# Chapter 7

# Summary and future directions

## 7.1 – Summary of thesis findings

In this thesis, I have used a combination of *in silico* and *in vitro* approaches to increase our understanding of the differences between pandemic and non-pandemic lineages of *Vibrio cholerae*. The initial thesis aims (section 1.5) were to study the evolution of 7PET longitudinally and across a large geography, and to characterise the non-7PET *V. cholerae* present in a country during a pandemic, alongside an external introduction of 7PET. From that baseline, the thesis then set out to investigate further the diversity of non-epidemic and non-7PET *V. cholerae*, and the functional differences between pandemic and non-pandemic lineages of *V. cholerae* O1. This involved the detailed investigation of single fully-assembled genome sequences for important bacterial isolates, the collective examination of large numbers of genomes, and the design and execution of transcriptomic experiments.

In Chapter 3, I presented a study of Argentinian *V. cholerae* which has increased our understanding of how 7PET evolves, and does not evolve, following its introduction into a country and population that was naïve to pandemic cholera caused by 7PET. These findings have also been actionable, contributing to changes in public health policies in Argentina (Chapter 3). Very minimal change in 7PET was observed, in terms of SNV accumulation, gene gain/loss, and recombination, though important conclusions were drawn from the observed variation in Inaba/Ogawa serotype and genotype, and the potentially-misleading conclusions that might be drawn from looking at phenotypic serotype variation alone. The clonality of 7PET contrasted dramatically with that of the non-7PET *V. cholerae* from Argentina that were included in this thesis research. Many of these bacteria were isolated from suspected cholera cases, and the roles of virulence determinants other than CTXφ, such as T3SS, remains to be investigated. These non-7PET *V. cholerae* present in Argentina throughout the cholera epidemic are proposed to represent the population of *V. cholerae* that are truly endemic to Argentina, which may or may not cause symptoms that resemble clinical cholera.

In Chapter 4, I illustrated the insights that can be gleaned from the study of high-quality closed genome assemblies. I presented novel and unique aspects of the genome biology of recently-isolated *V. cholerae* O139, such as the presence of multiple *ctxB* alleles in the same genome, and cross-chromosomal duplications of VSP-1. I also described a genome assembly for non-toxigenic *V. cholerae* O139 which occupies an unusual position in the *V. cholerae* phylogeny. Furthermore, as part of developing methods for work with *V. cholerae* at CL3, I assembled a

closed genome assembly for NCTC 30, which we believe to be the world's oldest publicly-available non-O1 *V. cholerae*. Using this genome sequence alongside complementary experimental methods, I confirmed the presence of a functional β-lactamase in this isolate (which pre-dates the introduction of antimicrobials as therapeutics), as well as the presence of a T3SS-2β element, and provided plausible explanations for the absence of flagella in this strain.

Following this detailed characterisation of closed *V. cholerae* genome sequences, I proceeded to analyse specific, relevant genotypes and phenotypes across a set of 646 highly-diverse *V. cholerae* (Chapter 5). These included virulence determinants, antimicrobial resistance genes, and determinants of *V. cholerae* O1 biotypes, which are relevant to clinicians, to molecular microbiologists, and to those who study *V. cholerae* genomics. These data reinforced a number of previous findings, including the fact that very few plasmid replicons were detectable amongst diverse *V. cholerae* as well as within 7PET, and that T3SS virulence determinants are increasingly common amongst these diverse isolates, both those of clinical and environmental origin. This work also led to the first sequence of an historical IncA/C2 plasmid devoid of antimicrobial or heavy-metal resistance genes. This backbone may represent an "ancestral" IncA/C2 plasmid, and future work to perform phylogenetics of the many IncA/C plasmids being sequenced from *V. cholerae* and other species would be worthwhile.

These data also highlighted inconsistencies amongst previous observations that had been considered to be dogma in *V. cholerae* biology. For instance, I identified numerous examples of *V. cholerae* pathogenicity islands present in non-pandemic isolates and lineages. Canonically, these elements have been described to be unique to epidemic and pandemic lineages, and this observation emphasises the fact that as we sequence additional members of this species, our definition of these truisms and our understanding of whether a mobile element is a 'marker' for a lineage will need to be refined. Similarly, by studying historical El Tor biotype *V. cholerae* O1, I have shown that the phenotypes associated with the classical biotype, and the mutations leading to these phenotypes, collectively only describe the Classical lineage – the El Tor biotype is broadly distributed amongst diverse *V. cholerae*, irrespective of serogroup. These results highlighted and emphasised that biotyping reactions are of limited utility in modern studies of *V. cholerae*, suggesting that the microbiological resources invested in biotyping might be better employed in other ways. However, this is wholly understandable in the context of how our understanding of *V. cholerae* and of cholera has developed and

evolved since the bacterium was first observed by Pacini. As our understanding has evolved, and as technologies with which to study *V. cholerae* have improved from biochemistry to whole-genome sequencing, the resolution at which we understand this pathogen has grown. These refinements to our understanding mean that current cholera control strategies, such as those of the GTFCC, now have the opportunity to effect meaningful change to global health.

In Chapter 6, I describe transcriptomic experiments designed and executed in order to characterise lineage-specific differences in gene expression amongst *V. cholerae* O1. These were pilot experiments, designed both to optimise transcriptomic methodologies for use in CL3, and to discern robust differences in gene expression across *V. cholerae* lineages. These experiments have provided novel results as well as recapitulating the results from previous studies, which used different bacterial strains. These include apparent lineage-specific differences in T6SS regulation, and lineage-specific temperature-mediated differential expression of virulence and pathogenicity genes. The work presented in this chapter also highlights the difficulties in working with heterogenous bacteria, and the need to balance optimal experimental design with realistic practical considerations.

### 7.2 – Future directions

Cholera is an ancient disease, which we have failed to control to date, although efforts to do so are ongoing. It is clear from this thesis research, as well as from a growing body of literature, that taking a broad approach to sequencing *V. cholerae* – *i.e.*, expanding focus from the exclusive sequencing of epidemics – is generating a wealth of genomic diversity for future analysis. Although the data in this thesis has advanced our understanding of the *V. cholerae* species as discussed above and in previous chapters, they also highlight open questions in this research area, and lay the groundwork by which to investigate these questions. In future projects, sequencing and functional research must not only focus on members of 7PET, but also on environmentally- and clinically-isolated *V. cholerae* irrespective of their serogroup and toxigenicity. Efforts should be invested in carrying out comprehensive studies of the pangenome of this species as a whole, alongside functional assays to determine the transcriptomic, biochemical, and metabolic differences amongst diverse *V. cholerae*. Particular attention should be paid to the clade of diverse isolates highlighted throughout this thesis (Chapters 4, 5). It is only by adopting such an holistic approach, aiming to capture all *V. cholerae* including those that do not appear to cause disease, that we will amass sufficient

genomic data and biological material with which to explore the question of what enables pandemic lineages to cause pandemics.

In future work, it will be important to determine the significance of the lack of genomic variation amongst 7PET. It is unknown whether this lack of variation be explained mechanistically, and whether this is due to a reduced ability, or reduced opportunities, to participate in HGT. It is also unknown whether the pandemic lineage has fixed advantageous phenotypes by reducing genomic variation. This is a particularly relevant consideration because the reasons why 7PET can spread rapidly and to cause global pandemics are still uncertain. Further functional research into this area will be required to explain this observation. Similarly, the apparent lack of plasmids amongst *V. cholerae*, even amongst diverse isolates (Chapter 5), should also be studied in more detail. It is unknown why IncA/C plasmids are one of the very few plasmids that this species appears to be able to maintain. It will also be important to pursue this research avenue without assuming that databases of plasmid replicons contain all known *Vibrio* plasmids – it might be that genomic analyses fail to identify *V. cholerae* plasmids because the Plasmidfinder database simply lacks the replicon sequences for plasmids found in this species. Coupling genome sequencing to wet-lab bacteriology, and functional studies of plasmid dynamics and diversity, will expand our knowledge of HGT mechanisms and potential within this species.

The role of T3SS in disease mediated by *V. cholerae* is also a research area with great potential. Although we found in this thesis that many clinically-isolated *V. cholerae* encoded T3SS in the absence of other canonical virulence factors, causal links between these systems and pathology in patients have not yet been proven. Considerable opportunity exists to exploit the genome sequences generated in this thesis to characterise in detail the diversity of T3SS systems *in silico*, as well as to avail of having access to live isolates of multiple T3SS$^{+}$ *V. cholerae* to study the regulation of these systems, their mobility amongst different *Vibrio* spp., and their functionality in models of pathogenesis.

The potential to discover new aspects of *V. cholerae* biology using this holistic approach – studying *V. cholerae* beyond 7PET – is exemplified by the transcriptomic studies set out in this thesis (Chapter 6). However, it is important to bear in mind that although it is logical and plausible that transcriptomic responses to stresses and signals would be more similar amongst members of the same lineage than across lineages, this remains an hypothesis and must be

tested *in vitro* and *in vivo*. The work in this thesis has set the stage for subsequent experiments to be refined and optimised, to test these hypotheses, and to expand our understanding of the *V. cholerae* species beyond that of 7PET and Classical lineages.