Exploring the genomic and phenotypic diversity of the *Vibrio cholerae* species

Matthew James Dorman

Wellcome Sanger Institute

and

Churchill College, University of Cambridge

July 2020

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the Biology Degree Committee (60,000 words).

Matthew J. Dorman
July 2020

Summary

Exploring the genomic and phenotypic diversity of the *Vibrio cholerae* species

Matthew James Dorman

Vibrio cholerae is the aetiological agent of cholera, an acute diarrhoeal disease which is estimated to result in up to 143,000 deaths *per annum*. Cholera is a considerable public health concern because it can spread rapidly in and explosive pandemics. Current pandemic cholera is caused by a highly-clonal phylogenetic lineage of *V. cholerae* serogroup O1, which spreads across the globe in periodic 'waves'. However, *V. cholerae* is a species rich in diversity, and although much is known about the population structure of the pandemic lineages, the biology and pathogenicity of non-pandemic and non-O1 *V. cholerae* has been comparatively neglected. In this dissertation, I have studied the biology, genome dynamics, and diversity of non-pandemic *V. cholerae*, in comparison to the current pandemic lineage.

I first present an analysis of the 1992-1998 cholera epidemic in Argentina, a country which had been free of pandemic cholera for nearly 100 years before 1992. I use the genome sequences of 490 *V. cholerae* from Argentina to study the micro-evolution of the pandemic lineage upon its introduction into a naïve population. I use these data to describe the progression of the Argentinian cholera epidemic using genomic epidemiology approaches, and to contrast this pandemic lineage to the non-epidemic *V. cholerae* that were present in Argentina at the same time as the pandemic lineage.

I then present a study of important recent and historical *V. cholerae* isolates, sequenced to completion using long-read technologies. I describe aspects of these genomes that could only be resolved using closed assemblies, and present functional validations of several *in silico* observations. Having performed this forensic, manual study of a small number of genomes, I then extrapolate those insights into a wider context, by mapping the distribution of key genetic determinants of important *V. cholerae* phenotypes across a phylogenetic tree of 651 highly-diverse *V. cholerae*. Finally, I integrate the knowledge gained in this research to make a rational selection of *V. cholerae* isolates for transcriptomic analysis, based on their phylogenetic position and gene content, to investigate whether differential gene expression might explain the stark differences between pandemic and non-pandemic *V. cholerae*.

The data presented here add substantially to our understanding of the diversity of *V. cholerae*. They emphasise the stark differences in genome flux and evolution between pandemic and non-pandemic lineages. They also show that many of the genetic and phenotypic markers of epidemic and pandemic lineages are misleading, and do not describe that which they were originally chosen to describe.

General contribution, copyright, and ethics statements

The nature of this research project, specifically the collation of a large collection of bacterial strains and DNA extracts, was highly collaborative. Sequencing libraries were generated by the staff of the Wellcome Sanger Institute core pipelines, except where indicated. Electron micrographs were captured by Claire Cormie and David Goulding. In all other cases, I performed the work and analyses described herein, and produced all figures, except where stated in contribution statements at the beginning of each chapter.

Several of the figures and results reported in this dissertation have been published post-peer-review, or are in press, under open-access copyright licences. Any data, figures, or text from manuscripts arising from this research that have been reproduced here are acknowledged and cited in line with the CC-BY 4.0 licence under which they were published. The maps in Chapter 3 rely on data from OpenStreetMap, which are made available under a CC-BY-SA licence which permits reuse with appropriate recognition and citation (see Chapters 2, 3). Figure 3.1 was drawn using publicly-available data from the World Health Organisation and the Pan American Health Organisation. All other figures and data were produced as part of this research and, to our knowledge, should not be affected by matters of external copyright.

All of the bacteria sequenced and handled as part of this PhD were obtained from publicly-accessible culture collections, or were originally collected by our collaborators as part of routine surveillance for public health purposes. Identifiable patient data were not solicited or made available to us at any stage of this research. No experimentation involving eukaryotic cell lines or animal models was performed. Accordingly, ethical approval for these projects was not required.

Publications

Publications arising directly from this work:

Dorman MJ, Domman D, Poklepovich T, Tolley C, Zolezzi G, Kane L, Viñas MR, Panagópulo M, Moroni M, Binsztein N, Caffer MI, Clare S, Dougan G, Salmond GPC, Parkhill J, Campos J, Thomson NR. Genomics of the Argentinian cholera epidemic elucidate the contrasting dynamics of epidemic and endemic *Vibrio cholerae*. *Nature Communications* **11** (1):4918. PMCID: PMC7530988. DOI: 10.1038/s41467-020-18647-7.

Dorman MJ, Kane L, Domman D, Turnbull JD, Cormie C, Fazal M-A, Goulding DA, Russell JE, Alexander S & Thomson NR (2019). The history, genome and biology of NCTC 30: a non-pandemic *Vibrio cholerae* isolate from World War One. *Proceedings of the Royal Society B* **286** (1900): 20182025. PMCID: PMC6501683. DOI: 10.1098/rspb.2018.2025.

Dorman MJ*, Domman D*, Uddin MI*, Sharmin S, Afrad MH, Begum YA, Qadri F & Thomson NR (2019). High quality reference genomes for toxigenic and non-toxigenic *Vibrio cholerae* serogroup O139. *Scientific Reports* **9** (1): 5865. PMCID: PMC6458141. DOI: 10.1038/s41598-019-41883-x. (* Joint first author)

Dorman MJ & Thomson NR (2020). Community evolution: Laboratory strains in the era of genomics. *Microbiology* **166** (3): 233-238. Invited “insight review” article. PMCID: PMC7376263. DOI: 10.1099/mic.0.000869.

All remaining data from this dissertation are in preparation for peer-reviewed publication.

Acknowledgements

I thank Nick Thomson very sincerely for supervising this PhD research, and for his advice, guidance, and patience throughout this project. I also thank my thesis committee members – George Salmond and Gordon Dougan, my co-supervisors, and Julian Parkhill, my graduate advisor – for their input and suggestions throughout this research.

I acknowledge formally the support of those who have contributed materials, strains, and genome sequences to this project, particularly Josefina Campos and the team at the Malbrán Institute, Sarah Alexander and the team at NCTC, Florian Marks (International Vaccine Institute), Claire Jenkins (Public Health England), and Firdausi Qadri (icddr,b). I am very grateful to have had the opportunity to train in Argentina with the *Enterobacterias* team at the Malbrán Institute, and to have been able to visit NCTC to collate metadata and observe the NCTC operation. I thank all of those named in each chapter's contribution section for their support, and I particularly thank Daryl Domman for his intellectual input into the design of this work, and both Leanne Kane and Charlotte Tolley for their help working at CL3.

The composition of Nick's research group has changed considerably and frequently over the last four years. Rather than risking forgetting to thank anybody by name and causing inadvertent offence, I will instead thank everybody that has been part of the group or in our offices since 2016, and at the Sanger Institute more broadly, who has been involved in this work and has made suggestions or comments throughout the process. I also thank those at Sanger who played an equally important role by carrying out DNA sequencing, and providing the administrative support, infrastructure, facilities and services necessary to see this work completed, particularly Sally Kay and Liz McMinn, all of the Pathogen Informatics group, our Programme's admin team including Danielle Walker, Joseph Woolfolk, and Kate Auger, and the Institute's Graduate Studies office.

Lastly, I acknowledge Wellcome, for financial support.

Table of contents

Preamble Page

Declaration ii

Summary iii

General contribution, copyright, and ethics statements iv

Publications v

Acknowledgements vi

Table of contents vii

List of figures xii

List of tables xvi

Abbreviations xvii

Chapter 1 – Introduction 1

1.1 – Initial overview: Cholera, cholera incidence, and case definitions 2

1.1.1 – Clinical presentation and incidence 2

1.1.2 – Treatment and vaccines 2

1.1.3 – Epidemiology and case definitions 3

1.1.4 – This chapter 4

1.2 – *Vibrio cholerae* 5

1.2.1 – *V. cholerae* microbiology 5

1.2.2 – Mechanism of bacterial pathogenesis in cholera cases 7

1.2.3 – Molecular genetics of the CTXφ bacteriophage 9

1.2.4 – Regulation of *ctxAB* and virulence gene expression 10

1.2.5 – Importance of horizontal gene transfer in *V. cholerae* biology and pathogenicity 13

1.3 – Pandemic and non-pandemic cholera 14

1.3.1 – Pandemic cholera 14

1.3.1.1 – History of cholera pandemics 15

1.3.1.2 – The cholera paradigm 16

1.3.1.3 – *V. cholerae* serogroups and serotypes 17

1.3.1.4 – Classical and El Tor biotypes 18

1.3.2 – Non-pandemic cholera 20

1.3.2.1 – *V. cholerae* O139 21

1.3.2.2 – *V. cholerae* O37 22
1.3.2.3 – Cholera on the Gulf Coast 23
1.3.2.4 – Non-O1/O139 *V. cholerae* infections and virulence determinants 24
1.4 – Insights from *V. cholerae* genomics 26
1.4.1 – Comparative *V. cholerae* genomics 26
1.4.2 – *V. cholerae* phylogenetics and genomic epidemiology 27
1.4.3 – Patterns of disease and local lineages 29
1.5 – Open questions and aims of this thesis 29

Chapter 2 – Methods 31
2.1 – Computational analyses 32
2.1.1 – Bacteria sequenced in the course of this PhD research 32
2.1.2 – Genome assembly 32
2.1.2.1 – Short-read data 32
2.1.2.2 – Long read data 32
2.1.2.3 – Hybrid assemblies and data visualisation 33
2.1.3 – Sequencing quality control 33
2.1.4 – Automated genome annotation 34
2.1.5 – Pangenome calculations 34
2.1.6 – Short-read mapping and SNV detection 34
2.1.7 – Detection of potentially-recombined chromosomal regions 35
2.1.8 – Phylogenetic analyses and tree-building 35
2.1.9 – Data clustering and lineage assignment 36
2.1.10 – Comparative genomics 36
2.1.11 – Pairwise SNV and ANI calculation 36
2.1.12 – Assessment of temporal signal in phylogenetic data 37
2.1.13 – Detection of antimicrobial resistance genes, *wbeT* and *ctxB* variants, plasmid replicons and virulence determinants 37
2.1.14 – Confirming the presence of genomic elements 38
2.1.15 – Construction of BLAST atlases 38
2.1.16 – Statistical tests 38
2.1.17 – Promoter sequence analysis 38
2.1.18 – RNA-seq data analysis and statistical cut-offs 39

2.1.19 – Additional data visualisation 39
2.2 – Experimental methods 40
2.2.1 – Strains, plasmids, and oligonucleotides 40
2.2.2 – Routine bacterial culturing 42
2.2.3 – Single-colony purification of strains received during this project 42
2.2.4 – Recovery of lyophilised NCTC isolates 42
2.2.5 – Extraction of gDNA from *V. cholerae* 43
2.2.6 – Next-generation DNA sequencing 44
2.2.7 – High-throughput analysis of growth kinetics 44
2.2.8 – Transmission electron microscopy 44
2.2.9 – Motility assays 45
2.2.10 – Amplicon sequencing of the *flrC* locus 45
2.2.11 – Chemical transformation of *E. coli* 45
2.2.12 – Molecular cloning of $bla_{CARB\text{-}like}$ 46
2.2.13 – Modified antimicrobial sensitivity assay 47
2.2.14 – Extraction of plasmids from *V. cholerae* 47
2.2.15 – Bacterial cultures for transcriptomic experiments 48
2.2.16 – RNA isolation and purification 48
2.2.17 – RNA integrity assessment 49
2.2.18 – rRNA depletion and RNA sequencing 49

Chapter 3 – Contrasting sources and behaviour of epidemic and endemic *Vibrio cholerae* in the Argentinian cholera epidemic, 1992-1998 50
3.1 – Overview 51
3.2 – Specific aims 55
3.3 – Ethical statement 55
3.4 – Results 56
3.4.1 – WHO/PAHO records 56
3.4.2 – The records of isolate receipt at INEI-ANLIS 56
3.4.3 – Selection of isolates to sequence for this study 60
3.4.4 – 7PET phylogeny 63
3.4.5 – LAT-1 phylogenetics 69
3.4.6 – Inaba and Ogawa serotype variation within LAT-1 76
3.4.7 – Plasmids and antimicrobial resistance in LAT-1 78

3.4.8 – Phylogenetic contextualisation of Argentinian non-7PET isolates 80
3.4.9 – Type III secretion systems in Argentinian non-7PET isolates 81
3.4.10 – Comparison of Argentinian LAT-1 and non-7PET pangenomes 82
3.5 – Discussion 88

Chapter 4 – Long-read sequencing of modern and historical *V. cholerae* 92
4.1 – Introduction 93
4.2 – Specific aims 95
4.3 – Results 96
4.3.1 – Closed genome assemblies for toxigenic and non-toxigenic *V. cholerae* O139 96
4.3.2 – Phylogenetic position of toxigenic *V. cholerae* O139 96
4.3.3 – CTXφ prophage sequences in toxigenic *V. cholerae* O139 97
4.3.4 – Genomic island complements of *V. cholerae* O139 isolates 101
4.3.5 – Antimicrobial resistance and accessory virulence determinants 105
4.3.6 – Phylogenetic position of non-toxigenic *V. cholerae* O139 107
4.3.7 – NCTC 30 genome sequencing and assembly 110
4.3.8 – NCTC 30 motility and flagellation defects 114
4.3.9 – Antimicrobial resistance in NCTC 30 119
4.3.10 – Virulence determinants in NCTC 30 121
4.3.11 – Phylogenetic position of NCTC 30 and distribution of T3SS-2β 124
4.4 – Discussion 127

Chapter 5 – The accessory genome - concordance and conflict between *V. cholerae* genomics and phenotypic dogma 130
5.1 – Overview 131
5.2 – Specific aims 133
5.3 – Results 134
5.3.1 – Expansion of the *V. cholerae* phylogeny 134
5.3.2 – Initial characterisation of diverse *V. cholerae* 136
5.3.3 – Virulence gene distribution across the *V. cholerae* phylogeny 139
5.3.4 – Serogroup assignment of isolates *in silico* 143
5.3.5 – Distribution of key pathogenicity islands amongst *V. cholerae* 144

5.3.6 – Plasmid and antimicrobial resistance gene distribution amongst *V. cholerae* 146
5.3.7 – Phylogenetic positions of historically-important NCTC isolates 149
5.3.7.1 – Pandemic NCTC isolates 151
5.3.7.2 – Non-pandemic NCTC *V. cholerae* O1 153
5.3.7.3 – NCTC 8457 155
5.3.8 – Biotype determinants 159
5.3.8.1 - Voges-Proskauer test and acetoin biosynthesis in *V. cholerae* 160
5.3.8.2 – Cholera toxin expression: an additional contrasting phenotype between classical and El Tor biotype isolates 162
5.3.8.3 – Polymyxin B sensitivity and haemolysis 167
5.4 – Discussion 169

Chapter 6 – Variation in gene expression in phylogenetically-selected *V. cholerae* 172
6.1 – Overview 173
6.2 – Specific aims 175
6.3 – Results 176
6.3.1 – Methods optimisation and initial transcriptomic studies 176
6.3.2 – RNA integrity and sequencing 177
6.3.3 – Identification of differentially-expressed genes in pilot data 177
6.3.4 – Effect of temperature on transcriptome of Classical and 7PET isolates 189
6.3.5 – Selection of isolates for transcriptomic experiments 192
6.3.6 – Assaying differential gene expression in eight strains 196
6.4 – Discussion 201

Chapter 7 – Summary and future directions 204
7.1 - Summary of thesis findings 205
7.2 – Future directions 207

References 210

Appendix 1 – List of *V. cholerae* sequenced for this PhD research 261

List of figures

1.1 – Scanning electron micrograph of *V. cholerae* 6

1.2 – Model of *V. cholerae* pathogenesis leading to the diarrhoea characteristic of cholera 8

1.3 – Summary of gene/protein interactions involved in regulating virulence gene expression 12

3.1 – Cholera cases reported to PAHO and WHO from Argentina, 1991-1999 56

3.2 – Origins of *V. cholerae* received by INEI, 1992-2002 57

3.3 – Dates of isolation for *V. cholerae* received by INEI, 1992-2002 59

3.4 – Locations from which the isolates analysed in this study were obtained 60

3.5 – Dates of isolation for the bacteria sequenced and analysed in this study 61

3.6 – Example of the application of an assembly length cut-off to SPAdes assemblies produced in this study 62

3.7 – Maximum-likelihood core-gene phylogeny of the 490 Argentinian *V. cholerae* contextualised with 675 additional genomes 63

3.8 – N16961 chromosome regions predicted to be recombined from the 7PET alignment. 64

3.9 – A maximum-likelihood phylogeny of 7PET 65

3.10 – Confirming the presence and absence of VPI-1 and CTXφ in isolates phylogenetically-related to F99/W 67

3.11 – Confirming the absence of CTXφ and VPI-1 from CCBT0194 68

3.12 – Confirming the presence of VSP-1 in F99/W isolate genomes 69

3.13 – A1552 genome regions predicted to be recombined using the LAT-1 dataset 70

3.14 – LAT-1 phylogeny and SNV distances from A1552 71

3.15 – Root-to-tip distance *versus* time for the LAT-1 phylogeny 72

3.16 – LAT-1 pangenome gene presence/absence matrix visualisation 73

3.17 – Gene discovery as LAT-1 genomes are added to the pangenome 74

3.18 – Antimicrobial resistance genes, plasmid replicons, and *wbeT* genotype variants within LAT-1 75

3.19 – Variation in *wbeT* genotype across the LAT-1 phylogeny 77

3.20 – Comparison of a fully-assembled IncA/C2 plasmid from Argentinian isolate CCBT0329 to published *V. cholerae* multidrug resistance plasmids 79

3.21 – Non-7PET *V. cholerae* phylogeny 80
3.22 – Comparison of T3SS elements detected in Argentinian *V. cholerae* against T3SS taken from reference sequences 82
3.23 – Map showing geographic origin for isolates used in this analysis 83
3.24 – Comparing the summary statistics for the LAT-1 and *V. cholerae* species pangenomes calculated in this study 83
3.25 – Visualisation of the *V. cholerae* pangenome gene presence/absence matrix 84
3.26 – Identification of new genes as genomes are added to the *V. cholerae* pangenome 85
3.27 – Summary statistics for LAT-1 and non-7PET *V. cholerae* pangenomes 86
3.28 – ANI values for genomes sequenced in this study relative to the A1552 reference sequence 87

4.1 – Maximum-likelihood phylogenetic tree of 7PET 97
4.2 – Comparison of the CTXφ region in assembly 48853_H01 and N16961 98
4.3 – Validating the presence of multiple CTXφ copies by mapping 99
4.4 – Alignment of *ctxB* and CtxB variants 100
4.5 – Reads corresponding to both *ctxB4* and *ctxB5* from toxigenic *V. cholerae* O139 map to N16961 101
4.6 – BLAST atlas illustrating the location, presence, and absence of key genomic islands in *V. cholerae* O139 assemblies relative to the N16961 reference sequence 103
4.7 – Presence of VSP-1 on both chromosomes of *V. cholerae* O139 104
4.8 – BLAST atlas comparing *V. cholerae* O139 assemblies to MO10 105
4.9 – A phylogeny of non-7PET *V. cholerae* 108
4.10 – O-antigen chromosomal loci in select *V. cholerae* 109
4.11 – The NCTC 30 check card 110
4.12 – A timeline of events in the history of NCTC 30 111
4.13 – Illustration of the NCTC 30 chromosome sequences 112
4.14 – A comparison between chromosome 1 of NCTC 30 and N16961 112
4.15 – Mapping reads across the inversion junctions 113
4.16 – Growth kinetics of NCTC 30 and NCTC 5395 114
4.17 – *V. cholerae* transmission electron micrographs 115
4.18 – NCTC 30 is non-motile when grown in soft agar 115

4.19 – Illustration of flagellar biosynthesis and regulatory hierarchy for *V. cholerae* 116
4.20 – Schematic of the predicted truncation of FlrC caused by the frameshift in *flrC* 117
4.21 – Confirmation of *flrC* frameshift using Sanger sequencing 118
4.22 – Strategy to clone *bla*$_{CARB\text{-}like}$ from NCTC 30 gDNA 120
4.23 – Ampicillin sensitivity phenotypes of *V. cholerae* and plasmid-harbouring *E. coli* 121
4.24 – Presence and absence of genomic islands and virulence genes in the NCTC 30 genome assembly 123
4.25 – Comparison of T3SS from NCTC 30 and other Vibrios 124
4.26 – A maximum-likelihood *V. cholerae* phylogeny including NCTC 30 125
4.27 – The distribution of T3SS-2β and *bla*$_{CARB\text{-}like}$ within the *V. cholerae* phylogeny 126

5.1 – A maximum-likelihood phylogeny of 646 *V. cholerae* and 5 *Vibrio* spp. 135
5.2 – Illustrating the iterative expansion of the *V. cholerae* phylogeny during this thesis research 136
5.3 – Comparing ANI values for *V. cholerae*, *Vibrio* spp., and the group of diverse sequences highlighted in Figures 5.1 and 5.2 137
5.4 – Summary statistics and visualisation of the gene presence/absence matrix for the expanded *V. cholerae* phylogeny 138
5.5 – Distribution of key virulence genes across the *V. cholerae* phylogeny 140-141
5.6 – The distribution of O1 antigen biosynthesis genes in the *V. cholerae* pangenome 143
5.7 – Distribution of genes encoded by canonical pathogenicity islands in the *V. cholera*e pangenome 145
5.8 – Distribution of AMR genes and plasmid replicons within the *V. cholerae* phylogeny 147
5.9 – The vast majority of *V. cholerae* isolates harbour two or fewer AMR genes 148
5.10 – A *V. cholerae* phylogenetic tree annotated with the names and IDs of NCTC isolates 152
5.11 – Visualisation of the De Brujin graph for the polished, rotated hybrid assembly for NCTC 3661 154
5.12 – Comparing pNCTC3661 to three *V. cholerae* IncA/C2 plasmids 156
5.13 - Visualisation of the De Brujin graph for the NCTC 8457 hybrid assembly 157
5.14 – Gel of plasmid preps from *V. cholerae* and *E. coli* 159

5.15 – Overview of the effects of P_{aphA} alleles on the expression of acetoin metabolism genes 161
5.16 – P_{aphA} motif generated from a Clustal Omega alignment of 632 P_{aphA} sequences extracted using *in silico* PCR 162
5.17 – Overview of the effects of P_{tcpPH} alleles on virulence gene expression 163
5.18 – Coordinate effects of Classical P_{tcpPH} and P_{aphA} alleles on virulence gene expression 164
5.19 – P_{tcpPH} motif generated from a Clustal Omega alignment of 258 P_{tcpPH} sequences extracted using *in silico* PCR 165
5.20 – Distribution of P_{aphA} and P_{tcpPH} allelic variants across the *V. cholerae* phylogeny 166
5.21 – Presence of intact and disrupted polymyxin B resistance genes and *hlyA* across the *V. cholerae* phylogeny 168
5.22 – Model of genetic determinants contributing to cholera 169

6.1 – Overview of RNA-seq experimental methodology for pilot experiment 176
6.2 – Volcano plots comparing gene expression in 7PET and Classical strains at 37 ˚C 185
6.3 – Volcano plot comparing genes expressed at 37 and 30 ˚C in Classical and 7PET strains 190
6.4 – Upregulation of sialic acid metabolisms genes in Classical *V. cholerae* at 30 ˚C relative to 37 ˚C 191
6.5 – Illustration of *ctxAB* transcript levels in Classical *V. cholerae* at 30 ˚C 192
6.6 – Phylogenetic position of the live isolates chosen for this chapter research 195
6.7 – PCA comparing all of the 24 sequenced samples in this experiment 197
6.8 – PCA of the 18 samples remaining after the exclusion of MJD1405 and MJD1409 198
6.9 – The *hcp* gene (*VC_A0017*) is upregulated in Gulf Coast and a related strain of *V. cholerae* (MJD1408) relative to all other strains in this experiment 200

List of tables

1.1 – Summary of *V. cholerae* O1 biotyping phenotypes 20
1.2 – A summary of the three patterns of disease caused by virulent *V. cholerae* 29

2.1 – Strains, plasmids and oligonucleotides 41

3.1 – Presence of select pathogenicity islands in F99/W isolate genomes 66

4.1 – Summary statistics for four closed *V. cholerae* O139 assemblies 96
4.2 – Presence and absence of select pathogenicity islands present in *V. cholerae* O139 assemblies 102
4.3 – Accessory virulence genes in *V. cholerae* O139 107
4.4 – Accessory virulence genes present in NCTC 30 122

5.1 – χ^2 contingency table 148
5.2 – NCTC *V. cholerae* sequenced for this thesis research 150

6.1 – Lane IDs and summary statistics for RNA sequenced in pilot experiment 177
6.2 – Genes upregulated in 7PET relative to Classical (*i.e.*, $\log_2 FC \geq 2$) at 30 ˚C 179
6.3 – Genes upregulated in Classical relative to 7PET (*i.e.*, $\log_2 FC \geq 2$) at 30 ˚C 182
6.4 – Genes upregulated in 7PET relative to Classical (*i.e.*, $\log_2 FC \geq 2$) at 37 ˚C 186
6.5 – Genes upregulated in Classical relative to 7PET (*i.e.*, $\log_2 FC \leq 2$) at 37 ˚C 187
6.6 – Strains used in transcriptomic experiments 193
6.7 – Summary numbers for differentially-expressed genes across 8-way experiment 196

Abbreviations

7PET	Seventh pandemic El Tor
AMR	Antimicrobial resistance
ANI	Average nucleotide identity
ANLIS	Administración Nacional de Laboratorios e Institutos de Salud
AST	Antimicrobial Sensitivity Test
ATCC	American Type Culture Collection
ATCSA	Anti-Terrorism, Crime and Security Act
ATP	Adenosine triphosphate
AWD	Acute watery diarrhoea
bp	Base pair
cAMP	Cyclic adenosine monophosphate
CDC	Centers for Disease Control and Prevention
CDS	Coding sequence
CFTR	Cystic fibrosis transmembrane conductance regulator
Chr1	Chromosome One
Chr2	Chromosome Two
CL3	Containment Level Three
Conc.	Concentration
CT	Cholera toxin
CTXφ	CTX bacteriophage
ddH_2O	Double distilled water
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ENA	European Nucleotide Archive
ESBL	Extended-spectrum β-lactamase
FDR	False discovery rate
GAVI	Global Alliance for Vaccines and Immunisation
gDNA	Genomic deoxyribonucleic acid
GPS	Global Positioning System
Gsp	General secretory pathway
GTFCC	Global Taskforce for Cholera Control

GTP	Guanosine triphosphate
HGT	Horizontal gene transfer
icddr,b	International Centre for Diarrhoeal Disease Research, Bangladesh
Inc	Incompatibility group
INEI	Instituto Nacional de Enfermedades Infecciosas
INIDEP	Instituto Nacional de Investigación y Desarrollo Pesquero
IVI	International Vaccine Institute
LAT	Latin American Transfer
LPS	Lipopolysaccharide
MARTX	Multifunctional autoprocessing RTX
MIC	Minimum Inhibitory Concentration
MSC	Microbiological Safety Cabinet
MSF	Médecins Sans Frontières
MSHA	Mannose-sensitive haemagglutinin
NEB	New England Biolabs
NCTC	National Collection of Type Cultures
OCV	Oral cholera vaccine
ORF	Open reading frame
ORS	Oral Rehydration Solution
PAHO	Pan American Health Organisation
PCA	Principal component analysis
PCR	Polymerase chain reaction
PE	Paired-end
PFGE	Pulse-field gel electrophoresis
PG	Pandemic Group
PHE	Public Health England
ppGpp	Guanosine tetraphosphate
(p)ppGpp	Guanosine pentaphosphate
RNA	Ribonucleic acid
RNA-seq	Ribonucleic acid sequencing
rRNA	Ribosomal ribonucleic acid
RTX	Repeats-in-toxin
SEM	Scanning electron micrograph
SXT (SXT/R391)	Sulfamethoxazole and trimethoprim resistant conjugative element

T1SS	Type I secretion system
T2SS	Type II secretion system
T3SS	Type III secretion system
T6SS	Type VI secretion system
TCBS	Thiosulfate-citrate bile salt
TCP	Toxin co-regulated pilus
TEM	Transmission electron microscopy
TIGR	The Institute for Genomic Research
USA	United States of America
V. cholerae	*Vibrio cholerae*
V. metoecus	*Vibrio metoecus*
WHO	World Health Organisation
WSI	Wellcome Sanger Institute
WW1	World War One
VPI-1	*Vibrio* pathogenicity island 1
VPI-2	*Vibrio* pathogenicity island 2
VSP-1	*Vibrio* seventh pandemic island 1
VSP-1	*Vibrio* seventh pandemic island 1