Chapter 2

Methods

2.1 – Computational analyses

2.1.1 – Bacteria sequenced in the course of this PhD research

A list of *V. cholerae* that were sequenced in the course of this study is provided in Appendix 1. Additional previously-published genomes were also collated for analyses in Chapters 3-5. If the short-reads for externally-sequenced genomes were available, these were included with the short-reads generated in this study for analysis. If these external genomes were only available as assemblies and short-reads were required (*e.g.*, for mapping analyses), paired-end 100 bp Illumina reads were simulated from assemblies using wgsim v.0.3.2 [1], part of SAMtools [313], and parameters '-e 0 -r 0 -X 0 -1 100 -2 100' (these settings assume a haploid genome and are appropriate for bacteria). Additional *V. cholerae* genomes were also downloaded from Refseq as assemblies.

2.1.2 – Genome assembly

2.1.2.1 – Short-read data

Throughout this PhD research, short-read sequencing data generated at WSI were assembled using SPAdes v3.8.2 [314] as part of a WSI high-throughput analysis pipeline [315]. Externally-generated sequencing reads were assembled using the same pipeline.

2.1.2.2 – Long read data

The *V. cholerae* O139 long-read assemblies described in Chapter 4 (section 4.3.1) were produced using HGAP v3 and SMRT analysis software v2.3.0 [316]. The fold coverage to target when picking the minimum fragment length for assembly was set to 30 and the approximate genome size was set to 3 Mbp. Assemblies were circularised using Circlator v1.1.3 [317] and the pre-assembled reads (also known as corrected reads). Circularised assemblies were polished using the PacBio RS_Resequencing protocol and Quiver v1 [316]. Pilon v1.19 [318] did not identify any single nucleotide variants (SNVs) in any of the PacBio

[1] https://github.com/lh3/wgsim

assemblies using the corresponding short-read data – accordingly, no short-read corrections were made to these assemblies.

The NCTC 30 genome (Chapter 4, section 4.3.7) was similarly assembled from PacBio RSII reads using HGAP v3 and the RS_HGAP_Assembly.2 protocol, performed using SMRT Portal and SMRT Analysis v2.3.0.140936.p5.167094 [316]. Minimum polymerase read quality and length were set to 0.8 and 100, respectively. A minimum length of 500 bases was used to filter sub-reads. The minimum seed read length for assembly was set to 6,000, with BLASR options '-noSplitSubreads -minReadLength 200 -maxScore -1000 -maxLCPLength 16', an expected genome size of 5 Mbp, and a target coverage of 30 X. Contigs were circularised using Circlator v1.5.3 [317] using both the assembly and the corrected reads, and a final assembly was obtained by using the circularised sequences as a reference to re-assemble the PacBio reads using the RS_Resequencing.1 protocol (minimum subread length = 50 bp, minimum polymerase read quality = 75%, minimum polymerase read length = 50 bp, BLASR maximum divergence = 30%, minimum anchor size = 12). This assembly was then corrected using Quiver v1. The finished assembly was covered to an average depth of 148.01 X. To check the accuracy of the PacBio assembly, the corresponding Illumina short-reads were mapped to the assembly using SMALT v0.5.8 [2] (maximum insert size = 1000, minimum insert size = 50). No SNVs were identified in the assembly upon mapping of these data.

2.1.2.3 – Hybrid assemblies and data visualisation

The hybrid assemblies described in Chapter 5 (section 5.3.7) were produced using both the long- and short-reads for each isolate and Unicycler v0.4.0 [319], with options '-t 4 --mode conservative'. The De Brujin graphs from these assemblies were visualised using Bandage v0.8.0 [320].

2.1.3 – Sequencing quality control

Contamination of genome sequences was assessed using several methods. Kraken [321] was used to assess whether sequences belonging to bacterial species other than the species of interest were present in the sequencing data for an isolate. Additionally, the length of genome

[2] http://www.sanger.ac.uk/science/tools/smalt-0

assemblies was inspected and summarised using assembly-stats v1.0.1 [3]. Assemblies that were substantially larger than expected for a *V. cholerae* genome were treated as potentially-contaminated (*i.e.*, assembly length was ≥ 5 Mbp; the *V. cholerae* reference genome is 4.1 Mbp in length [59]). Phylogenetic trees were also inspected at the beginning of each analysis, to identify stark outliers (*e.g.*, isolates on very long branches). Any sequences which were suspected to be of poor quality were excluded from downstream analyses.

2.1.4 – Automated genome annotation

All genome assemblies used in this research were uniformly annotated using Prokka v1.5 [322] and a genus-specific database for *Vibrio* [323]. This was either run as part of the WSI automated assembly pipeline (for all automatically-assembled genomes, for genomes downloaded from RefSeq, and for any assemblies that were produced from simulated reads), or manually (for hybrid and long-read-only assemblies generated in Chapters 4 and 5).

2.1.5 – Pangenome calculations

All pangenomes described in this thesis were generated using Roary v3.12.0 [324], run with options '-e --mafft -s -cd 97'. These settings prevent the splitting of paralogous genes into separate gene clusters to avoid the over-partitioning of orthologous genes into separate clusters, and impose a CD-HIT clustering cut-off of 97%. Prokka-annotated genome assemblies in GFF3 format were used as input for these calculations. Gene presence/absence matrices were visualised using roary_plots.py v0.1.0 [4].

2.1.6 – Short-read mapping and SNV detection

High-quality SNVs, including small indels, were identified as described by Harris *et al* [325] using SMALT v0.7.4 [5]. Briefly, variant sites was identified from mapped sequence data using SAMtools mpileup v0.1.19 [313] and parameters '-d 1000 -DSugBf', and bcftools v0.1.19 [313], to produce a BCF file of all variant sites from which high-quality SNVs were identified using previously-described quality thresholds [325]. For analyses of 7PET genomes, mapping

[3] https://github.com/sanger-pathogens/assembly-stats
[4] https://github.com/sanger-pathogens/Roary/tree/master/contrib/roary_plots
[5] http://www.sanger.ac.uk/science/tools/smalt-0

and SNV-calling were performed against the N16961 reference genome sequence (accession # LT907989/LT907990 [59]). For the analysis of LAT-1 described in Chapter 3 (section 3.4.5), mapping and SNV-calling were performed against the sequence of the LAT-1 isolate A1552 (accession # CP025936/CP025937 [326]).

2.1.7 – Detection of potentially-recombined chromosomal regions

Potentially-recombined regions of *V. cholerae* genomes were identified using Gubbins v1.4.10 [327] and a "pseudo-genome" multiple-sequence alignment of all variant and invariant sites relative to a reference genome sequence. This tool produced a variant-only alignment of SNVs which excluded all variants predicted to be located within recombined regions of the chromosome. This alignment was used for subsequent phylogenetic analyses, to avoid SNVs in recombined regions potentially interfering with the topology of phylogenetic trees.

2.1.8 – Phylogenetic analyses and tree-building

7PET and LAT-1 phylogenies were calculated from non-recombinant SNVs called against the N16961 and A1552 reference sequences by mapping, as described above (section 2.1.6 and 2.1.7). *V. cholerae* species phylogenies were calculated from the core-gene alignment generated by Roary (section 2.1.5); core-gene alignments were trimmed using trimAl v1.4.1 [328], and SNP-sites v2.4.1 and v2.5.1 [329] was used to extract the variant sites from this trimmed alignment. The sizes of alignments used for calculations are reported throughout the thesis.

The maximum-likelihood phylogenetic trees presented in this thesis were calculated using IQ-Tree v1.6.10 [330] unless otherwise stated in the text, and either the alignments of non-recombinant SNVs produced by Gubbins (section 2.1.7) or the SNV-only alignments from a core-gene alignment, described above. Trees were calculated under the general time reversible (GTR) and ascertainment bias correction (ASC) models [331]. In order to assess the robustness of the computed phylogenies, five thousand approximate likelihood ratio tests [332] and ultrafast bootstrap approximations [333] were performed for each tree calculation.

2.1.9 – Data clustering and lineage assignment

In order to identify clusters of similar sequences from multiple-sequence alignments, alignments of parsimony-informative SNVs were extracted from SNV-only alignments using extract_PI_SNPs.py [6] (*i.e.*, any SNVs that were private to a single isolate were removed from the alignment). For alignments that were derived from *V. cholerae* species core gene alignments, sequence data for the *Vibrio* spp. outgroup were deleted from the alignment prior to running extract_PI_SNPs.py.

Alignments of parsimony-informative SNVs were used for cluster analysis using the R implementation of Fastbaps v1.0.1 [334]. The Dirichlet Process Mixture model was used to cluster 7PET and LAT-1 genomes, whereas the Bayesian Hierarchical Clustering prior [335] was used to cluster sequences in *V. cholerae* species phylogenies.

2.1.10 – Comparative genomics

Synteny plots were generated using ACT v13 [336] and Easyfig v2.2.2 [337] to enable comparative genomics analyses. Both of these tools rely on BLASTn [338] comparisons between two genome assemblies. Cut-offs for Easyfig comparisons were set to a maximum *e*-value of 0.001 and minimum length of 0; the minimum BLASTn identity percentage is specified in each comparison figure throughout the thesis. BamView [339] was used to visualise the mapping of Illumina reads to a genome sequence using the BAM file of mapped sequencing data.

2.1.11 – Pairwise SNV and ANI calculation

SNV distance matrices were calculated from SNV-only alignments using snp-dists v0.4 [7]. Average nucleotide identity (ANI) values were calculated using FastANI v1.0 [340] and a reference sequence as stated in the text.

[6] https://gist.github.com/jasonsahl/9306cd014b63cae12154
[7] https://github.com/tseemann/snp-dists

2.1.12 – Assessment of temporal signal in phylogenetic data

Maximum-likelihood phylogenetic trees produced in section 2.1.8 were read into TempEST v1.5.1 [341] to assess whether a temporal signal could be detected within the phylogeny. The names of leaves on the phylogeny were amended to include the date of isolation, where known.

2.1.13 – Detection of antimicrobial resistance genes, wbeT *and* ctxB *variants, plasmid replicons and virulence determinants*

Antimicrobial resistance genes and plasmid replicons were detected in genomes using both ARIBA v2.12.1 [342] and ABRicate v1.0.1 [8]. ARIBA performs local assembly of sequence data against a reference database of genes of interest, and was therefore only used when short-reads were available for all genomes of interest. In contrast, ABRicate screens assembled contigs for genes of interest, and was used when assemblies were available for all genomes of interest. The ResFinder database of antimicrobial resistance genes [343], and the PlasmidFinder database of plasmid replicon sequences [344] were used for ARIBA and ABRicate analyses (both databases accessed on 23/06/2019 and 11/05/2020 for ARIBA and ABRicate analyses, respectively). The VFDB database (accessed on 11/05/2020) was used for identifying virulence genes of interest using ABRicate. Default cut-offs for identity percentage and length were used throughout this thesis unless otherwise specified in the text.

ARIBA was also used to assign *wbeT* and *ctxB* genotypes to *V. cholerae* isolates. This relied on a custom database consisting of the *ctxB* nucleotide sequence from N16961 (accession # LT907989/LT907990) and the intact *wbeT* sequence from NCTC 9420 [213] (accession # CP013319/CP013320). This *wbeT* sequence translates into a protein sequence which is 100% identical to the WbeT sequence from the Ogawa isolate VX44945 (AEN80191.1 [196]), which has been used for *wbeT* genotyping in previous studies [158, 189].

The NCTC 30 genome assembly was screened for antimicrobial resistance genes using the ResFinder web server [343] (database version 2018-02-19) using default cut-offs of 90% nucleotide identity and 60% minimum length.

[8] https://github.com/tseemann/abricate

2.1.14 – Confirming the presence of genomic elements

The presence of the WASA-1 genomic island, a marker characteristic of the LAT-1 sub-lineage [234, 235], was confirmed in genome assemblies by using BLASTn to query the assembly for the presence of the WASA-1 sequence, as well as by using pangenome gene presence/absence matrices to check for the presence of WASA-1 gene orthologues in an assembly of interest. Similarly, the presence and absence of the *V. cholerae* serogroup O1 biosynthesis operon was confirmed using the pangenome gene presence/absence matrix, and by testing for the presence of the O1 biosynthesis operon sequence using BLASTn.

2.1.15 – Construction of BLAST atlases

BLAST atlas figures were generated using annotated genome assemblies (EMBL format) and the GView web server [9], which relies on CGView [345]. The GC% and GC skew calculation for each comparison was also carried out using GView.

2.1.16 – Statistical tests

Statistical tests and p-value cut-offs are listed and described in the text of the thesis. Unless otherwise specified, statistical tests were performed using the R package, v3.5.1 [346]

2.1.17 – Promoter sequence analysis

The *tcpPH* and *aphA* promoter sequences were extracted from genome assemblies using *in silico* PCR, a custom script at WSI which searches for the sequences of 5' and 3' "primers" within an assembly and reports the sequence between those as output. To "amplify" P_{tcpPH}, the primer sequences 5'-`cgtaaggggcaaaatgtcacaggaa`-3' and 5'-`atagacttgattagtgcattcattc`-3' were used (maximum amplicon length 1,500 bp, minimum length 100 bp), and for P_{aphA}, 5'-`gaatgcgcaatactggttaa`-3' and 5'-`catcgacaggtttggcttgg`-3' (maximum length 2,000 bp; minimum length 100 bp). Up to three mismatches were permitted between primer and assembly sequences. The extracted promoter sequences were concatenated into a multi-FASTA file and aligned using Clustal

[9] https://server.gview.ca/

Omega [347] and Seaview v4.6.1 [348]. trimAl v1.4.1 [328] was used to extract 50 bp sequences from this alignment which were then used to identify motifs and produce nucleotide usage frequency figures using MEME v5.1.1 [349]

2.1.18 – RNA-seq data analysis and statistical cut-offs

The Rockhopper 2 package [350, 351] was used to calculate expression of genes using the sequence and annotation of the N16961 isolate as a reference. Genes were determined to be differentially expressed if the $\log_2$(fold change) of a gene's expression in one condition relative to a second condition was greater than or equal to 2, or less than or equal to -2. Additionally, the false discovery rate (FDR) q-value cut-off for each result was set to <0.01 – this was used rather than a p-value cut-off in order to account for multiple testing implicit in large RNA-seq comparisons. Principal component analyses (PCA) were calculated using prcomp in R v3.5.1 and a matrix of normalised expression values *per* gene for all replicates in an experiment.

2.1.19 – Additional data visualisation

Incidence data reported in Chapter 3 (*e.g.*, section 3.4.2) were visualised and annotated using Tableau Desktop 2018.31. Maps produced in Tableau made use of OpenStreetMap (© OpenStreetMap contributors) which is licenced under a CC-BY-SA licence [10]. Phylogenetic trees were visualised and annotated using Figtree v1.4.3 [11], iCANDY [12], iTOL v3 [352], and the Phandango web server [353]. Various graphs and plots were produced using Adobe Illustrator CC v23.0.4, or R v3.5.1 with the ggplot2 v3.1.1 package [354] and reshape v0.8.8 [355] packages. Various comparative genomics figures were generated using Artemis v16 [356], ACT v13 [336], DNAplotter v1.11 [357], and Easyfig v2.2.2 [337]. Where figures were edited manually, this was performed using Adobe Illustrator CC v23.0.4.

[10] https://www.openstreetmap.org/copyright
[11] http://tree.bio.ed.ac.uk/software/figtree/
[12] https://github.com/simonrharris/iCANDY

2.2 – Experimental methods

2.2.1 – Strains, plasmids and oligonucleotides

The bacterial strains, plasmids, and oligonucleotide primers used in thesis chapters involving molecular and genetic experiments (Chapters 4 and 6) are listed in Table 2.1. This list excludes isolates which were exclusively used for whole-genome sequencing – these are listed in Appendix 1.

Internal ID	Strain name	Genotype/Details	Source/Reference
Vibrio cholerae			
MJD382	NCTC 30	Non-O1/O139 (probably O2). Isolated in 1916; Alexandria, Egypt. AmpR	NCTC
MJD439		Second clone of NCTC 30. AmpR	
MJD367	NCTC 10732	Serotype O1 Inaba, classical biotype. Isolated in 1952; India.	NCTC
MJD1404			
MJD389	NCTC 5395	Serotype O1 Ogawa, El Tor biotype. Pre-seventh pandemic. Isolated in 1938; Iraq. AmpS	NCTC. Sequenced in 2016 [213]
MJD1402	NCTC 10256	O1 El Tor biotype. 7PET lineage. Toxigenic. Hong Kong, 1961.	NCTC [358]
MJD1403	MJD474	O1 El Tor biotype. 7PET lineage. Toxigenic. Traveller from Somalia, 2017.	Claire Jenkins
MJD1405	NCTC 5596	Classical lineage. O1 classical biotype. Pre-1939.	NCTC
MJD1406	A213	Gulf Coast. O1 El Tor biotype. Non-toxigenic. Georgia, 1984.	IVI [234]
MJD1407	A219	Gulf Coast. O1 El Tor biotype. Toxigenic. Georgia, 1986.	IVI [234]
MJD1408	MJD462	Non-O1. Related to MS-6 and Chinese non-7PET isolates. Traveller from Thailand, 2017.	Claire Jenkins
MJD1409	NCTC 9422	O1 El Tor biotype. Non-7PET. Non-toxigenic. Pre-1955.	NCTC
Escherichia coli			
MJD839	ER2420 pACYC184	K-12 cloning strain harbouring pACYC184. CmR TcR	Francesca Short/NEB
MJD841	NEB® 5-alpha	*fhuA2* Δ*(argF-lacZ)U169 phoA glnV44* Φ80 Δ*(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	NEB
MJD842	NEB® 5-alpha pUC19	K-12 cloning strain harbouring pUC19. AmpR	This study
MJD844	NEB® 5-alpha pACYC184	K-12 cloning strain harbouring pACYC184. CmR TcR	This study
MJD847	MJD847	NEB® 5-alpha harbouring pMJD61. AmpR CmR TcS	This study

Plasmid name	Genotype/Details	Source/Reference
pACYC184	Low-copy cloning vector. CmR TcR	[359]
pUC19	High-copy cloning vector, ampicillin resistance positive control. AmpR	[360]
pMJD61	pACYC184 Ω(*tet*::$bla_{CARB\text{-}like}$). AmpR CmR TcS	This study

Primer ID	Other name	Sequence 5'-3'
oMJD96	BamHI_blaCARB-like-NCTC30_orf_5	CCGGATCCGGTTTCAGTGCCTAATGCTTTAAGTTAAGATG
oMJD97	blaCARB-like-NCTC30_orf_SalI_3	CCGTCGACATCAACGCGACTGTGATGTATAAACTTCAA
oMJD88	blaCARB-like-NCTC30_int _5	TGGGGTCACATACATGAAGTCT
oMJD89	blaCARB-like-NCTC30_int _3	CAGCAATACTCCACTTCACTG
oMJD98	pACYC184_tet_seq_Pf	GTTAAATTGCTAACGCAGTC
oMJD99	pACYC184_tet_seq_Pr	GTGAATCCGTTAGCGAGGTG
oMJD135	VC_2135_check_Pf	GTCAGGCAGATAGCTCAAACT
oMJD136	VC_2135_check_Pr	CTCATTGCTACCTCTGATGCC

Table 2.1 – Strains, plasmids, and oligonucleotides. Restriction enzyme recognition sites are underlined. AmpR: ampicillin resistant; CmR: chloramphenicol resistant; TcR: tetracycline resistant. AmpS: ampicillin sensitive. TcS: tetracycline sensitive.

2.2.2 – Routine bacterial culturing

V. cholerae and *E. coli* were cultured routinely on LB media (10 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride), supplemented with agar (1.5%) and antibiotics as appropriate (100 µg/ml ampicillin, 10 µg/ml chloramphenicol, or 10 µg/ml tetracycline). Pre-poured thiosulfate-citrate bile salt (TCBS) agar plates (Oxoid # PO0194) was used for selective culture of *V. cholerae* (on this medium, *V. cholerae* typically grows as yellow colonies). Plasmids were maintained in strains by culturing on LB media supplemented with antibiotics where required. All manipulations with live *V. cholerae* were carried out at Containment Level 3 (CL3) within a Class II microbiological safety cabinet (MSC).

2.2.3 – Single-colony purification of strains received during this project

Throughout this project, live *V. cholerae* were received from multiple sources, typically shipped on agar stabs or slopes. Upon receipt, these isolates were sub-cultured immediately on to LB agar plates and grown overnight at 37 ˚C (passage 1). At least two distinctly-isolated and representative colonies were streaked onto both TCBS and LB agar plates and grown overnight at 37 ˚C (passage 2). If bacteria recovered well on TCBS media, single colonies were taken from this plate and purified once more onto LB media (grown overnight at 37 ˚C). Colonies from these plates (passage 3) were then taken and mixed in a cryotube containing a 1:1 mix of 50% v/v glycerol and sterile LB media. These were indexed and frozen at -80 ˚C. If growth on TCBS was poor on day 1, distinctly-isolated colonies were taken from the LB agar plates from day 1 and frozen similarly. At least two independent colonies were frozen per strain.

2.2.4 – Recovery of lyophilised NCTC isolates

V. cholerae were received from NCTC as lyophilised stocks, and were revived, sub-cultured, and frozen at -80 ˚C following the method published by Public Health England Culture Collections [13]. Ampoules containing lyophilised bacterial culture were opened in a Class II MSC, and the contents were rehydrated at room temperature with 0.5 ml LB broth. After five minutes incubation, this suspension was mixed and spread over LB agar plates, which were incubated overnight at 37 ˚C (passage 1). Three well-isolated colonies from these plates were

[13] https://www.phe-culturecollections.org.uk/

single-colony purified onto both LB and TCBS agar (passage 2). Colonies were taken from TCBS plates (or from LB plates if the strain grew poorly on TCBS plates) and were used to inoculate 3 ml LB liquid media in a disposable test-tube. This was incubated for 24 h with shaking (180 rpm) at 37 ˚C (passage 3). Cultures were mixed with glycerol (25% v/v final concentration) and frozen at -80 ˚C.

2.2.5 – Extraction of gDNA from V. cholerae

DNA extractions from *V. cholerae* were carried out using the Masterpure Complete DNA and RNA Purification kit (Epicentre, #MC85200), with modifications to the manufacturer's protocol. An LB agar plate was lawned using a single isolated colony from each strain to be processed. This plate was grown overnight (37 ˚C). Five loopfuls of bacterial culture were taken from this plate and mixed with 300 µl Tissue & Cell Lysis Solution containing Proteinase K, vortexed (10 s), and then incubated at 65 ˚C for the shorter of 20-25 min or until the suspension had fully cleared. Samples were vortexed every ~ 5 min during this incubation to ensure even lysis of bacterial cells. RNAse A was then added to each sample, which were incubated at 37 ˚C for 30 min to ensure complete digestion of residual RNA before samples were chilled on ice for ≥ 5 min.

Proteins were precipitated from each sample by adding MPC Protein Precipitation Solution (150 µl) and vortexing the sample immediately (10 s). Precipitated debris was collected by centrifugation (16,000 x *g*; 10 min; 4 ˚C). Additional MPC reagent was added to each cleared sample (30 µl) to remove residual protein contaminants. Samples were then vortexed and re-centrifuged (16,000 x *g*; 10 min; 4 ˚C). Genomic DNA (gDNA) was precipitated from the cleared supernatant in a fresh 1.5 ml Eppendorf tube using room-temperature isopropanol (500 µl). Precipitated gDNA was collected by centrifugation (16,000 x *g*; 10 min; 4 ˚C) and washed with room-temperature 70% v/v ethanol (2 x 1 ml). The pellet was dried and resuspended in 80 µl nuclease-free water – ethylenediaminetetraacetic acid (EDTA) was excluded from the resuspension solution to avoid interfering with PacBio sequencing chemistry. Samples were then tested for sterility: 2 µl was spotted onto an LB agar plate which was incubated overnight at 37 ˚C; the absence of bacterial growth was interpreted as the sample being sterile. Sterile DNA samples were removed from containment and, where appropriate, submitted for sequencing.

2.2.6 – Next-generation DNA sequencing

gDNA was sequenced using the Illumina X10 and the PacBio RSII platforms by the high-throughput and long-read teams at WSI. Briefly, DNA fragments of approximately 450 bp were produced from 0.5 µg gDNA and used for Illumina library creation. Libraries were sequenced on a 150 bp paired-end run (target coverage 30 X; *V. cholerae* were multiplexed with up to 384 samples per Illumina X10 lane). Approximately 10 µg gDNA was used for PacBio sequencing, using polymerase version P6 and C4 sequencing chemistry reagents. gDNA from NCTC 30 batch 4 was sequenced on the PacBio Sequel platform.

2.2.7 – High-throughput analysis of growth kinetics

Single colonies of *V. cholerae* were picked and suspended in 0.5 ml LB broth by vortexing (10 s). Two microlitres of this suspension were used to inoculate 150 µl LB in a 96-well flat-bottomed microtitre plate (Corning CoStar #3595) and a gas-permeable seal was applied to the plate. Plates were incubated at 37 ˚C in a BMG Fluostar Omega microtitre plate reader for 24 h, with shaking at 300 rpm using the 'meander corner well shaking' mode. Absorbance (λ = 600 nm) was measured every 15 min (cycle time: 900 s); four readings were taken *per* well *per* timepoint in a square matrix and averaged. This was to minimise the risk that debris or aggregated cells passing through the light beam would artificially increase absorbance readings. Three colonies *per* strain (biological triplicate) were used to inoculate six technical replicate cultures each.

2.2.8 – Transmission electron microscopy

Bacterial colonies were picked and suspended in 0.5 ml sterile water. The suspension (4 µl) was applied to a glow-discharged Formvar carbon film copper TEM grid (FCF-100-Cu) and mixed with ammonium molybdate solution (2.5% final concentration). Following initial tests to confirm that this treatment killed *V. cholerae*, inoculated grids were removed from CL3 for microscopic analysis. Images were acquired using a FEI Tecnai G2 Spirit BioTWIN.

2.2.9 – Motility assays

Clearly-isolated bacterial colonies were re-suspended in 0.5 ml LB media by vortexing in a Class II MSC. The suspension (2 µl) was used to inoculate motility LB agar plates (0.3% agar in 140 mm dishes). The pipette tip was pushed through the agar surface during inoculation. Plates were incubated face up at 37 ˚C for 18 h prior to being photographed. Each experiment was repeated in biological triplicate.

2.2.10 – Amplicon sequencing of the flrC *locus*

The *flrC* mutation was confirmed by amplifying *flrC* from *V. cholerae* gDNA using Phusion® high-fidelity DNA polymerase (NEB, # M0530S) and primers oMJD135 and oMJD136 (annealing temperature: 55 ˚C; extension time: 60 s, 30 cycles). The resultant amplicon was purified using the QIAquick PCR Purification kit (Qiagen, #28104) according to the manufacturer's instructions. The DNA concentration in each purified sample was estimated using a Nanodrop One instrument, and an appropriate amount of DNA and sequencing primer was sent for amplicon sequencing (GATC/Eurofins). All amplicons were sequenced in both directions (*i.e.*, with both oMJD135 and oMJD136).

2.2.11 – Chemical transformation of E. coli

Chemically-competent *E. coli* were either purchased from New England Biolabs (NEB, #C2987I), or cultures of *E. coli* were treated with calcium chloride ($CaCl_2$) to produce stocks of competent cells. Chemically-competent cells were produced from 100 ml cultures of *E. coli* which harboured no plasmids, cultured in LB media in a baffled 500 ml Erlenmeyer flask (220 rpm, 37 ˚C) for 5 hours. The culture was then split into 2 x 50 ml Falcon tubes, and cells were pelleted by centrifugation (3,900 x *g*, 10 min, 4 ˚C). The supernatant was discarded and pellets immediately resuspended in 40 ml 0.1 M $CaCl_2$ solution. These suspensions were incubated on ice for 30 min and were then centrifuged (3,900 x *g*, 10 min, 4 ˚C). The cell pellets were washed twice in 20 ml 0.1 M $CaCl_2$, centrifuging after both washes (3,900 x *g*, 10 min, 4 ˚C). Both pellets were resuspended in 2 ml of 0.1 M $CaCl_2$ + 25% v/v glycerol, and this suspension was aliquoted into ice-cold 1.5 ml Eppendorf tubes (50 µl) before being frozen at -80 ˚C immediately.

Chemically-competent *E. coli* were transformed by defrosting an aliquot of competent cells on ice. Plasmid DNA (≤ 5 µl) was added to these cells, mixed briefly and gently using the pipette tip used to add the DNA, and the cells were incubated on ice for ≥ 30 min. Cells were then heat-shocked in a water bath at 42 ˚C for exactly 30 s, and then returned to ice for 5 min. LB broth pre-warmed to 37 ˚C was added to each tube of heat-shocked cells, and the mixture was transferred to a disposable 30 ml Sterilin tube to be incubated at 37 ˚C (1 hr, 220 rpm). From this culture, 150 µl was spread onto LB media containing appropriate antibiotics to select for transformants. An appropriate positive control (1 µl of a stock of either pUC19 or pACYC184) and negative control (5 µl of buffer EB) was included in every transformation experiment. Inoculated plates were incubated overnight and putative clones were sub-cultured for further analysis.

2.2.12 – Molecular cloning of $bla_{CARB\text{-}like}$

To clone $bla_{CARB\text{-}like}$, primers oMJD96 and oMJD97 were used to amplify $bla_{CARB\text{-}like}$ from gDNA extracted from NCTC 30. The primers were designed to incorporate restriction enzyme sites and STOP codons as outlined in Figure 4.22. PCR was carried out using Phusion® high-fidelity DNA polymerase according to the manufacturer's instructions, and a 10 mM dNTP mix (Thermo Scientific, #R0191). Twenty-nine PCR cycles were performed (annealing temperature: 65 ˚C; extension time: 90 s). Reaction intermediates were purified using the QIAquick PCR Purification kit (Qiagen, #28104) according to the manufacturer's instructions.

Approximately 1 µg of both this insert and pACYC184 were digested using 20 units each of BamHI and SalI (NEB, #R3136S and #R3138S respectively) in CutSmart buffer for 15 min at 37 ˚C. One unit of rSAP (NEB, #M0371S) was added to the pACYC184 digest mixture in order to dephosphorylate the 5' ends of the digested vector. Both digestion reactions were continued for a further 30 min before digested DNA was purified as above.

Digested insert and vector were mixed in a molar ratio of approximately 3:1, and ligated using T4 DNA ligase (NEB, #M0202S) at room temperature for 150 min. A ligase-minus reaction was included as a negative control. Ligase activity was inhibited by heating the mixtures to 65 ˚C for 10 min. These ligation mixtures (5 µl) were then used to transform competent *E. coli*.

E. coli that exhibited resistance to both ampicillin and chloramphenicol upon transformation, and were also subsequently confirmed to be tetracycline-sensitive, were cultured and stored as glycerol stocks. Plasmids were prepared from these transformants using the QIAprep Spin Miniprep kit (Qiagen, #27104) according to the manufacturer's instructions. The presence of $bla_{CARB\text{-}like}$ in pMJD61 was checked by PCR using oMJD88 and oMJD89 (homologous to $bla_{CARB\text{-}like}$), and confirmed by Sanger sequencing (GATC/Eurofins) with oMJD98 and oMJD99 (homologous to the sequences flanking *tet* on pACYC184).

2.2.13 – Modified antimicrobial sensitivity assay

A single colony of each strain to be assayed was used to grow a lawn of culture on an LB agar plate overnight (37 ˚C). Plasmid-harbouring strains were cultured on appropriate selective media. Sections of the lawn were suspended in 1.0 ml LB medium and the OD_{600} of this suspension was normalised to 0.5. This suspension was well-vortexed and cotton swabs were used to inoculate an LB agar plate with these standardised suspensions. Plates were allowed to air-dry for 15 min before an MICEvaluator Ampicillin test strip (Oxoid, #MA0110F) was applied to the plate surface. Plates were then incubated for 20 h at 37 ˚C. Break points were determined using the manufacturer's instructions.

2.2.14 – Extraction of plasmids from V. cholerae

Plasmids were extracted from *V. cholerae* using the Qiagen Plasmid Midi kit (# 12143) with minimal modification to the manufacturer's instructions. Briefly, 25 ml cultures of *V. cholerae* were grown overnight at 37 ˚C in baffled 250 ml Erylenmeyer flasks (180 rpm). Cells were collected by centrifugation (3,900 x *g*, 5 min, 4 ˚C) and the supernatant was discarded. The pellet was resuspended in 4 ml buffer P1 containing RNAse A. Buffers P2 and P3 were added according to the manufacturer's instructions, and all samples were kept on ice for ≥15 min before centrifugation (3,900 x *g*, 45 min, 4 ˚C). Debris had not fully pelleted from the samples, and so the supernatant was transferred to a fresh 50 ml Falcon tube and re-centrifuged (3,900 x *g*, 15 min, 4 ˚C). The resultant clarified supernatant was added to a pre-equilibrated Qiagen-Tip 100 and was allowed to flow through the column. The column was then washed (2 x 10 ml Buffer QC) before impure plasmid DNA was eluted in 5 ml Buffer QF. DNA was precipitated using 3.5 ml room-temperature isopropanol and centrifuged (3,900 x *g*, 30 min, 4 ˚C). The supernatant was discarded and the pellet of DNA washed with 2 ml 70% v/v ethanol before

being air-dried and rehydrated with 500 µl buffer TE. These samples were sterility-tested and removed from the CL3 facility.

2.2.15 – Bacterial cultures for transcriptomic experiments

Transcriptomic experiments were carried out using cultures derived from three independent colonies of a bacterial strain of interest. Each strain to be assayed was streaked for single colonies from -80 ˚C stocks onto an LB agar plate, which was incubated overnight at 37 ˚C. Three well-isolated colonies *per* strain were used to inoculate 3 ml overnight cultures in LB media in 30 ml Sterilin tubes, which were grown overnight (37 ˚C, 180 rpm). The OD_{600} of each overnight culture was then determined.

Baffled 250 ml Erlenmeyer flasks were used as the culture vessels for transcriptomic experiments, and contained a pre-prepared culture medium appropriate for the experiment in question (LB media). All media used within an experiment were prepared as part of the same batch, and using the same base reagents, to minimise batch-effect variation amongst transcriptomic experiments. An appropriate volume of overnight culture was added to these flasks to a final OD_{600} of 0.01. The flasks were then loaded into a shaking incubator and grown for an appropriate period of time (either to a target OD_{600} or for a defined period of time post-inoculation). Once the endpoint for the experiment had been reached, the contents of each flask was transferred to a 50 ml Falcon tube and cells were sedimented immediately (3,900 x *g*, 5 min, room temperature). The supernatant was discarded and the resultant cell pellet was immediately frozen at -80 ˚C. All cell pellets pertinent to an experiment were batched together in order to ensure they were all processed on the same day, and under the same conditions, to minimise variability.

2.2.16 – RNA isolation and purification

The Masterpure Complete DNA and RNA Purification kit (Epicentre, #MC85200) was used to extract RNA from *V. cholerae*. Prior to processing samples, MSC surfaces, pipettes, plasticware *etc*., were decontaminated using RNaseZap to remove residual RNase contaminants. Cell pellets were thawed on ice, treated with 300 µl Tissue & Cell Lysis Solution containing Proteinase K, vortexed (10 s) and incubated at 65 ˚C for 20 min with intermittent vortexing. Pellets were cooled on ice. MPC Protein Precipitation Reagent (150 µl) was added

to each sample, tubes were vortexed (10 s) and the contents centrifuged (13,200 x *g*, 10 min, 4 ˚C). If the resultant pellet was unstable, an additional 25 µl MPC reagent was added and the sample re-vortexed and re-centrifuged.

Supernatants were transferred to clean 1.5 ml Eppendorf tubes and total nucleic acids were precipitated using 500 µl room-temperature isopropanol. Tubes were inverted 40 times and then centrifuged to collect precipitated nucleic acids (13,200 x *g*, 10 min, 4 ˚C). Supernatants were discarded and residual isopropanol was removed using a pipette. Each nucleic acid pellet was resuspended in 200 µl DNase buffer containing active DNase I, and incubated at 37 ˚C for 30 min. After this incubation, 200 µl 2 X Tissue and Cell Lysis Solution (TCLS) was added to each sample, followed by 200 µl MPC reagent, to precipitate and inactivate DNase. Samples were vortexed (10 s) and centrifuged (13,200 x *g*, 10 min, 4 ˚C). Supernatants were transferred to clean 1.5 ml Eppendorf tubes and DNA-free RNA was precipitated using 500 µl room-temperature isopropanol. Samples were mixed by inversion 40 times and then centrifuged (13,200 x *g*, 10 min, 4 ˚C). Precipitated RNA pellets were washed twice using 1 ml freshly-prepared 70% v/v ethanol. The pellets were then dried and resuspended in 40 µl nuclease-free TE buffer. RiboGuard RNase inhibitor (1 µl) was added to each sample, which were then tested for sterility before being immediately stored at -80 ˚C.

2.2.17 – RNA integrity assessment

Once sterile RNA was removed from the CL3 facility, it was only defrosted once before being sent for sequencing, at which point its integrity was tested and it was aliquoted into a 96-well plate for submission for sequencing. To assess RNA integrity, an aliquot of each sample (5 µl) was electrophoresed on a 1% agarose gel, to check for the presence of discrete 23S and 16S rRNA bands. Select samples were also analysed using the Agilent Bioanalyser 2100 and the RNA 6000 Nano kit (Agilent, CA USA).

2.2.18 – rRNA depletion and RNA sequencing

rRNA was depleted and sequencing libraries prepared using the Ribo-Zero rRNA Removal Kit (Illumina) by the WSI Bespoke sequencing teams. RNA libraries were strand-specific and sequenced on a 150 PE Illumina 4000 run, multiplexed at 24 samples *per* lane (aiming to produce a minimum of ten million reads *per* sample).