4. The genetic basis of serotype variation in *V. cholerae* sampled during a clinical trial in Kolkata, India

NOTE: All the clinical trial isolates were collected by our collaborators based in Kolkata, India. The isolates were collected from 2003-2010 but only 2003 to 2007 were made available and the information about the patients whether they were vaccinated or not was not available to include in the analysis. The DNA was sent to the Sanger Institute for sequencing by the sequencing pipeline teams and raw short read data was made available for the analyses. The work explained in this chapter details the phylogenetic analyses, which was done by me and therefore forms a part of my PhD thesis.

4.1 Introduction

V. cholerae LPS is composed of an antigenically variable O-antigen polysaccharide (PS) and a core-PS, including lipid A, that exhibits relatively limited variation. *V. cholerae* expressing antigenically distinct O-antigen PS can be classed using antibodies to O-antigen into different serogroups. Using this approach, more than 200 *V. cholerae* serogroups have been identified to date, although only O1 and O139 strains can cause epidemic cholera. In addition to biotyping (classical and El Tor), *V. cholerae* O1 and O139 isolates can be further divided into serotype Ogawa or Inaba (Figure 4.1) based on further antigenic properties of the O-PS.

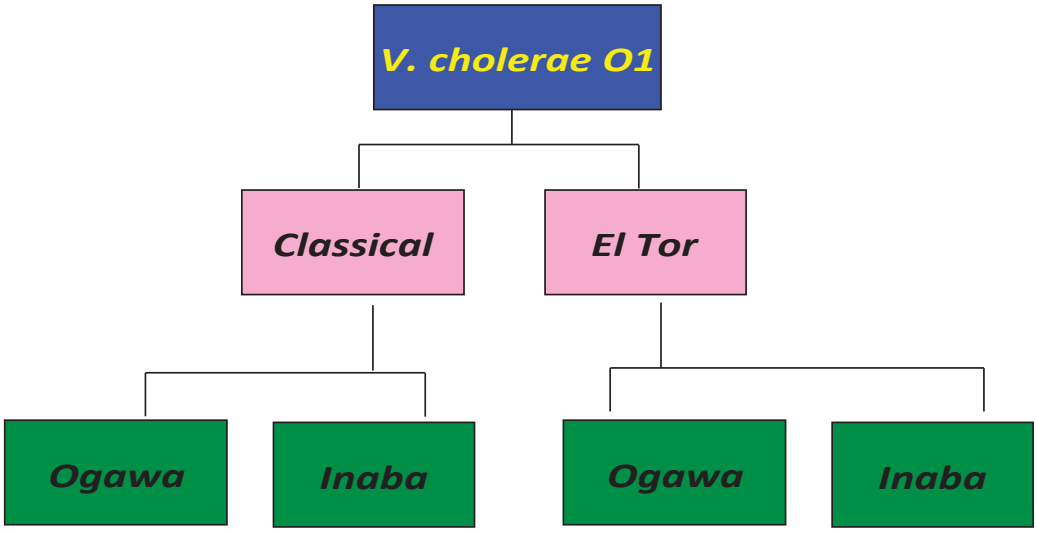


Figure 4.1: Classification of O1 *V. cholerae* into biotypes and serotypes. *V. cholerae* O1 has two biotypes Classical and El Tor and both biotypes can have two serotypes Inaba and Ogawa.

The presence or absence of a single methyl group on the terminal sugar of the O-PS can change the serotypic response of a *V. cholerae* isolate from Ogawa to Inaba. The proteins that drive the synthesis of the O1-antigen of *V. cholerae* are predominantly encoded by the '*wbe* region' of the genome, which is between 16-19 kb, varying between isolates. The *wbe*T gene, a methyl transferase, in the *wbe* operon encodes an enzyme responsible for the methylation of terminal 4-N-tetronylated-D-perosaminyl group (Chatterjee and Chaudhuri, 2003) on O-PS and the Inaba-Ogawa distinction has been correlated to the alteration in this gene (Stroeher, *et al.*, 1992). If *wbe*T is expressed in its wild type form, the resultant phenotype is Ogawa but if the gene is missing or does not drive the expression of the functional enzyme, the resultant phenotype is Inaba. Indeed, the introduction of a complete *wbe*T gene from an Ogawa into an Inaba isolate can mediate conversion to Ogawa (Stroeher, *et al.*, 1992). Interestingly, spontaneous conversion from Inaba to Ogawa occurs at a much lower relative frequency compared to Ogawa to Inaba conversions. One possible reason for this could be that while Ogawa to Inaba conversion would just require a mutation in *wbe*T, the Inaba to Ogawa conversion would need the parsimoniously much less likely event of mutation correction (for example by a recombination event). This fixation event is likely to be comparatively rare through natural evolutionary events

but could be favoured in a population if there is a strong selection for Ogawa in any specific environment.

The work described in this chapter is divided into two parts. First, the sequences of seventh pandemic O1 El Tor *V. cholerae* were screened informatically to identify mutations in *wbe*T that might influence the expression of the Ogawa and Inaba epitopes. Fortunately, for most of the sequenced isolates, there was phenotypic information available for serotype. The aims were to (a) identify common mutations within *wbe*T associated with the Ogawa to Inaba conversion; (b) determine if there were any examples where a phenotypic change in serotype did not correspond to a clear mutation in *wbe*T; (c) identify any likely examples of conversion of Inaba to Ogawa i.e. any correction of *wbe*T to wild type based on phylogenetic position; (d) identify isolates harbouring the wild type *wbe*T allele that are phenotypically Inaba. To achieve this, the *wbe*T gene of sequenced seventh pandemic *V. cholerae* were directly compared to the wild type El Tor Ogawa *wbe*T allele.

Evidence has accumulated that exposure of an individual to a *V. cholerae* Inaba challenge can provide some subsequent protection against both Inaba and Ogawa infection, whereas an Ogawa challenge only protects against an Ogawa infection (Longini, *et al.*, 2002). Consequently, the genetic basis of the Inaba-Ogawa variation was determined in a sample set of *V. cholerae* collected in a vaccine trial undertaken in Kolkata, India (Sur, *et al.*, 2009). The design of cholera vaccines has to take consideration of the serotype of the O1 *V. cholerae* used to manufacture the vaccine, particularly in the case of whole cell formulations. Consequently whole cell-based cholera vaccines currently on the market harbour killed *V. cholerae* cells as a mixture of both Inaba and Ogawa serotype strains (section 1.2.6), to get better overall protection. Thus, any knowledge about the serotypic composition of *V. cholerae* in a particular geographical region and the mechanisms by which serotypic switching might occur would be of practical value.

4.2 Results and discussion

4.2.1 *wbeT* sequence analysis

The genome sequences of each of the 1002 *V. cholerae* in the expanded seventh pandemic phylogeny (chapter 5) was assembled and the *wbe*T gene was determined to be of sufficient quality to analyse in detail in 777 of these sequences. The remaining 225 assemblies either had a contig break in *wbe*T or had poor coverage in this region. Of the 777 analysed *wbe*T gene sequences 244 were found to possess likely mutations in *wbe*T compared to the wild type allele. The types of mutations identified are shown as percentages in Figure 4.2. Non-synonymous single amino acid changes were the most frequent mutations in *wbe*T followed by frame shift mutations and insertions. The formation of stop codons, likely linked to premature termination, were more common than deletions.

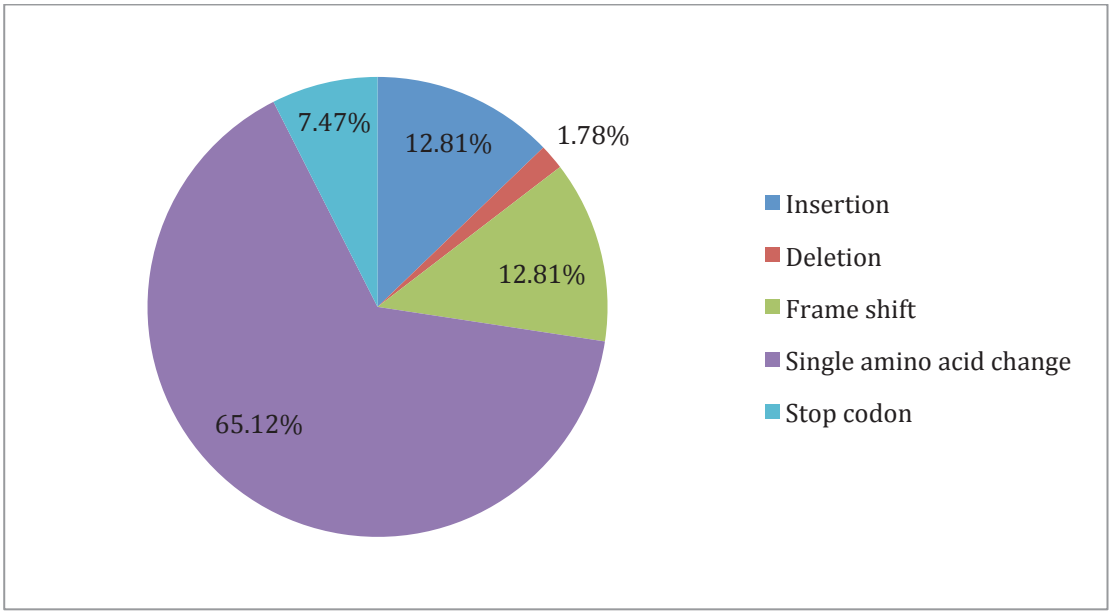


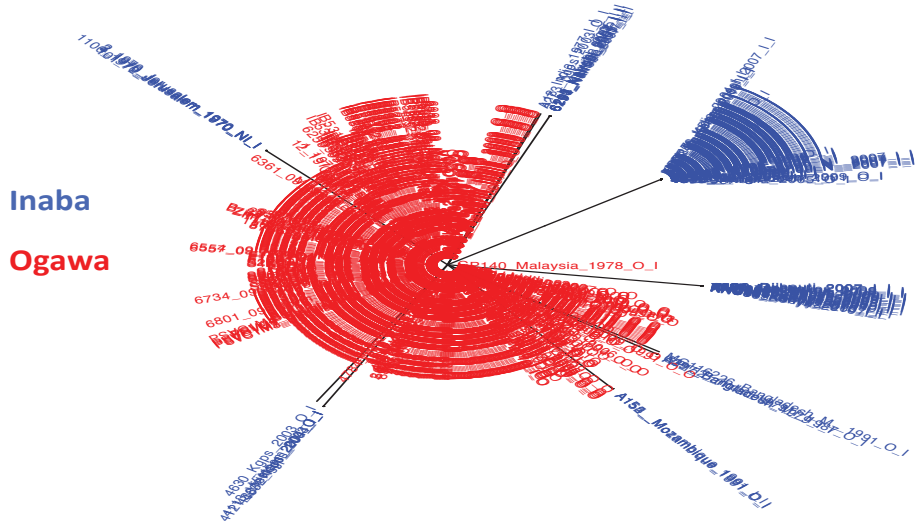
Figure 4.2: Pie-chart showing the percentage distribution of different types of mutations in the *wbeT* gene of 244 *V. cholerae* O1 El Tor sequences.

After identifying *V. cholerae* harbouring mutations in *wbe*T, a maximum likelihood phylogeny was constructed based on the *wbe*T gene alignment of all 777 analyzable sequences. This phylogenetic tree was used as a platform to visualize how well the phenotypic and genotypic characterization of serotype matched. As the reference *V. cholerae* N16961 was used to build the phylogeny and all the isolates with a wild type *wbe*T allele clustered together as one group whereas isolates with mutations formed

several small groups depending on the type of mutation (Figure 4.3A).

Since any inactivating mutation in the wild type Ogawa *wbe*T gene should result in a dysfunctional methyl transferase and therefore an Inaba conversion, serological phenotype should superimpose on this tree. However, although this was predominantly the case, there there were some phenotypically Ogawa isolates that mapped within clades with mutant *wbe*T and there were some phenotypically Inaba isolates located within clades of wild type *wbe*T (Figure 4.3B).

A.



B.

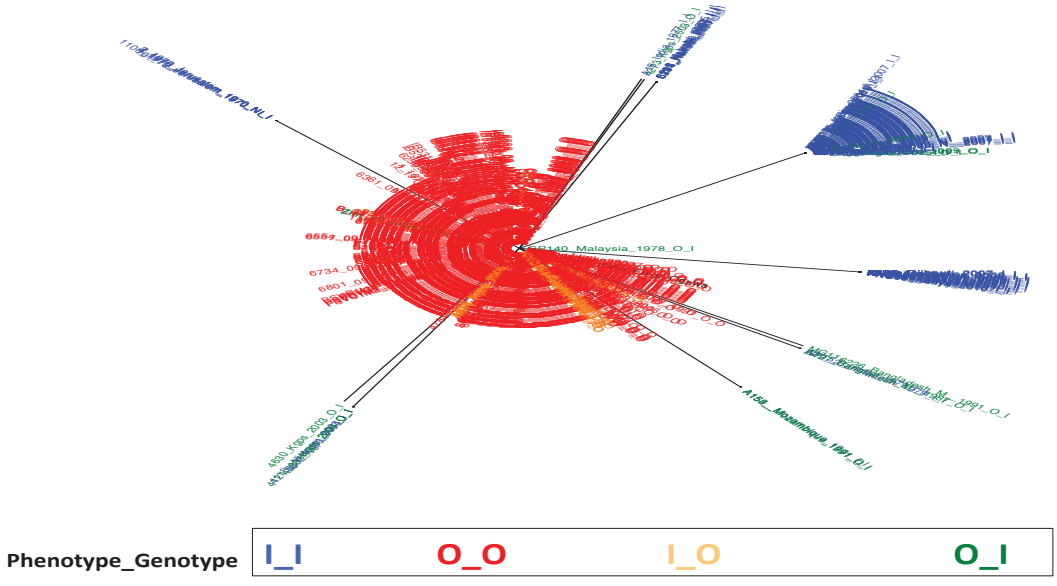


Figure 4.3: A) Maximum likelihood phylogenetic tree coloured as the Ogawa-Inaba strain arrangement based on genotype; B) The same phylogenetic tree showing how correlated phenotypic and genotypic data. I = Inaba; O = Ogawa.

Approximately ~90% of the isolates mapped onto the tree in accordance with both their genotype and phenotype, i.e. if they did not harbour any obvious mutation in the *wbe*T gene and were phenotypically Ogawa or if they did harbour a mutation in *wbe*T and they were phenotypically Inaba. However, ~10 % of the isolates did have a phenotype-genotype mismatch (Figure 4.3B, 4).

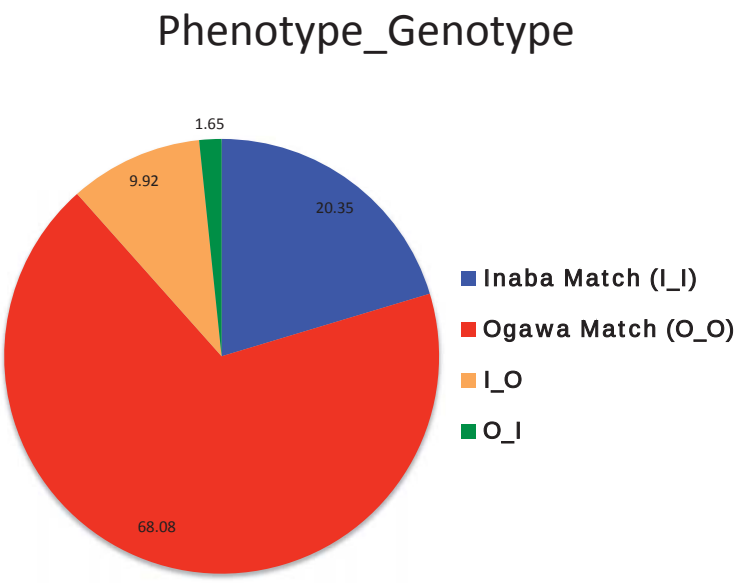


Figure 4.4: Pie-chart showing the percentage match and mismatch between the phenotypic and expected genotypic serotype. I = Inaba; O = Ogawa.

How can mismatches be explained? There were some mismatches where serotypically Inaba isolates harboured a wild type Ogawa *wbe*T allele. Such a combination of genotype and phenotype would be anticipated if the isolates harboured a mutation outside *wbeT*, which prevented the methyl transferase responsible for the methylation of terminal sugar from reaching the target, either through lack of expression or altered intracellular targeting. For example, a mutation could be in regions of the genome such as the promoter or a regulatory gene that influenced mRNA production or even

translation. Alternatively, the isolates could have been incorrectly serotyped, a relatively common phenomenon in routine serotyping laboratories (our unpublished observations).

There were a few isolates that harboured a mutated *wbe*T but were still phenotypically serotyped as Ogawa. This could be explained if such a mutation(s) did not have any effect on the expression of a functional enzyme or compensatory mutations were present elsewhere in the genome. Of course, mistakes in serotyping could also be an explanation here as well. Further work in the field will be required to investigate these possibilities in more detail.

4.2.2 Mapping *V. cholerae* from a vaccine trial performed in Kolkata to the global El Tor phylogeny

DNA from 405 *V. cholerae* O1 El Tor isolates collected from 405 cholera patients during a phase III field trail performed in Kolkata on a Shancol cholera vaccine were sequenced and analysed. This collection spanned five years, 2003-2007, with 2006 being the vaccination year (Sur, *et al.*, 2009). When these 405 *V. cholerae* were arranged in temporal order of their date of isolation and their phenotypically determined serotype was superimposed, an interesting pattern became apparent (Figure 4.5). In 2003 and 2004 Ogawa *V. cholerae* O1 El Tor isolates dominated but there was a relative increase in the number of isolates phenotypically Inaba in 2005, such that Ogawa isolates became the minority. Inaba isolates remained the dominant serotype in 2005 and 2006 but then there was an apparent decline in their population and Ogawa isolates re-emerged in 2007.

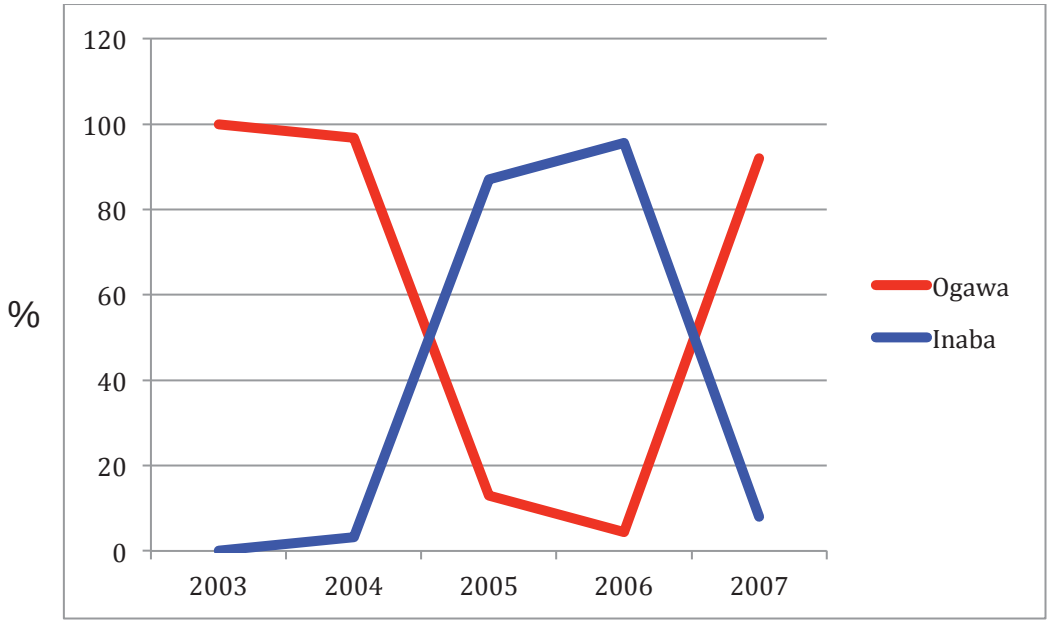


Figure 4.5: Distribution of Inaba and Ogawa serotype isolates during a clinical trial study in Kolkata, India. The data for isolates collected between 2007-2010 was not available.

Subsequently, a maximum likelihood phylogenetic tree was constructed to ascertain the position of these *V. cholerae* within the global seventh pandemic framework (not shown). All the isolates clustered in wave-3 of the seventh pandemic lineage and were distributed in both sub-clades 3a and 3b (explained in section 5.2.2.1). A further phylogenetic tree was constructed that included only the clinical trial isolates (Figure 4.6), using N16961 O1 El Tor as both reference and an outgroup, with the aim of obtaining a clearer understanding of their temporal and serotype distribution in a phylogenetic context. Interestingly, a clear temporal clustering of isolates was observed and the pattern of serotypic change from year to year corresponded well with the temporal clades in the tree.

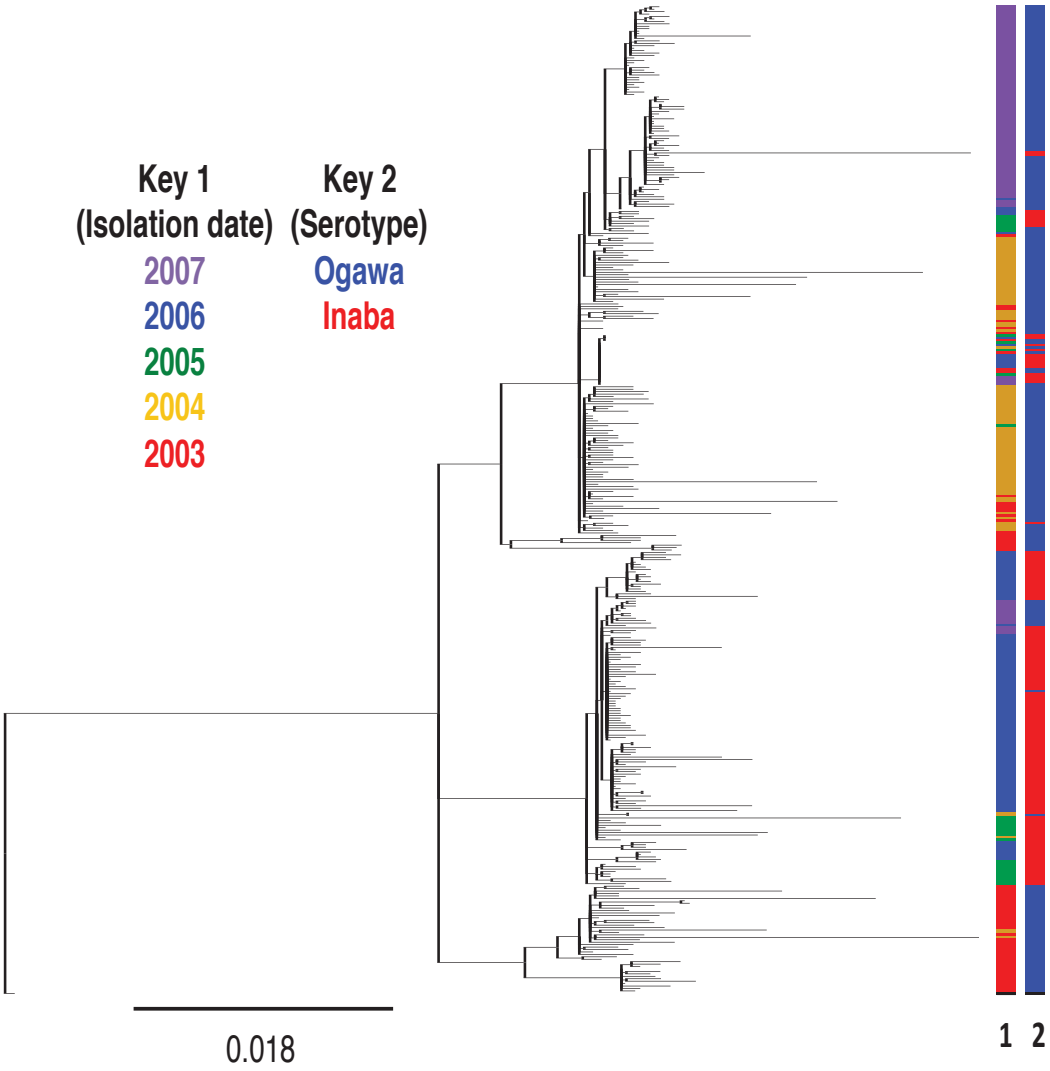


Figure 4.6: Maximum likelihood phylogeny of *V. cholerae* from the Kolkata vaccine trial with N16961 El Tor as both reference and root. Key 1 shows the year of isolation and Key 2 shows the serotype identified by antisera agglutination. The scale is given as substitution per variable site. The serotype switch from year to year is clear and there was visible temporal and serotypic correlation with the phylogeny.

4.3 Lessons learned and questions arising from this study

This study highlighted some aspects of the complexity of serotype variation in O1 *V. cholerae*. With the phenotypic agglutination test results and the genotypically predicted serotype matching for ~90% of the *V. cholerae* isolates, *wbeT* sequence appears to be a reasonably good marker for the genotypic classification of serotypes of *V. cholerae* O1. However, there were a few exceptions with a mismatch between

the phenotypically and genotypically determined serotype. *wbeT*, encoding the methyl transferase responsible for the methylation of the terminal sugar on O1-PS and consequently the Ogawa serotype of *V. cholerae* O1, harboured some mutations that were predicted to be null but did not apparently impact on serotype. There were also *V. cholerae* that harboured the wild type *wbeT* allele that reported as agglutinating with Inaba antisera. These data indicates that there may be mutations, for example polar mutations outside this gene, which influence the expression or the functionality of the final methyl transferase product.

Also, the data presented here prompts a number of potentially interesting questions including:

- Is a particular *wbeT* Inaba genotype prevalent in regional collections such as those from Kenya, Pakistan and Mexico ?
- Is a particular *wbeT* Inaba genotype found more frequently in the vaccine trial dataset ?
- Is there a particular mutated *wbeT* type which more fequently reverts back to the wild type Ogawa *wbeT*?
- Can we use similar data sets to predict if a selective pressure is operating in the field on the Inaba to Ogawa switch ?
- Is the change in serotype of *V. cholerae* population a switch driven by selection or is it simple strain replacement ?
- Is the mutation rate in *wbeT* gene the same as other genes of the wbe operon ?
- How different is the mutation rate in *wbeT* to the natural evolution rate of 3.3 SNPs/year in the seventh pandemic *V. cholerae* genomic backbone ?

Although these are all interesting questions there is insufficient data within the sample sets analysed within this study and further work must be planned to address these issues.