6. Conclusion and Future Directions

6.1 Conclusion

Even though cholera is regarded in many quarters as an old disease, the current global burden is between 3-5 million cases and more than 100,000 deaths are reported worldwide annually (*http://www.who.int/wer*). While many of these cholera cases are in the endemic regions of the Indian subcontinent, the increased incidence of cholera globally since 2007 has highlighted the need of more public health research on this potentially easy to treat disease. The etiological agent, *V. cholerae* is a genetically diverse species but out of more than 200 O-antigen serogroups only isolates of O1 and O139 can cause epidemic cholera. Of the two biotypes 'El Tor' and 'classical' of serogroup O1 *V. cholerae*, El Tor account for almost all of the cases in the current seventh pandemic.

This thesis utilizes the whole genome sequencing data generated from more than 1000 seventh pandemic *V. cholerae* El Tor as well as isolates from non-seventh pandemic lineages to prove that the classical and El Tor biotypes evolved independently and do not share a recent common ancestor. This data also clearly showed that currently successful lineage of El Tor isolates is monoclonal or monophyletic and evolves in a strict clock-like manner. The phylogeographical analysis indicates that in the seventh pandemic cholera has spread in the form of independent but overlapping waves from a source population in the Bay of Bengal to other regions of the world (chapter 2).

The total *V. cholerae* collection used in this study included sets of smaller collections obtained from different cities, regions or countries on which focused studies could be performed. In chapter 3, a series of case studies described evolutionary patterns identified within populations of *V. cholerae* O1 El Tor obtained from particular countries. Studies on *V. cholerae* isolated during the catastrophic 2010 floods in Pakistan identified two clear introductions of the disease into the country and provided insight into their spread within Pakistan. The study on *V. cholerae* collected over several decades in Mexico provided evidence for the persistence of wave-1 El Tor isolates since the 1990s. A Kenyan surveillance study shed light on the local

clonality of isolates and provided evidence for a recent common ancestor with South Asian isolates. Again, evidence was provided for independent entries of cholera into Kenya and the presence of two sub-clades.

The large amount of data generated from sequencing over 1000 *V. cholerae* was exploited to design SNP typing assays, in the form of kits that are suitable for use in public health and scientific laboratories in developing countries (chapter 5). Robust SNPs were selected, based on phylogenetic analysis, and the kits were designed so that they could rapidly and accurately identify *V. cholerae* associated with any outbreak.

The genotypic basis of the Ogawa to Inaba serotype conversion in *V. cholerae* O1 El Tor was studied in detail by analysing the *wbe*T sequence from 777 of the 1002 seventh pandemic *V. cholerae* analysed in this study. This analysis identified the mutations that underpinned the serotype switching and provided insight into mechanism. *V. cholerae* collected during a phase III vaccine trial in Kolkata, India were examined to identify any temporal and serotypic correlation in a phylogenetic context (chapter 4). This analysis provided evidence for regular sweeps of Inaba and Ogawa types spreading through the trial sites over successive cholera seasons. Such information is likely to have value for supporting future vaccine studies in the field.

In summary, the data described in this PhD thesis may facilitate future cholera surveillance performed as part of public health programs at a local or national level, facilitating quick and directed actions to contain the spread of an outbreak. The academic community will also benefit from this data, which is publically available to further research on the biology and epidemiology of *V. cholerae*.

6.2 Future Directions

6.2.1 Further expansion of the sequenced *V. cholerae* collection

Perhaps the most obvious way to extend this work is to expand efforts on the whole genome based phylogenetic analysis of the seventh pandemic collection by the

addition of new *V. cholerae* isolates from around the globe. Through the existing and new collaborations between the WTSI and partners based around the world, such an effort is already well in progress. Recent and historical strain collections from Pakistan, West Africa, India and Bangladesh are being sequenced to add finesse and detail to the structure of the global seventh pandemic phylogeny. With the addition of new isolates collected with detailed metadata, a more accurate prediction of the spread of *V. cholerae* could be made alongside an increased understanding of the overall phylogenetic framework. For example, a collection of O139 *V. cholerae* has been identified at the National Institute of Cholera and Enteric Diseases (NICED) in Kolkata and arrangements are being made to have these sequenced. The O139 *V. cholerae* appear to fall into a specific lineage within the El Tor phylogenetic tree and such isolates were predominantly isolated between 1992 and 2005, mainly in Kolkata and Bangladesh. The possible reasons behind the success of a lineage for just a few years before it disappeared completely are presently unknown; this project may shed light onto some of the contributing factors. For example, do all O139 isolates fall into the same lineage and do any isolates show evidence of more extensive recombination beyond the known O-antigen loci? This study should highlight the SNP variation that occurred in O139 isolates from the time when they were first causing outbreaks to the time they went extinct. This data may give some insight into any negative selection pressure that these strains may have encountered.

Further, analysis of *V. cholerae* that fall outside the seventh pandemic El Tor lineage could shed more light onto the *V. cholerae* pool circulating in the environment that is clearly interacting with the epidemic *V. cholerae* populations. Detailed analysis of new lineages identified in this PhD (for example, MLE-1 and MLE-2 in section 3.3.8) and previously known lineages such as the US-Gulf coast (section 2.3.1) could provide vital clues about the evolution of vibrios circulating in the environment.

6.2.2 Studies investigating the evolution of *V. cholerae* within cities, countries and continents

While the continued addition of sequenced *V. cholerae* isolates is expanding our understanding of the seventh pandemic, this data is also providing an opportunity to study regional level populations in greater detail. Some of the examples are:

- Post flooding *V. cholerae* continue to be collected in Pakistan and are being analysed to determine further patterns of evolution as well as assess if they are evolving at the same rate as vibrios in other branches of the the seventh pandemic tree;
- New isolates from across Africa are being added to the phylogeny to obtain a representative sample for a pan-continental study and to determine if the dynamic of cholera in Africa is similar to that observed in the representative country, Kenya;
- Studies on cholera in Latin America are being expanded with the inclusion of isolates from Argentina and Brazil. An aim here will be to determine if the persistence of waves in Mexico is common across the whole of Latin America or is just a local phenomenon.

6.2.3 A combined transcriptomics and proteomics study of intestinal tissues taken from mice at different stages of *V. cholerae* infection

In collaboration with researchers at the University of Gothenberg in Sweden, work is currently being conducted to investigate transcriptomics and proteomics patterns in the intestinal tissue of mice during a *V. cholerae* infection. An O1 serogroup *V. cholerae* El Tor Ogawa strain X25049 was used to infect infant mice orally at a dose of 10^6 viable bacteria. Infections were performed and tissues were collected in Sweden from groups of mice at 4 and 18 hours after infection, with uninfected mice acting as controls. RNA-seq and proteomic analysis is currently underway at the WTSI using these materials. The patterns of gene expression will be compared between the infected and un-infected mice and between the mice at different time points. Protein extracts will be analysed using a Liquid Chromatography-Mass Spectrometry (LC-MS) platform to compare the translation differences at the same time points as the RNA analysis. The data from these experiments would be crucial in highlighting aspects of the host response to *V. cholerae* infection. A comparison between transcriptomics and proteomics analyses results should provide a list of genes that could be important candidates in future cholera vaccine or drug design. A further study is also planned where similar mice will be challenged orally with

cholera toxin in order to compare the impact of *V. cholerae* infection with exposure to the cholera toxin.

6.2.4 A study designed to investigate household and community level spread of *V. cholerae*

A study has been instigated to exploit the high sensitivity and resolving power of SNP based phylogenetic analysis of *V. cholerae* genomes, to investigate transmission patterns involving household contacts and index cholera patients. To achieve this, a unique *V. cholerae* collection from Dhaka, Bangladesh will be sequenced at the whole genome level in an attempt to establish transmission chains, which could not be established using VNTR in a previously published study (Kendall, *et al.*, 2010). The *V. cholerae* isolates in this study were collected as follows: In Dhaka, if an individual reported to a hospital associated with the International Centre for Diarrheal Disease and Research (ICDDRB) and was subsequently found to be positive for *V. cholerae*, the patient was classed as the index patient and their household was investigated in order to recruit individuals who shared the same food and water sources. Daily rectal swabs were then taken for 10 days from implicated co-inhabitants. These swabs were then analysed for *V. cholerae* and culture positive samples were stored. With ~100 samples from the index patients and ~150 from the household contacts, this study will use the power of whole genome analysis and detailed metadata in an attempt to identify evidence for transmission within a house or between houses in a community.

In conclusion, the data and analysis provided and described in this thesis is underpinning a series of on going investigations into the biology of *V. cholerae* in both experimental laboratory and field settings.