Functional Genomics of the Rapidly Replicating Bacterium Vibrio Natriegens by CRISPRi

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

Henry H. Lee[1*,†], Nili Ostrov[1*], Brandon G. Wong[2], Michaela A. Gold[1], Ahmad S. Khalil[2,3], George M. Church[1,3,†]

Author affiliations

[1] Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

[2] Department of Biomedical Engineering and Biological Design Center, Boston University, Boston, Massachusetts 02215, USA.

[3] Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, USA.

[*] Equal contribution

[†] Correspondence: gchurch@genetics.med.harvard.edu, hhlee@genetics.med.harvard.edu

Main text

The fast growing Gram-negative bacterium *Vibrio natriegens* is an attractive microbial system for molecular biology and biotechnology due to its remarkable short generation time[1,2] and metabolic prowess[3,4]. 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 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 composed of putative essentials and growth-supporting genes which are enriched for respiratory pathways. We found 96% of core genes to be located on the larger chromosome 1, with growth-neutral duplicates of core gene located primarily on chromosome 2. Our screen also refines metabolic pathway annotations by distinguishing functional biosynthetic enzymes from those predicted based on 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 exploring laboratory culturing conditions which support rapid and consistent *V. natriegens* growth. We opted for Lysogeny Broth supplemented with 3% (w/v) final sodium chloride (LB3), which approximates ocean salinity and provides consistent growth (**Supplementary Figure 1, Supplementary Table 1)**. At 37°C, we measured *V. natriegens*' generation time to be 15.2 minutes in bulk culture and 14.8 minutes using single-cell, time-lapse microscopy (**Figure 1a-b, Supplementary Movie 1, Supplementary Figure 1, Supplementary Figure 2**). Longer generation time (30.8 minutes) was observed in minimal glucose media supplemented with 2% (w/v) sodium chloride (**Supplementary Figure 1**). These generation times are comparable with other studies[3–5].

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[7] (**RefSeq NZ_CP009977-8, Figure 1c, Supplementary Table 2, Supplementary Table 3, Methods**)[6]. We identified 4,578 open reading frames (ORFs), of which ~63% reside on chr1 and ~37% on chr2. Consistent with the broad metabolic capacity described for *Vibrio* spp.[8], nearly half of all annotated ORFs are involved in carbohydrates, RNA and protein metabolism (**Figure 1d**). *V. natriegens* genome carries 11 rRNA operons and 129 tRNA genes, more than *V. cholerae* N16961 (8 and 98, respectively)[9] or *E. coli* MG1655 (7 and 87, respectively)[10]. 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[11], and found a single bidirectional origin on each chromosome (**Figure 1e)** which displayed high sequence similarity to other *Vibrio* spp. (**Supplementary Figure 3**)[9].

We next sought to develop genetic tools to enable high-throughput assessment of gene functionality. While transposon mutagenesis is commonly used to generate single gene mutant libraries[12] , low insertion efficiency prohibited scalable saturation mutagenesis in *V. natriegens* (**Supplementary Table 8, Supplementary Discussion)**. Instead, we turned to CRISPR/Cas9 which has been demonstrated in well-studied model organisms[13–16] 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 GFP, and tested the *Streptococcus pyogenes* Cas9 nuclease with or without GFP-targeting guide RNA (gRNA) (plasmid pRSF-paraB-spCas9-gRNA, **Supplementary Figures 5-7, Supplementary Table 4**). No transformants were detected following electroporation of Cas9 nuclease with a GFP-targeting gRNA, even without induction of Cas9 expression (**Supplementary Figure 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 inhibition (CRISPRi) relies on the nuclease-deficient variant dCas9 to block RNA polymerase activity[15]. Based on design rules elucidated in other model bacteria[14,15], we designed guides targeting the template and non-template strand proximal to 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 (>13-fold, **Figure 2a**). As with Cas9 nuclease, induction of dCas9 was not required for activity and led to longer lag phase indicative of marginal toxicity (**Supplementary Figure 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 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 which target 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*$_{Vn}$, a homolog of an essential *E. coli* gene), one putative non-essential *V. natriegens* gene (*flgC*$_{Vn}$), and two control genes absent from the genome (GFP and *E.coli* gene *lptF*). All gRNAs plasmids (vector pCTX-R6K-gRNA, **Supplementary Figure 8**) were co-transformed into *V. natriegens* bearing dCas9 (plasmid pdCas9-bacteria[15]). Plasmids were extracted and sequenced to determine the abundance of each gRNA over three hours of competitive growth. We observed a significant reduction in abundance of the gRNA targeting the putative essential *lptF*$_{Vn}$ gene, while all other guides remained relatively unchanged (**Figure 2b, Supplementary Figure 10)**. These results demonstrate that gRNA abundance can be used as proxy for a gene's fitness in pooled competitive growth.

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[15]. 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 (**Figure 2c, Supplementary Table 5)**. Library screens were performed in LB3, M9-Glucose, or M9-Sucrose for 8 hours (**Figure 2d**, **Supplementary Figure 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)[17] (**Supplementary Figure 5**). A negative beta score indicates depletion of a gRNA under the tested condition.

Using sucrose as the sole carbon source, we identified fourteen unique genes with a significant negative beta scores (FDR≤0.05) (**Figure 3a, 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 in-frame deletion of both genes and measured each under non-competitive growth conditions. Consistent with CRISPRi results, we found the sucrose-6-phosphate hydrolase located on chr1 (PEG.1381, β_{suc} = -1.6907),

but not the one located on chr2 (PEG.3071, β_{suc} = 0.12), was required for growth in M9-sucrose media (**Figure 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≤0.05) enriched in RAST subcategories for amino acid and vitamin biosynthesis (**Figure 3c, Supplementary Figure 12a**). When mapped onto the respective pathways, this set can be used to distinguish functional genes from a larger set of computationally assigned biosynthetic enzymes[7,18]. For example, of all predicted L-leucine biosynthetic genes, only one of two acetolactate synthase complexes (PEGs 2236, 2237 and 2615, 2616) and one of two isopropylmalate synthases (PEGs 2243 and 3818) was identified as functional in our screen (**Figure 3d**). Similar results were obtained for tryptophan and isoleucine biosynthetic pathways. Moreover, we show functional enzymes are not always the most differentially expressed (**Figure 3d**, **Supplementary Figure 12b,c**). Thus, while transcriptional enrichment corroborates the importance of amino acid biosynthesis pathways in minimal media (**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 media (LB3). In our first experiment, we found 646 genes with significant negative beta score (**Figure 3a**). When repeated, the same screen resulted in a slightly larger number of significant genes (735), though further passaging of the culture yielded only 648 significant negative scoring genes. Of these, 587 (92.4%) overlap between the two experiment (**Figure 4a, Supplementary Figure 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' which are required for rapid growth in rich media (**Supplementary Table 6**).

We further analyzed the passaging experiments for evidence of CRISPRi escape by examining 100 genes which were identified as significant in the first, but not the last passage (**Supplementary Figure 13e, Supplementary Table 5**). We expect CRISPRi escape to result in increased beta score over passages. However, the majority of these genes (73%) displayed a decrease in beta score, suggesting that rather than escaping dCas9 inhibition, they likely lost significance due to increase in the total number of low scoring genes (**Supplementary Figure 13f**). Of the remaining 27 genes which increase in beta score over passaging, only seven were altered by more than 20% (**Supplementary Figure 14a**). For these, we found one of three gRNA was repeatedly elevated in the third passage (**Supplementary Figure 14b**), indicating that increase in beta score is due to an inefficient or promiscuous gRNA sequence rather than 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 for core genes

was 5.6-fold higher than 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 *Vibrio cholerae* genome[19]. We found 278 homologous genes, corresponding to 84.4% of the 329 mappable essential genes in *V. cholerae*, which we designate putative essential genes in *V. natriegens* (**Figure 4b**, **Supplementary Table 6**).

Similar to other *Vibrio* spp.[20], the vast majority of putative essentials are located on chr1 (272 of 278 genes). Putative essentials were statistically enriched for RAST categories including 'Protein metabolism' and 'Cell division and cell cycle', which include ribosomal proteins, tRNA synthetases, and DNA polymerase (**Figure 4c**, **Supplementary Figure 15**). For example, 49 of 55 total ribosomal proteins are included in the core set, of which 44 (89.8%) are in agreement with essentiality in *V. cholerae* and 37 (75.5%) are in agreement with essentiality in *E .coli*, as determined by transposon mutagenesis[19] and gene deletions[21], respectively (**Figure 4d**, **Supplementary Table 6)**. Other notable essentials include DNA adenine methylase *dam*[22] and *nadE*, a NAD+ synthase gene implicated in the conversion of chr2 from a megaplasmid into a critical component of the *Vibrio* genome[23]. Putative essential genes, which were all actively transcribed, showed lower beta scores and higher transcription levels than other core genes (**Figure 4e-f**, **Supplementary Figure 15a**), in line with reports of essential gene expression in other organisms[24].

Non-essential core genes (309 of 587 genes) were uniquely enriched for the RAST category 'Respiration' (**Figure 4c**, **Supplementary Figure 15b**). This set of 'growth-supporting' genes include the Rnf electron transport complex, cytochrome C oxidase biogenesis enzymes, and the Na+-NQR respiration complex implicated in aerobic respiration in halophilic bacteria[25] (**Supplementary Table 6**). These pathways are likely responsible for the high oxygen uptake and highly active electron transport chain activity[3] required to maintain rapid cell division. Others, such as quorum sensing regulator *hfq*[26], may determine the rapid growth rate observed in highly dilute cultures[4]. 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/EL, 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 Figure 16a**). Only five genes were found where both duplicates are 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 60.3% of

core genes with duplicate (53 of 84 genes) are located on chr1 with the duplicate located on chr2 (**Figure 4g**), contributing to the relatively large size of chr2 in *V. natriegens* compared to other *Vibrio spp*.[20]. Duplicated genes, previously reported in *V. cholerae*[19], may represent heterofunctional homologs or pseudogenes. Transcriptional profiling shows the median TPM for core genes is ~17-fold higher than that of duplicates (125.6 and 7.4, respectively), suggesting that expression of duplicates may be silenced under tested conditions (**Supplementary Figure 16b**). Further insight into the genesis and role of these gene duplications may benefit from studies of interchromosomal interactions[27] 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 for critical pathway enzymes[28]. 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[14]. 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, screen results may be affected by dCas9-specific toxicity, as observed for *recQ* helicase (**Supplementary Discussion, Supplementary Figure 17**). Additional studies using in-frame deletions are warranted to stringently evaluate the phenotype of single gene knockout.

Nevertheless, screening sensitivity and specificity may be further increased by improvements to the gRNA library and dCas9 protein. While 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 not captured by computational annotation. Additional design principles for CRISPRi libraries[16,29] may be used to improve gRNA binding specificity or reduce potential sequence-specific toxicity[29]. Furthermore, gene dosage effects can be explored by tuning transcriptional knockdown via engineering of *V. natriegens* promoters or synthetic dCas9 with reduced toxicity[30]. 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 the use of pooled genome-wide CRISPRi screens will enable rapid and high-throughput functional annotation of diverse bacteria.

Taken together, we provide initial genome-wide experimental annotations for *V. natriegens* which 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

provide valuable experimental guidance for construction of a fast growing minimal genome by bottom up construction or large-scale deletion[2]. With only 2.1% of putative essential genes located on chr2 (**Figure 4e**, **Supplementary Table 6**), spatial distribution of core genes presents intriguing opportunities for rational genome design, such as consolidating of functional genes to chromosome 1 and re-purposing of chromosome 2 as an artificial chromosome. Further investigation of this unique organism will facilitate its advancement as a versatile bacterial system for research and biotechnology.

Supplementary Materials

Supplementary Tables 1 – 8

Supplementary Figures 1 – 17

Supplementary Movie 1

Correspondence to gchurch@genetics.med.harvard.edu, hhlee@genetics.med.harvard.edu

Acknowledgements

We would like to acknowledge Javier Fernández Juárez, Jun Teramoto, Michael Mee, Andrew Camilli, and John Aach for helpful comments and discussions. Ahmad Khalil would additionally like to acknowledge Christopher Mancuso and Madeleine Joung. We thank Lyubov Golubeva for the pRSF plasmid, Brigid Davis and Matthew Waldor for *Vibrio cholerae* strains O395, BAH-2, and pCTX-Km and pCTX-Ap plasmids, Victor de Lorenzo for pBAM1 plasmid, D. Ewen Cameron and John Mekalanos for pTnFGL3 plasmid and Barry Wanner for BW29427 strain. This work was supported by Department of Energy Grant DE-FG02-02ER63445 (to G.M.C.), AWS Cloud Credits for Research program (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, performed experiments, analyzed data, and wrote the paper. B.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 financial interests

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

Materials and Methods

Strains used in this study

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

Growth Media

Unless denoted, LB3, Lysogeny Broth with 3% (w/v) final NaCl, is used as standard rich media. We prepare this media by adding 20 grams of NaCl to 25 grams of LB Broth - Miller (Fisher BP9723-500). Media are formulated according to manufacturer instructions and supplemented with 1.5% (w/v) final Ocean Salts (Aquarium System, Inc.) to make high salt versions of Brain Heart Infusion (BHIO), Nutrient Broth (NBO), and Lysogeny Broth (LBO). No additional salts were added to Marine Broth (MB). Minimal M9 media was prepared according to manufacturer instruction. For culturing *V. natriegens*, 2% (w/v) NaCl was added to M9 (in addition to the pre-included 0.5 g/L NaCl) based on the screen of NaCl concentrations in rich media. Carbon sources were added as indicated to 0.4% (v/v) final. SOC3 media is composed of 5 grams of yeast extract, 20 grams tryptone, 30 grams sodium chloride, 2.4 grams magnesium sulfate, and 0.4% (w/v) final glucose. Antibiotic concentrations used for plasmid selection in *V. natriegens*: ampicillin/carbenicillin 100 µg/ml, kanamycin 75 µg/ml, chloramphenicol 5µg/ml, spectinomycin 100 µg/ml. *E. coli* experiments were performed in standard LB media and M9.

Culturing and glycerol stock conditions

An inoculation of -80°C frozen stock of *V. natriegens* can reach stationary phase after 5 hours when incubated at 37°C. Prolonged overnight culturing (>15 hours) at 37°C may lead to an extended lag phase upon subculturing. Routine overnight culturing of *V. natriegens* was performed for 8-15 hours at 37°C or 12-24 hours at room temperature. Unless otherwise indicated, *E. coli* cells used in this study were K-12 subtype MG1655 cultured overnight (>10 hours) at 37°C. *V. cholerae* O395 was cultured overnight (>10 hours) in LB at 30°C or 37°C in a rotator drum at 150rpm. To prepare *V. natriegens* cells for -80°C storage, an overnight culture of *V. natriegens* is washed in fresh media before storing in glycerol. Cultures were centrifuged for 1 minute at 20,000 rcf and the supernatant was removed. The cell pellet was resuspended in fresh LB3 media and glycerol was added to 20% final concentration. The stock is quickly vortexed and stored at -80°C. Bacterial glycerol stocks stored in this manner are viable for at least 5 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 600nm taken every 2 minutes. Overnight cells were washed once in fresh growth media, then subcultured by 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 media, cells were cultured overnight in LB3 and LB respectively, and subcultured in the appropriate test media. Generation times were calculated by linear regression of the log-transformed OD across at least 3 data points when growth was in exponential phase. To avoid specious determination of growth rates due to measurement noise, the minimal OD considered for analysis was maximized and the ODs were smoothed with a moving average window of 3 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 "grofit" R package.

Construction of microfluidics device or single cell imaging

Microfluidic 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/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 in a monolayer [31]. Different trapping heights of 0.8 μm and 1.1 μm were used for *E. coli* and *V. natriegens,* respectively. Microfluidic devices were fabricated with polydimethylsiloxane (PDMS/Sylgard 184, Dow Corning) using standard soft lithographic methods. Briefly, microfluidic devices were fabricated by reverse molding from a silicon wafer patterned with two layers of photoresist (one for the cell trap, another for flow channels). First, the cell trap layer was fabricated by spin coating SU-8 2 (MicroChem Corp.) negative resist at 7000 RPM and 6800 RPM for *E. coli* and *V. natriegens,* respectively, and patterned using a high resolution photomask (CAD/Art Services, Inc.). Next, AZ4620 positive photoresist (Capitol Scientific, Inc.) was spun onto the silicon wafer and aligned with another photomask for fabrication of ~8 μm tall flow channels (same for both organisms). Reverse-molded 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 0.1 OD_{600} from an overnight culture at optimal growth conditions and allowed to grow for an hour in the corresponding media conditions (e.g. temperature, salt concentration) before loading onto the device. Next, cells were loaded and grown on the device in the corresponding environmental conditions until the cell trap chambers filled. Temperature was maintained with a Controlled Environment Microscope Incubator (Nikon Instruments, Inc.). Media flow on device was maintained by a constant pressure of 5 psi over the course of the experiment after cell loading. During the experiment, phase contrast images were acquired every minute with a 100x objective (Plan Apo Lambda 100X, NA 1.45) using an Eclipse Ti-E inverted microscope (Nikon Instruments, Inc.), equipped with the "Perfect Focus" system, a motorized stage, and a Clara-E charge-coupled device (CCD) camera (Andor Technology). After the experiment, individual cells were segmented from the image time course using custom

MATLAB (Mathworks, Natick, MA) software. Doubling time of cells was scored well before the density of the chamber impacted tracking and growth of cells. Results from repeat experiments on different days and devices were consistent (Data not shown).

Genome sequencing by Pacific Bioscience sequencing, *de novo* assembly, and annotation

V. natriegens (ATCC 14048) was cultured for 24 hours 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 minute movies on 3 SMRTCells. SMRTanalysis v2.1 on Amazon Web Services was used to process and assemble the sequencing data. The mean read length, after default quality filtering, was 4,407bp. HGAP3 with default parameters was used to assemble the reads which yielded 2 contigs. The contigs were visualized with Gepard and manually closed. The two closed chromosomes annotated using RAST under ID 691.12. The annotated genome is deposited in NCBI under Biosample SAMN03178087, GenBank CP009977-8, RefSeq NZ_CP009977-8. Base modification detection was performed on SMRTanalysis v2.1 with default setting 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, 1mL of each culture was collected for genomic DNA extraction. For exponential phase samples, each culture was subcultured and grown to OD_{600} ~0.4 and 10mL of each was collected for genomic DNA extraction. Genomic DNA was purified (Qiagen Puregene Yeast/Bact. Kit B). To maximize read length, ~1µg of genomic DNA for each sample was used as input. 1D sequencing libraries were prepared, barcoded (SQK-RAD002 and SQK-RBK001), and sequenced on the MinION with SpotON R9.4 flow cells for 48 hours. 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 computed using the iRep package.

Transcriptome profiling

Triplicate *V. natriegens* cultures were grown overnight from -80°C stocks for each condition to be assayed: 30°C in LB3, 37°C in LB3 and 37°C in M9 high-salt media supplemented with 2% (w/v) final sodium chloride and 0.4% (w/v) glucose. Each culture was subcultured in the desired conditions and grown to exponential phase (OD_{600} 0.3-0.6). To collect RNA, 10mL of each culture was stabilized with Qiagen RNAprotect Bacteria Reagent and frozen at -80°C. RNA extraction was performed with Qiagen RNeasy Mini Kit and rRNA 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 (RIN) was above 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 v3 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 [12,32]. Our conjugative suicide mariner transposon plasmid was propagated in BW29427, an *E. coli* with diaminopimelic acid (DAP) auxotrophy. BW29427 growth requires 300μM of DAP even when cultured in LB. Importantly, BW29427 does not grow in the absence of DAP, which simplifies counterselection of this host strain following biparental mating with *V. natriegens*. For conjugation from *E. coli* to *V. natriegens*, 24mL of each strain was grown to OD 0.4, spun down, resuspended and plated on LB2 plates (Lysogeny Broth with 2% (w/v) final of sodium chloride) and incubated at 37°C for 60 minutes (see **Supplementary Figure 4** for optimization of conjugation conditions). This conjugation time was chosen to minimize clonal amplification, based on optimization experiments using 100uL of each strain. The cells are recovered from the plate in 1mL of LB3 media. The resulting cell resuspension is washed once in fresh LB3, resuspended to a final volume of 1mL, and plated on 245mm x 245mm kanamycin selective plates (Corning). Plates were incubated at 30°C for 12 hours to allow the formation of *V. natriegens* colonies. Colonies were scraped from each plate with 3mL of LB3, gently vortexed, and stored as glycerol stock 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 media at all steps.

For analysis of transposon mutant library, genomic DNA was extracted (Qiagen DNeasy Blood & 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 1.2kb 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 1x50bp on a MiSeq. Since properly prepared amplicons contain 16 or 17bp of genomic DNA and 32 or 33bp of the ligated adapter, only those sequencing reads with the presence of the adapter were further analyzed. 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 $1.1x10^6$ transconjugants, prepared Tn-Seq fragments as previously described, and analyzed by MiSeq [12]. We obtained $6.9x10^6$ 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 $8.6x10^5$ mutants. We obtained $6.1x10^6$ 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 insertions were 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 media, and diluted 1:100 into growth media. Cells were harvested at OD_{600}~0.4 (1 hour growth when incubated at 37°C at 225rpm) and pelleted by centrifugation at 3500rpm for 5min at 4°C. The pellet is washed three times using 1ml of cold 1M sorbitol and centrifuged at 20,000rcf for 1 minute at 4°C. The final cell pellet was resuspended in 1M sorbitol at a 200-fold concentrate of the initial culture. For long term storage, the concentrated competent cells were aliquoted in 50µL shots in pre-chilled tubes, snap frozen in dry ice and ethanol, and stored in -80°C for future use. To transform, 50ng of plasmid DNA was added to 50µL of concentrated cells in 0.1mm cuvettes and electroporated using Bio-Rad Gene Pulser electroporator at 0.4kV, 1kΩ, 25µF and recovered in 1mL LB3 or SOC3 media for 45 minutes at 37°C at 225rpm, and plated on selective media. Plates were incubated at least 6 hours at 37°C or at least 12 hours 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. See **Supplementary Table 4** for plasmids used in this work. Plasmid maps are provided in **Supplementary Figure 5**. We used pRSF for the majority of our work since it carries all of 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, comprised of 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[33]. Our payload, the transposon DNA, consisted solely of the minimal kanamycin resistance gene required for transconjugant selection. We next 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 was found to replicate only in either BW29427 or EC100D pir^+/pir-116 *E. coli* cells. Induction systems were cloned onto the pRSF backbone. For CRISPR/Cas9 experiments, a single RSF1010 plasmid carried both *Streptococcus pyogenes* Cas9 and guide RNA. dCas9 was cloned under the control of *E. coli* arabinose induction genes and the guide RNA under control of the constitutive J23100 promoter.

DNA yield

pRSF-pLtetO-gfp was transformed via electroporation into *E. coli* MG1655 and *V. natriegens*. *E. coli* plates were incubated at 37°C and *V. natriegens* were incubated at room temperature for an equivalent time to yield approximately similar colony sizes. Three colonies from each plate was picked and grown for 5 hours at 37°C in 3mL of selective liquid culture (LB for *E. coli* and LB3 for *V. natriegens*) at 225rpm. Plasmid DNA was extracted from 3mL 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 150rpm at 30°C. Virions were purified from cell-free supernatant (0.22μm 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 1:1000 in LB and LB3 respectively and mixed gently with ~10^6 virions. After static incubation for 30 minutes at 30°C, the mixture was plated on selective media and incubated overnight for colony formation. Replicative forms were electroporated into host strains using described protocols.

Targeted gene perturbation by Cas9

All Cas9 experiments were performed using a single pRSF plasmid (pRSF-paraB-Cas9-gRNA) carrying Cas9 gene under the control of arabinose promoter, with or without GFP-targeting guide RNA. All plasmids carry carbenicillin selective marker. Wild-type *V. natriegens* or strain carrying genomically integrated GFP construct were grown at 37°C overnight (LB3 or LB3+100μg/ml kanamycin, respectively) and transformed with 50ng of plasmid DNA using the optimized transformation protocol described above. Following 1-hour recovery in LB3 at 37°C, cells were plated on LB3+100μg/mL carbenicillin plates and incubated overnight at 37°C. No arabinose induction was used for Cas9 experiments, as we observed low level of baseline expression using 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 cultures 1:1000 in fresh media supplemented with or without 1mM arabinose. We kinetically measured OD_{600} and fluorescence of each culture over 12 hours in a microplate with orbital shaking at 37°C (BioTek H1 or H4). Under these conditions, all cultures grew equivalently by OD_{600}.

Pooled CRISPRi screen - five-member gRNA library

We used dCas9 (plasmid pdCas9-bacteria was a gift from Stanley Qi; Addgene plasmid # 44249) under the control of tetracycline promoter. Guide RNA was expressed under 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 $lptF_{Vn}$ and $flgC_{Vn}$; three control targets that do not exist in the host: *E.coli* gene $lptF_{Ec}$ and two GFP guides. All guides were designed to target the non-template strand. An equal mix of all five plasmids, each 20ng, was co-transformed into a dCas9 expressing *V. natriegens* strain. The transformation was recovered in 1mL SOC3 media for 45 minutes at 37°C at 225rpm and plated on 245mm x 245mm plates (Corning) with appropriate antibiotics. After 13 hours at 37°C, colonies were scraped in LB3. Growth competition was performed by subculturing this library 1:1000 in LB3 at 37°C for 3 hours in baffled 250mL flasks (Corning). At each time point, gRNA plasmid was extracted from 3mL of culture (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 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, 20bp sequences with a terminal Cas9 NGG motif on the reverse complement strand were selected. Up to 3 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 oligo library. The OLS pool was amplified by cycle-limited real-time PCR, and assembled into the pCTX-R6K-gRNA backbone (NEBuilder HiFi) at 5-fold molar excess with 18bp overlap arms. 6µL of the assembled product was mixed with 300µL TransforMax EC100D pir+ *E. coli* (Epicentre) and 51µL aliquots of this mix was electroporated in 0.1mm cuvettes with a Bio-Rad Gene Pulser electroporator at 1.8kV, 200Ω, 25µF. These *E. coli* transformants were recovered in 6x 1mL SOC media for 60 minutes at 37°C at 225rpm, and plated on 245mm x 245mm spectinomycin selective plates (Corning). After 13

hours at 37°C, ~1.4x10^6 colonies were scraped and plasmid DNA extracted (Qiagen HiSpeed Plasmid Maxi).

The plasmid gRNA library was transformed into wild type *V. natriegens,* or a wild type strain expressing dCas9 (plasmid pdCas9-bacteria). No inducer was added for dCas9 expression, as previous results indicate leaky expression of dCas9 is sufficient for inhibition (**Supplementary Figure 9**) Briefly, ~600ng of the plasmid library was mixed with 300µL of electrocompetent cells and 53.5µL of this mix was electroporated in 0.1mm cuvettes with a Bio-Rad Gene Pulser electroporator at 0.4kV, 1kΩ, 25µF. Each transformation was recovered in 1mL SOC3 media for 45 minutes at 37°C at 225rpm and plated on 245mm x 245mm plates (Corning) with appropriate antibiotics. After 13 hours at 37°C, colonies were scraped in LB3 and stored at -80°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 25mL of media in baffled 250mL flasks (Corning). For media experiments, plasmids were extracted after 8 hours 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 hours of growth at 37°C in LB3 (Passage 1) and the culture is then diluted 1:1000 in 25mL of fresh LB3, yielding ~10^6 cells, and grown at 37°C for 4 hours (Passage 2). Subsequent passaging were performed in 4 hour increments. Plasmids were extracted from >3mL 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. Sequencing was used to verify coverage of the gRNA library. For media experiments, the initial no dCas9 and dCas9 library contains 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 contains 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 number for the initial no dCas9 and dCas9 media library is 70.3 and 62.3 reads per million, respectively. The median gRNA copy number for the initial no dCas9 and dCas9 passaging library is 70.1 and 62.1 reads per million, respectively. These counts were input into MAGeCK and analyzed using the 'mle' module with default settings[17] 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[19,21,34]. Enrichments are calculated as a one-sided hypergeometric test assuming independence between functional groups with multiple-hypothesis correction with FDR (Benjamini-Hochberg).

<u>Construction of in-frame genomic knockouts by natural competency</u>

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 500bp homology arms upstream and downstream, was created by overlap extension PCR. Generation of deletion mutants are performed as previously described [35]. Briefly, the strain carrying tfoX is grown overnight in 1mM final IPTG in a rotating drum at 30°C, then subcultured 1:100 in 350µL of 2XIO (28g/L Ocean Salts) with 1mM IPTG with >100ng of the deletion cassette. The culture is incubated statically at 30°C for 5 hours, then recovered with 1mL of LBv2[2] at 30°C for 2 hours with shaking (225-250rpm). Following recovery, cultures are plated on antibiotic-selective plates. Resulting transformants are screened by PCR to confirm gene deletion.

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.

Code Availability

Custom code is available at https://github.com/citizenlee/vnat_glib or will be made available upon request.

Figure legends

Figure 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, respectively). Designated cell trapping heights were used to keep each species growing in a monolayer. Scale bar 10μm (N=3 independent experiments). (b) Single-cell growth measurement at 37°C (LB3 and LB, respectively). Data shown are mean±SD (N=18 *E. coli* cells and N=21 *V. natriegens* cells). (c) *V. natriegens* chromosomes. From outside inward: two outer circles represent protein-coding genes on the plus and minus strand, respectively. Color coding by RAST annotation, as in panel d. The third circle represents G+C content relative to mean G+C content of the respective chromosome, using a sliding window of 3,000 bp. tRNA and rRNA genes are shown in the fourth and fifth circles, respectively. (d) Fraction of genes in each RAST category on chr1 (dark colors) 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. Origin (red) and terminus (blue) are denoted.

Figure 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 induction (gray) of dCas9. Data are shown as mean±SD (N=4 biological replicates). (b) Small-scale pooled CRISPRi screen. gRNAs include putative essential gene $lptF_{Vn}$ (solid black line), putative neutral genes $flgC_{Vn}$ (dashed black line), GFP (dashed gray line), and *E. coli* gene $lptF_{Ec}$ (solid gray line, target gene absent from genome). 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 (gray bars) and percentage (dotted line) of gRNAs across open reading frames are shown. (d) Schematic overview of pooled genome-wide CRISPRi screen.

Figure 3. CRISPRi inhibition 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 media; gray dots represent all other genes in the population. Histogram (gray) shows distribution of all targeted genes. Inset: The number of genes with significantly negative core in each media category. (b) Single gene knock-out of sucrose-6-phosphate hydrolase located on chr1 (ΔPEG.1381) or its paralog located on chr2 (ΔPEG.3017) in LB3 (gray), M9-glucose (black) or M9-sucrose (red). Data are shown as mean±SD (N=4 biological replicates) (c) Top ten RAST categories enriched in genes that are uniquely depleted in minimal media (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: Co-A Biosynthesis. Statistics are derived from one-tailed hypergeometric test for enrichment

assuming independence between subsystems. False discovery rate (FDR) controlled with Benjamini-Hochberg. (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 media are shown in red, other computationally assigned genes are shown in grey. Leucine biosynthetic genes numbered as follow: 1: acetoacetate synthase; 2: ketol-acid reductoisomerase; 3: dihydroxy-acid dehydratase; 4: isopropylmalate synthase; 5: 3-isopropylmalate dehydratase; 6: 3-isopropylmalate dehydrogenase; 7: aminotransferase.

Figure 4. Analysis of *V. natriegens* core genes. (a) Distribution of beta scores for all significantly depleted genes in rich media experiments (FDR≤0.05). Core genes, shown in red, display consistently negative beta score in both experiments. Negative scoring genes which were not included in core set are shown in blue (media experiment) and green (passage experiment). (b) Core genes were divided into putative essentials (yellow) and growth-supporting genes (blue). All other genes shown in gray. (c) Distribution of all targeted genes across RAST categories. Single genes, sorted by beta score, are displayed as gray lines in the respective category. Core genes shown in red. Histogram displays over-represented (filled bars) or under-represented categories (empty bars) within all core gene set (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. False discovery rate (FDR) controlled with Benjamini-Hochberg. N=587 core genes. Rast category names as shown in Figure 1b. (d) Ribosomal proteins essentiality as identified in *V. natriegens* (Vn) screen. Black square denotes essential gene. Empty square denone non-essential gene. Essentiality shown for *E.coli* (Ec)[21] and *V. cholerae* (Vc)[19]. LSU: Large Subunit; SSU: Small Subunit. (e) Beta score as calculated from CRISPRi screen for all gene (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 (transcripts per million) for all genes (grey) and for the set of core genes, as measured at 37°C in LB3 medium. Putative essentials (PE) shown in yellow, growth-supporting genes (GS) 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 is on chr1 and duplicate on chr2. Green lines indicate core gene on chr2 and duplicate on chr1.

References

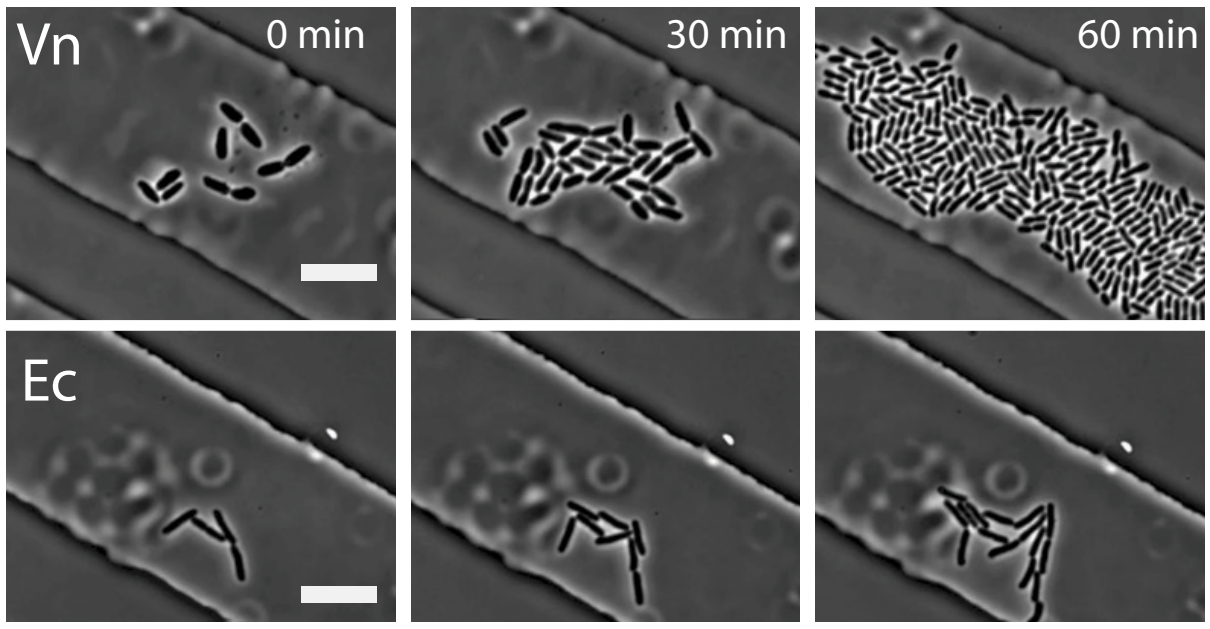
1. Eagon, R. G. Pseudomonas natriegens, a marine bacterium with a generation time of less than 10 minutes. *J. Bacteriol.* **83**, 736–737 (1962).
2. Weinstock, M. T., Hesek, E. D., Wilson, C. M. & Gibson, D. G. Vibrio natriegens as a fast-growing host for molecular biology. *Nat. Methods* **13**, 849–851 (2016).
3. Long, C. P., Gonzalez, J. E., Cipolla, R. M. & Antoniewicz, M. R. Metabolism of the fast-growing bacterium Vibrio natriegens elucidated by (13)C metabolic flux analysis. *Metab. Eng.* **44**, 191–197 (2017).
4. Hoffart, E. *et al.* High substrate uptake rates empower Vibrio natriegens as production host for industrial biotechnology. *Appl. Environ. Microbiol.* (2017). doi:10.1128/AEM.01614-17
5. Aiyar, S. E., Gaal, T. & Gourse, R. L. rRNA promoter activity in the fast-growing bacterium Vibrio natriegens. *J. Bacteriol.* **184**, 1349–1358 (2002).
6. Maida, I. *et al.* Draft Genome Sequence of the Fast-Growing Bacterium Vibrio natriegens Strain DSMZ 759. *Genome Announc.* **1**, e00648–13–e00648–13 (2013).
7. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**, D206–D214 (2013).
8. Thompson, J. R. & Polz, M. F. Dynamics of Vibrio populations and their role in environmental nutrient cycling. in *The biology of vibrios* 190–203 (American Society of Microbiology, 2006).
9. Heidelberg, J. F. *et al.* DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae. *Nature* **406**, 477–483 (2000).
10. Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**, D93–7 (2009).
11. 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).
12. van Opijnen, T. & Camilli, A. Genome-wide fitness and genetic interactions determined by Tn-seq, a high-throughput massively parallel sequencing method for microorganisms. *Curr. Protoc. Microbiol.* **Chapter 1**, Unit1E.3 (2010).
13. Rock, J. M. *et al.* Programmable transcriptional repression in mycobacteria using an orthogonal CRISPR

interference platform. *Nat Microbiol* **2**, 16274 (2017).

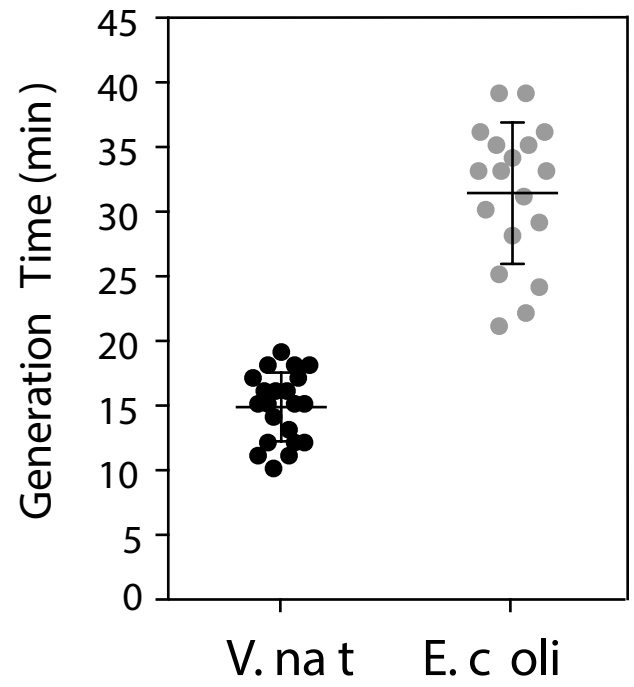
14. Peters, J. M. *et al.* A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. *Cell* **165**, 1493–1506 (2016).
15. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
16. Liu, X. *et al.* High-throughput CRISPRi phenotyping identifies new essential genes in Streptococcus pneumoniae. *Mol. Syst. Biol.* **13**, 931 (2017).
17. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
18. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624 (2016).
19. 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).
20. Okada, K., Iida, T., Kita-Tsukamoto, K. & Honda, T. Vibrios commonly possess two chromosomes. *J. Bacteriol.* **187**, 752–757 (2005).
21. Yamazaki, Y., Niki, H. & Kato, J.-I. Profiling of Escherichia coli Chromosome database. *Methods Mol. Biol.* **416**, 385–389 (2008).
22. Julio, S. M. *et al.* DNA Adenine Methylase Is Essential for Viability and Plays a Role in the Pathogenesis of Yersinia pseudotuberculosis andVibrio cholerae. *Infect. Immun.* **69**, 7610–7615 (2001).
23. 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).
24. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
25. Schleicher, L. *et al.* Vibrio natriegens as Host for Expression of Multisubunit Membrane Protein Complexes. *Front. Microbiol.* **9**, 2537 (2018).
26. 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).
27. Val, M.-E. *et al.* A checkpoint control orchestrates the replication of the two chromosomes of Vibrio cholerae. *Sci Adv* **2**, e1501914 (2016).

28. Mee, M. T., Collins, J. J., Church, G. M. & Wang, H. H. Syntrophic exchange in synthetic microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2149–56 (2014).
29. Cui, L. *et al.* A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. *Nat. Commun.* **9**, 1912 (2018).
30. Zhang, S. & Voigt, C. A. Engineered dCas9 with reduced toxicity in bacteria: implications for genetic circuit design. *Nucleic Acids Res.* **46**, 11115–11125 (2018).
31. Vega, N. M., Allison, K. R., Khalil, A. S. & Collins, J. J. Signaling-mediated bacterial persister formation. *Nat. Chem. Biol.* **8**, 431–433 (2012).
32. 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).
33. 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 (2008).
34. Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* (2017). doi:10.1093/molbev/msx148
35. Dalia, T. N. *et al.* Multiplex Genome Editing by Natural Transformation (MuGENT) for Synthetic Biology in Vibrio natriegens. *ACS Synth. Biol.* (2017). doi:10.1021/acssynbio.7b00116

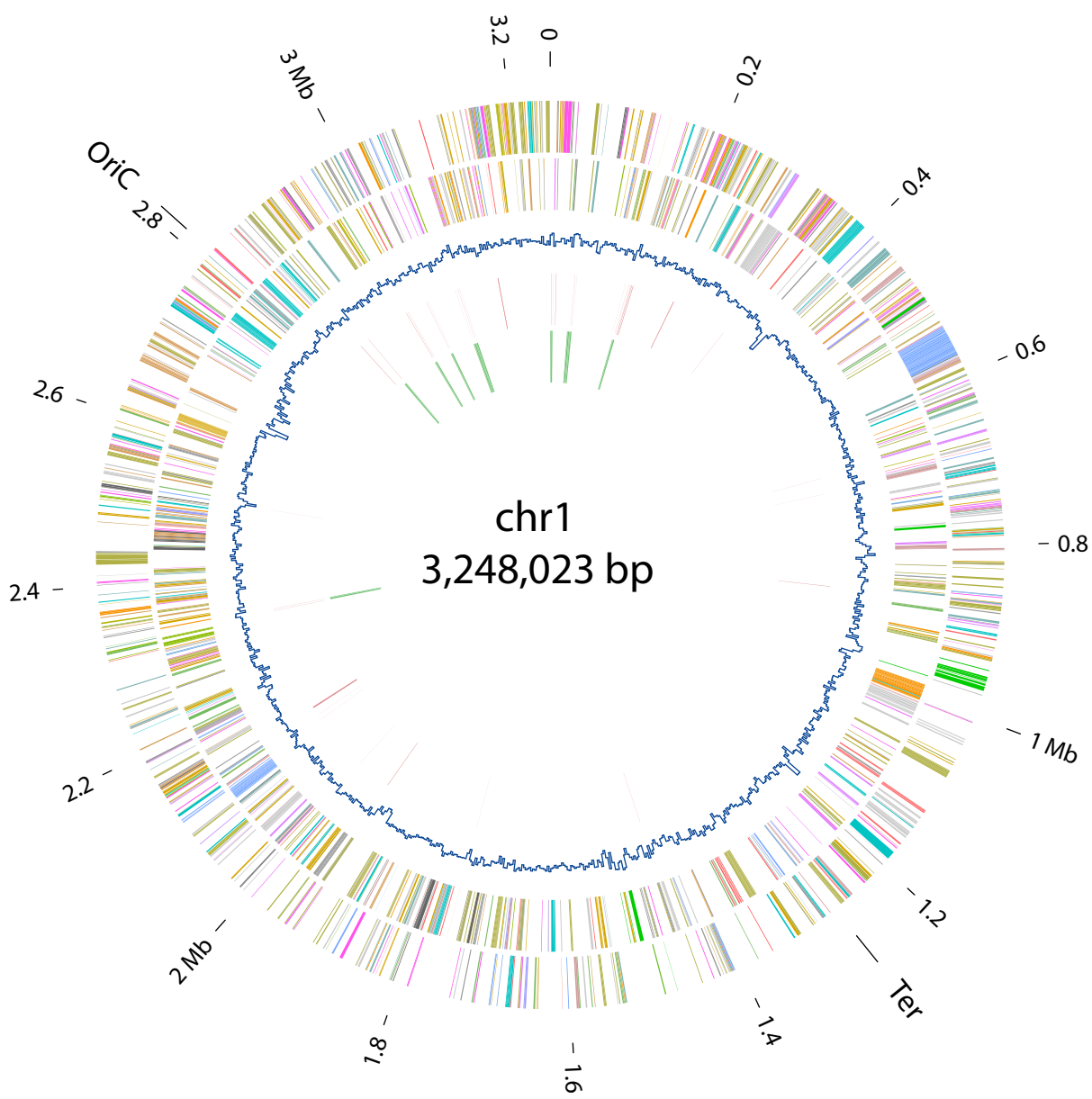
a

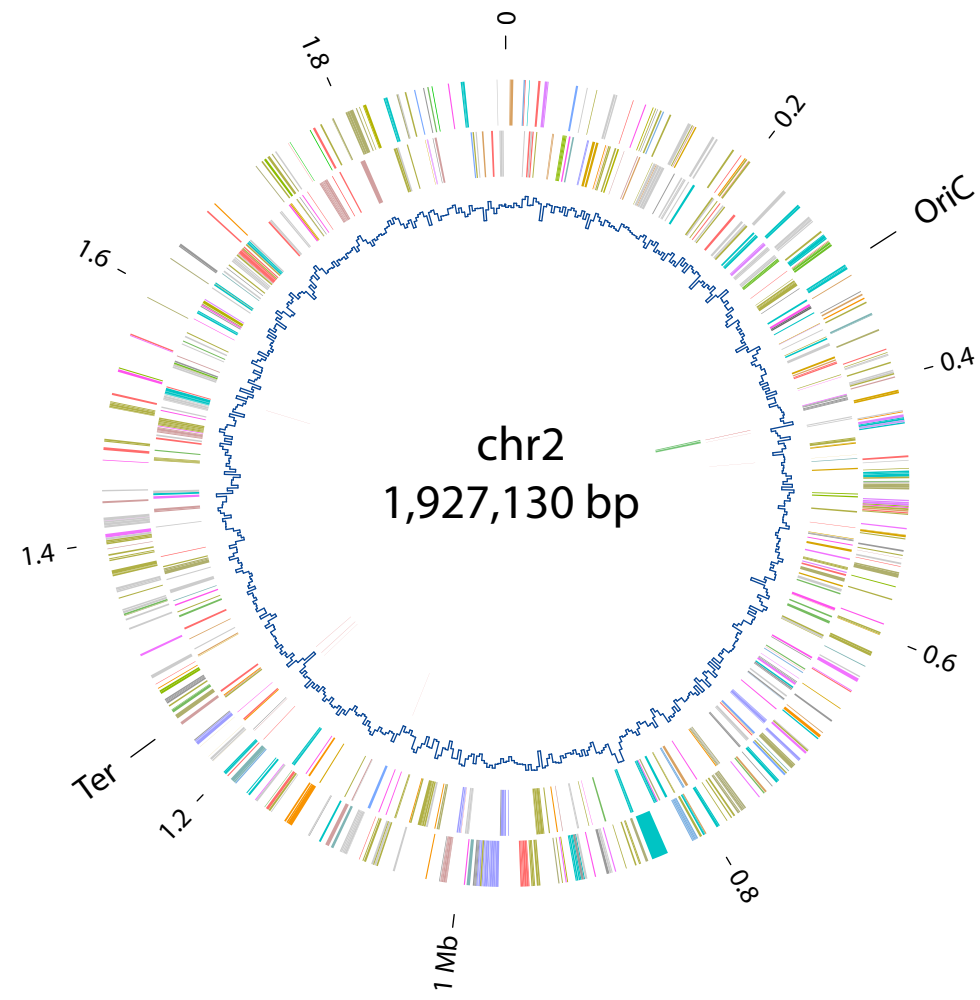

Vn
0 min
30 min
60 min
Ec

b

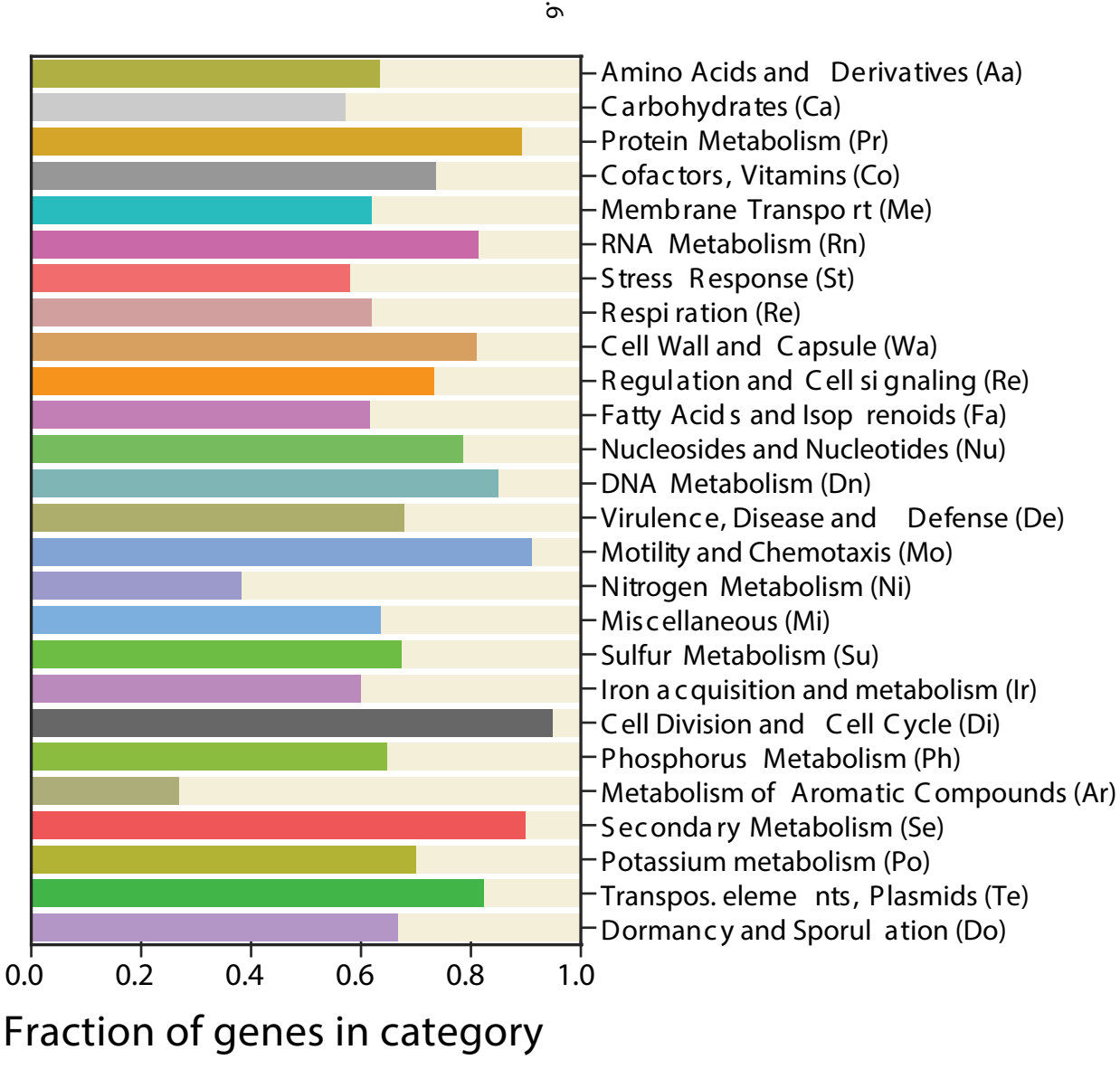

Generation Time (min)
45
40
35
30
25
20
15
10
5
0
V. na t
E. c oli

c

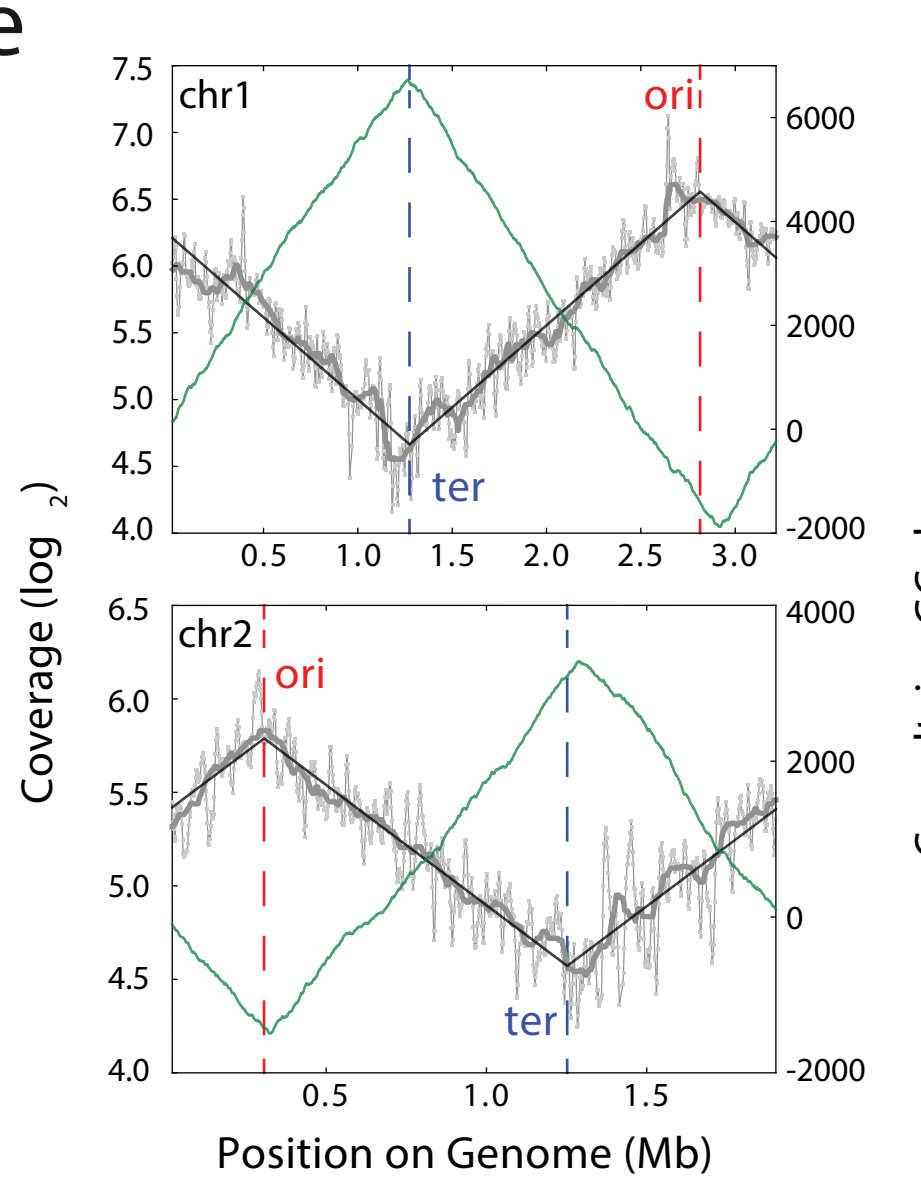

chr1
3,248,023 bp
OriC
Ter
0
0.2
0.4
0.6
0.8
1 Mb
1.2
1.4
1.6
1.8
2 Mb
2.2
2.4
2.6
2.8
3 Mb
3.2


chr2
1,927,130 bp
OriC
Ter
0
0.2
0.4
0.6
0.8
1 Mb
1.2
1.4
1.6
1.8

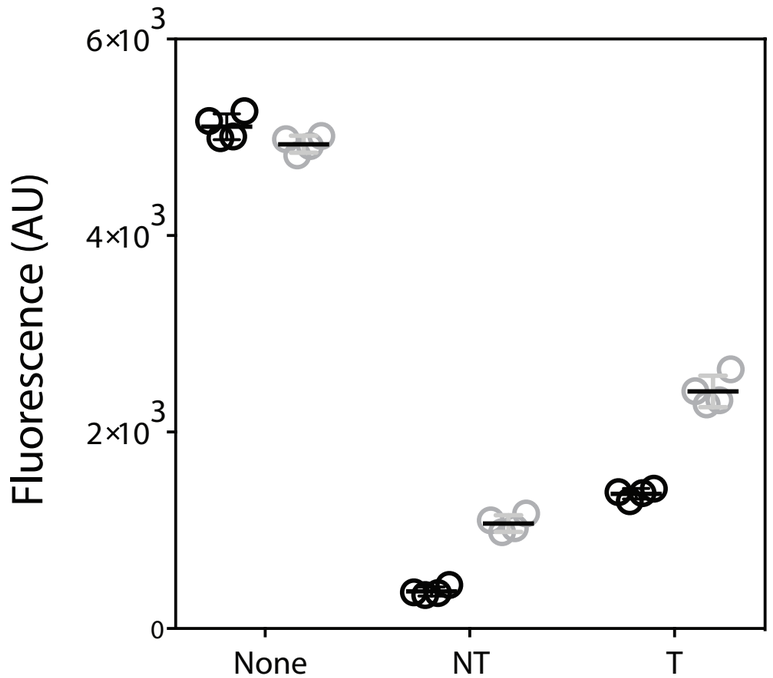
d


Amino Acids and Derivatives (Aa)
Carbohydrates (Ca)
Protein Metabolism (Pr)
Cofactors, Vitamins (Co)
Membrane Transpo rt (Me)
RNA Metabolism (Rn)
Stress Response (St)
Respi ration (Re)
Cell Wall and Capsule (Wa)
Regulation and Cell si gnaling (Re)
Fatty Acids and Isop renoids (Fa)
Nucleosides and Nucleotides (Nu)
DNA Metabolism (Dn)
Virulence, Disease and Defense (De)
Motility and Chemotaxis (Mo)
Nitrogen Metabolism (Ni)
Miscellaneous (Mi)
Sulfur Metabolism (Su)
Iron acquisition and metabolism (Ir)
Cell Division and Cell Cycle (Di)
Phosphorus Metabolism (Ph)
Metabolism of Aromatic Compounds (Ar)
Secondary Metabolism (Se)
Potassium metabolism (Po)
Transpos. eleme nts, Plasmids (Te)
Dormancy and Sporul ation (Do)
0.0
0.2
0.4
0.6
0.8
1.0
Fraction of genes in category

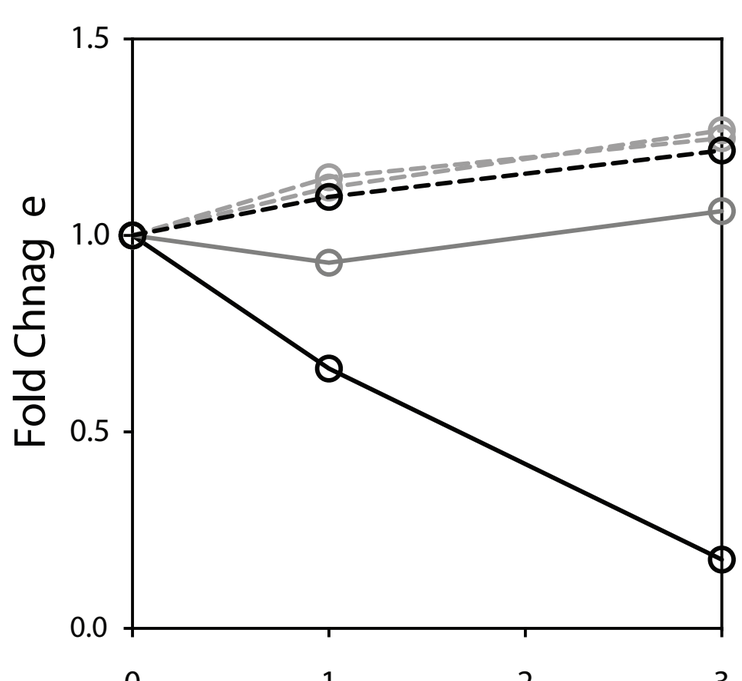
e


chr1
ori
ter
chr2
ori
ter
Coverage (log 2)
Cumultaive GC-skew
Position on Genome (Mb)
0.5
1.0
1.5
2.0
2.5
3.0
4.0
4.5
5.0
5.5
6.0
6.5
7.0
-2000
0
2000
4000
6000

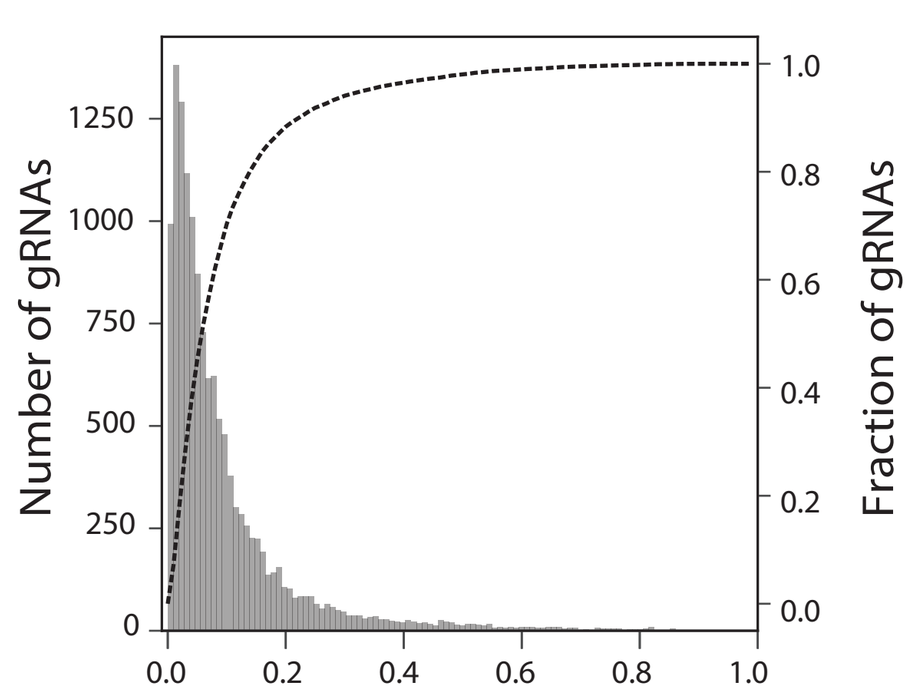
a


6×10^3
4×10^3
2×10^3
0
Fluorescence (AU)
None
NT
T

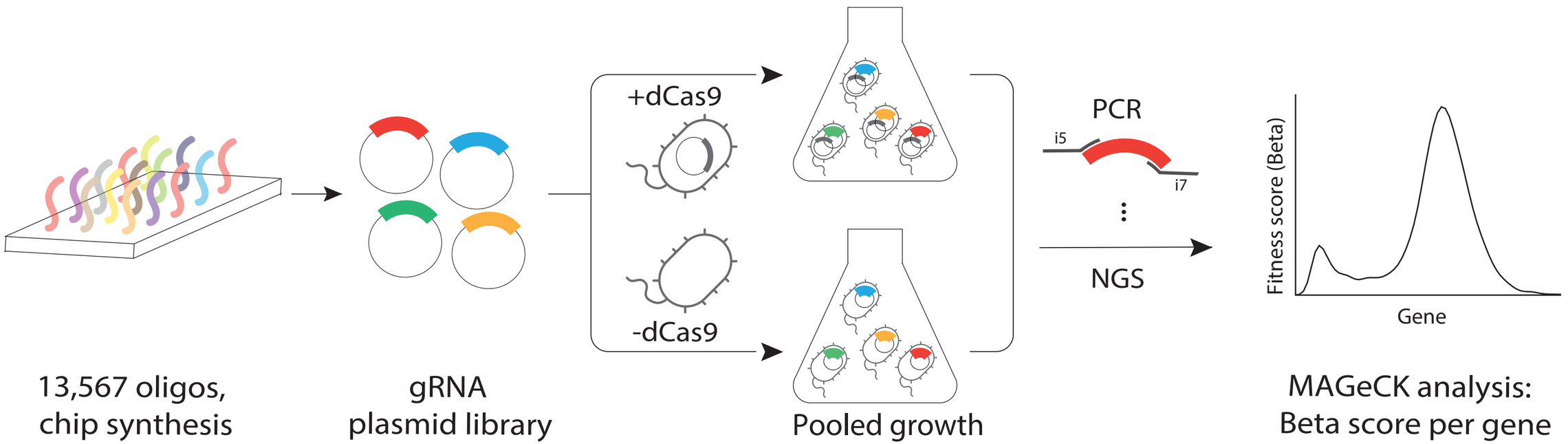
b


1.5
1.0
0.5
0.0
Fold Chnag e
0
1
2
3
gRN A
Time (hours)

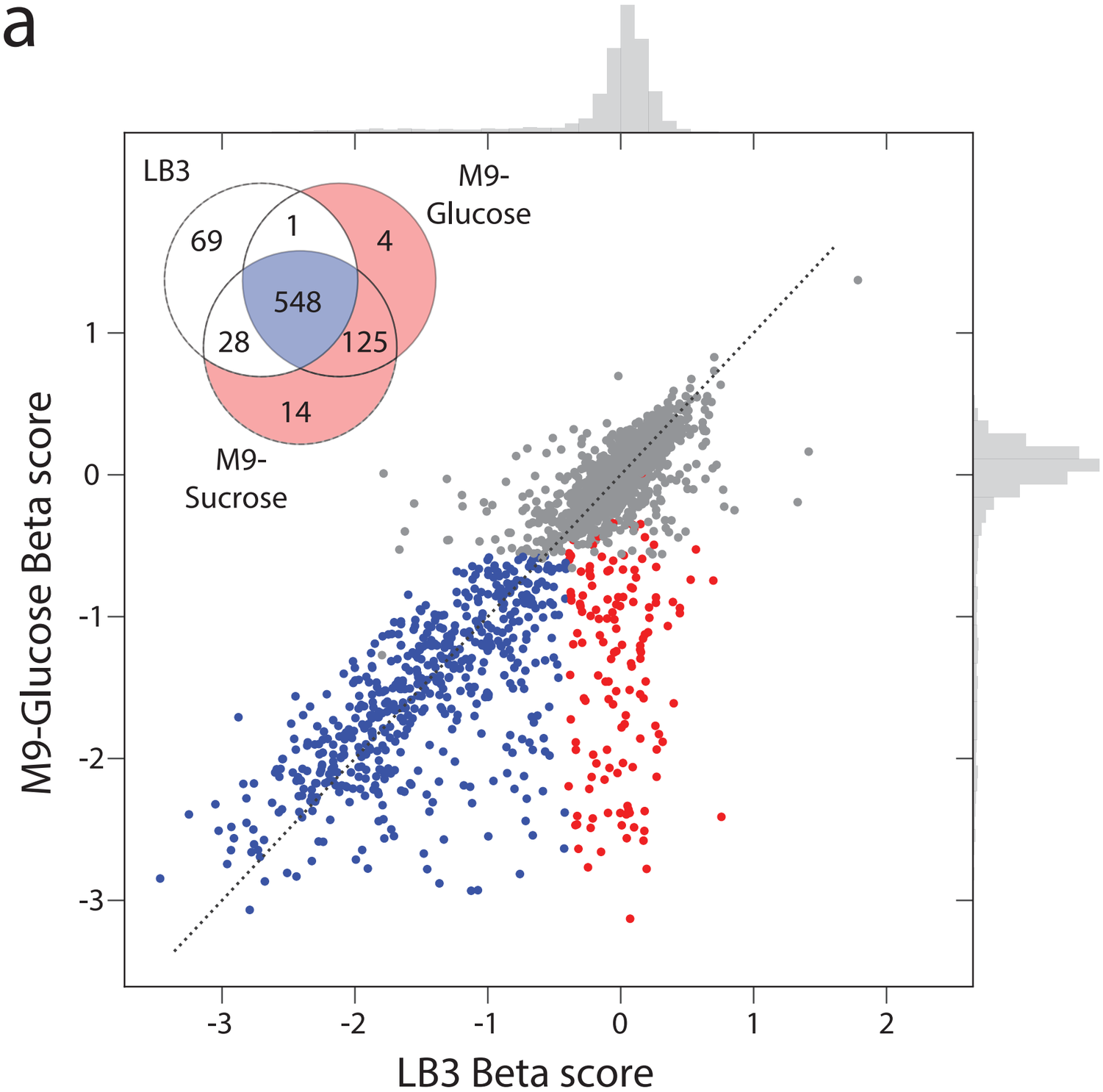
c


1250
1000
750
500
250
0
Number of gRNAs
1.0
0.8
0.6
0.4
0.2
0.0
Fraction of gRNAs
0.0
0.2
0.4
0.6
0.8
1.0
Fraction of ORF

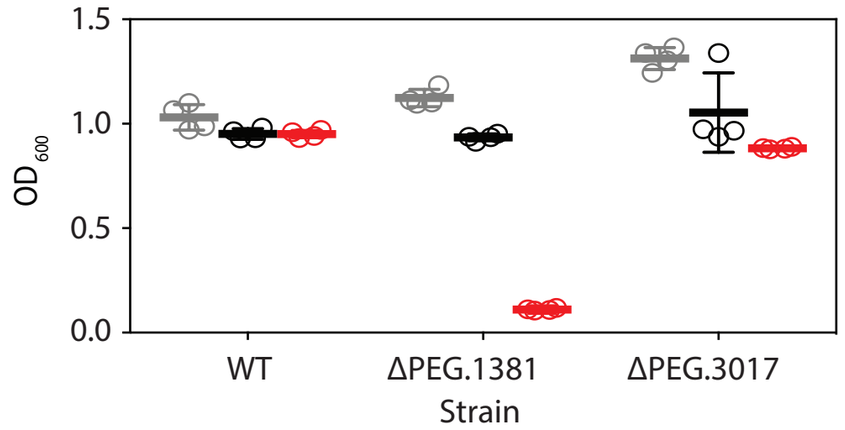
d


+dCas9
-dCas9
PCR
i5
i7
NGS
Fitness score (Beta)
Gene
13,567 oligos,
chip synthesis
gRNA
plasmid library
Pooled growth
MAGeCK analysis:
Beta score per gene

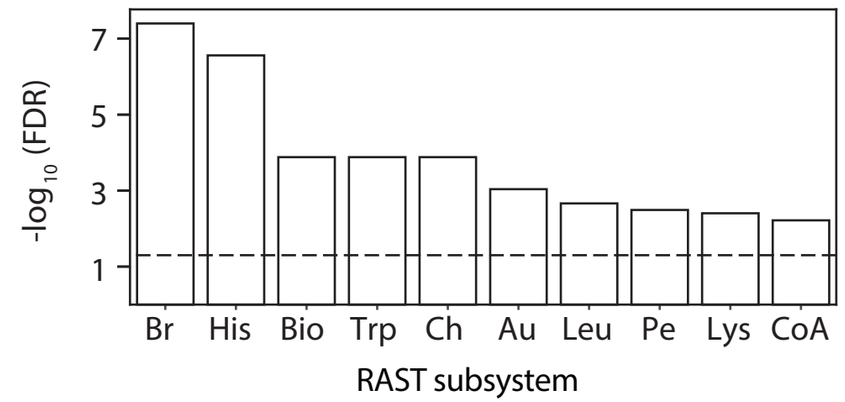
a


LB3
M9-
Glucose
M9-
Sucrose
69
1
4
548
28
125
14
M9-Glucose Beta score
LB3 Beta score
1
0
-1
-2
-3
-3
-2
-1
0
1
2

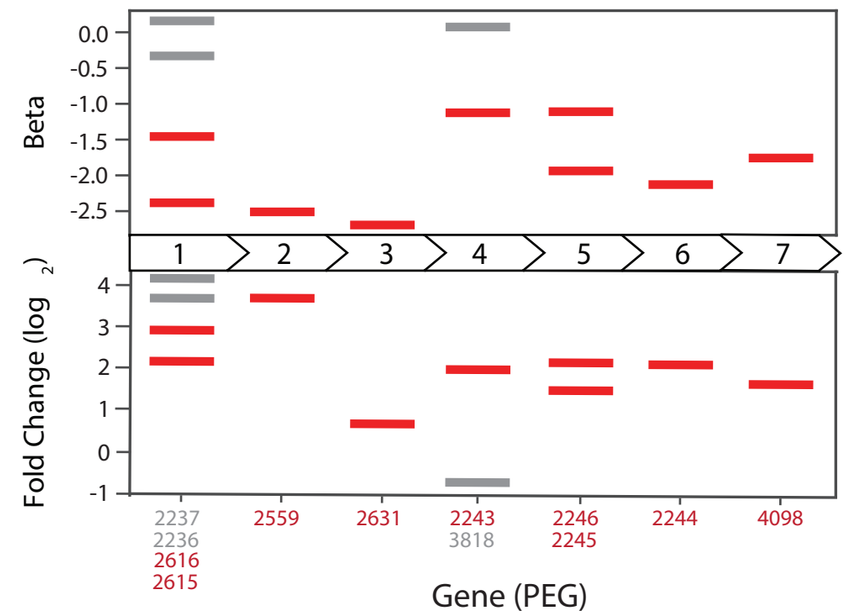
b

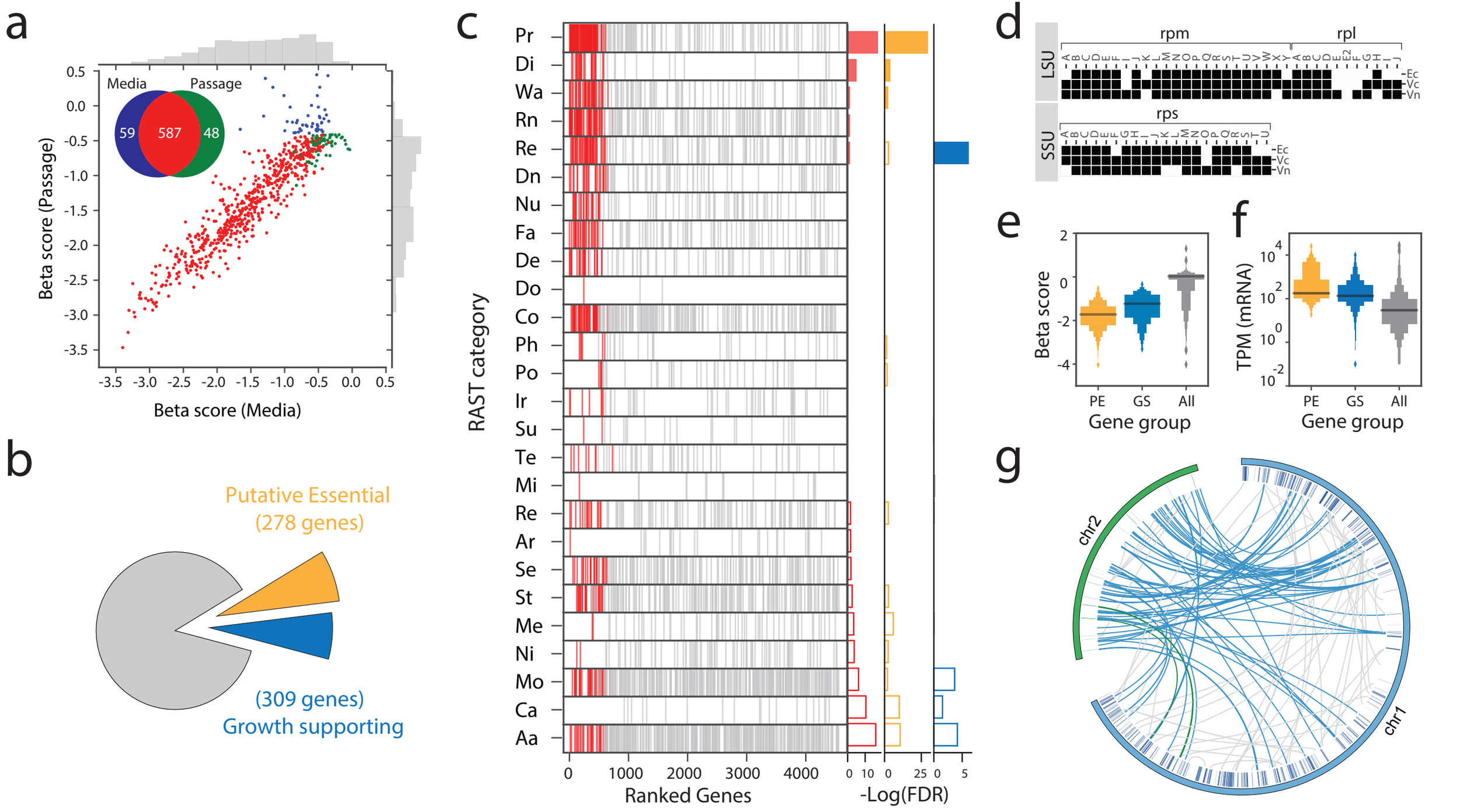

OD600
1.5
1.0
0.5
0.0
WT
ΔPEG.1381
ΔPEG.3017
Strain

c


-log10 (FDR)
7
5
3
1
Br
His
Bio
Trp
Ch
Au
Leu
Pe
Lys
CoA
RAST subsystem

d


Beta
0.0
-0.5
-1.0
-1.5
-2.0
-2.5
1
2
3
4
5
6
7
Fold Change (log 2)
4
3
2
1
0
-1
2237
2236
2616
2615
2559
2631
2243
3818
2246
2245
2244
4098
Gene (PEG)


a
Media
Passage
59
587
48
Beta score (Passage)
Beta score (Media)
b
Putative Essential
(278 genes)
(309 genes)
Growth supporting
c
RAST category
Pr
Di
Wa
Rn
Re
Dn
Nu
Fa
De
Do
Co
Ph
Po
Ir
Su
Te
Mi
Re
Ar
Se
St
Me
Ni
Mo
Ca
Aa
Ranked Genes
-Log(FDR)
d
LSU
SSU
rpm
rpl
rps
Ec
Vc
Vn
e
Beta score
PE
GS
All
Gene group
f
TPM (mRNA)
PE
GS
All
Gene group
g
chr2
chr1