were altered by more than 20% (Supplementary Fig. 14a). For these, we found one of three gRNAs was repeatedly elevated in the third passage (Supplementary Fig. 14b), indicating that the increase in beta score was due to an inefficient or promiscuous gRNA sequence rather than CRISPRi escape.

Core genes comprise 12.8% of 4,565 predicted coding genes in *V. natriegens*, and the vast majority are located on chr1 (96.0%, 564 genes) (Supplementary Table 6). Only 16 core genes (2.7%) were annotated as hypothetical proteins. We found the median expression of core genes was 5.6-fold higher than that of non-core genes (160.71 and 28.56 median transcripts per million (TPM), respectively). To assess essentiality, we compared the core set to genes previously identified by high-density transposon mutagenesis in the well-studied *V. cholerae* genome<sup>19</sup>. We found 278 homologous genes, corresponding to 84.4% of the 329 mappable essential genes

in *V. cholerae*, which we designated putative essential genes in *V. natriegens* (Fig. 4b and Supplementary Table 6).

Similar to other *Vibrio* spp.<sup>20</sup>, a large majority of putative essential genes are located on chr1 (272 of 278 genes). Putative essential genes were statistically enriched for RAST categories including ‘protein metabolism’ and ‘cell division and cell cycle’, which include ribosomal proteins, tRNA synthetases, and DNA polymerase (Fig. 4c and Supplementary Fig. 15). For example, 49 of 55 total ribosomal proteins were included in the core set, of which 44 (89.8%) were in agreement with essentiality in *V. cholerae* and 37 (75.5%) are in agreement with essentiality in *E. coli*, as determined by transposon mutagenesis<sup>19</sup> and gene deletions<sup>21</sup>, respectively (Fig. 4d and Supplementary Table 6). Other notable essential genes include DNA adenine methylase (*dam*<sup>22</sup>) and *nadE*, a NAD<sup>+</sup> synthase gene implicated in the conversion of chr2 from a megaplasmid into a critical component of the *Vibrio* genome<sup>23</sup>. Putative essential genes, which were all actively transcribed, showed lower beta scores and higher transcription levels than other core genes

(Fig. 4e,f and Supplementary Fig. 15a), in line with reports of essential gene expression in other organisms<sup>24</sup>.

Non-essential core genes (309 of 587 genes) were uniquely enriched for the RAST category ‘respiration’ (Fig. 4c and Supplementary Fig. 15b). This set of ‘growth-supporting’ genes include the Rnf electron transport complex, cytochrome C oxidase biogenesis enzymes and the Na<sup>+</sup>-NQR respiration complex implicated in aerobic respiration in halophilic bacteria<sup>25</sup> (Supplementary Table 6). These pathways are likely to be responsible for the high oxygen uptake and highly active electron-transport-chain activity<sup>2</sup> required to maintain rapid cell division. Others, such as quorum-sensing regulator *hfq*<sup>26</sup>, may determine the rapid growth rate observed in highly dilute cultures<sup>4</sup>. Further investigation of growth-supporting genes will refine our understanding of the mechanisms underlying rapid *V. natriegens* growth.

Interestingly, we found that 14.3% of core genes (84 of 587 genes)—including tyrosyl-tRNA synthetase, heat shock proteins GroES and GroEL, DNA polymerase I and chaperone DnaK—share identical annotation with one or more genes. Duplicates are distributed similarly between essential and growth-supporting core genes (39 and 45, respectively) and most have non-significant negative beta scores, suggesting they are unlikely to be functionally redundant (Supplementary Fig. 16a). Only five genes were found where both duplicates were independently identified in our screen, including EF-Tu, geranylgeranyl pyrophosphate synthetase, MotA/TolQ/ExbB proton channel family protein, ferric iron ABC transporter ATP-binding protein and a benzoquinol hydroxylase (Supplementary Table 6). We found that 60.3% of core genes with duplicates (53 of 84 genes) are located on chr1 with the duplicate located on chr2 (Fig. 4g), contributing to the relatively large size of chr2 in *V. natriegens* compared to other *Vibrio* spp.<sup>20</sup>. Duplicated genes, previously reported in *V. cholerae*<sup>19</sup>, may represent hetero-functional homologues or pseudogenes. Transcriptional profiling shows the median TPM for core genes is about 17-fold higher than that for duplicates (125.6 and 7.4, respectively), suggesting that expression of duplicates may be silenced under the tested conditions (Supplementary Fig. 16b). Further insight into the genesis and roles of these gene duplications may benefit from studies of inter-chromosomal interactions<sup>27</sup> in *Vibrio* genomes.

Interpretation of pooled CRISPRi results should consider potential caveats. First, metabolite cross-feeding, such as amino acid exchange, may obscure the effect of inhibition of critical pathway enzymes<sup>28</sup>. Second, as CRISPRi relies on reducing RNA polymerase elongation, inhibition of upstream genes in an operon may result in erroneous determination of essentiality for downstream genes<sup>14</sup>. Identification of transcriptional start sites and operon structures, currently unavailable for *V. natriegens*, will help to better address polar effects. Third, changes to the statistical significance threshold may alter the list of functionally relevant genes. Lastly, results of screens may be affected by dCas9-specific toxicity, as observed for *recQ* helicase (Supplementary Discussion and Supplementary Fig. 17). Additional studies using in-frame deletions are needed to stringently evaluate the phenotype of single-gene knockouts.

Nevertheless, screening sensitivity and specificity may be further increased by improvements to the gRNA library and dCas9 protein. Whereas three gRNAs were used to mitigate inefficient gene knockdown, we cannot exclude the possibility that inhibition levels vary between genes, since the knockdown efficiency for each gRNA was not measured. Future studies may utilize individually optimized gRNAs to ensure equal inhibition between genes or increase the number of gRNAs per gene for robust statistical inference. The library can also be expanded to target non-coding elements, such as RNA or intergenic regions, to reveal functional genomic features that are not captured by computational annotation. Additional design principles for CRISPRi libraries<sup>16,29</sup> may be used to improve gRNA binding specificity or reduce potential sequence-specific

toxicity<sup>29</sup>. Furthermore, gene-dosage effects can be explored by tuning transcriptional knockdown via engineering of *V. natriegens* promoters or synthetic dCas9 with reduced toxicity<sup>30</sup>. Overall, this pooled approach can be utilized to screen for genotypes underlying any growth-linked phenotype to accelerate our understanding of this organism. More broadly, we envision that the use of pooled genome-wide CRISPRi screens will enable rapid and high-throughput functional annotation of diverse bacteria.

Genome-wide experimental annotations of *V. natriegens* can inform targeted studies of its fast growth and metabolic capacity. These findings may also provide a starting point for large-scale genome engineering. For example, core genes can be used as a basis for probing the limits of codon reassignment. Furthermore, beta scores can offer valuable experimental guidance for construction of a fast-growing minimal genome by bottom-up construction or large-scale deletion<sup>2</sup>. With only 2.1% of putative essential genes located on chr2 (Fig. 4e and Supplementary Table 6), spatial distribution of core genes presents intriguing opportunities for rational genome design, such as consolidating of functional genes to chr1 and re-purposing of chr2 as an artificial chromosome. Further investigation of this unique organism will facilitate its advancement as a versatile bacterial system for research and biotechnology.

## Methods

**Strains used in this study.** Wild-type *V. natriegens* (ATCC 14048) was used in this study. The GFP reporter strain was constructed by genomic integration of a pLtetO-gfp-kanamycin cassette using the mariner transposon system described in Methods.

**Growth media.** Unless noted, LB3 was used as standard rich medium. This was prepared by adding 20 g NaCl to 25 g of LB Broth, Miller (Fisher BP9723-500) per litre of medium. Media were formulated according to the manufacturer’s instructions and supplemented with 1.5% (w/v) final Ocean Salts (Aquarium System) to make high-salt versions of brain heart infusion, nutrient broth and LB. No additional salts were added to marine broth. Minimal M9 medium (BD Difco, cat. no. 248510) was prepared according to the manufacturer’s instructions. For culturing *V. natriegens*, 2% (w/v) NaCl was added to M9 (in addition to the pre-included 0.5 g l<sup>-1</sup> NaCl) on the basis of the screen of NaCl concentrations in rich medium. Carbon sources were added as indicated to 0.4% (v/v) final. SOC3 medium comprised 5 g l<sup>-1</sup> yeast extract, 20 g l<sup>-1</sup> tryptone, 30 g l<sup>-1</sup> NaCl, 2.4 g l<sup>-1</sup> MgSO<sub>4</sub> and 0.4% (w/v) final glucose. Antibiotic concentrations used for plasmid selection in *V. natriegens*: ampicillin–carbenicillin 100 µg ml<sup>-1</sup>, kanamycin 75 µg ml<sup>-1</sup>, chloramphenicol 5 µg ml<sup>-1</sup>, spectinomycin 100 µg ml<sup>-1</sup>. *E. coli* experiments were performed in standard LB and M9 media.

**Culturing and glycerol stock conditions.** An inoculation of –80 °C frozen stock of *V. natriegens* can reach stationary phase after 5 h when incubated at 37 °C. Prolonged overnight culturing (>15 h) at 37 °C may lead to an extended lag phase on subculturing. Routine overnight culturing of *V. natriegens* was performed for 8–15 h at 37 °C or 12–24 h at room temperature. Unless otherwise indicated, *E. coli* cells used in this study were K-12 subtype MG1655, cultured overnight (>10 h) at 37 °C. *V. cholerae* O395 was cultured overnight (>10 h) in LB at 30 °C or 37 °C in a rotator drum at 150 r.p.m. To prepare *V. natriegens* cells for –80 °C storage, an overnight culture of *V. natriegens* was washed in fresh medium before storage in glycerol. Cultures were centrifuged for 1 min at 20,000g and the supernatant was removed. The cell pellet was resuspended in fresh LB3 medium and glycerol was added to 20% final concentration. The stock was quickly vortexed and stored at –80 °C. Bacterial glycerol stocks stored in this manner are viable for at least five years.

**Bulk measurements of generation time.** Growth was measured by kinetic growth monitoring (Biotek H1, H4 or Eon plate reader) in 96-well plates with continuous orbital shaking and optical density (OD) measurement at 600 nm (OD<sub>600nm</sub>) taken every 2 min. Cells grown overnight were washed once in fresh growth medium, then subcultured with at least 1:100 dilution. To assay *V. natriegens* growth in different rich media, cells were cultured overnight from frozen stock into the respective media. To assay *V. natriegens* and *E. coli* growth in minimal medium, cells were cultured overnight in LB3 and LB respectively, and subcultured in the appropriate test medium. Generation times were calculated by linear regression of the log-transformed OD<sub>600nm</sub> across at least three data points when growth was in exponential phase. To avoid spurious determination of growth rates due to measurement noise, the minimal OD<sub>600nm</sub> considered for analysis was maximized and the OD<sub>600nm</sub> values were smoothed with a moving-average window of three data points for conditions that were challenging for growth. Maximal growth rates

were computed from generation times. Apparent lag times were estimated with a fitted model-free smoothed spline using the *grofit* R package.

**Construction of microfluidics device for single-cell imaging.** Microfluidics devices were used as tools to measure and compare growth rates of *E. coli* and *V. natriegens*. In these devices, cells are grown in monolayer and segmented and tracked in high temporal resolution using time-lapse microscopy. The cells are constricted for imaging using previously described Tesla microchemostat device designs, in which cell traps have heights that match the diameters of the cells, minimizing movement and restricting growth to a monolayer<sup>31</sup>. Different trapping heights of 0.8  $\mu\text{m}$  and 1.1  $\mu\text{m}$  were used for *E. coli* and *V. natriegens*, respectively. Microfluidic devices were fabricated with polydimethylsiloxane (PDMS) (PDMS/Sylgard 184, Dow Corning) using standard soft lithographic methods. In brief, microfluidic devices were fabricated by reverse moulding from a silicon wafer patterned with two layers of photoresist (one for the cell trap and another for flow channels). First, the cell trap layer was fabricated by spin coating SU-8 2 (MicroChem) negative resist at 7,000 r.p.m. and 6,800 r.p.m. for *E. coli* and *V. natriegens*, respectively, and patterned using a high resolution photomask (CAD/Art Services). Next, AZ4620 positive photoresist (Capitol Scientific) was spun onto the silicon wafer and aligned with another photomask for fabrication of  $\sim 8\text{-}\mu\text{m}$ -tall flow channels (same for both organisms). Reverse-moulded PDMS devices were punched and bonded to no. 1.5 glass coverslips (Fisher Scientific).

**Time-lapse microscopy and image analysis.** Cells were diluted down to  $\text{OD}_{600\text{nm}} = 0.1$  from an overnight culture in optimal growth conditions and allowed to grow for an 1 h in the corresponding medium and conditions (for example, temperature and salt concentration) before loading into the microfluidics device. Next, cells were loaded and grown in the device under the corresponding environmental conditions until the cell trap chambers were filled. Temperature was maintained with a Controlled Environment Microscope Incubator (Nikon Instruments). Medium flow in the microfluidics device was maintained after cell loading by a constant pressure of 5 psi over the course of the experiment. During the experiment, phase-contrast images were acquired every minute with a  $\times 100$  objective (Plan Apo Lambda  $\times 100$ , numerical aperture 1.45) using an Eclipse Ti-E inverted microscope (Nikon Instruments), equipped with the Perfect Focus system, a motorized stage and a Clara-E charge-coupled device camera (Andor Technology). After the experiment, individual cells were segmented from the image time course using custom MATLAB (Mathworks) software. Doubling time of cells was scored well before the density of the chamber had an impact on tracking and growth of cells. Results from repeat experiments on different days and in different microfluidics devices were consistent (data not shown).

**Genome sequencing by Pacific Biosciences sequencing, de novo assembly and annotation.** *V. natriegens* (ATCC 14048) was cultured for 24 h at 30°C in Nutrient Broth with 1.5% NaCl according to ATCC instructions. Genomic DNA was purified (Qiagen Puregene Yeast/Bact. Kit B) and sequenced on a Single Molecule Real Time (SMRT) Pacific Biosciences RS II system (University of Massachusetts Medical School Deep Sequencing Core) using 120 min movies on three SMRTCells. SMRTAnalysis v.2.1 on Amazon Web Services was used to process and assemble the sequencing data. The mean read length, after default quality filtering, was 4,407 bp. HGAP3 with default parameters was used to assemble the reads which yielded two contigs. The contigs were visualized with Gepard and manually closed. The two closed chromosomes were annotated using RAST under ID 691.12. The annotated genome was deposited in NCBI under Biosample SAMN03178087, GenBank CP009977-8 and RefSeq NZ\_CP009977-8. Base-modification detection was performed on SMRTAnalysis v.2.1 with default settings and the closed genome as reference. Codon usage was calculated using EMBOSS cusp.

**Mapping chromosomal origins and termini by Oxford Nanopore sequencing.** *V. natriegens* was cultured in LB3 and *E. coli* was cultured in LB. Both cultures were grown overnight at 37°C. For stationary phase samples, 1 ml of each culture was collected for genomic DNA extraction. For exponential phase samples, each culture was subcultured and grown to  $\text{OD}_{600\text{nm}} \sim 0.4$  and 10 ml of each was collected for genomic DNA extraction. Genomic DNA was purified (Qiagen Puregene Yeast/Bact. Kit B). To maximize read length,  $\sim 1\ \mu\text{g}$  of genomic DNA for each sample was used as input. One-dimensional sequencing libraries were prepared, barcoded (SQK-RAD002 and SQK-RBK001), and sequenced on the MinION with SpotON R9.4 flow cells for 48 h. Cloud base-calling and sample demultiplexing was performed on Metrichor 1.4.5 and FASTQ files prepared from FAST5 HDF files with a custom Python script. Sequences were aligned to the reference genome using GraphMap 0.5.1. Coverage was computed with BEDTools 2.26.0 and GC skew was computed using the *iRep* package.

**Transcriptome profiling.** Triplicate *V. natriegens* cultures were grown overnight from  $-80^\circ\text{C}$  stocks for each condition to be assayed: 30°C in LB3, 37°C in LB3 and 37°C in M9 high-salt medium supplemented with 2% (w/v) final NaCl and 0.4% (w/v) glucose. Each culture was subcultured in the desired conditions and grown to exponential phase ( $\text{OD}_{600\text{nm}} = 0.3\text{--}0.6$ ). To collect RNA, 10 ml of each culture was stabilized with Qiagen RNeasy Protect Bacteria Reagent and frozen at  $-80^\circ\text{C}$ .

RNA extraction was performed with Qiagen RNeasy Mini Kit and rRNA was depleted with Illumina Ribo-Zero rRNA Removal Kit (Bacteria). Samples were spot-checked for RNA sample quality on an Agilent 2100 RNA 6000 Nano Kit to ensure that the RNA integrity number was  $>9$ . Sequencing libraries were prepared with the NEXTFlex Rapid Directional qRNA-Seq Kit. Each sample was barcoded and amplified with cycle-limited real-time PCR with KAPA SYBR FAST. Resulting libraries were sequenced with MiSeq v.3 150 to obtain paired-end reads.

Sequences were trimmed with cutadapt. Transcripts were quantified with Salmon 0.8.1 and counts were summarized with tximport for differential expression analysis with DESeq2.

**Construction and analysis of transposon mutant libraries.** To facilitate transposon mutagenesis, we engineered a suicide mariner-based transposon vector modified for insertion mapping by high-throughput sequencing<sup>12,32</sup>. Our conjugative suicide mariner transposon plasmid was propagated in BW29427, an *E. coli* with diaminopimelic acid auxotrophy. BW29427 growth requires 300  $\mu\text{M}$  diaminopimelic acid even when cultured in LB. Importantly, BW29427 does not grow in the absence of diaminopimelic acid, which simplifies counterselection of this host strain following biparental mating with *V. natriegens*. For conjugation from *E. coli* to *V. natriegens*, 24 ml of each strain was grown to  $\text{OD}_{600\text{nm}} = 0.4$ , spun down, resuspended and plated on LB2 plates (LB with 2% (w/v) final NaCl concentration) and incubated at 37°C for 60 min (see Supplementary Fig. 4 for optimization of conjugation conditions). This conjugation time was chosen to minimize clonal amplification, on the basis of optimization experiments using 100  $\mu\text{l}$  of each strain. The cells were recovered from the plate in 1 ml LB3. The resulting cell resuspension was washed once in fresh LB3, resuspended to a final volume of 1 ml, and plated on 245 mm  $\times$  245 mm kanamycin-selective plates (Corning). Plates were incubated at 30°C for 12 h to allow the formation of *V. natriegens* colonies. Colonies were scraped from each plate with 3 ml LB3, gently vortexed, and stored as glycerol stocks as previously described. No colonies were detected in control experiments with only BW29427 donor cells. A similar protocol was used to generate an *E. coli* transposon mutant library, except LB was used as the medium in all steps.

For analysis of the transposon mutant library, genomic DNA was extracted (Qiagen DNeasy Blood and Tissue Kit), and digested with MmeI. To enrich for the fragment corresponding to the kanamycin-transposon fragment, the digested genomic DNA was electrophoresed on a 1% TAE gel and an area of the gel corresponding to approximately 1.2 kb was extracted. The resulting DNA fragment was sticky-end ligated to an adapter. PCR was used to selectively amplify the region around the transposon mosaic end and to add the required Illumina adapters. These amplicons were sequenced ( $1 \times 50$  bp) on a MiSeq. Since properly prepared amplicons contain 16 or 17 bp of genomic DNA and 32 or 33 bp of the ligated adapter, only those sequencing reads with the presence of the adapter were analysed further. All adapters were trimming and the resulting genomic DNA sequences were aligned to the reference genome with Bowtie. Statistical enrichment of RAST categories were computed with the hypergeometric test and resulting *P* values were adjusted with Benjamini–Hochberg correction. Transposon mutagenesis data is provided in Supplementary Table 8.

For the *E. coli* Himar1 mutant library, we isolated  $1.1 \times 10^6$  transconjugants, prepared Tn-Seq fragments as previously described, and analysed them by MiSeq<sup>12</sup>. We obtained  $6.9 \times 10^6$  total reads, of which 1.6% mapped to the transposon plasmid; 98.3% of filtered reads were mapped to the genome. These insertions represent 107,723 unique positions, where  $>10$  unique insertions were present in 3,169 out of 4,917 features. For the *V. natriegens* Himar1 mutant library we isolated  $8.6 \times 10^5$  mutants. We obtained  $6.1 \times 10^6$  reads, of which 36.4% mapped to the transposon plasmid; 97.2% of filtered reads were mapped to the genome. These insertions represent 4,530 unique positions, proportionally distributed between the two chromosomes where  $>1$  unique insertion was found in 2,357 out of 4,940 features.

**Electroporation protocol for DNA transformation of *V. natriegens*.** An overnight *V. natriegens* culture was pelleted, washed once in fresh medium, and diluted 1:100 into growth medium. Cells were harvested at  $\text{OD}_{600\text{nm}} \sim 0.4$  (1 h growth when incubated at 37°C at 225 r.p.m.) and pelleted by centrifugation at 3500 r.p.m. for 5 min at 4°C. The pellet was washed three times using 1 ml cold 1 M sorbitol and centrifuged at 20,000g for 1 min at 4°C. The final cell pellet was resuspended in 1 M sorbitol as a 200-fold concentrate of the initial culture. For long term storage, the concentrated competent cells were aliquoted in 50- $\mu\text{l}$  shots in pre-chilled tubes, snap frozen in dry ice and ethanol, and stored at  $-80^\circ\text{C}$  for future use. For transformation, 50 ng plasmid DNA was added to 50  $\mu\text{l}$  of concentrated cells in 0.1-mm cuvettes and electroporated using a Bio-Rad Gene Pulser electroporator at 0.4 kV, 1 k $\Omega$  and 25  $\mu\text{F}$  and recovered in 1 ml LB3 or SOC3 medium for 45 min at 37°C at 225 rpm, and plated on selective medium. Plates were incubated at least 6 h at 37°C or at least 12 h at room temperature.

**Plasmid construction.** Routine cloning was performed by PCR of desired DNA fragments, assembly with NEB Gibson Assembly or NEBuilder HiFi DNA Assembly, and propagation in *E. coli* unless otherwise indicated. Plasmids used are listed in Supplementary Table 4. Plasmid maps are provided in Supplementary Fig. 5.

We used pRSF for the majority of our work since it harbours its own replication machinery and should be minimally dependent on host factors. For transformation optimizations, we constructed pRSF-pLtetO-gfp, which constitutively expresses GFP due to the absence of the tetR repressor in both *E. coli* and *V. natriegens*. We engineered pCTX-R6K shuttle plasmid by fusing the pCTX-Km replicon, comprising genes RstR and RstA, with the pir-dependent conditional replicon R6K. After electroporation of this shuttle plasmid into *V. natriegens*, we were able to extract plasmid from a standard miniprep, demonstrating plasmid replication. To construct the conjugative suicide mariner transposon, we replaced the Tn5 transposase and Tn5 mosaic ends in pBAM1 with the mariner C9 transposase and the mariner mosaic ends from pTnFGL3<sup>33</sup>. Our payload, the transposon DNA, consisted solely of the minimal kanamycin-resistance gene required for transconjugant selection. Next, we performed site-directed mutagenesis on both transposon mosaic ends to introduce an MmeI cut site, producing the plasmid pMarC9, which is also based on the pir-dependent conditional replicon R6K. We also constructed a transposon plasmid capable of integrating a constitutively expressing GFP cassette in the genome by inserting pLtetO-GFP with either kanamycin or spectinomycin in the transposon DNA. All plasmids carrying the R6K origin were found to replicate only in BW29427 or EC100D *pir*<sup>+</sup> and *pir*<sup>-</sup> 116 *E. coli* cells. Induction systems were cloned onto the pRSF backbone. For CRISPR-Cas9 experiments, a single RSF1010 plasmid carried both *S. pyogenes* Cas9 and gRNA. dCas9 was cloned under the control of *E. coli* arabinose-induction genes and the gRNA under control of the constitutive J23100 promoter.

**DNA yield.** pRSF-pLtetO-gfp was transformed into *E. coli* MG1655 and *V. natriegens* by electroporation. *E. coli* plates were incubated at 37°C and *V. natriegens* plates were incubated at room temperature for an equivalent time to yield similar colony sizes. Three colonies were picked from each plate and grown for 5 h at 37°C in 3 ml selective liquid culture (LB for *E. coli* and LB3 for *V. natriegens*) at 225 r.p.m. Plasmid DNA was extracted from 3 ml of culture (Qiagen Plasmid Miniprep Kit).

**CTX vibriophage infection.** *V. cholerae* O395 carrying the replicative form of CTX, pCTX-Km, was cultured overnight in LB without selection in a rotator drum at 150 r.p.m. at 30°C. Virions were purified from cell-free supernatant (0.22- $\mu$ m filtered) of overnight cultures. Replicative forms were extracted from the cells by standard miniprep (Qiagen). To test infectivity of the virions, naive *V. cholerae* O395 and *V. natriegens* were subcultured at 1:1,000 in LB and LB3, respectively and mixed gently with  $\sim 10^6$  virions. After static incubation for 30 min at 30°C, the mixture was plated on selective medium and incubated overnight for colony formation. Replicative forms were electroporated into host strains using the electroporation protocols described above.

**Targeted gene perturbation by Cas9.** All Cas9 experiments were performed using a single pRSF plasmid (pRSF-paraB-Cas9-gRNA) carrying the *cas9* gene under the control of the arabinose promoter, with or without GFP-targeting gRNA. All plasmids carried a carbenicillin-selective marker. Wild-type *V. natriegens* or strain carrying genomically integrated GFP construct were grown at 37°C overnight (LB3 or LB3 with 100  $\mu$ g ml<sup>-1</sup> kanamycin, respectively) and transformed with 50 ng plasmid DNA using the optimized transformation protocol described above. Following 1 h recovery in LB3 at 37°C, cells were plated on LB3 with 100  $\mu$ g ml<sup>-1</sup> carbenicillin and incubated overnight at 37°C. No arabinose induction was used for Cas9 experiments, as we observed a low level of baseline expression using the arabinose promoter.

**Repression of chromosomally encoded GFP with dCas9.** We transformed this engineered *V. natriegens* strain with a CRISPRi plasmid (pRSF-paraB-dCas9-gRNA) carrying dCas9 under arabinose promoter and gRNA targeting GFP. To test the repression of the chromosomally encoded GFP with CRISPRi, we subcultured an overnight culture 1:1,000 in fresh medium with or without 1 mM arabinose. We kinetically measured OD<sub>600nm</sub> and fluorescence of each culture over 12 h in a microplate with orbital shaking at 37°C (BioTek H1 or H4). Under these conditions, all cultures grew equivalently according to OD<sub>600nm</sub> measurements.

**Pooled CRISPRi screen using a five-member gRNA library.** We used dCas9 (pdCas9 bacteria were a gift from S. Qi; Addgene plasmid # 44249) under the control of tetracycline promoter. gRNA was expressed under the constitutive J23100 promoter (plasmid pCTX-R6K-gRNA). Five pCTX-R6K plasmids (spectinomycin-selective marker), each carrying a gRNA, were used for targeted inhibition of the following genes: *V. natriegens* targeting genes *lptF<sub>vi</sub>*, and *flgC<sub>vi</sub>*; and three control targets that are not present in the host: *E. coli* gene *lptF<sub>ec</sub>* and two *gfp* targets. All guides were designed to target the non-template strand. An equal mix of all five plasmids, 20 ng each, was co-transformed into a dCas9 expressing *V. natriegens* strain. The transformation was recovered in 1 ml SOC3 medium for 45 min at 37°C at 225 r.p.m. and plated on 245 mm  $\times$  245 mm plates (Corning) with appropriate antibiotics. After 13 h at 37°C, colonies were scraped in LB3. Growth competition was performed by subculturing this library 1:1,000 in LB3 at 37°C for 3 h in baffled 250 ml flasks (Corning). At each time point, gRNA plasmid was extracted from 3 ml of culture using a Qiagen Plasmid Miniprep Kit. Barcoded

Illumina sequencing libraries were prepared by cycle-limited PCR with real-time PCR and sequenced with MiSeq v3 150. Resulting sequences were trimmed for the promoter and gRNA scaffold and the count of each guide sequence was first normalized by the number of sequences per time point, then expressed as a fraction of the sequence before growth competition.

**Construction, testing and analysis of a genome-wide gRNA library.** A custom Python script was used to select gRNA sequences targeting the non-template strand of each RAST-predicted protein-coding gene. Starting at the 5' end of the gene, 20-bp sequences with a terminal Cas9 NGG motif on the reverse complement strand were selected. Up to three targets were selected for each RAST-predicted gene features; each guide sequence was prefixed with a promoter and suffixed with part of the gRNA scaffold. This sequence was synthesized by the OLS process (Agilent Technologies) as an oligonucleotide library. The OLS pool was amplified by cycle-limited real-time PCR, and assembled into the pCTX-R6K-gRNA backbone (NEBuilder HiFi) at fivefold molar excess with 18-bp overlap arms. Six microlitres of the assembled product was mixed with 300  $\mu$ l TransforMax EC100D *pir*<sup>+</sup> *E. coli* (Epicentre) and 51- $\mu$ l aliquots of this mix were electroporated in 0.1 mm cuvettes with a Bio-Rad Gene Pulser electroporator at 1.8 kV, 200  $\Omega$  and 25  $\mu$ F. These *E. coli* transformants were recovered in 6  $\times$  1 ml SOC medium for 60 min at 37°C at 225 r.p.m., and plated on 245 mm  $\times$  245 mm spectinomycin-selective plates (Corning). After 13 h at 37°C,  $\sim 1.4 \times 10^6$  colonies were scraped and plasmid DNA was extracted (Qiagen HiSpeed Plasmid Maxi).

The plasmid gRNA library was transformed into wild-type *V. natriegens*, or a wild-type strain expressing dCas9 (pdCas9 bacteria). No inducer was added for dCas9 expression, as the earlier results indicated that leaky expression of dCas9 was sufficient for inhibition (Supplementary Fig. 9). In brief,  $\sim 600$  ng of the plasmid library was mixed with 300  $\mu$ l of electrocompetent cells and 53.5  $\mu$ l of this mix was electroporated in 0.1 mm cuvettes with a Bio-Rad Gene Pulser electroporator at 1.4 kV, 1 k $\Omega$  and 25  $\mu$ F. Each transformation was recovered in 1 ml SOC3 medium for 45 min at 37°C and 225 r.p.m. and plated on 245 mm  $\times$  245 mm plates (Corning) with appropriate antibiotics. After 13 h at 37°C, colonies were scraped in LB3 and stored at  $-80^\circ\text{C}$  as library master stocks. Pooled screens for control and test libraries (with or without the dCas9 plasmid) were performed in duplicate, starting with an initial population of  $<10^7$  cells in 25 ml medium in baffled 250 ml flasks (Corning). For media experiments, plasmids were extracted after 8 h of growth at 37°C in LB3, M9 + 0.4% (w/v) glucose, or M9 + 0.4% (w/v) sucrose. For serial passaging experiments, plasmids were extracted after 8 h of growth at 37°C in LB3 (passage 1) and the culture was then diluted 1:1,000 in 25 ml fresh LB3, yielding  $\sim 10^6$  cells, and grown at 37°C for 4 h (passage 2). Subsequent passaging was performed in 4-h increments. Plasmids were extracted from  $>3$  ml of culture (Qiagen Plasmid Miniprep Kit). Barcoded Illumina sequencing libraries were prepared by cycle-limited real-time PCR and sequenced with NextSeq v2 High Output 500/550. Resulting sequences were trimmed for the promoter and 5'-end of the gRNA scaffold. Sequencing was used to verify coverage of the gRNA library. For media experiments, the initial no dCas9 and dCas9 library contained 15,539 (99.65%) and 13,579 (99.9%) of 13,587 total guides, respectively. For serial passaging experiments, the initial no dCas9 and dCas9 library contained 13,567 (99.94%) and 13,584 (99.98%), 13,587 total guides, respectively. After sequencing all samples, the count of each guide sequence was normalized by the number of reads per sequencing library to yield reads per million (RPM) per guide. The median gRNA copy numbers for the initial no dCas9 and dCas9 media libraries were 70.3 and 62.3 RPM, respectively. The median gRNA copy number for the initial no dCas9 and dCas9 passaging libraries were 70.1 and 62.1 RPM, respectively. These counts were input into MAGeCK and analysed using the mle module with default settings<sup>17</sup> to obtain a fitness value (beta score and FDR-adjusted *P* value) for each gene. Essential genes from *E. coli* and *V. cholerae* were mapped to *V. natriegens* via bactNOG or COG using eggNOG-mapper based on eggNOG 4.5 orthology data<sup>19,21,34</sup>. Enrichments were calculated as a one-sided hypergeometric test assuming independence between functional groups with multiple-hypothesis correction with FDR (Benjamini-Hochberg).

**Construction of in-frame genomic knockouts by natural competency.** The endogenous *tfoX* gene (PEG.1425) was subcloned on to the RSF1010 backbone plasmid with an *E. coli* IPTG-inducible promoter. For each gene target, a deletion cassette, consisting of an antibiotic resistance marker with 500-bp homology arms upstream and downstream, was created by overlap-extension PCR. Generation of deletion mutants was performed as previously described<sup>35</sup>. In brief, the strain carrying *tfoX* was grown overnight in medium containing 1 mM IPTG in a rotating drum at 30°C, then subcultured 1:100 in 350  $\mu$ l 2XIO (28 g l<sup>-1</sup> Ocean Salts) in medium with 1 mM IPTG with  $>100$  ng of the deletion cassette. The culture was incubated statically at 30°C for 5 h, then recovered with 1 ml LBv2<sup>2</sup> (LB Miller broth supplemented with 200 mM NaCl, 23.14 mM MgCl<sub>2</sub> and 4.2 mM KCl) at 30°C for 2 h with shaking at 225–250 r.p.m. Following recovery, cultures were plated on antibiotic-selective plates. Resulting transformants were screened by PCR to confirm gene deletion.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

Genome sequences are available from NCBI (GenBank: CP009977, CP009978; RefSeq: NZ\_CP009977, NZ\_CP009978). Sequencing data for gRNA counts are available at NCBI Sequence Read Archive under BioProject PRJNA511728 (SRR8369136, SRR8369137, SRR8369138, SRR8369139) and transcriptome data is available at the Gene Expression Omnibus under accession number GSE126544 (GSM3603279, GSM3603280, GSM3603281, GSM3603282, GSM3603283, GSM3603284). All other data are available in the Supplementary Information or upon request from the corresponding authors.

**Code availability**

Custom code is available at [https://github.com/citizenlee/vnat\\_glib](https://github.com/citizenlee/vnat_glib) or will be made available upon request.

Received: 14 December 2017; Accepted: 26 February 2019;

Published online: 8 April 2019

**References**

- Eagon, R. G. *Pseudomonas natriegens*, a marine bacterium with a generation time of less than 10 minutes. *J. Bacteriol.* **83**, 736–737 (1962).
- Weinstock, M. T., Heseck, E. D., Wilson, C. M. & Gibson, D. G. *Vibrio natriegens* as a fast-growing host for molecular biology. *Nat. Methods* **13**, 849–851 (2016).
- Long, C. P., Gonzalez, J. E., Cipolla, R. M. & Antoniewicz, M. R. Metabolism of the fast-growing bacterium *Vibrio natriegens* elucidated by <sup>13</sup>C metabolic flux analysis. *Metab. Eng.* **44**, 191–197 (2017).
- Hoffart, E. et al. High substrate uptake rates empower *Vibrio natriegens* as production host for industrial biotechnology. *Appl. Environ. Microbiol.* **83**, e01614-17 (2017).
- Aiyar, S. E., Gaal, T. & Gourse, R. L. rRNA promoter activity in the fast-growing bacterium *Vibrio natriegens*. *J. Bacteriol.* **184**, 1349–1358 (2002).
- Overbeek, R. et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* **42**, D206–D214 (2013).
- Maida, I. et al. Draft genome sequence of the fast-growing bacterium *Vibrio natriegens* strain DSMZ 759. *Genome Announc.* **1**, e00648–13 (2013).
- Thompson, J. R. & Polz, M. F. in *The Biology of Vibrios* (eds Thompson, F. L. et al.) 190–203 (American Society of Microbiology, 2006).
- Heidelberg, J. F. et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**, 477–483 (2000).
- Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**, D93–D97 (2009).
- Brown, C. T., Olm, M. R., Thomas, B. C. & Banfield, J. F. Measurement of bacterial replication rates in microbial communities. *Nat. Biotechnol.* **34**, 1256–1263 (2016).
- van Opijnen, T., Lazinski, D. W. & Camilli, A. Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms. *Curr. Protoc. Microbiol.* **106**, 7.16.1–7.16.24 (2010).
- Rock, J. M. et al. Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR interference platform. *Nat. Microbiol.* **2**, 16274 (2017).
- Peters, J. M. et al. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell* **165**, 1493–1506 (2016).
- Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
- Liu, X. et al. High-throughput CRISPRi phenotyping identifies new essential genes in *Streptococcus pneumoniae*. *Mol. Syst. Biol.* **13**, 931 (2017).
- Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
- Tatusova, T. et al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624 (2016).
- Chao, M. C. et al. 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 (2013).
- Okada, K., Iida, T., Kita-Tsukamoto, K. & Honda, T. Vibrios commonly possess two chromosomes. *J. Bacteriol.* **187**, 752–757 (2005).
- Yamazaki, Y., Niki, H. & Kato, J.-I. Profiling of *Escherichia coli* chromosome database. *Methods Mol. Biol.* **416**, 385–389 (2008).

- Julio, S. M. et al. DNA adenine methylase is essential for viability and plays a role in the pathogenesis of *Yersinia pseudotuberculosis* and *Vibrio cholerae*. *Infect. Immun.* **69**, 7610–7615 (2001).
- Egan, E. S., Fogel, M. A. & Waldor, M. K. Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Mol. Microbiol.* **56**, 1129–1138 (2005).
- Wang, T. et al. Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
- Schleicher, L. et al. *Vibrio natriegens* as host for expression of multisubunit membrane protein complexes. *Front. Microbiol.* **9**, 2537 (2018).
- Lenz, D. H. et al. The small RNA chaperone Hfq and multiple small RNAs control quorum sensing in *Vibrio harveyi* and *Vibrio cholerae*. *Cell* **118**, 69–82 (2004).
- Val, M.-E. et al. A checkpoint control orchestrates the replication of the two chromosomes of *Vibrio cholerae*. *Sci. Adv.* **2**, e1501914 (2016).
- Mee, M. T., Collins, J. J., Church, G. M. & Wang, H. H. Syntrophic exchange in synthetic microbial communities. *Proc. Natl Acad. Sci. USA* **111**, E2149–E2156 (2014).
- Cui, L. et al. A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9. *Nat. Commun.* **9**, 1912 (2018).
- Zhang, S. & Voigt, C. A. Engineered dCas9 with reduced toxicity in bacteria: implications for genetic circuit design. *Nucleic Acids Res.* **46**, 11115–11125 (2018).
- Vega, N. M., Allison, K. R., Khalil, A. S. & Collins, J. J. Signaling-mediated bacterial persister formation. *Nat. Chem. Biol.* **8**, 431–433 (2012).
- Martínez-García, E., Calles, B., Arévalo-Rodríguez, M. & de Lorenzo, V. pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiol.* **11**, 38 (2011).
- 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. USA* **105**, 8736–8741 (2008).
- Huerta-Cepas, J. et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
- Dalia, T. N. et al. Multiplex genome editing by natural transformation (MuGENT) for synthetic biology in *Vibrio natriegens*. *ACS Synth. Biol.* **6**, 1650–1655 (2017).

**Acknowledgements**

We acknowledge J. F. Juárez, J. Teramoto, M. Mee, A. Camilli and J. Aach for comments and discussions; C. Mancuso and M. Joung; Lyubov Golubeva for the pRSF plasmid; B. Davis and M. Waldor for *V. cholerae* strains O395 and BAH-2, and the pCTX-Km and pCTX-Ap plasmids; V. de Lorenzo for the pBAM1 plasmid; D. E. Cameron and J. Mekalanos for the pTnFGL3 plasmid and B. Wanner for the BW29427 strain. This work was supported by Department of Energy Grant DE-FG02-02ER63445 (to G.M.C.), AWS Cloud Credits for Research programme (to H.H.L.) and a National Science Foundation CAREER Award MCB-1350949 (to A.S.K.).

**Author contributions**

H.H.L. and N.O. designed and performed experiments, analysed data and wrote the paper. B.G.W. and A.S.K. designed and performed single-cell microfluidics experiments and provided input on the paper. M.A.G. contributed to the electroporation experiments and formulated recovery media. G.M.C. supervised the project.

**Competing interests**

H.H.L., N.O. and G.M.C. have filed patents related to this work.

**Additional information**

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41564-019-0423-8>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to H.H.L. or G.M.C.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2019

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- |                                     |                                     |  |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A description of all covariates tested   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/>            | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated   |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

N/A

Data analysis

All publicly available software and their versions are detailed in the Materials and Methods. Custom code is available in the GitHub repository indicated under Data Availability or by request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Genome sequences are available in NCBI (GenBank: CP009977, CP009978; RefSeq: NZ\_CP009977, NZ\_CP009978). Sequencing data for gRNA counts are available in NCBI SRA under BioProject PRJNA511728 (SRR8369136, SRR8369137, SRR8369138, SRR8369139) and transcriptome data is available via GSE126544 (GSM3603279, GSM3603280, GSM3603281, GSM3603282, GSM3603283, GSM3603284). All other data are available in the Supplementary Information, or by request.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	<input type="text" value="sample size selection is not relevant to our study."/>
Data exclusions	<input type="text" value="no data were excluded."/>
Replication	<input type="text" value="replicates were successful for all experiments and are included in manuscript data."/>
Randomization	<input type="text" value="sample randomization is not relevant in our study."/>
Blinding	<input type="text" value="blinding is not relevant to our study."/>

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- | n/a                                 | Involvement in the study                             |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines       |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data               |

### Methods

- | n/a                                 | Involvement in the study                        |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |