Investigation of *Vibrio cholerae* pathogenesis using a novel human intestinal model of infection

Leanne Kane

Thesis submitted in accordance with the requirements for the degree of

Doctor of Philosophy
of the
University of London

September 2022

Department of Infection Biology

Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE

Funded by The Wellcome Trust

LSHTM supervisor: Prof. Brendan Wren

WSI supervisor: Prof. Nicholas Thomson

Declaration

The work presented in this thesis is my own unless otherwise stated. It was carried out under the supervision of Prof. Nicholas Thomson of the Wellcome Sanger Institute (WSI), Hinxton, Cambridgeshire, UK and Prof. Brendan Wren of the London School of Hygiene and Tropical Medicine, London UK. The work was carried out at the Wellcome Sanger Institute.

All experiments and analysis were carried out by me with the exclusion of the following: Transmission electron microscopy and high pressure freezing of the bacterial isolates was carried out by David Goulding (WSI). The short read library creation and DNA sequencing was performed by the DNA pipelines group (WSI). Assembly and annotation of the short read genomes were carried out by the Pathogen Informatics group (WSI). RNA seq. pipeline was run by Oliver Lorenz of the Pathogens Informatics group (WSI).

This thesis does not exceed 100,000 words in length, including tables, figures and footnotes in accordance with the school regulations.

Abstract

Cholera is an acute diarrhoeal disease caused by infection with the Gram-negative bacterium *Vibrio cholerae*. Infection occurs upon ingestion of this bacterium via contaminated food or water, and if left untreated can be life-threatening. In terms of incidence cholera, remains a global health concern with the World Health Organisation estimating between 1.3 to 4 million cases of cholera worldwide each year, equating to between 21,000 and 143,000 deaths. The major virulence factor for pathogenic strains of *V. cholerae* is cholera toxin, the presence of which is believed to illicit the profuse diarrhoea synonymous with cholera. Recent genomic data has shown that different phylogenetic lineages of *V. cholerae* are linked to different burdens and patterns of disease ranging from geographically restricted lineages responsible for low level sporadic cases, we term endemic disease. To a single lineage, termed 7PET, that is linked to all the major epidemics in the ongoing pandemic 7, which began in the 1960's.

The use of animals has been invaluable in understanding key virulence factors associated with epidemic disease, and to date, much of the cholera research field has been based upon animal models for bacterial colonisation and disease. Despite the wealth of research using animal models, the only known natural host of *V. cholerae* are humans and no studies have framed differences in phylogenetic position to pathogenicity and outcome. The main objective of my PhD thesis is to add to the existing knowledge and understanding of *V. cholerae* infection dynamics by developing and validating a novel human model of infection. Human induced pluripotent stem cell (HuiPSC) derived intestinal organoids were developed to understand and differentiate disease potential between defined *V. cholerae* lineages occupying distinct positions within the species phylogeny. I confirm that all cell types present within the small intestine are represented in the organoids cultured, that the complex architecture comprising of a basal lateral surface and an inner lumen are recreated in this organoid system. During this investigation I observed two technical anomalies of the organoid system. Firstly, two morphologically distinct organoid variants are formed during the aging and passaging process that importantly respond differently to challenge with toxin. Secondly, cholera toxin administered externally to the basolateral surface or through

microinjection into the organoid lumen and so through the apical membrane elicited a similar response. During the validation of the model, I found key similarities in how both animal models and the human intestine respond to cholera toxin. For example, confocal staining and imaging shows degranulation of goblet cells, leading to release of mucus that has been shown to be a response that rabbit ilium illicit following exposure to cholera toxin. Furthermore, transcriptomic analysis reveals biomarkers linked to administration of cholera toxin which have previously been shown to be differentially regulated during natural infection in cholera patients.

To test the model in *V. cholerae* strains representative of the species, I selected three distinct pathogenic isolates including the archetypal reference 7PET strain N16961, which has been included in the majority of *V. cholerae* studies to date, a classical strain which pre-dates the seventh pandemic and a strain that harbours a type three secretion system - all strains situated on different lineages. All genome differences between these representative strains were catalogued and explored for phenotypic differences in the intestinal organoid model. Immunofluorescent staining and electron microscopy revealed differences in early colonisation patterns of the representative strains with the classical strain showing distinct patterns of association with fucose residues within the intestinal organoids.

Having established the model, future work would include microinjection of further pathogenic and non-pathogenic isolates and ultimately comparing both the host and pathogen response within the HuiPSC derived intestinal organoid.

Acknowledgements

Somewhere, something incredible is waiting to be known

- Carl Sagan

I would like to start by thanking Gordon Dougan for giving me the opportunity to undertake this part-time PhD and thank you to Simon Clare for allowing me the time required to read papers and set up the project in the beginning. Thank you to Nick Thomson and Brendan Wren for being my supervisors over the course of this thesis. It has been a long haul but thank you for sticking by me and tirelessly reading through my chapters, and for the direction and clarity when I needed it most. To Chris Hale, you have not only been an amazing lab companion but also a supportive friend over my years within the lab. I will miss working with you very much, but I wish you all the best with your well-earned retirement! Matthew Dorman, thank you for your patience with me while I attempted to learn bioinformatics during the Covid lockdown and for answering the multiple questions I have had along the way! To all if the members of team 15, and 216 and everybody who has given be helpful insights during my research journey, thank you.

Of course, an enormous amount of thanks goes to my family, Mum, Dad and my brother Anthony, you have always encouraged me and given me the strength on difficult days, your belief in me has been unwavering. Everything I am, I owe to you!

My husband Michael, you have only ever known me while I have been undertaking this thesis, thank you for your support and telling me, “Just get it done this weekend!” You have always believed in me and for that I will be forever thankful. You have supported me when I doubted myself. I could not imagine my life without you.

To my unborn child, thank you for giving me the inner strength and determination I did not know I had to see the final part of this journey through. I cannot wait to meet you and this thesis is dedicated to you. I hope that one day you will be proud of your mummy.

Contents

List of Figures

Chapter 1:

Introduction

1 Cholera

In developing countries, diarrhoeal disease is one of the most important classes of infectious disease responsible for an elevated level of mortality, particularly in children under five years of age [1]. Despite cholera being an infamous historical disease in the UK, associated with the Victorian era, it remains one of the leading causes of diarrhoeal disease globally with between 1.3 to 4.0 million cases and between 21,000 and 143,000 deaths worldwide every year [2]. Cholera is a severe and potentially lethal diarrhoeal disease that is a consequence of infection with the bacterium *Vibrio cholerae.* Cholera patients produce copious amounts of watery diarrhoea with the look and consistency of water produced when boiling rice, hence the general description of it being 'rice water stool'. Patients can lose up to a litre of fluid per hour and so can quickly become severely dehydrated if these fluids are not replaced. This, through the process of organ failure by hypovolemic shock, can lead to death [3].

The ground-breaking work of John Snow in 1854 was the first report linking cholera to drinking contaminated water, disputing the pervasive and previously held 'miasma' or 'bad air' theory for the spread of cholera and other infectious diseases. Studying a particular outbreak of cholera in 1854, John Snow showed that cholera deaths in London were linked to drawing water from a specific public water pump located in Broad Street, now called Broadwick Street. After correlating the incidence of disease to a map of the area, which can be seen in Figure 1.1 , he noticed that cholera cases clustered around the water pump and importantly discovered that some people who contracted cholera but did not live near or around the pump still drew their water from it, either because they preferred the taste or they passed by the pump on the way to school [4]. Snow had the handle of the pump removed so that it could no longer be used resulting in the significant decrease in the number of cases of cholera disease in this part of London. Snow himself reported that 'two or three days after the use of the water was discontinued, the number of fresh "attacks" became very few.' [4]. The work of John Snow is greatly recognised and appreciated by many as the foundations of modern-day epidemiology [5].

1a)

1b)

Figure 1.1 a. John Snow's map showing incidences of cholera in London in 1854. The red circles indicate locations of water pumps, while the black dashes illustrate a death at the corresponding address. Figure 1.1b shows a magnified view of the Broad Street pump in the middle of the map with numerous black dashes on the surrounding roads. Modified by myself but taken from a report by Dr. John Snow. Wellcome Collection.

Although many of the first reports of cholera are attributed to the seminal work of John Snow, Filippo Pacini was the first to identify the infectious agent responsible for cholera in 1854 [6]. He identified significant numbers of curved bacteria in the intestinal contents of cholera patients at autopsy, he called this curved bacteria *Vibrio cholerae* [7]. It was not until after his death that the wider scientific community recognised his work. Furthermore in 1883 Robert Koch demonstrated that the disease cholera was instigated by infection with a comma shaped bacteria [7]. Koch called the bacteria *Kommabazillen* later re-termed *Vibrio comma,* ultimately following the recognition of Pacini's earlier work the bacterium was renamed *Vibrio cholerae,* as it is known today.

1.1 Vibrio cholerae

Vibrio cholerae is a Gram-negative rod-shaped bacterium that possess a polar flagellum, which can be seen in the transmission electron micrograph shown in Figure 1.2. The bacterium primarily respires aerobically, but is capable of switching to anaerobic respiration in the absence of oxygen [8].

V. cholerae is also known as an environmental bacterium that is associated with zooplankton, such as copepods possessing chitin exoskeletons, found attached to aquatic plants and present in river or estuarine sediments [9]. A single copepod can bear up to 10^4 *V. cholerae* bacterial cells [10], consistent with studies of estuarine water samples showing *V. cholerae* to be five times higher in samples that contained copepods compared to water samples that had been filtered to remove them [11]. Studies in Bangladesh showed the use of sari cloth to filter water, removed most of the plankton, also resulted in lower incidences of cholera [12].

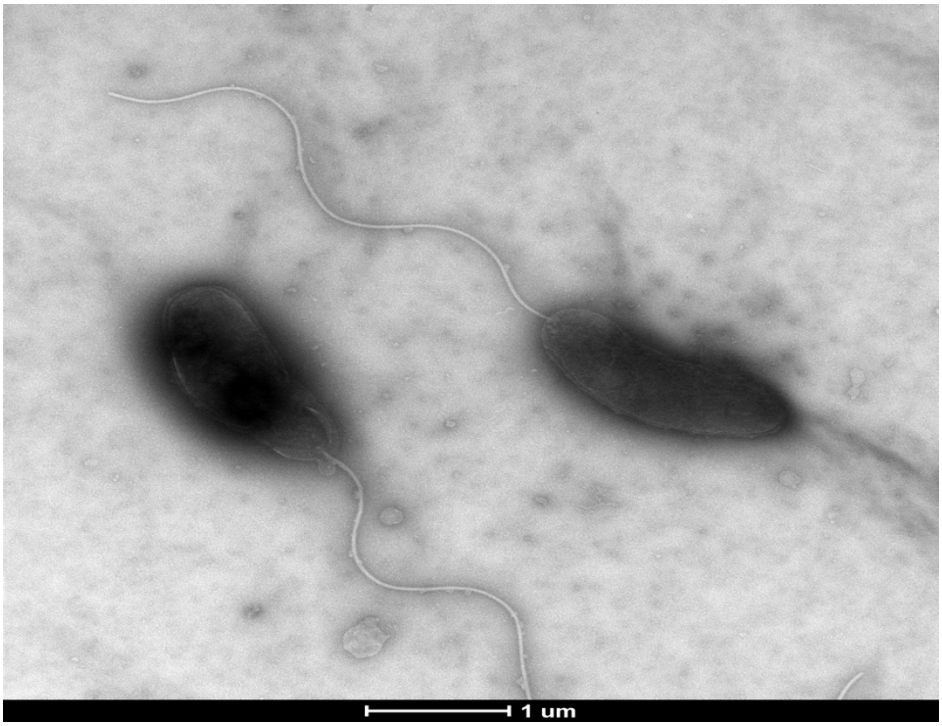


Figure 1.2. Transmission electron micrograph of the Vibrio cholerae *reference strain N16961 taken by Claire Cormie at Wellcome Sanger Institute Microscopy Facility for the purpose of this project.*

Over 200 serogroups of *V. cholerae* have been defined [13]. Serogrouping is based on the variability in the composition of the O side chain of the lipopolysaccharide (LPS) molecules [14]. However, only two of these serogroups, O1 and O139, are known to cause epidemics [15].

The O1 serogroup is further divided into two major serotypes, Ogawa and Inaba, which are associated with the O antigen of the LPS and are capable of unequal reciprocal interconversion [16]. Initial reports showed the switching of Ogawa serotypes to Inaba by growth in the presence of anti-Ogawa serum; the reverse of which failed to be demonstrated. The switching of Ogawa serotypes to Inaba serotypes involves the *wbeT* gene (formerly known as *rfbT* gene), which represents a gene that encodes a methyltransferase, responsible for the expression of the B determinant of the O-antigen that is required for the expression of Ogawa- specific serotype [16-18]. Both Ogawa and Inaba serotypes can cause cholera outbreaks with no obvious differences in pathology [19]. The epidemiological significance of the shift remains unknown; however, it has been suggested that the serotype conversion might be driven by the acquisition of serotype-specific immunity within the host population

[20]. The serogroup of *V. cholerae* is determined by the antigens present within the isolates, Inaba are known to express only the A and C antigens, Ogawa express the A and B antigens [21, 22]. There is a third serotype- Hikojima however it is rare and unstable [16]. Hikojima strains express all three antigens- A, B and C [21, 22]it has been suggested that Hikojima represents strains that undergo conversion at an elevated frequency [22].

To add an additional layer of complexity, *V. cholerae* can be further subdivided into two biotypes, *V. cholerae* biotype classical and *V. cholerae* biotype El Tor (described as classical and El Tor from now on). The biotype is based on the result of the following; haemolysis of sheep erythrocytes, agglutination on chicken erythrocytes, Voges-Proskauer reaction, sensitivity to polymyxin B and different sensitivity patterns to lytic bacteriophage [23],[24]. *V. cholerae* strains that are found within the environment most commonly lack the primary virulence determinant, the cholera enterotoxin or simply cholera toxin (CT) [25], and are considered to not cause epidemic cholera disease, but can cause gastroenteritis [26]. CT will be discussed in greater depth in section 1.6.1.

Although cholera is endemic in many countries, the global incidence of disease has fluctuated over time. Figure 1.3 shows global cases of cholera and mortality caused by cholera reported between 1989 and 2017. The peaks of incidences correspond with cholera epidemics, for example, in 2010 when an earthquake struck Haiti, causing catastrophic destruction of infrastructure and displacement of the population. Haiti had previously no reported incidences of cholera since the 1960s [27]. UN peacekeepers were sent to Haiti to aid the recovery mission in the aftermath of the earthquake. The UN has since admitted that peacekeepers from Nepal carried the disease with them to Haiti [28], leading to thousands of cholera cases and deaths. This incident has highlighted how fast cholera can spread in countries or regions either lacking or with damaged infrastructure and poor sanitation. Sadly, Haiti currently still reports incidences of cholera annually. Another more recent peak can be seen in 2017 during the Yemen outbreak, which has been the worst epidemic witnessed since records began. It is important to recognise that although these numbers are helpful in showing disease trends, they are most likely an underestimation [2], as many people, particularly in endemic regions, are infected and die without treatment from healthcare professionals, or there is no active monitoring system in place to record cases of disease.

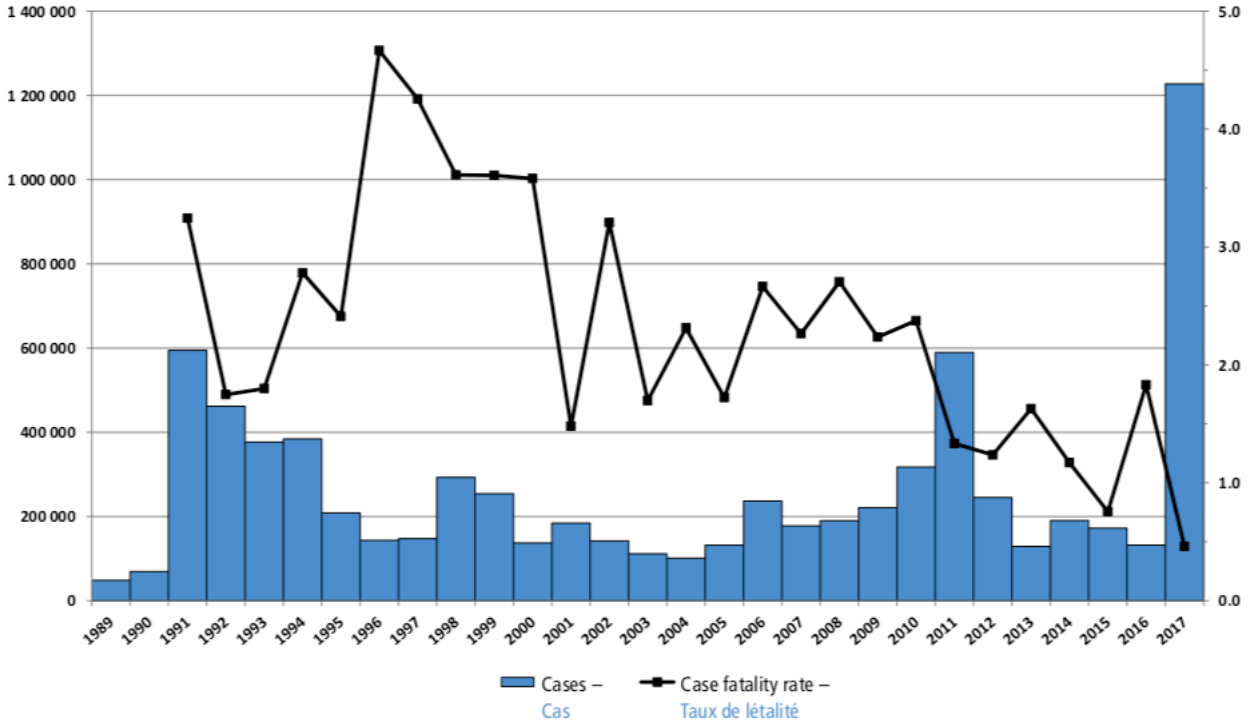


Figure 1.3. Illustrating annual cholera cases and mortality reported by year between 1989 and 2017. Blue bars indicate cholera cases, in which peaks can be seen during known outbreaks e.g., Haiti in 2010. The black line corresponds to the case fatality rate. Credit to WHO weekly epidemiological record 2018.

1.2 Cholera reporting

The CDC (Centers for Disease Control and Prevention) define cholera as 'An illness characterized by diarrhea and/or vomiting; severity is variable', in which the laboratory criteria for diagnosis are isolation of toxigenic (i.e., cholera toxin-producing) *Vibrio cholerae* O1 or O139 from stool or vomitus, serologic evidence of recent infection. Critically 'Illnesses caused by strains of *V. cholerae* other than toxigenic *V. cholerae* O1 or O139 should not be reported as cases of cholera. The etiologic agent of a case of cholera should be reported as either *V. cholerae* O1 or *V. cholerae* O139. Only confirmed cases should be reported to National Notifiable Diseases Surveillance System (NNDSS) by state health departments' [29].

The WHO (World Health Organization) case definition of cholera is 'a case of cholera should be suspected when: in an area where the disease is not known to be present, a patient aged 5 years or more develops severe dehydration or dies from acute watery diarrhoea; in an area where there is a cholera epidemic, a patient aged 5 years or more develops acute watery diarrhoea, with or without vomiting. A case of cholera is confirmed when *Vibrio cholerae* O1 or O139 is isolated from any patient with diarrhoea. In children under 5 years of age, several pathogens can produce symptoms like those of cholera, such as rice-water diarrhoea. To maintain specificity, therefore, children under 5 are not included in the case definition of cholera'.

Hence, both CDC and WHO case definitions rely heavily on the detection of O1 or O139 serogroup *V. cholerae*. These definitions are used as surveillance systems as cholera is a notifiable disease, both in the U.S. and in the field. The diagnosis framework is a tool used by healthcare workers to uniform the diagnosis and treatment of cholera. It is also used for outbreak reporting, in order to record the severity and spread of outbreaks.

1.2.1 The history of cholera pandemics

To date, there have been seven recorded cholera pandemics, the first of which occurred from 1817 to 1823. Figure 1.4 shows a schematic of the pandemic timelines dating back to 1817 [30]. The second pandemic spanned 1829 to 1851 coinciding with John Snow's research on the London epidemic. Remarkably, Devault et al. were able to sequence *V. cholerae* from a preserved section of intestine that was removed from a patient that died from cholera in 1849, the group concluded from the genomic analysis that the cholera sequenced was 95-97 % similar to the classical strain O395 [31]. The third pandemic took place between 1852 and 1859, followed by the fourth from 1863 to 1879, fifth from 1881 to 1896 and the sixth from 1899 to 1923. From 1926 to 1960, many people believed that another cholera pandemic was unlikely to occur as water supplies had been greatly improved [32]. Many parts of the world did become free of cholera, however the current seventh pandemic began in 1961. It has been postulated that the first six cholera pandemics were caused by the serogroup O1 cholera toxin positive, classical biotype. El Tor strains circulated and were associated with sporadic disease during the sixth pandemic [7], however it was not until 1961 that the El Tor biotype caused the seventh pandemic and became recognised to have epidemic potential [7]. Classical and El Tor biotypes not only differ genetically and phenotypically, but also in their infection patterns within humans [8]. El Tor strains in the past have been associated with more frequent asymptomatic infections and fewer fatalities, better survival in the human host and in the environment and more efficient host-to-host transmission compared to their classical counterparts [33].

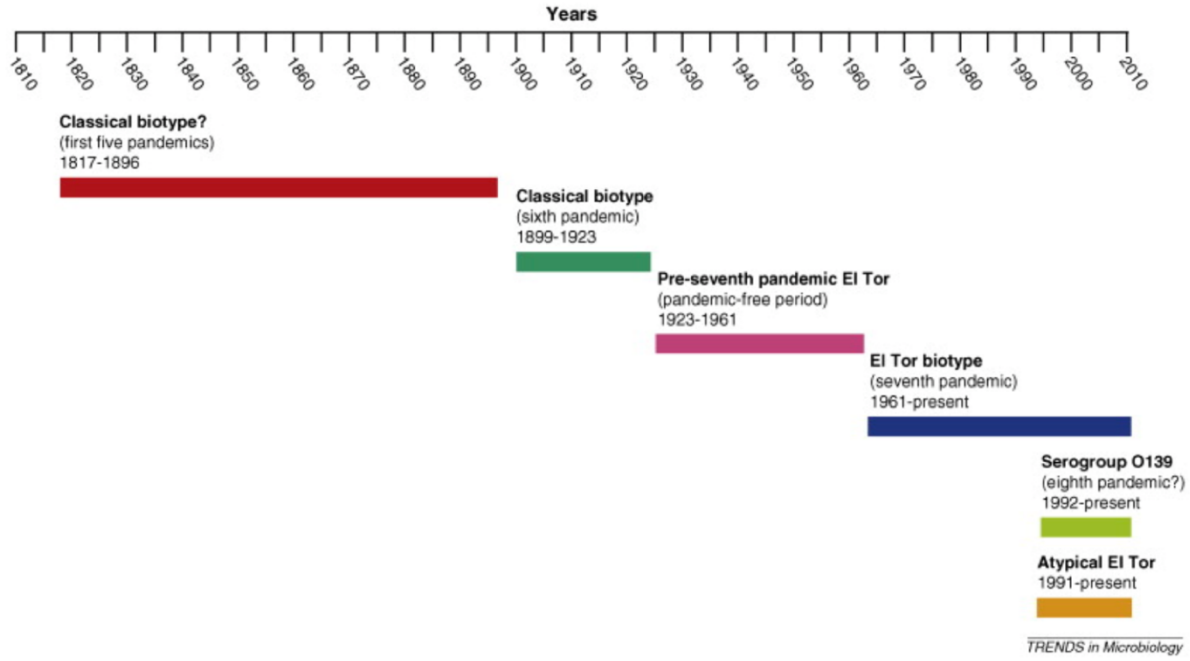


Figure 1.4. A schematic illustrating the timeline of cholera pandemics, from 1817 to 2010, and the biotype that caused the corresponding pandemic taken from Safa et al. [30].

1.2.2 Modern cholera

We are currently experiencing the 7th pandemic of cholera. The beginning of pandemic 7 is officially recorded as 1961 where it began in Sulawesi in Indonesia [34]. A timeline of cholera pandemics is illustrated in Figure 1.4, the epidemic subsequently spread globally, in particular to South Asia in 1963, entering Africa in the 1970's, Latin America in 1991 and the Caribbean (i.e., Haiti) in 2010 [35]. Between 2008 and 2012 the global burden of cholera was estimated at 1.4 million to 4 million cases annually, with 21,000 to 143,000 deaths [2]. The only continent to have not fallen victim to any of the cholera pandemics is Antarctica [36], with the continent most affected being Africa [37]. According to the World Health Organization, 45,159 cases and 3,488 deaths in ten African nations were reported up to 23rd July 1991 [32]. The bacterium responsible for the current 7th pandemic is *V. cholerae* O1 serogroup, cholera toxin positive, El Tor biotype. Whilst the 7th pandemic officially began in 1961, El Tor biotypes replaced classical biotypes in the current seventh pandemic between 1817 and 1923 [7], [38] [39]. The El Tor biotype is named after the Egyptian city where the first El Tor biotype isolate was reported in 1905, cultured from six pilgrims who were quarantined there after returning from the Hajj in Mecca by Gotschlich [40]. The reason that El Tor biotype have replaced the classical biotype is largely a mystery. It is widely accepted that strains belonging to the

classical biotype produce a larger amount of toxin, potentially explaining the more severe disease caused by classical biotypes as opposed to El Tor biotypes. One other possible explanation is that El Tor survives longer in the environment than classical [40]. A third explanation as to why the El Tor biotype has replaced the classical biotype in the context of pandemic cholera is offered by Yoon and Mekalanos [41], who report carbohydrate metabolism differences between the two biotypes. In the study [41], growth of the classical biotype *V. cholerae* strain O359 with the addition of glucose leads to production of organic acids, with a drop in pH of the growth media and ultimately growth inhibition. Whereas the El Tor *V. cholerae* strain N16961 produces 2,3-butanediol which allows the bacteria to grow to much higher densities as opposed to accumulating the organic acids which posed detrimental to their classical counterparts. 2,3-butanediol fermentation is a highly conserved fermentation pathway present in many bacterial species [42]. Yoon and Mekalanos hypothesise that this difference in carbohydrate metabolism likely reflects the ability of the bacterial strains to grow on chitin affecting their growth in the environmental niche allowing for the replacement of the classical biotype with El Tor [41].

In India, El Tor has been isolated from numerous natural water sources, usually in the absence of disease in the neighbouring community, however classical strains have never been able to be isolated from such water sources in the absence of cases of disease [43, 44].

The duration of the El Tor carrier state in asymptomatic contacts is approximately one week [40]. However, cases of long-term carriage have been reported, 'Cholera Dolores' was infected with El Tor cholera in 1962 and continued to intermittently secrete vibrios of the 'same characteristics as the original isolate' up until 1966, a total of nine years after the first infection [45]. Duodenal intubation proved that the bacteria were lodged in her biliary tract, most likely in her gall bladder [40].

Globally, the 7th pandemic *V. cholerae* El Tor (7PET) isolates are genetically homogenous and associated to a single source, the Bay of Bengal in South Asia [46]. The genomics of *V. cholerae* will be expanded on later in this chapter (See section 1.5 A Genomic view of Cholera). In the late 1980s it was thought that the classical biotype had disappeared globally [47], however, in a recent study in Thailand, classical strains were isolated from samples collected as recently as 2000 [48], indicating that the classical biotype persisted longer than previously thought and may very well have been circulating alongside El Tor for some time.

In 1992 it was thought that the 8th Pandemic of cholera had begun in Bangladesh and India where isolates with the serogroup O139 began to cause epidemics [49]. From November 1992 to July 1993, 95.6% of *V. cholerae* isolates submitted to the National Institute of Cholera and Enteric Diseases from 28 locations in India were of serogroup O139 [50]. However, isolates of this serogroup failed to spread globally, appearing to have been restricted to Bangladesh and India. They have subsequently largely disappeared, although a small number of cases continue to be reported annually [51]. The O139 serogroup is still included in most cholera diagnostic tests and although there are sporadic reports of severe disease they tend to be isolated cases and are not associated with epidemics currently [51].

The O139 serogroup unlike O1 strains possess a capsule, with the capsular layer being distinct from the LPS antigen [52]. The O139 serovar is identified by; the absence of agglutination in O group 1 specific antiserum, the agglutination in O group 139 specific antiserum and by the presence of a capsule [53]. The O139 serogroup is believed to have evolved from the O1 El Tor biotype with major changes in the O-antigen biosynthetic genes [54-58].

The Yemen outbreak of cholera had the first reported cases in September 2016 is regarded as the largest outbreak of the disease in recent history, being responsible for 1.1 million cases and 2,300 deaths [35]. Interestingly, the disease-causing bacteria strain had an unusual antimicrobial resistance profile unlocked by genomic analysis of Yemeni bacterial isolates (resistance to nalidixic acid and susceptibility to ciprofloxacin) It was in fact the breakdown in infrastructure rather than bacterial fitness that was responsible for the outbreak [59].

1.3 Treatment

The most effective treatment for cholera remains the apparently archaic, but highly effective 'cholera cot' or 'Bangkok bed' which can be seen in Figure 1.5, combined with oral rehydration therapy. The cholera cot is important because it allows the 'rice water stool' diarrhoea to be collected in a container below an opening in the bed. The advantage of using these collecting vessels is twofold, on one hand it allows healthcare workers to measure the volume and rate of fluid lost from patients and to be able to administer the correct volume of rehydration, it is also very effective at infection control, containing the contaminated stool and limiting re-infection to others.

Figure 1.5. Typical 'cholera cot' used in the treatment of cholera showing the collecting vessel beneath the bed to monitor fluid loss and contain contaminated stool to aid infection control. https://causticsodapodcast.com/2012/01/16/cholera/

Rehydration therapy involves the administration of oral rehydration salts for less severely dehydrated patients. More severe dehydration is treated by intravenous Ringer's lactate solution- the contents of which include sodium, chloride, potassium, calcium and sodium lactate [60]. The administration of rehydration therapies, if delivered in a timely fashion and with suitable volumes, will reduce fatalities to under 1% of all patients [61].

Antibiotics are used in conjunction with rehydration therapies for more severely dehydrated individuals, the most used being, doxycycline in adults and azithromycin in children and pregnant women. Antibiotics limit the extent of the episode, reducing the volume of stool by up to 50% and decreasing the length of shedding of *V. cholerae* to just one to two days [62]. Mass antibiotic treatment is not recommended as front-line treatment for cholera and should not be used as a preventative measure. In addition, there is insufficient evidence to suggest that antibiotic treatment can decrease secondary transmission of cholera within households [63], [64]. Resistance to antibiotics has been recorded in both endemic and epidemic settings [65, 66], so the use of antibiotics and the emergence of resistant strains is very carefully monitored. In fact, the increase of antibiotic resistance can be seen if we examine the genomic data: During the recent cholera outbreak in Zimbabwe in 2018/2019 only azithromycin could be used to treat cholera disease, as the strain in circulation was resistant to all other front-line antibiotics. This is one example of how frightening the use of antibiotics can be, ultimately limiting the effective treatments of the disease in which rehydration therapies would have been sufficient.

There are currently three WHO pre-qualified oral cholera vaccines (OCV) available: Dukoral® Shanchol™, and Euvichol-Plus®. All three of these vaccines include *V. cholerae* strains that have either been heat or formalin inactivated, they require two doses for full protection, and require the addition of clean drinking water for administration, which may be difficult to find in an epidemic setting.

In addition to these inactivated vaccines, VAXCHORA™, which is licenced in the U.S.A. is a live attenuated vaccine, consisting of live *V. cholerae* that has been engineered to have an inactivate copies of the *ctxA* gene, while still possessing the non-toxigenic, immunogenic *ctxB* gene. VAXCHORA™ has a marker inserted into the hemolysin gene (*hlyA*) to enable the vaccine strain to be differentiated from any wild type *V. cholerae* O1 circulating strain. (https://www.fda.gov/media/98688/download).

More than 30 million doses of oral cholera vaccines have been distributed and used in endemic regions. The effectiveness of timely treatment can be seen in Figure 1.3, whereby in 2017 during the Yemen outbreak, although the number of cholera cases were high, case

fatality rate remained remarkably low with thanks to prompt distribution of OCV in addition to rehydration therapies.

On October 2017, The Global Task Force on Cholera Control (GTFCC) endorsed a call for commitment from all stakeholders to support cholera-affected countries and align energies, efforts and resources to end cholera transmission. This taskforce has a target to reduce cholera deaths by 90% and to eliminate cholera in as many as 20 countries by 2030. The GTFCC strategy can be seen in Figure 1.6

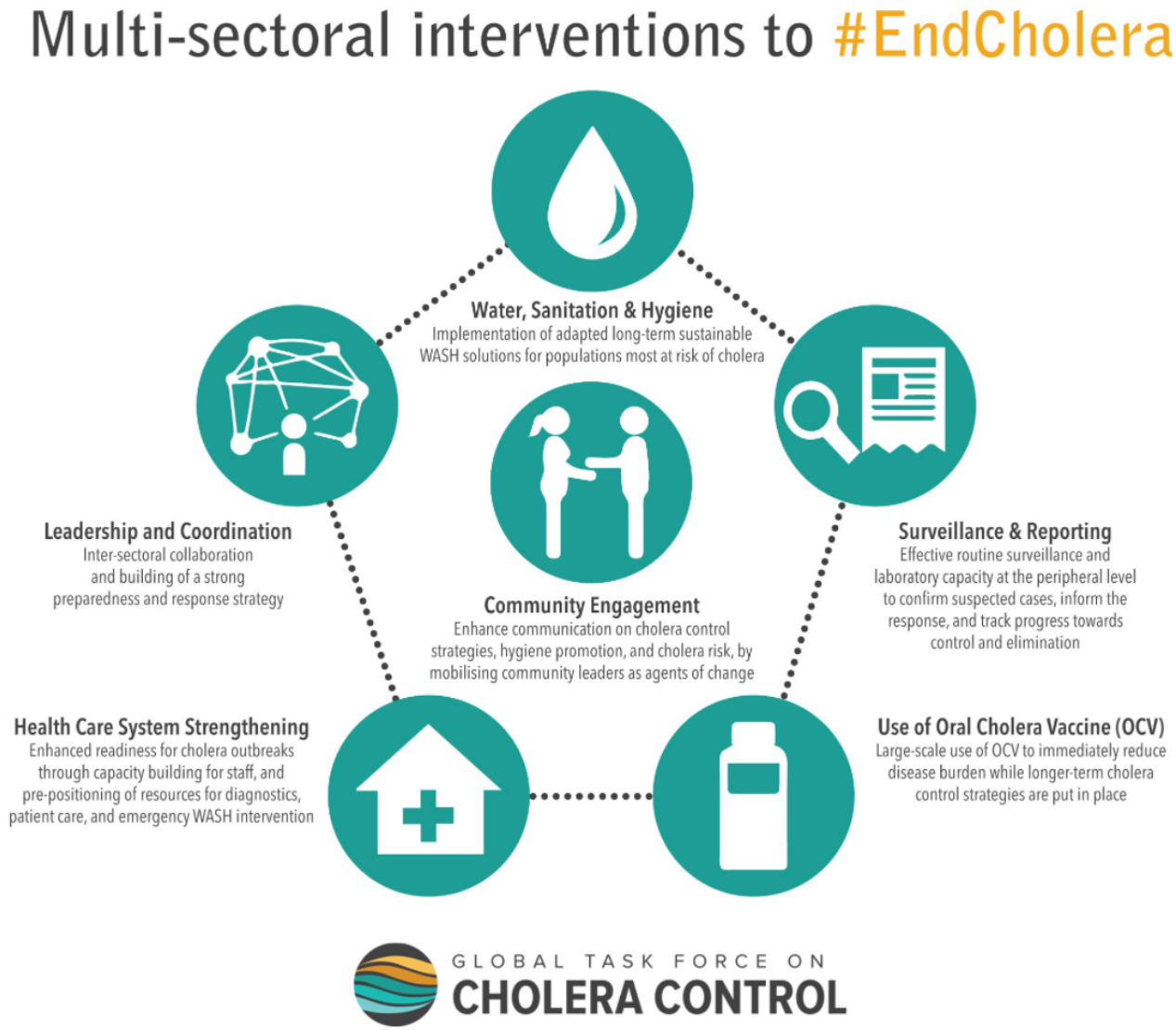


Figure 1.6. The GTFCC strategy to end cholera which includes implementing WASH solutions for populations that are most at risk of cholera, in addition to the large-scale use of OCVs to reduce disease burden as well as community engagement to highlight the importance of hygiene, and effective disease surveillance and reporting to monitor the progress of the eradication mission.

1.4 The Cholera Paradigm Theory

There have been several theories to explain the patterns of global cholera disease. Perhaps the most accepted and strongly perused is the cholera Paradigm Theory developed in 1996, which suggests that cholera epidemics have been directly driven by climate-induced variations on coastal aquatic reservoirs of *V. cholerae* [67], such as, sea surface temperature and height, as well as plankton blooms. Research focusing on cholera outbreaks in coastal Africa showed that coastal epidemics constitute only a minor part of the continental cholera burden [67]. Most of the coastal cholera foci are located near estuaries and lagoons, however outbreaks often originate in coastal cities, where cholera is more likely to be imported from distant areas [67]. This study also showed that frequent seasonality of cholera incidence appeared to be driven by the rainfall-induced contamination of unprotected water sources through latrine overflow, as well as the periodicity of human activities such as fishing or travelling, not by climate variation. In addition to this, Colwell theorised that cholera epidemics were caused by El Niño events subsequently leading to plankton blooms [32]. The warming of the surface waters by 1 °C as a result of an El Niño in the tropical Pacific from 1990 to June 1995 coincidentally correlated with a cholera outbreak in Peru. This was the longest El Niño event since the 1870s, lasting for more than three years [32]. Warmer ocean surface temperature has been held responsible for other cholera outbreaks, in the Bay of Bengal, rises in ocean surface temperatures measured by satellite monitoring have been linked to rises in cholera incidences (unpublished data- Beck, Wood. Huq and Colwell).

However, El Niño events have been occurring with some regularity for many years and have not been previously associated with cholera outbreaks. Moreover, in 1991, following the U.S. Environmental Protection Agency suggesting there might be a slight cancer risk due to the chlorination of water, Peruvian authorities took the decision not to chlorinate the country's drinking water possibly contributing to the cholera epidemic, as it is well known that chlorination of water is not only effective at cholera prevention, but also other waterborne diseases. Therefore, there must be another explanation for these explosive epidemics rather than the mild warming of the ocean surface waters.

1.5 A Genomic view of Cholera

More recently, as the cost of whole genome sequencing has reduced, we have had our first insights into the whole genome relationships between bacterial isolates causing the current 7th Pandemic. Utilising whole genome sequence (WGS) data and single nucleotide polymorphisms (SNPs), Mutreja et al. were able to construct a high-resolution phylogeny for the species and link these specific lineages of *V*. *cholerae* with their geographic location: phylogeography [46] Figure 1.7 shows the phylogeny in addition to a geographical map illuding to the spread of cholera in three distinct waves. Using isolates collected between 1957 and 2010 they showed that the Bay of Bengal was the modern-day global source, equivalent to John Snows 'Broad Street pump', (Figure 1.7b), whereby all 154 strains analysed from the seventh pandemic originated from a common ancestor in the 1950s within this region of South Asia. From Figure 1.7a, it is clear that the *V. cholerae* phylogeny consists of a single strongly defined branch (referred to as the backbone of the tree) which tracks the evolution of the source population, continually evolving over time (coloured black on the tree). The multiple minor branches are shown in a variety of colours corresponding to the country of isolation for the isolate. This shows that lineages of *V. cholerae* evolve away from the source population and then move into different geographic regions where they can cause outbreaks. It can be clearly seen that when a strain is isolated within a certain country or region, that lineage becomes isolated and eventually goes extinct. This is seen by the fact that lineages causing disease more recently branch from the backbone source population higher up the backbone replacing those from earlier expansions in those regions. The earlier lineages then are not seen to cause disease again in that region, or generally anywhere else in the world. In addition to this, looking at all the expansions of lineages into different regions, there was evidence of three major global expansions or radiations of El Tor, denoted Wave 1, 2 and 3. These waves spread in globally successive waves with the latest wave of isolates eventually replacing isolates from the previous wave in those regions. Currently cholera causing isolates are almost exclusively linked to Wave 3. This changed the way we think about *V. cholerae* evolution and showed that the cholera paradigm theory did not explain the patterns of

disease reported globally and that disease spread was in fact caused by population movement rather than weather events or linked to local marine or brackish water environments.

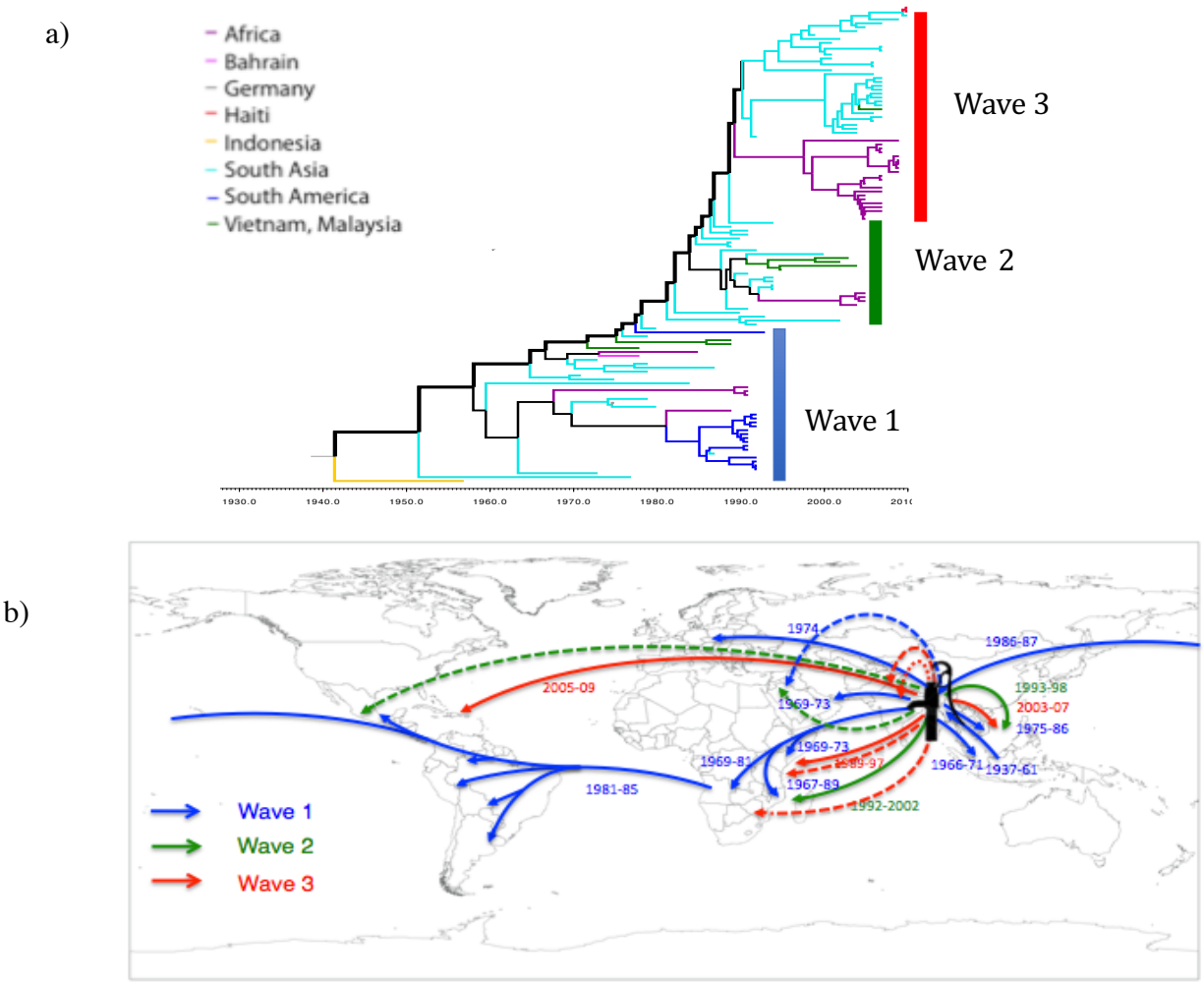


Figure 1.7. a) Mutreja et al. showing phylogenetic differences between wave 1, 2 and 3 V. cholerae *isolates. The colours on the phylogenetic tree corresponding to country or region of isolation, it can be seen that the outbreaks occur, expand, and become extinct within that region until the re-introduction of a new isolate that is an ancestor of the previous epidemic strain. (b) The genomic evolution can be geographically mapped to show worldwide spread of disease. doi:10.1038/nature10392.*

Using the resolution of WGS, it is not only possible to track isolate spread on a global scale, but it can also take place nationally, for example within strains collected during severe flooding in Pakistan in 2010 [68]. During these floods, Shah et al. collected and sequenced 38 *V. cholerae* O1 El Tor isolates from affected areas. All isolates were classified into two subclades (PSC-1 and PSC-2) showing clear differences from *V. cholerae* that had been

isolated elsewhere in the region [68]. The group established that PSC-1 isolates geographically originated in the south of Pakistan whereas PSC-2 isolates came from inland areas in the north that were flooded by the Indus River [68]. These findings demonstrated that the outbreak in Pakistan was in fact an epidemic within an epidemic and that there had been two separate introductions of *V. cholerae* into Pakistan. Genetic differences between PSC-1 and PSC-2 were investigated and it transpired that PSC-1 isolates are resistant to tetracycline while PSC-2 strains are susceptible. All Pakistan flood strains fall within Wave 3 of pandemic seven within the global collection.

More recent studies in Africa and Latin America have reinforced this view of cholera spread and evolution. In Africa Weill et al. demonstrated that each introduction of 7PET resulted in mid- to long-term spread of cholera [35] and one export from Africa to Peru was responsible for the Latin American epidemic in the 1990's [69]. Prior to this outbreak, Latin America had not witnessed a cholera outbreak for almost 100 years. Once in Peru, the outbreak spread rapidly to nearly every country in Latin America, causing 1.2 million disease cases and 12,000 deaths by 1997 [70].

In Latin America, Domman et al. showed vast diversity in Latin American strains compared with 665 global isolates [69]. Their phylogeny has revealed that the 1991 Latin American cholera epidemic was the result of two independent intercontinental introductions of 7PET *V. cholerae* [69]. See Figure 1.8.

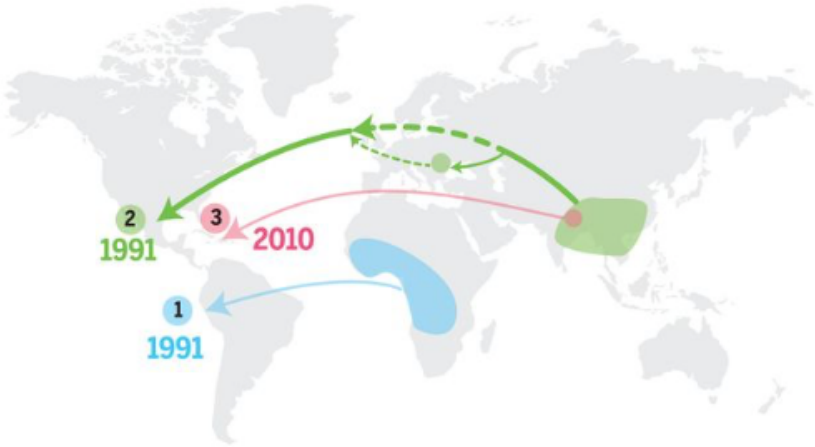


Figure 1.8. Map showing the two independent introductions of V. cholerae to Latin America taken from Domman et al. *2017 [69]. The blue lines represent the first introduction in 1991, the green signify the second in 1991 and red the third introduction of V. cholerae into Latin America in 2010.*

1.6 Virulence factors

V. cholerae is transmitted by the faecal-oral route of transmission. It is generally accepted that infection occurs through the consumption of contaminated food or water. In cholera endemic countries, poor sanitation continues to be the main reason for the high incidence rates.

From studies in the 1960s using human volunteers, it is know that the infectious dose of *V. cholerae* in an un-buffered human is approximately 10^{11} colony forming units (CFU), however this infectious dose can be reduced to 10^{6} upon administration of a sodium bicarbonate buffer to neutralise gastric acid [71]. Once ingested, *V. cholerae* passes through the stomach to the small intestine, the main site of infection. Prior to attaching to the intestinal epithelial cells, *V. cholerae* must overcome a range of host defence mechanisms such as many antimicrobial peptides including, but not limited to, bile in the stomach followed by intestinal mucus in the form of mucins secreted by goblet cells of the epithelium.

It is believed that there were two critical evolutionary steps marking the development of epidemic *V. cholerae*. First, the ancestral *V. cholerae* strain acquired the TCP (toxin co-regulated pilus) pathogenicity island by lateral gene transfer. Secondly, a TCP positive strain was infected by the bacteriophage CTXϕ carrying the cholera toxin (see section 1.6.1 below) [72]. CTXϕ is a bacterial virus (bacteriophage) with a 6.9 kb circular genome. This bacteriophage is capable of its own transmission and carries genes which are essential for regulation, replication and integration into the host chromosome, this process of stable integration is known as lysogeny [73]. The *ctxA* and *ctxB* genes encode the alpha and beta subunits of CT [74] and are shown with red arrows in Figure 1.9. They are located within, are linked to and transferred with the CTXϕ. In addition to *ctxA* and *ctxB* the CTXϕ is composed of 8 genes that have been shown to be involved in phage reproduction, packaging, secretion, integration and regulation.

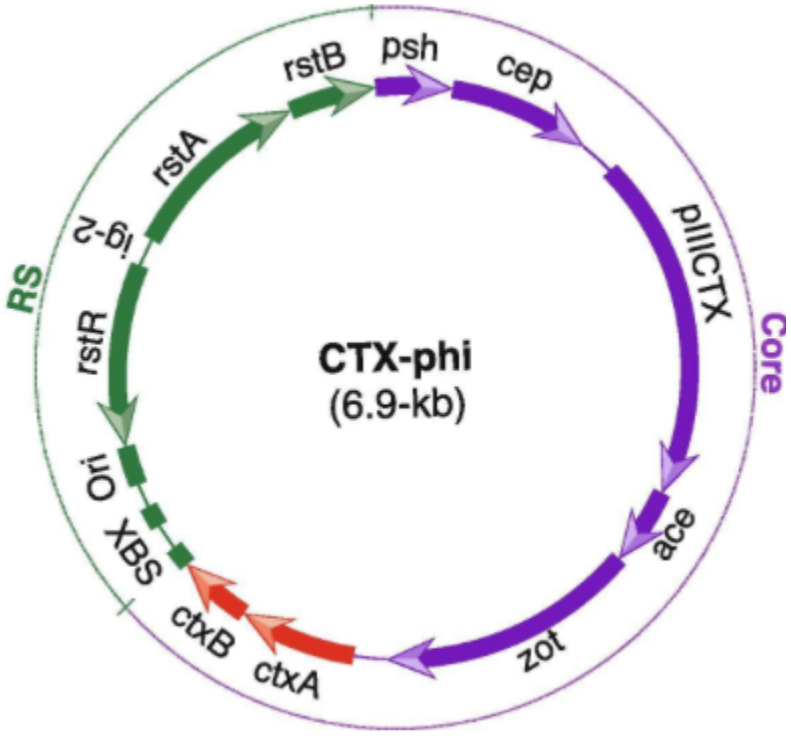


Figure 1.9. A schematic representation - not to scale, of the CTXφ genome. Genes important for replication, integration and regulation of the transcription are shown with green arrows, the ctxAB *genes encoding CT are indicated by red arrows. Boxes represent DNA sequences essential for phage replication and chromosomal integration. Taken from [75].*

TCP encoded by the *tcpA,* genes is a type IV pilus that has been shown to be essential for intestinal epithelial attachment and colonisation by *V. cholera*e [76]. TCP is encoded within a pathogenicity island [77] called *Vibrio* Pathogenicity Island 1 (VPI-1) which encodes 29 potential proteins and is 41.2 kb in size in both classical and El Tor biotypes [78, 79]. Of note TCP also acts as the external cell surface receptor for the CTXϕ, whereby isolates lacking TCP are not susceptible to CTXϕ infection and therefore lysogeny. TCP has also been shown to tether bacterial cells together via a filament or matrices that can protect the bacterium from antimicrobial compounds such as bile within the host [80]. TCP is an essential *V. cholerae* intestinal colonisation factor [76, 81]. TCP production is induced within the intestine, whereas production in other environments appears to be minimal [82]. It is primarily strains of O1 and O139 serotypes that encode the *tcp* genes [83, 84].

1.6.1 Cholera Toxin- The primary virulence determinant

In 1884 Robert Koch proposed that the agent responsible for the disease cholera was a 'special poison' [85], which acts upon the epithelium and suggested that the cells were being poisoned. We now know today that this 'special poison' is CT.

CT is an A-B_5 type exotoxin that is responsible for the profuse watery diarrhoea characteristic of cholera [72]. CT is composed of two subunits, A and B, encoded by the genes *ctxA* and *ctxB,* respectively [86]. The structure of CT can be seen in Figure 1.10. The B subunit forms a ring shaped pentamer that actively binds to the GM_1 ganglioside receptor on the surface of the host small intestine [87], once bound to the GM_1 receptor, the whole toxin is endocytosed by the host epithelial cells and is trafficked to the Endoplasmic Reticulum (ER) via a retrograde pathway [88]. Within the ER, the A and B subunits of the toxin dissociate [88] . The A subunit of the toxin is comprised of two peptides, A_1 and A_2, the disulphide bond linking these two peptides (highlighted by the black arrow in Figure 1.10) is reduced, the enzymatic A_1 peptide activates adenylate cyclase, this activation in turn increases the concentration of cAMP within the cell ultimately leading to protein kinase A phosphorylating the cystic fibrosis transmembrane regulator (CFTR) resulting in an increased flow of Cl^- ions into the lumen of

the small intestine [89]. The activation of adenylate cyclase by the A_1 peptide of CT takes between 15-60 minutes after initial uptake of the toxin, this is enough time for the A_1 peptide to translocate to the cell membrane and to be delivered to its substrate [90]. The outline of this process is shown in a schematic diagram in Figure 1.11.

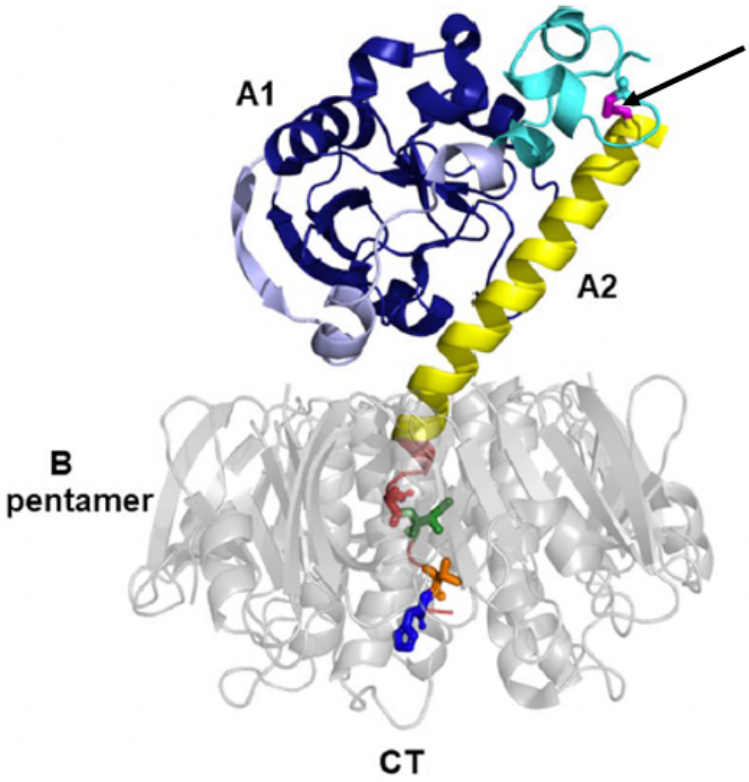


Figure 1.10. A schematic showing the structure of CT. The pentameric B-subunits form a halo in which the A2 subunit anchors itself, in turn connecting the B-subunits to the enzymatic A1 peptide. The disulphide bond linking the A1 and A2 peptides is shown in magenta and highlighted with the black arrow (adapted from [91]).

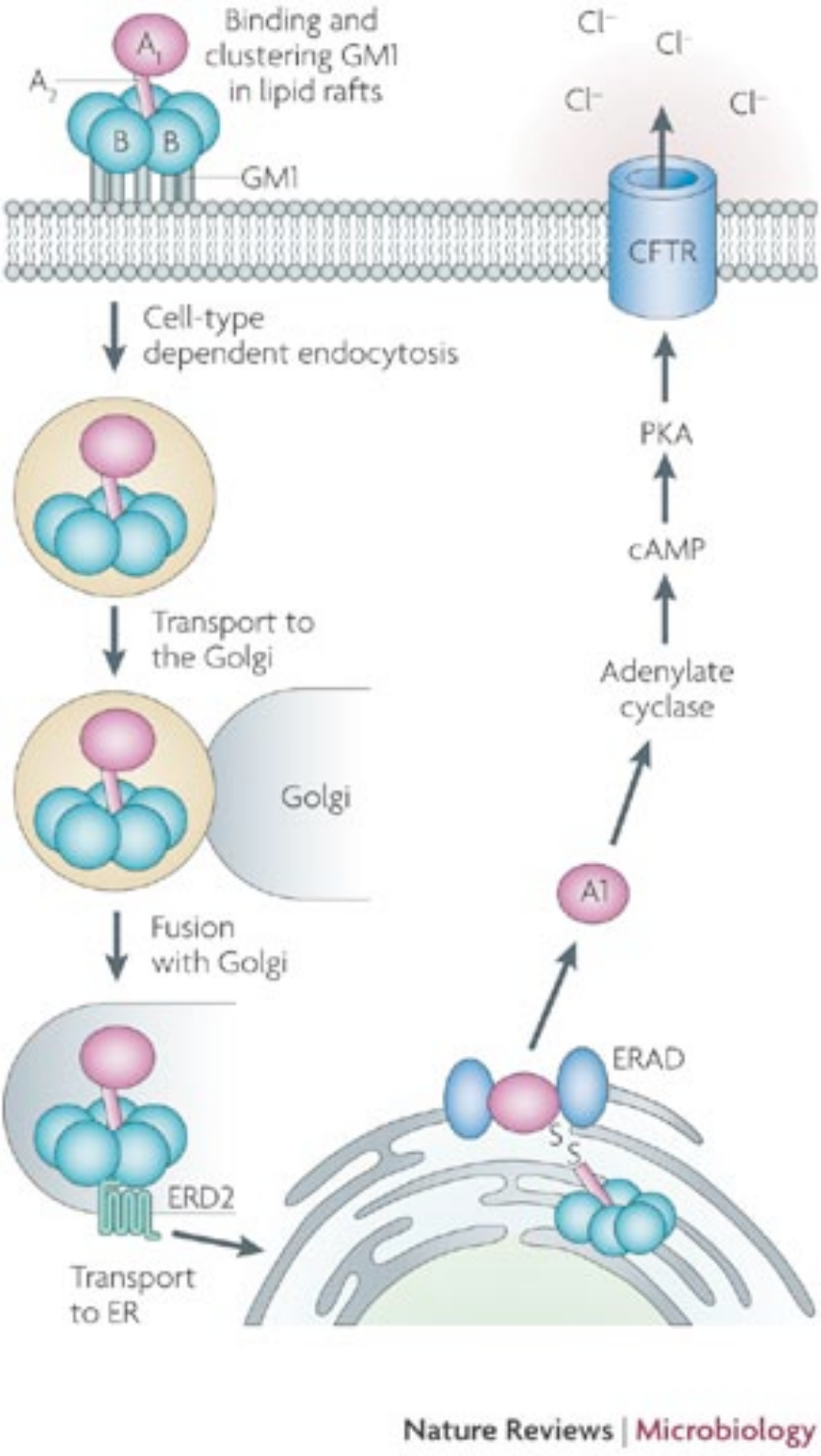


Figure 1.11. Schematic diagram to illustrate the effect of cholera toxin on host intestinal epithelial cells [89].

1.6.2 Cholera toxin receptor- GM1

As described in 1.6.1, the most well accepted and well characterised receptor for CT binding is the GM1 ganglioside which is present on the human intestinal epithelium. Holmgren et al. in the early 1970s discovered that GM1 had a strong binding affinity for CT. Subsequently, research groups in the 1980s demonstrated that cells lacking GM1 do not bind CT [92, 93].The strong binding affinity of CT to GM1 can be explained if we examine at the structure of GM1.

In 1994 Merritt et al. [94] published the crystal structure of cholera toxin B- pentamer bound to GM1 pentasaccharide. Although having a high binding affinity for one another, the Holmgren research group also showed that human small intestine had a surprisingly low concentration of GM1 ganglioside when compared to beef and porcine small intestine (0.1 nmol/g in human, 2.0 nmol/g in pig and 43 nmol/g in beef) [87]. This low concentration of receptor is surprising as *Vibrio cholerae*, is only capable of infecting and causing disease to humans, yet the receptor for its toxin is more abundantly expressed in other animal species. The question remains as to whether this low concentration of GM1 is sufficient to cause this profuse watery diarrhoeal disease. The presence of the GM1 ganglioside within the iHO model described in this dissertation will be explored in Chapter 3 of this thesis. Although GM1 is the historical and most studied of the cholera toxin receptors, there is a mounting number of publications and growing evidence that there may be alternative receptors that can bind cholera toxin- this will be discussed in Chapter 3 of this thesis.

1.7 Alternative virulence factors present within *V. cholerae* isolates

1.7.1 Type three secretion system (T3SS)

I have previously discussed the two major virulence factors (CT and TCP) required by *V. cholerae* to cause pandemic cholera, however non-pandemic lineages of *V. cholerae* are equipped with alternative virulence determinants to cause disease. One of these alternative virulence factors is a type three secretion system (T3SS)- further information will be discussed in Chapter 5 of this thesis. In brief, T3SSs were first characterised in V. cholerae by Dziejman et al. in 2005 [95] in a non-O1/O139 strain known as AM-19226, which lacks both CT and TCP and instead, encodes a T3SS- which is genetically similar to the T3SS2 gene cluster found in a pandemic clone of *Vibrio parahaemolyticus* Dziejman et al. concluded that the genes that encode T3SS (in AM-19226) were present within some clinical and non-O1/ non-O139 and at least one clone was globally distributed [95] .T3SS are particularly interesting when studying host-pathogen interactions; by using these systems, bacteria are able to directly inject their effector proteins into the host cell across both bacterial and host membranes- where they can manipulate host cell function [96]. The most well characterised T3SS effector protein studied in *V. cholerae* is VopF, which has been associated with actin cytoskeletal organisation and been shown *in vitro* to influence the integrity of tight junctions in cultured mammalian monolayers [97]. Interestingly, VopF is conserved in all *Vibrios* encoding a T3SS alpha and is positioned in the core, conserved region of the system [95, 98, 99]. The assembly of the T3SS involves over 20 different proteins [100-103] and are structurally and genetically related to the bacterial flagellum [104]. Subsequent studies in *V. cholerae* (AM-19226) have shown that the T3SS is essential for intestinal colonisation in the infant mouse [97].

1.8 Critical considerations for this PhD project: *V. cholerae* is a Schedule 5 agent in the UK

Within the U.K, *V. cholerae* has bioterrorism status through The Anti-Terrorism, Crime and Security Act 2001. This was set up to ensure good security of dangerous substances that may be targeted or used by terrorists. The list of dangerous pathogens and toxins (biological agents) that fall within the scope of the Act are contained within Schedule 5. The security needed to handle *V. cholerae* within the laboratory adds significant restrictions on how work can be conducted with this organism. Working hours are reduced and the need for companion working means that all experiments must be carefully considered before being undertaken. A designated laboratory adhering to specific anti-bioterrorism guidelines had to be built and this is the only designated place that experiments with live *V. cholerae* can take place. Access to this designated laboratory can only be granted after successful completion of extensive training and competency, in addition to a rigorous background security screen.

Due to these government restrictions for *V. cholerae*, I had to show competency in all experiments that were conducted using a non- schedule 5 designated organism before they could be repeated with *V. cholerae.* Therefore, I chose Enterotoxigenic *E. coli* (ETEC) strain H10407 to act as a necessary surrogate for *V. cholerae*. Although they differ in the way they cause disease there are some common features (discussed in more detail below). Both *V. cholerae* and *E. coli* fall within the γ- proteobacteria classification so share some evolutionary similarities in their genetic makeup [72]. ETEC encodes either, or both, a heat-stable toxin (ST) and a heat-labile toxin (LT), the latter of which bears many structural and functional similarities to CT. CT and LT-I in particular share approximately 80% amino acid sequence identity in the A and B subunits [105]. Both ETEC LT-I toxin and CT belong to the AB_5 class of bacterial toxins, specifically, both hold the AB_5 structure formed of one A and five B subunits the structure of which can be seen in Figure 1.12. The pentameric B subunit of both toxins attach to the GM_1 ganglioside receptor to allow delivery of the toxin and the potential to cause disease as previously mentioned in more detail for CT (see section 1.6). The main difference between CT and LT-I is that CT will bind with strong affinity to GM_1 gangliosides

only, whereas LT-I has been shown to also have affinity for other gangliosides, polyglycosylceramides and glycoproteins [106], [107] and [108].

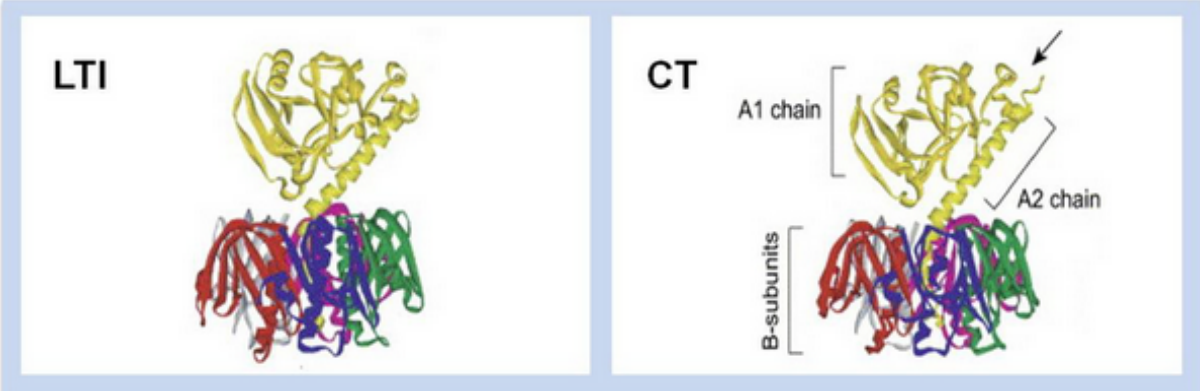


Figure 1.12. Showing structural similarities between ETEC LT-I and V. cholerae *CT. doi:10.3390/toxins5112009.*

1.9 *In vivo* differences in disease potential and infection models

Using our genomic data, we have hypothesised that there are different lineages of *V. cholerae* that have different disease potential in humans. This PhD sets out to develop a human model to characterise differences in representative strains within these lineages using *in vivo* infection models.

The most used and well-defined models to understand disease are the rabbit and infant mouse model. Although *V. cholerae* is only known to naturally infect and colonise humans, one key difference between animal models and human infection is that *V. cholerae* does not naturally colonise adult mice or rabbits due to their mature host defences and the presence of intestinal microbiota [109]. However *V. cholerae* can infect and colonise infant animals, germfree [109] and antibiotic pre-treated adult animals [110]. Whilst these models have been incredibly useful in understanding the basic science of *V. cholerae* infection, there are limitations for the use of these models. The advantages and disadvantages of both will be discussed below.

1.9.1 Infant or Sucking mouse model

Unlike *V. cholerae* isolates in adult mice, isolates orally administered to 3- to 5-day-old mice colonise the intestine and cause fluid accumulation [111]. Similarly, to humans, the main location of colonisation of suckling mice by *V. cholerae* is the middle small bowel [112]. Mice are used as a model for many infections due to their small size, ease of handling, housing, and relatively quick gestation period. Currently the suckling mouse is the most well characterised and widely used disease model for cholera. The infant mouse model has been used to identify many important virulence factors in *V. cholerae* infection, including the accessory colonisation factors [113], a haemagglutinin [114], and several metabolic proteins [115-118]. Many research groups that have utilised this model have shown that the suckling mouse bears important similarities with human infection; with TCP being necessary for the colonisation of the mouse small intestine by *V. cholerae* [80]. These findings are consistent with volunteer studies [76]. *V. cholerae* has been found to be associated with clouds of mucus within the intestinal lumen of infant mice [119]. Mucins have also been linked with limiting *V. cholerae* intestinal colonisation, particularly in the proximal small intestine of infant mice, where there appears to be a more abundant mucus layer [120].

One of the major limitations of using the suckling mouse model for *V. cholerae* are practical problems. The infant mice are very small and difficult to work with (see Figure 1.13). Infant mice cannot be separated from their mother as they are still suckling, increasing the risk of being killed by the mother which can be a result of transferring human odour to the infant mice via handling. It is also sometimes difficult to predict how many pups will be born and survive to the start of the experiment, affecting experimental design and making working with infant mice highly problematic.

Figure 1.13. A researcher performing an oral gavage on a suckling mouse demonstrating one of the difficulties faced by researchers working with suckling mice. Photo ID: WRAIR 281 Source collection: OHA 375: WRAIR Korean War Photographs, Repository: National Museum of Health and Medicine, Otis Historical Archives.

Once these obstacles have been overcome, the next potential drawback worth considering is that the infant mice do not survive long enough after infection for scientists to be able to perform vaccine studies due to the immature development of their immune system [121]. This point may not be seen as a limitation if the focus of the research is centred on innate immunity, however this does limit the ability to investigate acquired immunity. Infant mice also lack the ability to produce watery diarrhoea upon infection with *V. cholerae*, another factor worth considering if the emphasis of research is on the development and mechanism of the production of this characteristic diarrhoea. Although adult mice cannot be readily infected with *V. cholerae*, germ free mice and mice pre-treated with streptomycin can be colonised with *V. cholerae* [109]. However, it has been reported that the survival and growth of *V. cholerae* does not require TCP in this system [122], so its validity as an infection model has been questioned.

1.9.2 Rabbit model

Metchnikoff was one of the first to experiment using infant rabbits as a model for cholera. He believed that the absence of intestinal microbiota of new-born animals would allow colonisation of *V. cholerae* [123]. There are two rabbit models that have been used to understand *V. cholerae* pathogenesis, these include the reversible ileal-tie adult rabbit diarrhoeal disease model (RITARD) which was developed over 70 years ago [124]. This continues to be the most widely used adult intact model of disease [8]. The caecum of adult rabbit is temporarily ligated, and bacteria are administered into the tied intestine. The ligature is then removed, and the animals are monitored for signs of diarrhoea.
The RITARD model was pioneered in 1981 [125], not only for *V. cholerae* but also ETEC and has been successful in helping researchers to understand the distribution of bacteria within the gut during infection with *V. cholerae* and also the role that the mucus layer plays in host defence [126]. It was found that disease from ETEC, although severe, was not as 'explosive' as that from *V. cholerae* infections [125].

The second rabbit model for disease, developed more recently by Ritchie et al. demonstrates that orogastric inoculation of *V. cholerae* into three day old rabbits pre-treated with cimetidine (to inhibit gastric acid production), leads to watery diarrhoea in almost every animal and that the diarrhoea produced is of the same chemical composition as that shed by cholera patients [127]. This study identifies that goblet cell mucin depletion is reliant on CT; this is one factor that has interested us and that we have explored in human intestinal epithelium over the course of this project. The infant rabbit model also highlights the importance of CT and TCP in *V. cholerae* infection, demonstrating that, as with humans, *V. cholerae* strains that do not produce CT and TCP do not cause cholera-like disease in rabbits [127].

1.9.3 Human volunteer studies

Human volunteer studies have also been conducted for cholera and have been invaluable in allowing scientists to identify the infectious dose needed to produce consistent results in healthy humans. However, even using the same bacterial *V. cholera* strain (O139 AI1837) host variation was highlighted. Cohen et al. [128] documented that North American volunteers required a dose containing 10^6 *V. cholerae* to get an 82% incidence of disease, whereas in Thai volunteers Pitisuttithum et al. (unpublished) reported that 10^8 bacterial inoculum only caused disease in less than 20% of volunteers [129]. This is an important factor to consider during vaccine development.

Interestingly, both groups also confirmed an increase in susceptibility to *V. cholerae* in patients with blood group O. Related to this, the Ganges delta has the lowest prevalence of blood group O individuals worldwide [130], it is plausible that cholera, at least historically, has exerted significant selective pressure by causing mortality before reproductive age, thus contributing to the low levels of blood group O individuals in cholera endemic regions [131]. Blood group O does not affect the likelihood of being infected by *V. cholerae.* However, blood group O is associated with more severe disease [132].

Human studies have also allowed us to further investigate the pathogenesis of *V. cholerae* within its natural host. Levine et al. was the first to report the critical role of CT in disease using human volunteers [133], suggesting that the activity of the enterotoxin primarily accounts for the clinical manifestations of *V. cholerae* infection. However, the use of human volunteers has decreased in recent time, the most common models for cholera to date are the suckling mouse and rabbit models.

1.10 An alternative model of infection: Human Organoids

A compromise between the Human volunteer and animal models is perhaps the use of human organoids. The Collins dictionary defines an organoid as 'an artificially grown mass of cells or tissue that resembles an organ' [134]. In 2007, Takahashi et al. described the induction of pluripotent stem cells, called induced pluripotent stem cells (iPSCs) from adult human fibroblasts via the retroviral transduction of four defined transcription factors: human OCT3/4, SOX2, KLF4, and c-MYC. The consequence of the successful transduction results in induced pluripotent stem cells, ready to be differentiated into almost any tissue within the human body.

Organoids are a relatively new technology with the first publication in 2008 from Sasai et al. reporting that stem cells could be differentiated into balls of neural cells that self-organise into distinctive layers [135]. The following year, Hans Clevers' research group discovered that single $LGR5^+$ stem cells generate crypt-villus structures in which all differentiated cell types are present [136]. The Spence group were able to derive small intestinal organoids and showed that cell types present in these small intestinal organoids include goblet cells, Paneth cells, enterocytes and enteroendocrine cells [137], all the cells needed to maintain homeostasis within the intestinal epithelium. Since this discovery only ten years ago, scientists all over the world have been working on and developing organoids that mimic a wide variety of human tissue, including, but not limited to, lung, pancreas, breast, cerebral, thyroid, epithelial, kidney, cardiac, gut and intestinal organoids. This vast array of organoids that can be developed and grown in a laboratory setting has given rise to more specific disease modelling, where it is now possible to look at certain cell types of particular interest without any surrounding cells distorting any immune or phenotypic response that would be present in current animal models.

1.10.1 Development of Human Organoids

A literature search for the word 'organoid' returns a vast number of results. It is very important to note that the term 'organoid' is used to cover both iPSC derived organoids and primary derived organoids. Figure 1.14 shows a simplistic view of the different differentiation techniques used by laboratories worldwide, comparing iPSC derived organoids and primary derived organoids. Ethical issues are raised in the production of organoids derived from embryonic stem cells as the early embryo must be destroyed in order to attain the stem cells. The organoids in this project have been derived from adult skin fibroblasts obtained from a dermal puncture, in which consent was given so any ethical issues have been resolved.

Primary derived organoids are produced from, in this case, a biopsy from a patient's intestinal tissue, the stem cells are isolated from the base of the intestinal crypts and organoids grown thereafter. Once organoids are obtained from primary tissue, a secondary stage of differentiation is necessary to obtain mature epithelial cells. Commonly, biopsies are taken from patients with gastrointestinal disease, for example inflammatory bowel disease (IBD), the biopsies taken for primary organoids are from patients that do not show any sign of disease after exploratory examination, or from a 'healthy' area of the intestine. This must be taken with caution as 'healthy' may not always mean healthy and there could be an underlying cause of the symptoms that the patient is presenting with.

iPSCs derived organoids provide a relatively unlimited supply of intestinal organoids, whereas once the organoids are grown from the biopsy or embryos, it becomes more challenging, if not impossible to retrieve additional samples from the patient in future. The use of characterised iPSC cell lines also limits host genetic variation, coming from one original source that has been sequenced as opposed to potentially multiple patients which is generally with the case of primary tissue derived organoids.

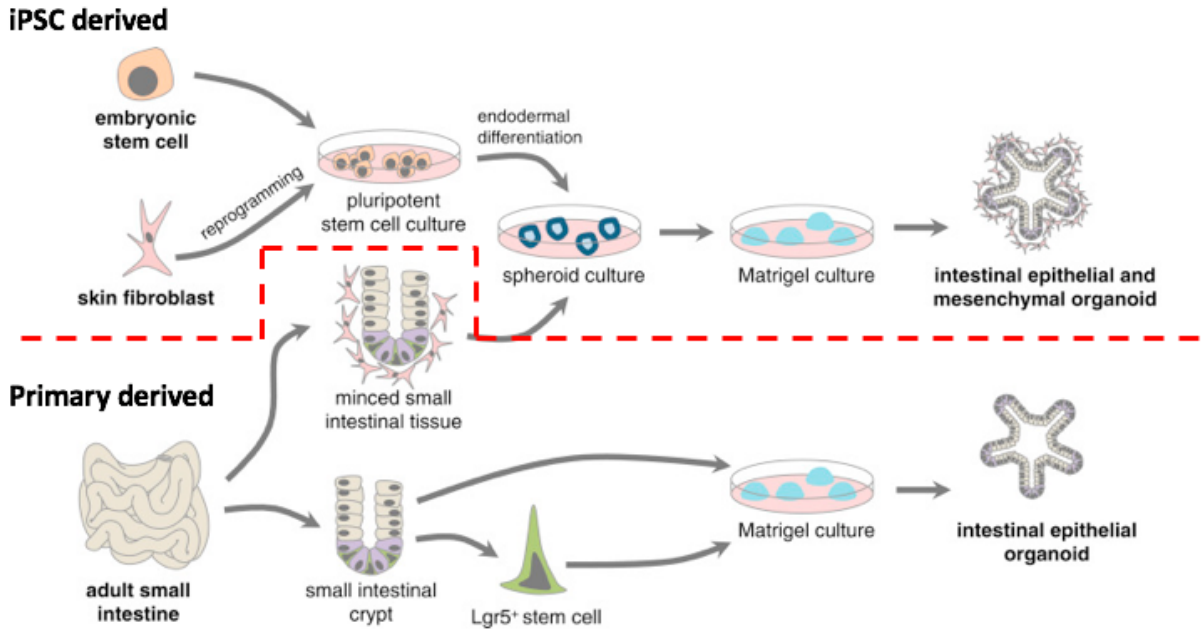


Figure 1.14. A schematic illustrating the different methods of organoid isolation. The iPSC derived organoids are reprogrammed back to pluripotency then differentiated into intestinal epithelial and mesenchymal cells whereas the primary derived organoids are derived from a biopsy of small intestine, the stem cells at the case of the crypts are isolated and grown in culture to form epithelial organoids, Credit to [138].

As previously mentioned, organoids can be developed using two different approaches, first, from isolated patient biopsies and secondly from induced pluripotent stem cells (iPSCs). The differentiation method used in this project was based on the iPSC protocol. The differentiation of hIPSCs into induced intestinal organoids mimics intestinal development in a dish shown in Figure 1.15.

The principle is that stem cells can be isolated and reprogrammed from dermal fibroblasts, following a previously published differentiation protocol [139] into highly organised spheres of cells mimicking small intestinal tissue that additionally contain a luminal centre, allowing for the micro-injection of bacteria, toxins or other chemicals to measure the effect that these have on the intestinal cells of the organoid. Representative images of the stages of differentiation from iPSC to iHO can be seen in Figure 1.15. The full protocol followed can be found in Chapter 2 Materials and Methods section 2.1.2.

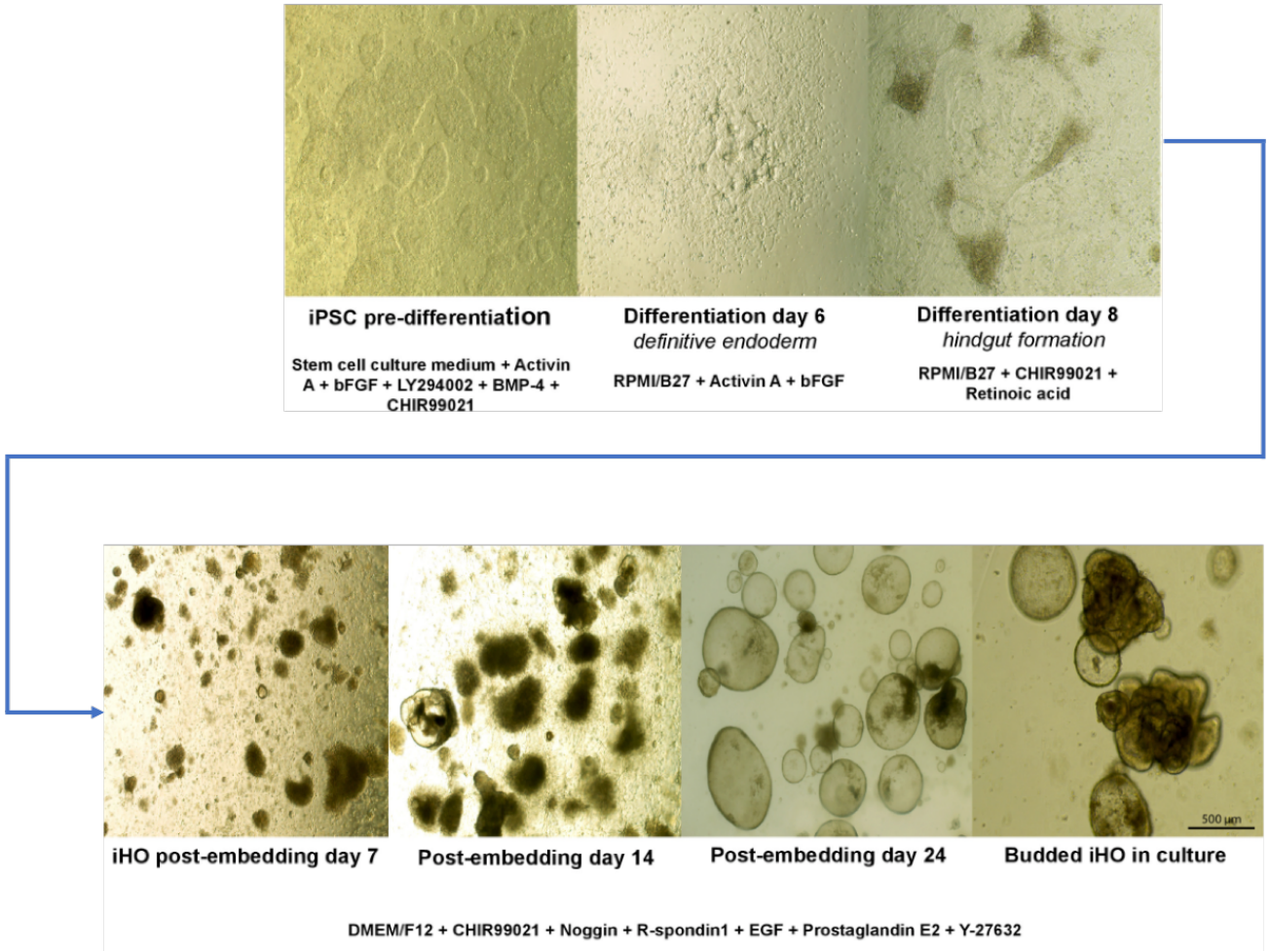


Figure 1.15. Light microscopy images showing representative stages of pluripotent stem cell to small intestinal organoid differentiation. Details of the media conditions and supplemented growth factors will be explained in chapter 2 section 2.1.2. Image taken and adapted from [140].

The hIPSC cell line that the small intestinal organoids are derived from in this project is named Kolf2, it was acquired through the Human Induced Pluripotent Stem Cell Initiative Consortium (HipSci; www.hipsci.org) [141]. The Kolf2 stem cells were obtained from dermal fibroblasts and were reprogrammed back to pluripotency using methods previously described [137]. With thanks to the Wellcome Sanger Institute high throughput mutagenesis project, Kolf2 cells will also be used to generate CRISPR host gene knock out mutations. The use of a defined cell line will allow the control of genetic background and minimise inter-line variation, which has been seen in different cell lines in the past.

Subsequently, the potential to create host cell mutations raises the possibility of targeting genes of interest within the host in addition to the bacteria provide great future potential for this project and future work within the field of organoid research.

Organoid models have the advantage of being more complex than immortalised cell lines, in that they possess a variety of cell types within one structure, adding a further depth and realism to the research by the array of cell types present, truly mimicking real human intestinal tissue.

1.10.2 Previous studies with Human Organoids

Organoids have been successfully used to model numerous bacterial infections, including iPSC derived organoids to model *Salmonella* Typhimurium *(S.* Typhimurium*) and Clostridiodes difficile* infection. Primary organoids derived from tissue biopsies have been used to model *Helicobacter pylori* infections. Summaries of which are outlined below.

iPSC derived small intestinal organoids have been used to model *Salmonella* Typhimurium infection Forbester et al*.* showed large-scale transcriptional changes within the organoid in response to *Salmonella* Typhimurium*.* The organoid model mirrors the human intestinal pathology in response to *S.* Typhimurium infection, showing the ability of *S.* Typhimurium to invade the intestinal epithelial cells, this can be seen in Figure 1.16, an electron micrograph showing the bacteria inside a *Salmonella* containing vacuole (SCV) within the intestinal epithelium of the organoid.

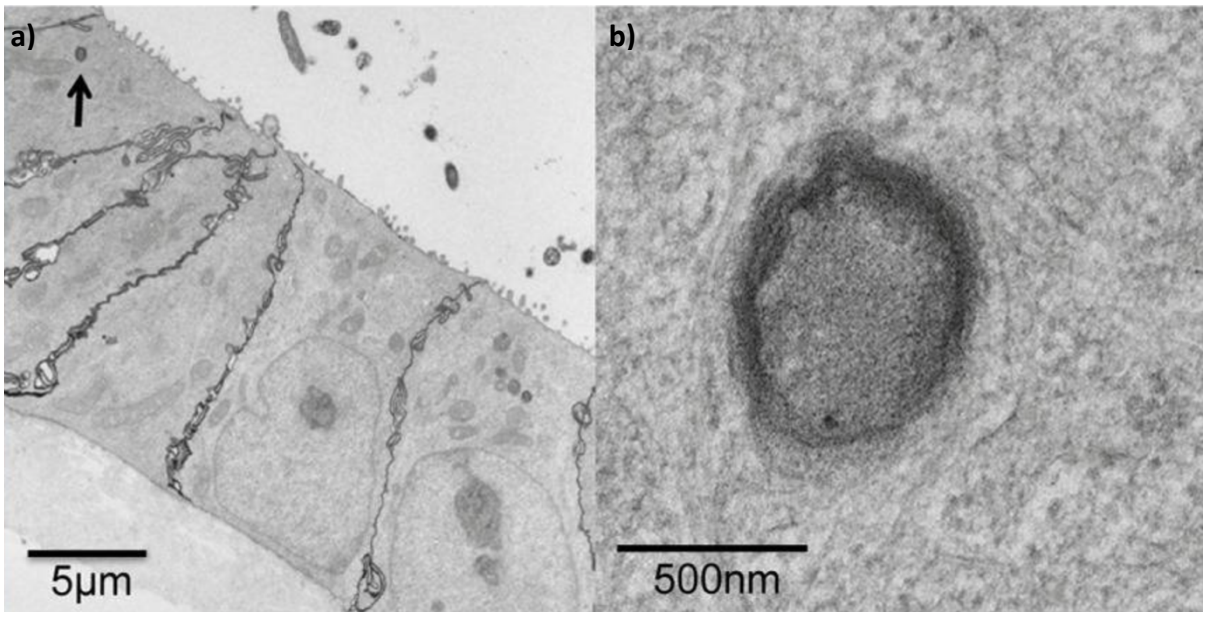


Figure 1.16. Electron micrographs showing the interaction of Salmonella Typhimurium with the small intestinal organoid. A) The arrow points to the Salmonella containing vacuole B) a magnified view of the Salmonella containing vacuole, showing real pathology that is witnessed in human infection can be reproduced in intestinal organoids [142].

Clostridiodes difficile infection has been modelled using induced pluripotent stem cell derived organoids [143]. Microinjection of the anaerobic bacteria allowed the Spence research group to monitor colonisation of the bacteria. It was discovered that a toxigenic strain of *C. difficile* caused damage to the organoid epithelium whereas a non-toxigenic strain of the bacteria did not. The same study also reported that administration of toxin impaired the barrier function of the organoids compared to a control group of organoids.

Cultivation of primary gastric organoids has allowed research to be conducted by the Clevers research group on the pathogen *Helicobacter pylori* [144], in which infection has been linked to the development of gastric cancer. In this study, biopsies were taken from patients and stem cells isolated to enable the gastric organoids to form and grow. *H. pylori* was micro-injected into the lumen of the gastric organoids and host responses were measured by microarray analysis. The gastric organoids displayed upregulation of human cytokines known to play a role in *H. pylori* infection, confirming that these gastric organoids are a model to develop further in the future.

Aims

The aims of this study are to:

- Successfully differentiate pluripotent Kolf-2 stem cells into small intestinal organoids.

- Successfully microinject cholera toxin and live *V. cholerae* into the lumen of the small intestinal organoids and to test and measure the host response.

- Understand how representative *V. cholerae* strains vary in their ability to cause infection and colonise our intestinal epithelial model.

- Investigate bacterial strain differences within selected isolates that represent the diversity of the species.

One of the aims of this research project will be to investigate if the genomic differences between bacterial strains, other than antibiotic resistance, relate to phenotypic difference that can be measured in the newly developed human intestinal organoid model. The ultimate aim of this research is to validate the intestinal organoid model for cholera toxin and *V. cholerae* infection.

Chapter 2:

2 Materials and Methods

2.1 Ethical considerations

The Kolf-2 cell line used for this project was one of the HipSci cell lines, it is an open access cell line. (https://www.hipsci.org/lines/#/lines/HPSI0114i-kolf_2). Patient consent for the skin biopsy was taken. Work carried out for the purpose of this project is covered under HMDMC (Human Materials and Data Management Committee) number 15_074.

2.2 Growth and Differentiation of hiPSCs into iHOs

2.2.1 Culture and passage of induced pluripotent stem cells

Passaging of the iPSCs took place when the cells were between 80 % - 90 % confluent. One hour before passage, 10cm tissue culture-treated dishes were coated with Vitronectin 10 μg/mL (Stemcell Technologies 7180) diluted in Dulbecco's phosphate- buffered saline (DPBS) without calcium or magnesium, to tissue culture treated plates. The volumes for coating are dependent on the size of the plate, see manufacturer's instructions. While the plates are being coated with Vitronectin, allow the E8 Flex stem cell culture medium to warm to room temperature (r.t.).

The media was removed from the iPSCs ready for passage and the cells were washed twice with DPBS without calcium or magnesium. Versene solution (Life Technologies) was added to the plates ensuring the entire surface of the plate was coated, the plate was then incubated at r.t. for 5-8 mins. The Versene solution was aspirated and discarded when holes started to appear in the centre of the iPSC colonies. Stem cell culture medium was added to the plated and iPSC colonies were dislodged gently by washing the medium over the surface of the colonies a few times. The iPSCs were passaged as clumps as opposed to single cells. The displaced iPSC colonies were transferred to a 15 mL falcon tube.

The Vitronectin was removed from the Vitronectin coated plates and replaced with stem cell culture medium. The conical tube containing the iPSCs was rotated a few times to ensure the cells were in suspension as opposed to settled at the bottom of the tube. The iPSCs were diluted to ensure a 1:10 dilution of cells was used to seed the vitronectin coated plates. These ratios were adjusted depending on stem cell growth rate. The Vitronectin coated plates were rocked a few times to ensure the iPSCs were evenly distributed on the surface of the plates and then incubated at 37 °C, 5 % CO_2. iPSCs were fed the day after passage with fresh complete E8 Flex medium.

This protocol was developed by the supplier (https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2FMAN0013768__hPSCs_Essentl8Med_Essentl8FlexMed.pdf)

2.2.2 Differentiation of iPSC to iHO

On Day 0 of the differentiation, once the iPSCs have grown to 80 % - 90 % confluence on a 10 cm tissue culture-treated dish, pre-coated with Vitronectin XF, the E8 Flex stem cell medium was replaced with 10 mL E8 Flex stem cell medium (Gibco) containing activin A (10 ng/mL) and basic fibroblast growth factor (bFGF; 12 ng/mL).

On day 1, the growth medium was replaced with 10 mL fresh E8 Flex medium containing activin A (10 ng/mL) and bFGF (12 ng/mL).

On day 2, the differentiation commenced by changing the medium to 10 mL E8 Flex stem cell medium containing: activin A (100 ng/mL), bFGF (100 ng/mL), bone morphogenic protein 4 (BMP – 4; 10 ng/mL), phosphoinositol 3- kinase inhibitor LY294002 (10 μM), and glycogen synthase kinases (GSK3) inhibitor CHIR99021 (3 μM).

On day 3, the medium was replaced with 10 mL E8 Flex medium containing activin A (100 ng/mL) and bFGF (100 ng/mL), BMP – 4 (10 ng/mL) and LY294002 (10 μM). Note, these supplements promote endoderm specification, changes to the iPSCs should be seen over the next 24 hours.

On day 4, The medium was changed to 10 mL RPMI/ B27 medium containing activin A (100 ng/mL) and bFGF (100 ng/mL).

On day 5, the medium was changed to 10 mL RPMI/ B27 containing activin A (50 ng/mL).

On day 6, to begin patterning the posterior endoderm to the hindgut, the medium was changed to 10 mL RPMI/ B27 containing CHIR99021 (6 μM) and retinoic acid (3 μM).

On days 7, 8 and 9, the day 6 protocol was repeated. Note, over the course of these few days, the hindgut formation should become apparent, covering the surface of the plate.

On day 10, the hindgut should be embedded in basement membrane matrix (Matrigel).

This protocol was developed by Lees et.al [140] in conjunction with this thesis.

2.2.3 Embedding the hindgut

Intestinal human organoid base growth media (iHO BGM) was made according to Table 2-1. Once prepared, the medium was then filter sterilised. The medium was removed from the plate containing the hindgut and the plate was washed with DPBS without calcium and magnesium. 5 mL of collagenase (Life Technologies 17104019- reconstituted to 0.1 %) (see Table 2-3 for constituents) was added to the plate and incubated at 37 °C, 5 % CO_2 for 5 mins. Collagenase solution was made by adding 500 mg of collagenase IV powder to 400 mL advanced DMEM/ F12. 100 mL of serum replacement, 5 mL L-glutamine (200 mM), and 3.5 μL of 2- mercaptoethanol. The solution was filter sterilised once the collagenase powder had fully dissolved. The solution can be stored at – 20 °C for up to 6 months in aliquots.

After the 5 mins incubation of the collagenase solution, the reaction was inactivated by adding 5 mL of iHO base growth medium (iHO BGM) as defined above. Cells were removed from the plate using a cell scraper and collected in a 15 mL conical tube. The cells were centrifugated at 240 x *g* for 1 min and the supernatant removed. 10 mL of iHO BGM was added and pipetted gently to break up large clumps of hindgut, cells were then centrifugated at 95 x *g* for 1 min. The cells were washed twice in iHO BGM by resuspending the cells and centrifugation. The cells were resuspended in 300 – 500 μL iHO BGM then, approximately 100 μL of this disaggregated cell suspension iHO BGM mixture was added to 1.5 mL of thawed basement membrane matrix. The basement membrane matrix remained on ice during the whole of this protocol to avoid solidifying too early. A 24-well plate was warmed to 37 °C on a plate heater within the class II MSC and 60 μL of the cell basement membrane matrix mix was spotted gently into the centre of the wells. The spotted matrix was allowed to set on the 37 °C plate heater and the density was subsequently checked under a microscope. Once all of the hindgut basement membrane matrix mixture was plated, the 24-well plate was incubated at 37 °C for 10 mins to allow the basement membrane matrix to set fully. Once set, 800 μL of iHO BGM containing; R- Spondin (500 ng/mL), Noggin (100 ng/mL), epidermal growth factor (EGF- 100 ng/mL), CHIR99021 (3 μM), prostaglandin E2 (2.5 μM) and Y – 27632 dihydrochloride monohydrate (ROCK inhibitor- 10 μM). The iHO BGM was changed every 3-4 days or immediately if the media began to discolour. After the initial seeding, distinct spheres should begin to be seen after 3-4 days in culture.

This protocol was developed by Forbester [142] and Lees [140] in conjunction with this thesis.

2.2.4 Generation of hiPSC derived iHO (Media and Growth Factors- iHO)

Table 2-1. Components of iHO Base Growth Medium (iHO BGM)

Component	Source	Amount
Advanced DMEM/F12	Invitrogen	500 mL
B27 serum-free supplement	Life Technologies	10 mL
N2 serum-free supplement	Life Technologies	5 mL
HEPES 1 M	Life Technologies	5 mL
L-glutamine 200mM	Life Technologies	5 mL

Table 2-2. Growth factors required for supplementation of iHO BGM

Component	Source	Amount
Recombinant human R-spondin 1	R&D	500 ng/mL
Recombinant human Noggin	R&D	100 ng/mL
Epidermal growth factor (EGF)	R&D	100 ng/mL
Prostaglandin E_2	Sigma	2.5 μM
CHIR99021	Abcam	3 μM
Y-27632 dihydrochloride monohydrate	Sigma	10 μM

Table 2-3. Components of collagenase solution

Component	Source	Amount
Collagenase IV powder	Life Technologies	500 mg
Advanced DMEM/F12	Gibco	400 mL
KnockOut Serum Replacement	Gibco	100 mL
L-glutamine 200 mM	Life Technologies	5 mL
2-Mercaptoethanol	Sigma	3.5 μL

Table 2-4. A list of components with their catalogue number and supplier used over the course of this thesis.

Name	Catalogue number	Source	Additional information
2- Mercaptoethanol	M6250-10ML	Sigma- Aldrich	
Advanced DMEM/F12	12634010	Gibco	
B-27 Supplement (50 X), serum free-10 mL	17504044	Gibco	
BMP -4 recombinant human protein	PHC9534	R&D	Stock concentration 10 μg/ mL. final concentration 10 ng/ mL
Cell Recovery Solution	354253	Corning	
CHIR99021	Ab120890-5 mg	Abcam	Stock concentration 3 mM, final concentration 3 μM
Collagenase type IV powder	17104019	Life Technologies	Reconstitute at 0.1 %
Costar TC treated 24 well culture plates	CLS3527	Corning	
Dulbecco's PBS (without Calcium and Magnesium)	14190-144	Life Technologies	
Epidermal growth factor	236-EG-200	R&D	Stock concentration 100 μg/mL, final concentration 100 ng/mL
Gentamicin	G1272-10ML	Sigma- Aldrich	Stock concentration 10 mg/mL, final concentration 0.1 mg/mL
HEPES 1 M	15630056	Life Technologies	
KnockOut Serum Replacement (serum replacement)	10828010	Gibco	
L-Glutamine	A2916801	Life Technologies	Stock concentration 200 mM, final concentration 2 mM
LY294002	V1201	Promega UK	Stock concentration of 50 mM, final concentration 10 μM
Matrigel	356231	Corning	Growth factor reduced; phenol red free
N-2 Supplement (100 X)-5 mL	17502048	Gibco	Stock concentration 100 x, final concentration 1 x

Phenol Red	P0290-100mL	Sigma- Aldrich	
Prostaglandin E_2	P0409-1MG	Sigma	Stock concentration 2.5 mM, final concentration 2.5 mM
Recombinant Human EGF	236-EG-200	R&D	
Recombinant human FGF basic	233-FB-025	R&D	Stock concentration 100 μg/mL, final concentration 100 ng/mL
Recombinant Human Noggin	6057-NG-100	R&D	Stock concentration 100 μg/mL, final concentration 100 ng/mL
Recombinant human R-spondin1	4645-RS-025	R&D	Stock concentration 25 mg/mL, final concentration 500 ng/mL
Recombinant human/ mouse/ rat Activin A	338-AC-050	R&D	Stock concentration 100 μg/mL, final concentration 100 ng/mL
Recovery cell culture freezing medium	12648010	Gibco	
Retinoic acid	R2625-50MG	Sigma- Aldrich	Stock concentration 3 mM, final concentration 3 μM
Vitronectin XF	7180	Stemcell Technologies	
Y-27632 dihydrochloride	Y0503-1MG	Sigma	

2.2.5 Passage of iHOs

iHOs were passaged every 4-7 days, depending on growth rate. The medium was removed and replaced with Cell Recovery Solution and incubated at 4 °C for one hour. The supernatant was removed and replaced with iHO BGM (as defined in Table 2-1). iHOs were allowed to settle, supernatant removed and washed for a second time with iHO BGM. The supernatant was removed, leaving approximately 300 μl of media surrounding the iHOs. The iHOs were mechanically broken up into smaller cellular clumps by pipetting three to four times with a P200 pipette (Eppendorf). Between 150 and 300 μl of iHO suspension was added to pre-thawed Matrigel (Corning) and pipetted to evenly distribute the cells in the Matrigel. The Matrigel containing iHOs was spotted (60 μl/spot) evenly on a 6-well plate (Corning PLPL0011). The 6-well plate was incubated at 37 °C for 15 minutes to allow the Matrigel to solidify before adding the Base Growth Media containing growth factors 0.01 % R-spondin 1 (R&D), 0.001 % CHIR 99021 (Sigma), 0.001 % Noggin (R&D), 0.001 % Prostaglandin E_2 (Sigma), 0.001 % Epidermal growth factor (R&D) and 0.0033 % y-27632 (Sigma).

Organoid cells are grown at 37 °C with 5 % CO_2.

If the luminal cavity filled with debris/ dead cells, the media turned yellow due to fast growing iHOs, the Matrigel starts to break down or if the iHOs start to grow out of the Matrigel then the iHOs were re-passaged as described above.

This protocol was developed by Forbester [142] and Lees [140] in conjunction with this thesis.

2.2.6 Cryopreservation of iHOs

The media was removed and discarded from the organoids and 500 μl of BD Cell Recovery solution (BD 354253) added. Cells with recovery solution were incubated at 4 °C for 45 minutes. After 45 minutes, the cells were transferred to a 15 ml falcon tube (352097) and centrifuged at 800 rpm for 5 minutes. 500 μl per 6-well dish of BD Recovery Cell Culture Freezing Medium (12648-010) were added to the pellet of cells and gently re-suspended, taking care not to break up the iHO. 700 μl of the cell and freezing medium mix was added to pre-labelled cryo-tubes and placed into a freezing unit in a -80 °C for a few days and finally transferred to liquid nitrogen for long-term storage.

This protocol was developed by the supplier (https://assets.fishersci.com/TFS-Assets/LSG/manuals/12648010_cell_culture_freeze_media_PI.pdf).

2.2.7 Recovery of cryopreserved iHOs

1000 μl of Matrigel with the appropriate growth factors was prepared ready for the cells.

Base Growth Medium was warmed to room temperature.

A vial of iHO in freezing medium was removed from the liquid nitrogen store and immediately placed into a 37 °C water bath to thaw quickly. Once thawed, the iHO cells were added to 10 ml of base growth medium in a 15 ml falcon tube. The cells were centrifuged at 800 rpm for 2 minutes and the supernatant was discarded to remove the freezing medium. The cell pellet was re-suspended in 10 ml base growth medium to allow a second wash and pelleted once more at 800 rpm for 2 minutes. The supernatant from the final wash was removed and replaced with 100 μl base growth medium containing growth factors, again taking care not to dissociate the iHOs. The iHO suspension was added to the prepared Matrigel and spot onto a 6-well plate in 60 μl domes. The plate was placed in the incubator at 37 °C for 15 minutes to allow the Matrigel to solidify before adding the base growth medium supplemented with growth factors.

This protocol was developed by Forbester [142] and Lees [140] in conjunction with this thesis.

2.3 Development of 2-D organoids

Two-dimensional iHOs were generated directly from three-dimensional iHO samples, 2-D iHOs are effectively three-dimensional iHOs that have been dissociated and cultivated on a 2-D plane or monolayer.

96-well plates or 8-well glass slides were coated with Collagen 1, Rat tail (A10403-01) by adding 85 μl of collagen to 5 ml Acetic Acid 0.02 M and incubated for 1 hour at room temperature.

After treatment with Cell Recovery Solution and centrifugation (steps in passage as described above) the supernatant was removed and the pellet containing iHOs was re-suspended in 10 ml ice-cold PBS. Once the iHOs had settled to the bottom of the 15 ml falcon tube, 6 ml of PBS was removed and replaced with fresh PBS. iHOs were centrifuged at 800 rpm for 2 minutes. The cell pellet was re-suspended in 4 ml TripLE Express (ThermoFisher 12604013). The suspension was incubated at 37 °C and vortexed intermittently until the solution became cloudy (Approximately 5 minutes). The enzymatic reaction of TripLE was stopped by adding base growth media to the cells. The suspension was then centrifuged again, and the resulting pellet was suspended in Base Growth Media containing the same concentration of growth factors used in the passage protocol (see Table 2-2).

The number of cells was determined using a counting chamber and trypan blue (Sigma T8154). Between 10^4 and 10^5 cells were then added to a 96-well plate (Corning PLPL0102) or to an 8-well glass slide (MilliCell EZ slide PEZGS0816) previously coated with collagen and washed with PBS three times to remove any residual Acetic acid. Media was changed every two days until the desired cell density was reached.

This protocol was developed as part of this thesis.

2.4 Phenotyping of iHOs

2.4.1 Immunostaining of iHOs

During passage, iHO samples to be immunostained were seeded onto Millicell EZ 8-well slides (Merck) for approximately 5 days. Samples were then fixed with 4 % paraformaldehyde (Alfa Aesar) in DPBS (no Ca^{2+}or Mg^{2+}) for either 20 minutes (2-D cell layers) or 60 minutes (3-D culture) at room temperature. Fixed samples were blocked and permeabilised with 2 % Triton X-100, 5 % FCS/PBS (Sigma) for two hours at room temperature. Primary antibodies with their appropriate dilution (see Table 2-5) were added to 0.25 % Triton X-100, 5 % FCS/PBS, overlaid on the samples and incubated for 2 hours at room temperature or overnight at 4 °C with gentle shaking. Nuclei were counterstained with DAPI in PBS (0.5 μg/ml, Sigma) for 20 minutes at room temperature. Cell membranes were occasionally stained with DiD (2.5 μg/ml Life Technologies) for 90 minutes at room temperature. 3-D iHO were immersed in FocusClear™ (elExplorer) for three hours at room temperature. FocusClear™ was removed and all plastic caging was taken off the slide. All samples were then mounted in DAPI Prolong Gold Diamond Antifade Mountant (Life Technologies).

This protocol was devised in the Dougan laboratory and built upon over the course of this project.

Table 2-5 Antibodies used for Immunostaining of iHOs

Target	Catalogue number	Host	Source	Dilution
DAPI dilactate		-		
MUC2	Ab11197	Mouse	Abcam	1:200
LYZ	Ab2408	Rabbit	Abcam	1:50
VIL1	Ab130751	Rabbit	Abcam	1:50
CHGA	Ab36997	Mouse	Abcam	1:50
GM1	Ab23943	Rabbit	Abcam	1:100
UEA-1			VectorLabs	1:100
Goat anti- rabbit IgG Alexa fluor 546	A-11010	Goat	Life Technologies	1:100
Goat anti-mouse IgG FITC		Goat		

2.4.2 Confocal imaging of organoids

Once the immunostaining method had taken place (method 2.3.1), the slides/dishes are ready to be imaged by confocal microscopy. The Leica SP8 confocal microscope was used in this thesis.

Briefly, the microscope, computer and lasers were turned on and the Leica Application Suite X software prepared. The colours of the dyes used was chosen from the menu in the Leica software to correspond to the laser needed to capture the image e.g., DAPI- blue 405, FitC- green 488, TritC- red 543 etc. The slide/ dish was loaded onto the imaging stage and the eyepieces were used to locate the samples of interest before acquiring the image using the corresponding lasers and the Leica software. This was repeated in order to capture the necessary images with the replicates for each experiment.

This protocol was developed by Leica but adapted for use in this project for the 3-D organoid cells used.

2.4.3 Electron microscopy sample preparation of iHOs

3-4 days prior to injection, while passaging, the disaggregated 3-D iHOs were resuspend in 200 μl of Matrigel and plated onto glass bottomed Willco Wells microinjection dishes. The cells were overlaid with 3 mL iHO BGM (with growth factors).

The day before injection the overnight culture of selected bacteria was set up to the desired OD and prepared as mentioned previously.

If only toxin was administered, the required concentration was prepared and either microinjected into the lumen of the organoids or overlaid with the media. iHOs were incubated at 37 °C for the necessary timepoints, usually no more than 48 hours. After injection/ toxin treatment, the medium was removed and replaced with 3 mL of primary fixative (2 % PFA with 2.5 % glutaraldehyde in 0.1 M sodium cacodylate buffer) and incubated for 1 hour at room temperature. The samples were washed three times for 10 minutes each in sodium cacodylate buffer with added chlorites (see Table 2-6). The samples were then treated with 1 % osmium tetroxide in sodium cacodylate buffer at room temperature for one hour. The samples were washed again three times for 10 minutes each in sodium cacodylate buffer. The buffer solution was removed from the plates and sections of Matrigel were cut and placed into glass vials. 1 % tannic acid was added to the glass vials containing the iHOs for 30 minutes. After 30 minutes, the samples were rinsed with 1% sodium sulphate for 10 minutes. The samples were dehydrated through a series of ethanol concentrations, 30 %, 50 %, 70 %, 90 %, followed by three 100 % for 20 minutes per concentration. The 100 % ethanol was replaced by propylene oxide for 2 washes of 15 minutes each. The propylene oxide was removed and replaced with a 1:1 mix of propylene oxide and Epon resin (Epoxy Embedding Mediumkit, Sigma) for at least an hour. The mix was removed and replaced with neat Epon resin overnight.

The following day, fresh Epon resin was made, and the samples were embedded into a flat moulded tray. The samples were cured at 65 °C for 48 hours. 500 nm sections were cut using a Leica UCT ultramicrotome and stained with toluidine blue on a microscope slide. Images were taken of the toluidine blue stained samples on a Zeiss Axiovert CCD camera and areas selected for 50 nm ultrathin sectioning. 50 nm sections were collected onto copper grids and

contrasted with uranyl acetate and lead citrate before viewing on a FEI 120kV Spirit BioTWIN TEM and recorded CCD images on a F4.15 Tietz charge-coupled device camera.

This protocol was devised by David Goulding in the Microscopy facility within The Wellcome Sanger Institute.

Table 2-6 Components required to make sodium cacodylate buffer

Component	Source	Amount
Distilled H_20	MilliQ	1 litre
Sodium cacodylate trihydrate	Sigma	21.4 g
Magnesium chloride	Sigma	1 g
Calcium chloride	Sigma	0.5 g
Hydrochloric acid	Fisher Scientific	For use in adjusting pH to 7.42

2.4.4 Fucose staining of organoids

Stocks of Invitrogen Click-iT® fucose alkyne (tetra-acetylfucose alkyne cat. No. C10264) were aliquoted and kept at – 20 °C at 50 mM concentrations.

During passage of the organoids, while the media containing growth factors was being made up, the fucose alkyne was supplemented to the media to give a final concentration of 50 μM. Organoids were then seeded into an 8-well glass slide (MilliCell EZ slide PEZGS0816) and over laid with the cocktail of media, growth factors and fucose alkyne.

The organoid cells were placed back into the 37 °C incubator with 5 % CO_2 for 48 hours.

After 48 hours, the media was removed, and the cells were bathed in 4 % paraformaldehyde in PBS for one hour at room temperature. The fixed samples were blocked and permeabilised with 2 % Triton X-100, 5 % FCS/PBS (Sigma) for two hours at room temperature. The cells were washed once in 1 % BSA in PBS. The Click-iT® reaction cocktail (to total 5 ml) was prepared by combining 1 X Click-iT® cell reaction buffer (4.4m l) with $CuSO_4$ (100 μl), Click-iT ® cell buffer additive (500 μl) and finally the azide-modified molecule (50 μl). The samples were incubated for 30 minutes at room temperature and protected from light. The cells were washed once with 1 % BSA in PBS. In order to visualise which cells had taken up the fucose, the detection reagent used was Alexa Fluor® 488 azine (cat no A10266). Continue with immunostaining protocol for staining targets of choice following fucose labelling (Section 2.4.1).

This protocol was developed by Christine Hale within the Thomson laboratory.

2.5 Infection Assays

2.5.1 Microinjection with bacteria and toxin

3-4 days prior to injection, while passaging, the disaggregated iHOs were resuspended in 200 μl of Matrigel and plated onto glass bottomed Willco Wells microinjection dishes (Willco Wells HBSB-5040) or MatTek (P50G-0-14-F), depending on the planned experiment. The cells were overlaid with 3 mL iHO complete base growth medium (with growth factors).

The day before injection the overnight culture of selected bacteria was set up by streaking the desired bacterial isolate on ampicillin LB plates.

On the day of injection, the microinjection microscope chamber was set to 37 °C, 5 % CO_2 for approximately one hour prior to use. The bacterial inoculums were prepared by diluting the bacteria in PBS to the required OD (optical density), spinning 1 ml of the bacteria for 5 minutes at 13 r.p.m, discarding the supernatant and resuspending with 25 μl Phenol Red or Trypan Blue and 25 μl PBS (as previously reported by by Forbester [142] and Lees [140]).

The injection arm of the system was moved to the right to allow suitable space to manoeuvre the microinjection dish within the chamber. The microinjection dish containing iHOs to be injected was loaded into the environmental chamber of the microscope carefully and their lid was removed to allow access to the iHOs. The injector arm and control station were turned on to allow calibration. The injection pressure was set to 600 kPa with an injection time of 0.5 secs.

Eppendorf transferMan microinjection tips (Piezo drill tip 5195000087) 6 μm, 25 ° were loaded with 10 μl inoculum and the needle were very carefully inserted into the Eppendorf TransferMan NK2- Femtojet express microinjection system. The injector arm was pulled into position over the dish for injecting. The arm was then slowly lowered onto the dish using the joystick controller. The needle was moved into position while looking down the microscope, using the joystick to lower the needle slowly and the microscope objective to keep the tip of the needle in focus. Once the tip of the needle was aligned in position and lowered into the selected organoid, the organoids were injected by pressing ‘inject’ on the injection controller.

After injection, the needle was carefully raised out if the injected organoid. The process was repeated until the necessary number of organoids were injected. Once sufficient organoids were injected, the injector was raised to the top position and moved to the right to allow safe removal of the dish and safe disposal of the used injector needle. The injected organoids were incubated at 37 °C, 5 % CO_2 for the required timepoint. When live bacteria were injected to the organoid, the base growth media was replaced after 20 mins with base growth media containing gentamicin to allow any bacteria not contained within the organoid lumen to be killed.

This protocol was developed by Forbester [142] and Lees [140] in conjunction with this thesis.

2.5.2 Basolateral administration of toxin

3-4 days prior to the start of the experiment, while passaging, the disaggregated iHOs were resuspended in 200 µl of Matrigel and plated onto glass bottomed Willco Wells microinjection dishes (Willco Wells HBSB-5040) or MatTek (P50G-0-14-F)- depending on the planned experiment. The cells were overlaid with 3 mL iHO complete base growth medium (with growth factors) and allowed to form for 3-4 days prior to the start of the experiment.

On the day of toxin administration, the iHO complete base growth medium was removed and replaced with complete growth medium (without growth factors) and the desired concentration of toxin dependant on the experiment was added to the cells. The toxin treated iHO were then incubated for the desired time at 37 °C, 5 % CO_2 and either imaged over the course of the experiment or fixed and stained for imaging. Immunostaining of iHOs.

This protocol was developed as part of this thesis.

2.5.3 Toxin concentration calculations

Basolateral toxin administration:

Toxin was delivered in 1 mg vials. 1 ml of water was added- 1 mg/ml (stock concentration).

For basolateral experiments 1 µl was added to 1000 µl of media to give 1 µg/ml in the most used toxin concentration.

2.6 RNA isolation from organoids

At the selected timepoint, BD cell recovery solution was added to the plates containing organoids and incubated at 4 °C for one hour. After incubation, cells were pelleted by centrifugation and the supernatant was discarded. The Quiagen RNA Easy Kit (74104) was used for RNA extraction. 600 μl RLT buffer was pipetted onto the pelleted organoids, flicked to release the pellet and to homogenise the cells. 600 μl 70 % ethanol was added to the tube containing the cells and mixed by pipetting. 700 μl of the mixture was added to the RNeasy spin column which was placed into a 2 ml collection tube and centrifuged for 15 s at 10,000 rpm. The flow through was discarded. This step was repeated as my samples had a volume greater the 700 μl to start with. 700 μl RW1 was added to the spin column, centrifuged for 15 s at 10,000 rpm and the flow through discarded. 500 μl of RPE buffer was added to the spin column and centrifuged for 15 s at 10,000 rpm and flow through discarded. Another 500 μl of RPE buffer was added to the spin column, centrifuged for 15 s at 10,000 rpm and the flow through discarded again. The spin column was placed into a new 1.5 ml collection tube, 50 μl RNAse free water was added directly to the membrane of the spin column, centrifuged for 1 minute at 10,000 rpm. Another 50 μl of RNAse free water was added directly to the spin column and centrifuged for 1 minute at 10,000 rpm. The quantity of RNA in the final 100 μl in the collection tube was recorded using Thermo Scientific NanoDrop One. Samples were labelled and stored at -20 °C for subsequent analysis.
This protocol was developed by Quiagen.

2.7 Host response methods

2.7.1 Cytokine measurement using the Luminex

Five hundred microliter samples of supernatant harvested from organoids which had been either left as matched controls or exposed to toxin baso-laterally for 4, 24 and 48 hours in 24 well plates were filtered through 13 mm 0.2 µM syringe filter units and stored at -80 ^{o}C until required. On thawing, 25 µl aliquots of the supernatants were added to 25 µl volumes of the relevant mixes of anti-cytokine specific magnetic beads for a set period of time. The beads were washed with the kit wash buffer utilising a 96 well magnet and then exposed first to the anti-cytokine Biotinylated reagent and then bound biotinylated antibody was detected with the streptavidin-Phycoerythrin (PE) conjugate. After several washes with the kit wash buffer, the samples were analysed on a Luminex Flex-Map 3D. The data was plotted using GraphPad Prism to show amounts of each cytokine detected in pg/ml.

List 2-7 shows the 48 cytokines that were assayed using the supernatants of cholera toxin treated organoids for this project.

This protocol was developed by Christine Hale in conjunction with this thesis.

List 2-7. A list of cytokines used in this project in the 48-plex Bio-Rad Bio-Plex Pro Human Cytokine Assay Kit.

Cytokine	**Cytokine**	**Cytokine**
CTAK	Il-6	MIF
Eotaxin	IL-7	MIG
FGF- basic	IL-8	MIP-1α
G-CSF	IL-9	MIP-1β
GM-CSF	IL-10	β-NGF
GRO-α	IL-12 (p70)	PDGF-BB
HGF	IL12- (p40)	RANTES
IFN-α2	IL-13	SCF
IFN-γ	IL-15	SCGF-β
IL-1α	IL-16	SDF-1α
IL-5	M-CSF	IL-1β
IL-1RA	Il-18	TNF-β
IL-2	IP-10	TRAIL
IL-2Rα	LIF	VEGF-A
IL-3	MCP-1 (MCAF)	
IL-4	MCP-3	
IL-17A	TNF-α	

2.7.2 RNA Seq method

Organoids were passaged in accordance with section 2.2.5 and placed into microinjection dishes. After 3-4 days of growth, budded and cystic organoids were microinjected with cholera toxin and trypan blue, in addition to budded and cystic organoids not being injected (as per the experimental design outlined in chapter 4.2 of this thesis). Selected organoids were isolated and removed from the Matrigel at 4, 24 and 48 hours post injection and placed in to RLT plus buffer. The samples were then handed to the Research and Development team within DNA pipelines at the Wellcome Sanger Institute for downstream processing.

Once the data had been quality controlled and processed through the Wellcome Sanger Institute Pathogens Pipeline, the reads were aligned using nfcore [145] rnaseq pipeline https://github.com/nf-core/rnaseq using "star_salmon" alignment method [146, 147]. Differential analysis was conducted using edgeR [148] with the Benjamini-Hochberg method of adjustment to account for multiple testing. Estimates of dispersion were performed (estimate of variance- expect a negative binomial distribution for gene expression). The samples were normalised against each other to eliminate bias between samples. The Generalized linear model likelihood ratio test (GLMRT) for differential expression was used. Pathway analysis was conducted using pathfindR library [149].

This protocol was developed by the Wellcome Sanger Insitiute Research and Development team and The Wellcome Sanger Institute Pathogens pipeline for the purpose of this project.

The filters for differential expression were:

Log2foldchange > abs(1) (greater than 1 or less than -1 log fold change)

P value < 0.05

2.8 Bacterial Methods

2.8.1 Construction of fluorescent bacteria

The bacterial strain of interest was streaked onto LB agar plates and incubated at 37 °C overnight. The following day one colony of bacteria was picked and mixed with 1000 μl LB broth, vortexed to mix and break up the bacteria. The vortexed mixture was then poured into a baffled flask and topped up with LB broth, placed into the shaking 37 °C incubator and left to grow for a few hours. While the cells were growing in the incubator, 10 % glycerol and the plasmid of choice were placed on ice to chill. The centrifuge was set to 4 °C and allowed to cool down prior to use. The bacterial cells were harvested by centrifugation at 4000 rpm for 10 minutes at 4 °C. After centrifugation, the cells were left on ice. The supernatants were poured off and the bacterial pellets were resuspended in 10 ml glycerol. The cells were left on ice for approximately 20 minutes. The bacteria were centrifugated at 4000 rpm for an additional 10 minutes to pellet the cells. A 1.7 ml tube was placed on ice to chill. The bacterial pellet was resuspended in 1 mL glycerol by using a p1000 pipette tip that had been cut at the end and transferred to the chilled 1.7 ml tube. The bacteria were centrifugated at 6000 rpm for 5 minutes to pellet the bacterial cells once more. The spin and resuspend steps were followed two more times and after the final spin, the bacterial pellet was resuspended in 200 μl glycerol. 1 μl of the chosen plasmid was mixed with 40 μl of bacterial cells in a cuvette and then electroporated with the BioRad micropulser with the setting 'Ec1' (1.80kV for 5ms). 1 ml of LB broth was added to the cuvette and left at 37 °C for one hour to allow the bacterial cells to recover. 100 μl of the electroporated mixture was plated onto pre- warmed Ampicillin plates and incubated at 37 °C overnight. Bacteria were checked for growth and counted the following morning to confirm resistance to Ampicillin and successful placement of the plasmid. Bacterial stocks were made from the successfully electroporated colonies. For this thesis, pDiGc was transfected into the bacterial isolates and selected for on LB plates containing ampicillin in order to select colonies containing green fluorescence protein (GFP). The plasmid map of pDiGc can be seen in Figure 2-1.

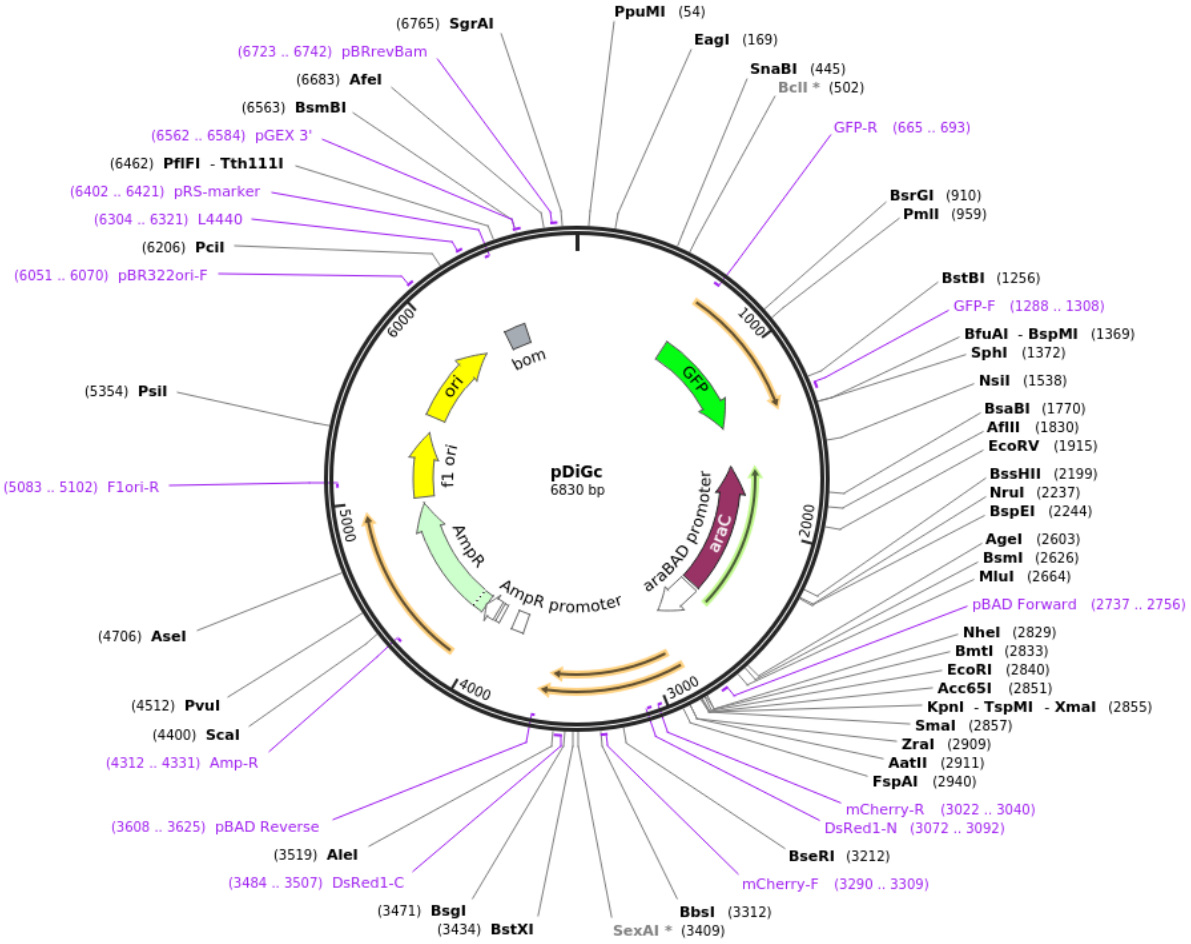


Figure 2-1. A plasmid map for pDiGc that was used to transfect N16961, NCTC10732 and VC0009 for this thesis. Following transfection, the bacteria are selected on ampicillin plates and the resultant bacteria express the green fluorescence protein (GFP).

2.8.2 DNA extraction of *V. cholerae*

DNA extractions were carried out using the Masterpure Complete DNA and RNA Purification kit (Epicentre #MC85200), with modifications to the manufacturer's protocol.

The -80 °C stock of the bacterial strain of interest was streaked out on LB agar plates and incubated overnight at 37 °C. The following day, the strain was then lawned onto a LB agar plate using a single colony and incubated over night at 37°C. Five loopfuls of the lawned bacteria were taken and mixed with 300 μl Tissue & Cell Lysis Solution containing Proteinase K. This mixture was then vortexed for 10 seconds and incubated at 65 °C for 20-25 minutes or until the suspension had fully cleared. The samples were then vortexed every 5 minutes during the 65°C incubation to ensure even lysis of the bacterial cells. RNAse A was then added, and the samples were incubated at 37°C for 30 minutes to ensure complete digestion of residual RNA. Samples were then chilled on ice for approximately 5 minutes.

Proteins were precipitated from the sample by adding 150 μl MPC Protein Precipitation Solution and immediately vortexing for 10 seconds. The precipitated debris was collected by centrifugation (16,000 x *g* for 10 minutes at 4 °C). An additional 30 μl MPC reagent was added to each cleared sample to remove any possibly residual protein contaminants. The samples were then vortexed and recentrifuged (16,000 x *g* for 10 minutes at 4 °C). Genomic DNA was precipitated from the cleared supernatant in a fresh 1.5 ml Eppendorf tube using 500 μl room-temperature isopropanol. Precipitated genomic DNA was collected by centrifugation (16,000 x *g* for 10 minutes at 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 interference with the PacBio sequencing chemistry. Samples were tested for sterility by spotting 2 μl onto an LB agar plate which was incubated overnight at 37 °C – the absence of bacterial growth deemed the samples to be sterile and so able to be removed from the laboratory to be processed for sequencing.

2.8.3 Next-generation DNA sequencing

Genomic DNA (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 to create an Illumina library. Libraries were sequenced on a 150 bp paired-end run. Approximately 10 μg gDNA was used for PacBio sequencing, using polymerase version P6 and C4 sequencing chemistry reagents.

Oxford Nanopore MinION technology:

Using the high molecular weight genomic DNA extracted using method 2.7.2. The Rapid Barcoding Sequencing kit form Nanopore technologies (SQK-RBK004) was used. Briefly, the components for the kit were thawed, spun briefly and mixed by pipetting as indicated in the table below (Table 2-7).

Table 2-8. Table of components that make up the rapid barcoding sequencing kit (SQK-RBK004) from Nanopore technologies and information on how to treat them prior to use.

Reagent	Thaw at R.T.	Briefly spin again	Mix well by pipetting
Fragmentation mix RBq1-12	Not frozen	yes	yes
Rapid Adapter (RAP)	Not frozen	yes	yes
Sequencing buffer (SQB)	yes	yes	yes
Loading beads (LB)	yes	yes	Mix immediately before use
Flush buffer (FLB)	yes	yes	yes
Flush tether (FLT)	yes	yes	yes

Once the reagents were prepared according to table 2-7, the genomic DNA was adjusted to 10 μl with nuclease free water, the tube mixed by flicking and spun down. The tube containing the genomic DNA was incubated for 1 minute at 80 °C and briefly put on ice to chill. The beads were pelleted on a magnet until the eluate was clear and colourless. 10 μl of eluate was removed and placed into a clean 1.5 ml Eppendorf DNA LoBind tube. 1 μl of RAP was added to the 10 μl of barcoded DNA, the tune was mixed by gently flicking the tune and then spun down. The reaction was then incubated for 5 minutes at room temperature. The sequencing buffer, loading buffer, flush tether and flush buffer was thawed at room temperature and then mixed by vortexing and then spun down at room temperature. The MinION flow cell was opened under the clip and the priming port cover was slid clockwise to open the priming port, checking for air bubbles carefully. A P1000 pipette was set to 200 μl, the tip was inserted into the priming port and the wheel on the pipette dial was turned until it showed 220-230 μl you until a small volume of buffer could be seen entering the pipette. 30 μl of thawed and mixed flush tether was added directly to the tube of thawed and mixed flush buffer, the mixture was mixed and vortexed at room temperature. 800 μl of the priming mix was loaded into the flow cell via the priming port- carefully to avoid the introduction of air bubbles followed by a 5-minute wait to allow time to prepare the library for loading. A new tube was prepared containing 34 μl of sequencing buffer, 25.5 μl loading beads (mixed immediately before use), 4.5 μl nuclease free water, 11 μl DNA library to make a total volume of 75 μl within the tube. The SpotON sample port cover was gently lifted to enable the sample to be loaded. 200 μl of the priming mix was added into the flow cell via the priming port- carefully to avoid the introduction of air bubbles. The prepared library was mixed gently by pipetting up and down just prior to loading, 75 μl of sample was added to the flow cell via the SpotON sample port in a dropwise fashion- ensuring that the previous drop had entered the port before dropping the next. The SpotOn sample port cover was replaced ensuring the bung entered the port, the priming port was closed, and the lid of the flow cell was replaced. The flow cell was then attached to the computer and the sequencing run started.

2.8.4 Electron microscopy of *V. cholerae*- High Pressure Freezing

Vibrio cultures were sampled selecting a single colony from each strain, mixing with hexedecene and rapidly freezing in a Bal-Tec HPM010 high pressure freezer followed by freeze-substitution in a Leica EM AFS at -90 °C in acetone containing 0.1% tannic acid and 0.5% glutaraldehyde sequentially with acetone containing 1% osmium tetroxide and 0.1% uranyl acetate followed by room temperature embedding in TAAB 812 resin. 70 nm ultrathin sections were cut on a Leica EM UC6 and contrasted with lead citrate and uranyl acetate and imaged on an FEI Spirit Biotwin 120 kV transmission electron microscope using a Tietz 4.18 CCD.

This protocol was developed by David Goulding.

2.9 Bioinformatic methods

2.9.1 Phylogenetic trees

Illumina sequencing reads were assembled using SPAdes [150] as part of a high-throughput analysis pipeline [151], and annotated using Prokka [152]. External publicly available sequences were similarly assembled from raw sequencing reads- where these were not available, assemblies were downloaded from Genbank and annotated using Prokka to allow for uniformity within the dataset.

Roary v3.12.0 was run on the assembled genomes using the options '-e-maft -s -cd 97' which prevents the splitting of paralogous genes into separate gene clusters to avoid over-partitioning of orthologous genes into separate gene clusters. The CD-HIT clustering cut off was 97%. Core-gene alignments were trimmed using trimAl v1.4.1 [153], SNP-sites v2.5.1 [154] was used to extract the variant sites from the trimmed alignment. When running Roary, core genes are defined as genes present within 97%- 100% of the genomes that were included in the phylogenetic tree, with a minimum 95% amino acid sequence similarity (based on blastp)

IQ-TREE V 1.6.10 was used to build the maximum-likelihood phylogenetic tree [155] in chapter 6 of this thesis. The general time reversible (GTR) and ascertainment bias correction (ASC) models used [156]. Five thousand approximate likelihood ratio tests were conducted to assess the robustness of the computed phylogenies [152] and ultrafast bootstrap approximations took place for each tree calculation [157]. The phylogenetic tree was then visualised using Phandango [158]. The phylogenetic tree was rooted using *Vibrio metoecus*- the most closely related species- to provide an outgroup.

Settings for IQ tree were: (settings iqtree -s /lustre/scratch118/infgen/team266/lk4/genomesfortree/roaryfile3_1633698467/core_gene_alignment_trimmed.snps.fasta -nt 16 -m GTR+ASC -bb 5000 -alrt 5000) which used ultrafast bootstrap. A list of genomes used to create the phylogeny can be found in Appendix AA.

Chapter 3:

Development of the iHO model for *Vibrio cholerae* infection

Aim:

To assess the utility of a human organoid model for gaining a better understanding of *Vibrio cholerae* infection dynamics.

As previously mentioned, (sections 1.9.1 and 1.9.2 in the Introduction chapter), the current models of infection used to study cholera are all animal models and include the suckling mouse and infant rabbit as well as the RITARD (reversible ileal-tie adult rabbit diarrhoea) model. It is important to remember that mice and rabbits are not naturally colonised with or infected by *Vibrio cholerae.* Neither infant nor adult mouse models produce the watery diarrhoea that is characteristic of cholera disease. As discussed in Introduction Chapter section 1.9.1 Early infection within the mouse model is considered to be 2-7 hours post infection. The infant mouse model has been adapted to look for genes involved in the later stages of the infection, namely when the bacteria start to detach from the epithelial surface and are released back into the lumen of the intestine [159].

Humans are the only known susceptible host of *V. cholerae.* Therefore, one of the aims of this project was to develop an infection model that more closely represents infection of the human small intestinal epithelium, the site of infection for *V. cholerae.* The significant advantage of using human iPSC derived organoids is that they are human cells and that organoids from defined human cell lines can be maintained indefinitely and assayed repeatedly through passage. Hence, I aimed to develop the organoid model to focus on the primary site of infection and to understand the key dynamics of early infection of the human intestinal epithelium by *V. cholerae.* In establishing this model for cholera, I will also confirm how well the iHO model mimics the small intestine by looking for the presence of the major human intestinal epithelial cell types, the organisation of the epithelial layer itself, for example the presence of tight junctions, and assay for key features important for *V. cholerae* infections such as the presence of the GM1 ganglioside receptor (see below for a more detail).

This chapter focusses on development of the model, chapter 4 will examine the dynamics of cholera toxin within the intestinal organoids. Subsequently chapter 5 will investigate the comparative genomics of the selected *V. cholerae* isolates and finally, chapter 6 will explore the regions of differences between the bacterial isolates and delve into phenotypes of these isolates with respect to the newly developed intestinal organoid model.

3 Choosing between 3-D and 2-D iHO models

Previous work has shown that it is possible to derive many tissue types from pluripotent stem cells, given the correct management and supplementation, in this instance, small intestinal cells (see Chapter 1 Section 1.10.1 for a brief description). iPSCs can also be used to mimic several human tissues, including, but not limited to, lung, pancreas, breast, cerebral, thyroid, epithelial, kidney and cardiac.

2-dimensional cells have been used as the major cell culturing methods since the early 1900s [160], these 2-D cell culture methods have vastly contributed to our knowledge and understanding of many scientific fields including development, infection and cell signalling involved in cancer biology. However, they are not without limitation, 2-D cell culture does not accurately represent tissue cells *in vitro* [161]. More recently, the development of 3-D cell culture has meant that we are now able to mimic tissue *in vitro*, with cells forming a more complex architecture and morphology that bears striking resemblance to the tissue it is modelling. 2-D cell culture generally produces cells that are flat and of an elongated shape as the cells can only naturally grow along a two-dimensional plane, however the natural cell shape is maintained when cells are cultured in accordance with 3-D culturing protocols [161, 162]. Another important factor to consider is that cell junctions within 2-D cultures are less common and do not accurately represent real cell junctions, whereas the junctions are more common within 3-D culture and the cells are well differentiated [161-165].

Over the course of this project, I trialled the differentiation and use of 2-D intestinal organoids, however these cells did not grow to confluency; the cells did not grow enough to cover the bottom of the culture dishes, the cells did not hold any architecture; they grew in a

very sporadic and unorganised fashion and did not provide a useful tool for this project. Further work would be required to optimise the 2-D organoid cells for use in projects such as this. Henceforward, 3-D organoids were used for the rest of the experiments in this thesis.

Figure 3.1a. shows a schematic of an intestinal human organoid showing the structural invaginations and the presence of the different cell types characteristic of the human small intestinal. Figure 3.1b. is a microscopy image taken using an EVOS XL core to show the morphology of the organoids used during this project, similarities of which can be seen in the schematic in Figure 3.1a. To date, we are not aware of any modelling of *V. cholerae* or cholera toxin in the iPSC derived small intestinal organoid.

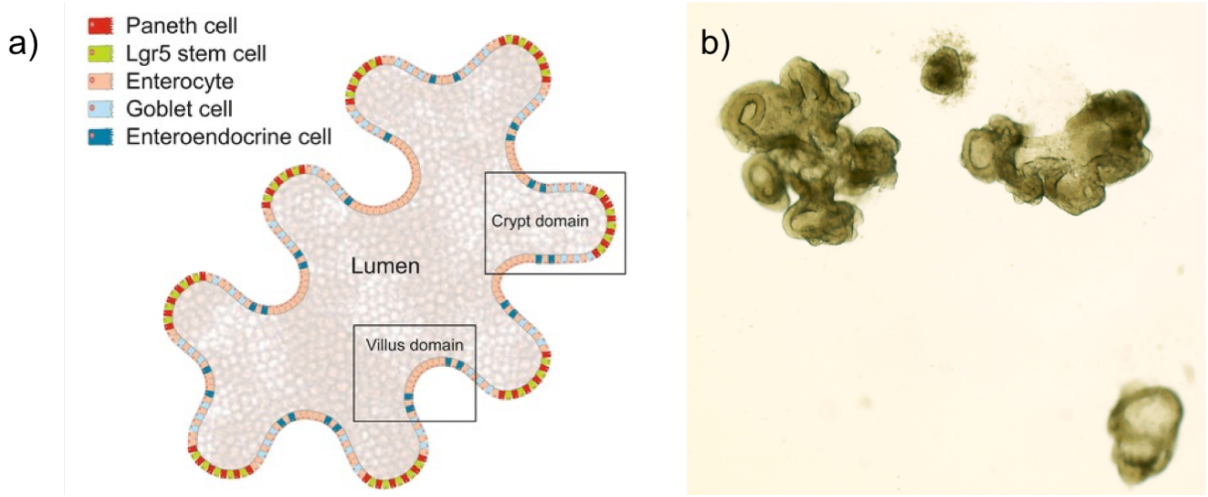


Figure 3.1a) A schematic of a small intestinal organoid showing the location of the cell types present and the crypt villus structure surrounding a luminal cavity in the centre (taken from Roeselers G, Ponomarenko M, Lukovac S, Wortelboer HM (2013) Ex vivo systems to study host–microbiota interactions in the gastrointestinal tract. Best Pract Res Clin Gastroenterol 27:101–113). Figure 3.1) b) shows an organoid culture from Kolf-2 organoids used in this project showing a similar structure with the crypt villus formation surrounding a hollow luminal cavity in the centre. (Taken with EVOS XL core).

3.1 Generation of 3D iHOs: confirmation and identification of cell types

Here, I used this existing protocol to derive 3D-small intestinal organoids from pluripotent stem cells. See Chapter 2 section 2.1.4.

To confirm that the successful differentiation of Kolf-2 iHOs took place, immunostaining (see methods chapter section 2.3.1) followed by confocal microscopy (methods chapter section 2.3.2) of the major cellular components of the human small intestine were conducted. Figure 3.1b shows the presence of crypt and villus structures in addition to luminal cavities within our Kolf-2 organoids. Next, I sought to confirm the presence of the four major human intestinal epithelial cell types within these iHOs. These include the mucus-producing Goblet cells stained with the mucus marker MUC2 (Figure 3.2a), hormone-producing enteroendocrine cells (Figure 3.2b) stained with chromogranin A (CHGA), absorptive enterocytes (Figure 3.2c) stained with Villin (VIL), and antimicrobial molecule-producing Paneth cells stained with Lysozyme (LYZ; Figure 3.2d). Of note, the distribution of these cellular markers was also representative of the distribution that can be seen in the schematic in Figure 3.1a, for instance, the enterocytes (stained with VIL Figure 3.2c) are of the same proportion of that expected and seen in the schematic of Figure 3.1a in addition to the lower frequency of Paneth cells (stained by LYZ Figure 3.2d) and enteroendocrine cells (stained with CHGA Figure 3.2b).

For quality control staining and imaging was repeated with every stem cell differentiation protocol and every time a vial of iHOs were thawed to confirm that the iHOs were representative over the course of the whole project.

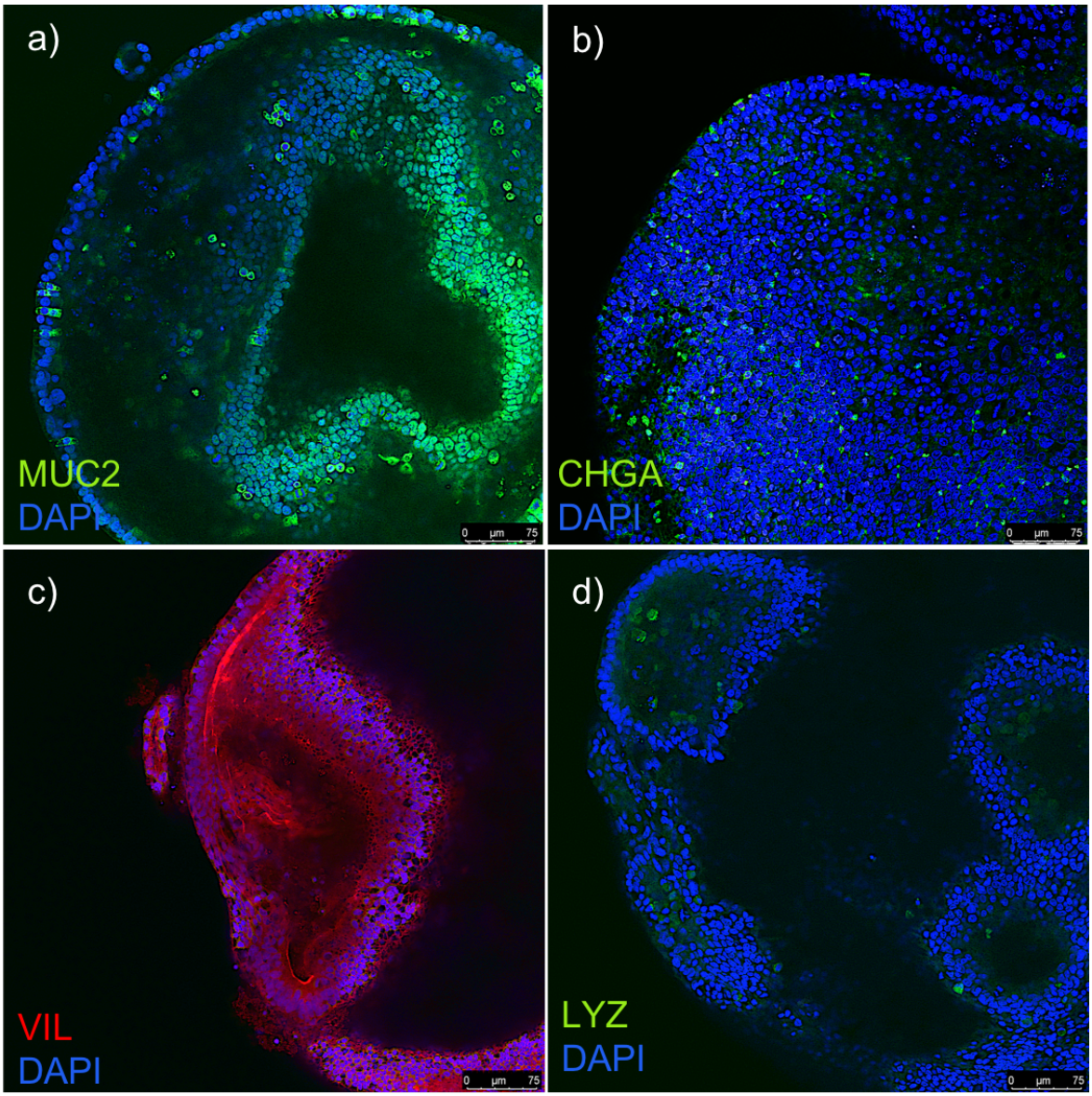


Figure 3.2 Confocal microscopy of Kolf-2 iHOs used in this project. A) shows the presence of Goblet cells, stained using MUC2 FITC pictured in green, B) shows the presence of Enteroendocrine cells stained with CHGA FITC pictured in green, C) Enterocytes are shown with staining for VIL, captured in red and D) Paneth cell staining using LYZ FITC pictured in green. DAPI was used in all experiments to stain nuclei.

In addition to the identification of the major cellular components of the human small intestine present in the iHO system, we were also able to observe fine structural components such as the tight junctions. Figure 3.3b shows a transmission electron micrograph of the Kolf-2 iHOs differentiated in this project. In a healthy gut, tight junctions are critical in regulating the mucosal barrier [166, 167]. The main function of the small intestine is the absorption of nutrients, tight junctions aid this process by regulating the paracellular uptake of water, nutrients and electrolytes [168, 169]. This regulation of multiple components and processes is a very complex procedure, maintaining homeostasis and when we consider that the human gut has approximately 8-9 litres of fluid ingested every day, and all except 100-200 ml are reabsorbed under normal conditions [170]. This balance within the intestinal epithelium can be disrupted by many external factors which in turn can influence the integrity of the tight junction, for example, Changes in pH, osmotic load or cytoskeleton function [171].

The integrity of the tight junction is regulated by the arrangement of actin and the interaction between integral transmembrane and peripheral membrane proteins [172]. One of these proteins, occludin (shown in Figure 3.3a) is thought to be important in the assembly and maintenance of tight junctions. Occludin, present at tight junctions is phosphorylated in comparison to occludin found within the cytoplasm of the cell. It is this phosphorylation that appears to control its interaction with other tight junctional proteins e.g., ZO-1 [173]. The interaction between occludin and ZO-1 is essential for tight junction integrity [172]. Previous studies have in fact shown that *V. cholerae* strains that possess a zinc binding metalloprotease called haemagglutinin protease (HA/P) can disrupt tight junctions by cleaving occludin, altering ZO-1 and by the rearrangement of actin [170].

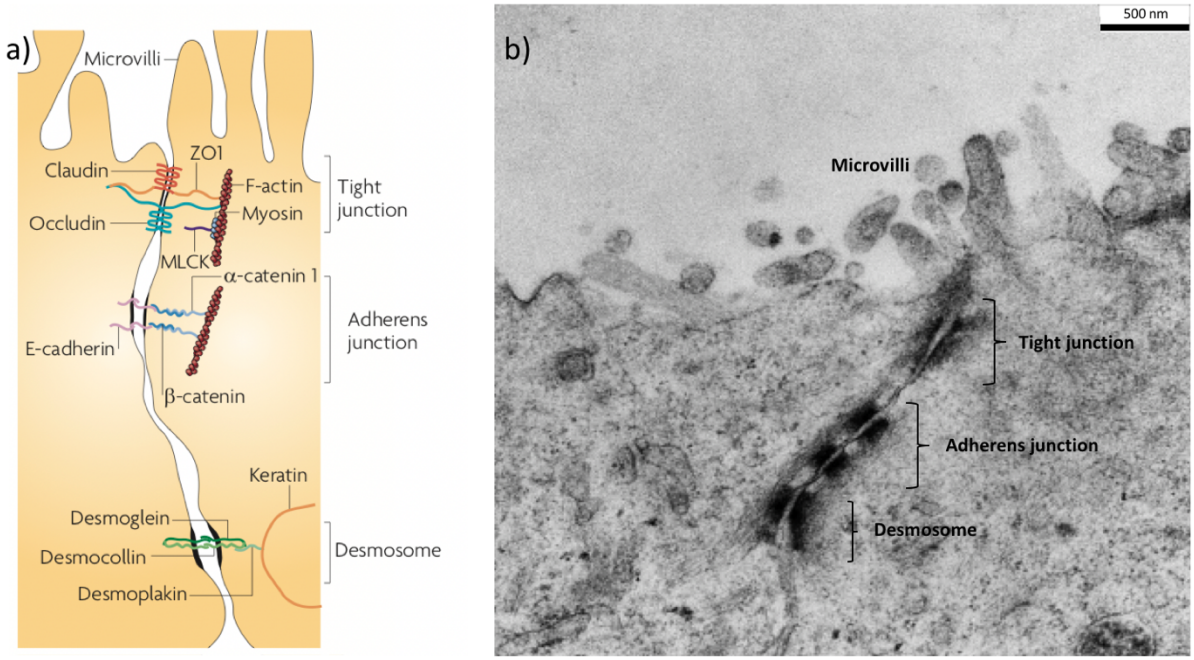


Figure 3.3 iHO tight junction. A) A schematic of a tight junction showing all components from the apical surface of the tight junction complex followed by a paracellular space between adjoining cells then the adherens junction followed by a desmosome (Taken from Turner JR. 2009. Intestinal mucosal barrier function in health and disease. Nat Rev Immunol 9: 799-809 [16]). B) A transmission electron micrograph of the small intestinal iHOs generated in this project. The tight junction complex can clearly be seen with multiple structures towards the baso-lateral side of the adjoining iHO cells.

3.2 Microinjection of 3-D organoids

To gain access to the lumen of 3-D organoids, microinjection is essential. Over the course of this project, I utilise microinjection as a tool to administer purified cholera toxin and live *V. cholerae* bacteria to the luminal surface of the organoids. A photograph of the microinjection setup is shown in Figure 3.4- the microinjection dish is loaded onto the stage through the chamber doors, the stage can be moved and controlled by the joystick labelled. The microinjection needle is placed into the needle holder which is then lowered onto the injection dish and controlled by the grey joystick. Pressure and speed of the injection are controlled by the Femtojet labelled on the photograph. Detailed methods for microinjection can be found in Chapter 2 section 2.4.1 but also available to view in video format in the Jove article authored by myself and colleagues in the laboratory at the time, https://www.jove.com/v/59478/using-human-induced-pluripotent-stem-cell-derived-intestinal

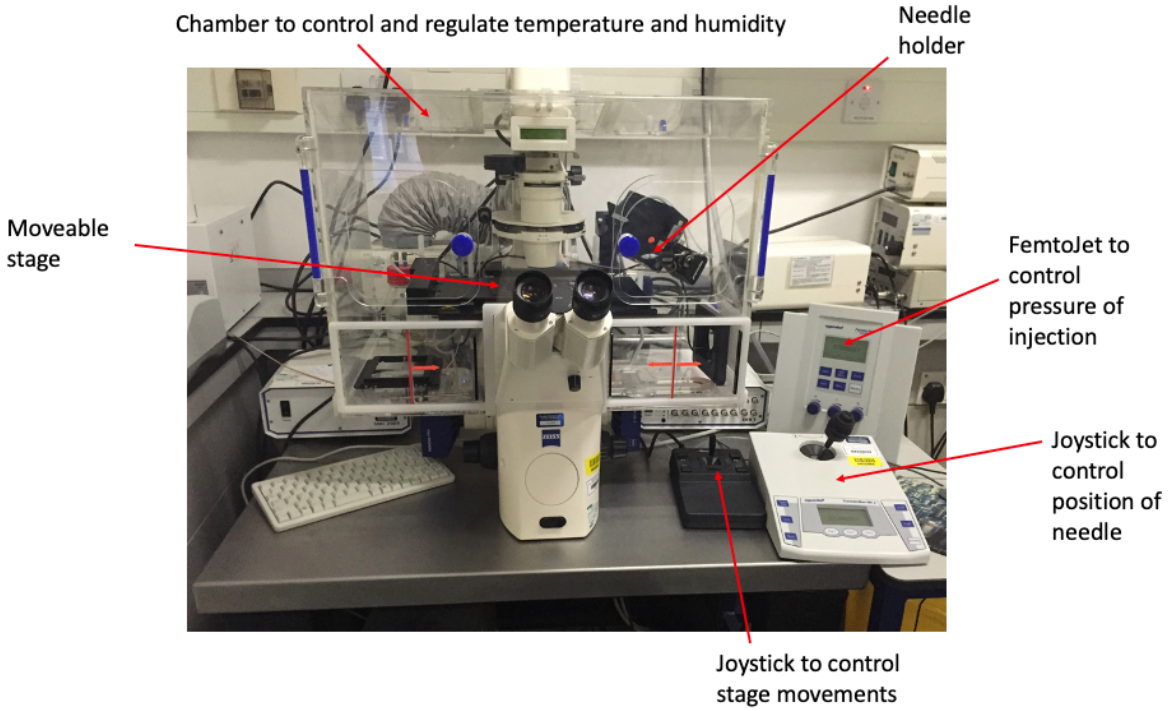


Figure 3.4 A photograph of one of the microinjection set-ups used over the course of this project. The microinjection dish is loaded onto the stage through the chamber doors, the stage can be moved and controlled by the joystick labelled. The microinjection needle is placed into the needle holder which is then lowered onto the injection dish and controlled by the grey joystick. Pressure and speed of the injection are controlled by the Femtojet labelled on the photograph.

Phenol red or trypan blue were injected into the organoids in addition to the desired conditions (CT, *V. cholerae* or PBS) to enable visible detection of microinjected organoids compared with those that were not injected.

3.3 Passage of organoids

The small intestinal organoids need to be passaged once a week and growth medium changed every 3-4 days. A detailed method of the passaging process can be found in Chapter 2 section 2.1.5 of this thesis, similarly, a video explaining the method used can be found in the Lees 2019 Jove publication [174].

In brief, the organoids are dissociated from the Matrigel, manually broken up into smaller pieces using a pipette, then re-mounted into fresh Matrigel with fresh medium containing the necessary growth factors overlaid. It is at this stage that the organoids can be placed into different vessels depending on subsequent experiments e.g., placed into microinjection dished for microinjection, or chamber slides to allow immune staining and future confocal imaging.

3.4 GM1

Now that I had selected the 3-D iHO model it was important to identify additional markers known to be important for *V. cholerae* infection. As briefly outlined in chapter 1 section 1.6.2. The most characterised receptor for cholera toxin (CT) is the monosialotetrahexosylganglioside (GM1) [87]. The strong interaction between CT and GM1 led researchers to propose that the GM1 ganglioside is the sole receptor for CT. Gangliosides are highly abundant in the nervous system and play a critical role in the modulation of membrane proteins and ion channels, the loss of gangliosides in the brain result in severe neurodegenerative disorders [175]. However, GM1 receptors are rare in the human GI (Gastrointestinal tract) [87, 176].

To confirm the presence of the GM1 ganglioside within the Kolf2 organoids, I used immuno-staining for GM1 using both a directly conjugated antibody as well as an un-labelled primary antibody in addition to a fluorescently labelled secondary antibody (for detailed methods see Chapter 2 Section 2.3.1). Both directly conjugated and unlabelled antibodies were used at different stages of the project to ascertain if the directly conjugated tag had any impact on binding. The staining was conducted on three vials of iHOs. The successful staining was then visualised using confocal microscopy on the Leica SP8 microscope and at least 5 images were taken following each time the cells were stained. Figure 3.5 shows the presence and localisation of GM1 within the iHOs. The distribution of this key receptor is located exclusively on the apical surface of the organoids in a punctate distribution. The staining process was also repeated without permeabilising the cells to confirm that the staining was ‘real’ and not an artefact of the staining procedure/ protocol.

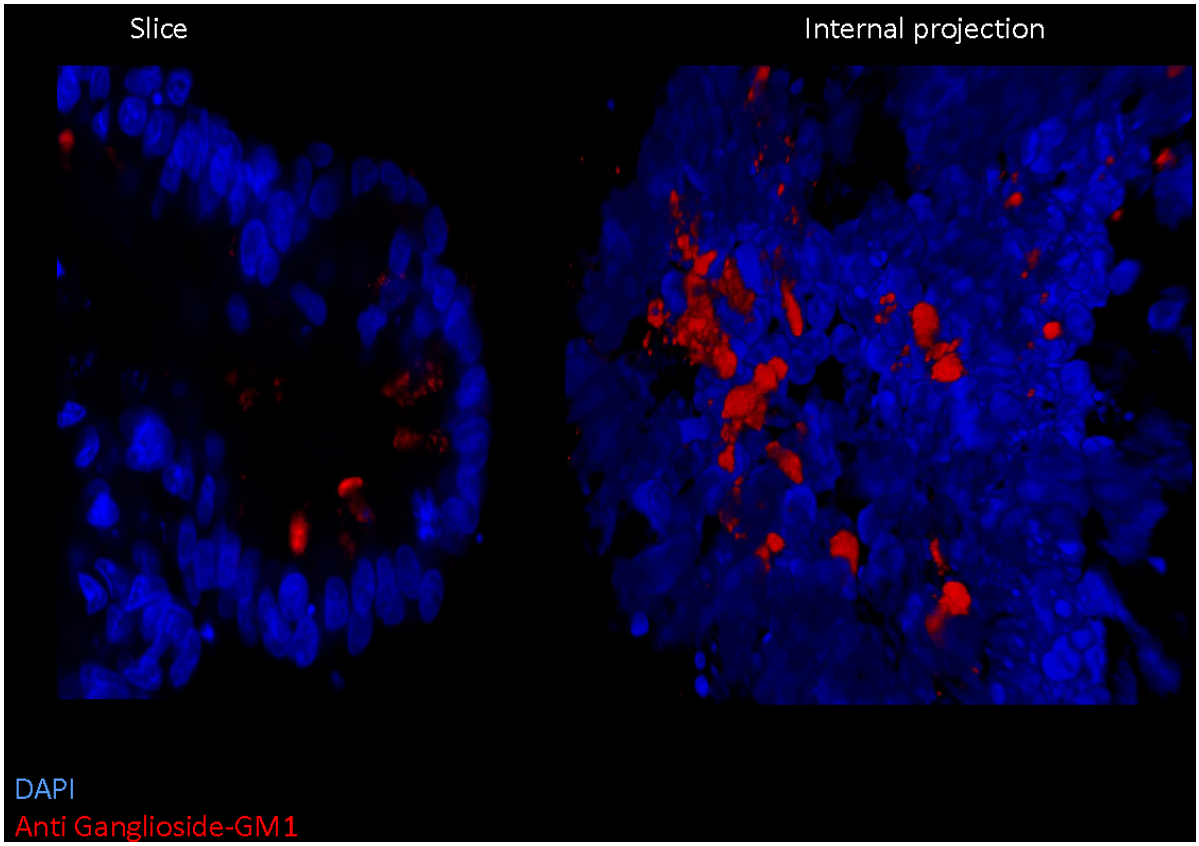


Figure 3.5. Confocal imaging of Kolf-2 iHO showing presence of GM1 ganglioside stained with TritC and 4',6-diamidino-2-phenylindole (DAPI) dilactate used as a nuclear marker. Localisation of GM1 ganglioside was observed solely on the apical side of the organoid in a punctuated fashion. Magnification 63X.

Once the presence of GM1 ganglioside had been established on the Kolf2 iHOs we then set out to determine whether if cholera toxin bound to the GM1 as predicted. As described and the structure shown in Chapter 1 (section 1.6.1) cholera toxin is comprised of two subunits CTA and CTB. CTA is the enzymatic subunit of the toxin, formed of two peptides, that is responsible for the profuse watery diarrhoea seen in cholera disease. CTB is the pentameric subunit that forms a doughnut shape that binds to the GM1 gangliosides that are present on the small intestinal epithelium. The CTB on binding to epithelial cells is a critical step in the mode of action of the toxin, allowing the whole toxin to be endocytosed in a cell type dependent manner, into cell to be transported to the endoplasmic reticulum and ultimately chloride ions and water being expelled into the intestinal lumen. The commercial cholera toxin used in subsequent chapters of this study consist of whole toxin purified from *V. cholerae* Inaba strain 569B.

In order to visualise any co-localised, and therefore any implied interactions between GM1 and CTB, iHOs were immuno-stained with fluorescently labelled cholera toxin B subunit (CTB). This would also allow me to assess if cholera toxin binds solely to the GM1 ganglioside or to other receptors which has been suggested previously [176, 177].

Figure 3.6 shows immuno-stained Kolf-2 organoids labelled with CTB and GM1. The images show a punctuated staining pattern for GM1, as described above (seen in red), however FitC-labelled CTB (green in Figure 3.6) appeared to be more widely distributed across the surface and membranes of the organoids. In these data there was minimal evidence of co-staining, areas of the organoids that showed staining for both CTB and GM1. Therefore, based on the experiments to date, I propose that GM1 may not be the sole receptor for CTB, at least in our small intestinal model.

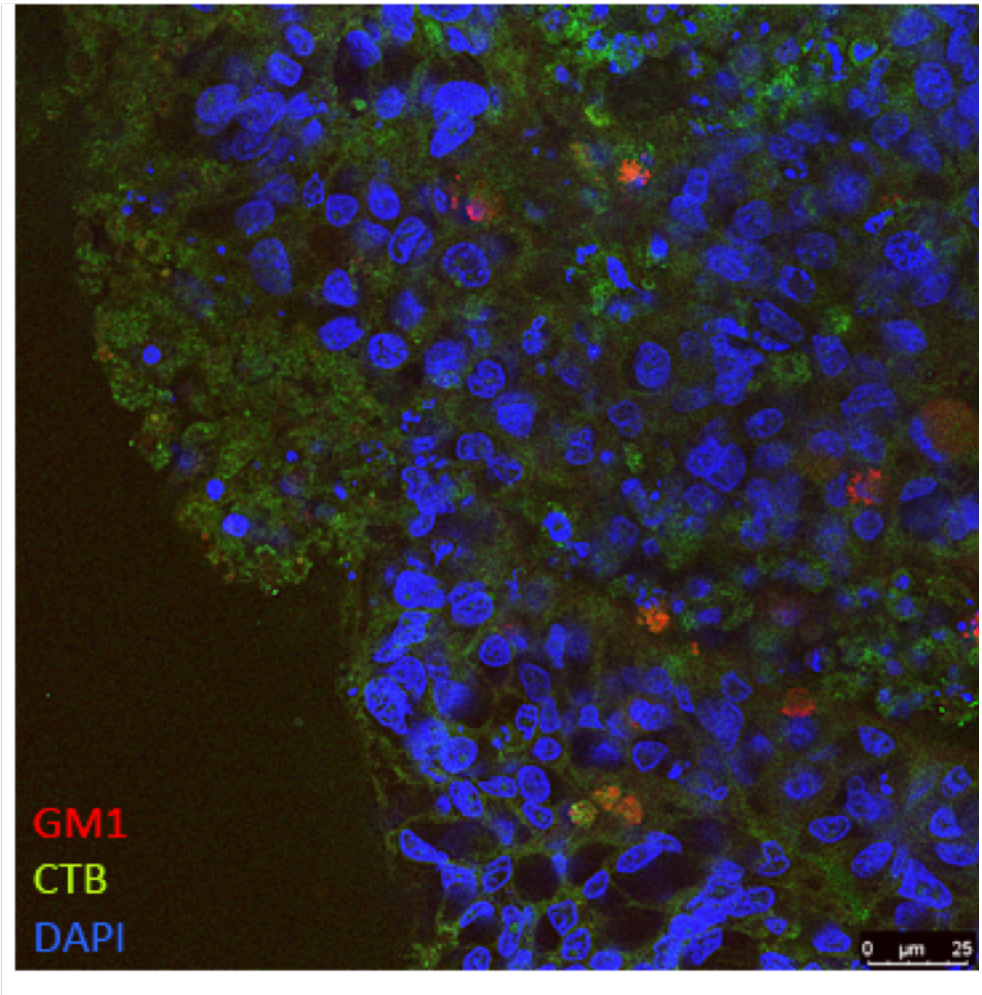


Figure 3.6. Confocal microscopy image of an iHO organoid showing cell nuclei stained blue with 4',6-diamidino-2-phenylindole (DAPI) dilactate, GM1 ganglioside stained in red with a primary and secondary antibody TritC and cholera toxin B subunit (CTB) stained in green using FitC 63X. Co-localisation of CTB and GM_1 is indicated by a mixture of green and red showing yellow.

3.4.1 CTB cell type co-localisation

In this chapter, I have shown that although GM1 is present on the iHOs, there is no exclusive GM1 CTB co-localisation that we can visualise. Next, I wanted to explore the CTB binding in the Kolf2 iHOs further, asking whether CTB preferentially bound to a particular iHO cell type. Figure 3.7 shows immuno-staining captured by confocal microscopy of each of the major cellular components in our iHOs as described in section 3.1.1 of this chapter. The Goblet cells were identified by staining for MUC2 Figure 3.7a, enteroendocrine cells were identified by staining with CHGA Figure 3.7b. VIL was used as a marker for enterocytes Figure 3.7c and LYZ for Paneth cells Figure 3.7d (see section 2.3.1 of the Methods for a detailed description of the methods, see table 2-5 for specific antibodies used over the course of this project).

From Figure 3.7, we can see some evidence of co-localisation of CTB with MUC2 and CHGA which suggests that there is some preferential interaction with goblet and enteroendocrine cells. However, CTB staining is not exclusively associated to either of these cell types and appears to bind almost ubiquitously across all cells in the Kolf2 iHO system.

These results are consistent with recent publications which suggest GM1 is not the sole receptor for cholera toxin binding to epithelial cells [176, 177].

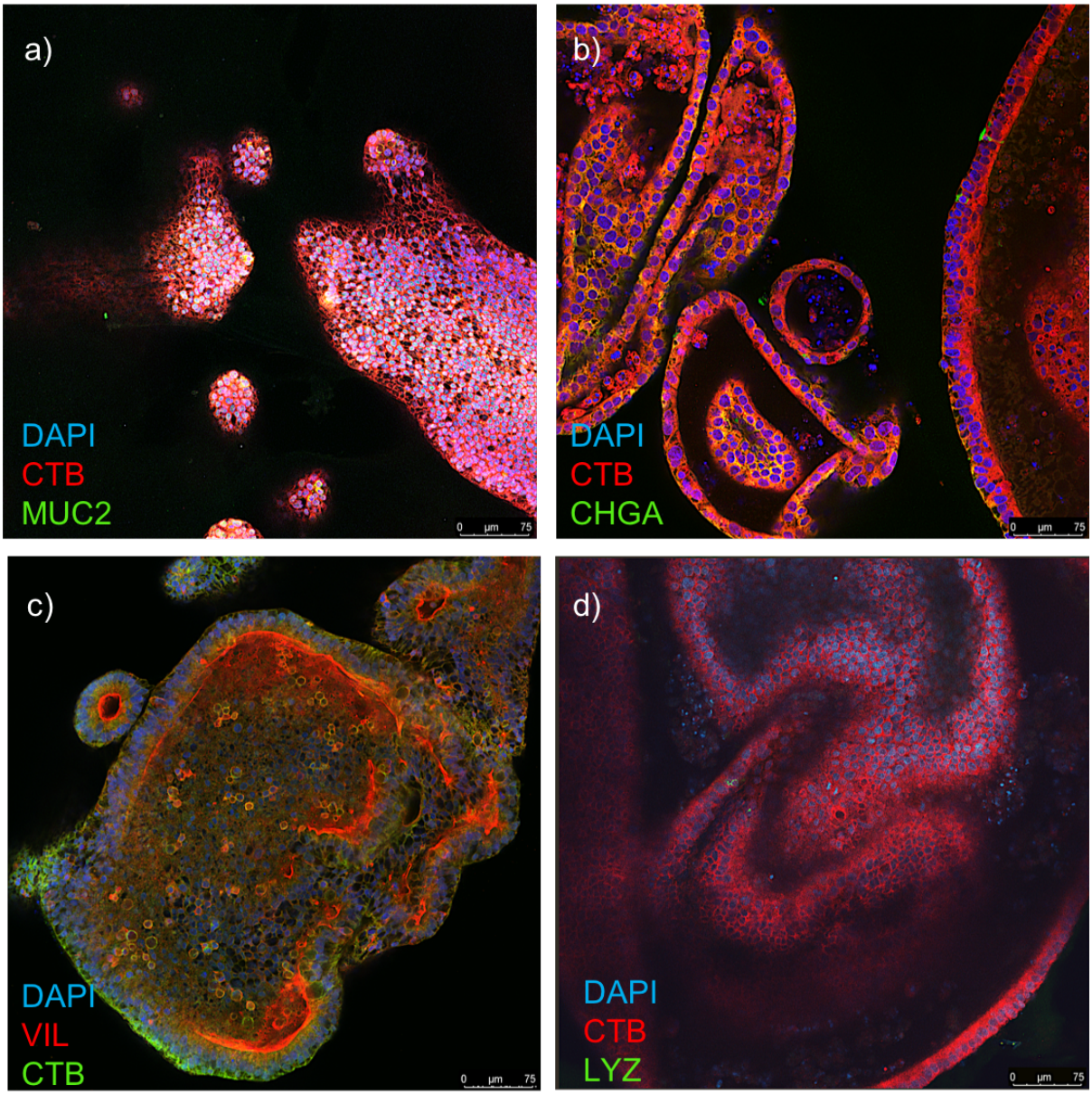


Figure 3.7. Confocal images of immuno-staining to show cell type and CTB colocalization. Figure 8a) MUC2 staining by FitC allows visualisation of goblet cells in green, CTB is stained with AlexaFlour 647-red, and nuclei can be seen with DAPI staining in blue. Figure 8b) CHGA FitC allows imaging of enteroendocrine cells, CTB is stained with AlexaFlour 647-red, nuclei with DAPI. Figure 8c) CTB is visualised with AlexaFlour 488- green while VIL- red enables imaging of villus cells, DAPi illustrates nuclei. Figure 8d) CTB is stained with AlexaFlour 647 while LYZ FitC staining shows Paneth cells, DAPI highlights nuclei. Images taken using… 20X magnification.

3.4.2 The changing dogma - Fucose as an alternative receptor for CT

Following on from the work above that suggested that GM1 may not be the sole receptor for CTB. Recently, it has been reported that GM1 independent CTB binding to fucosylated glycans may be responsible for CT binding to intestinal cells [178]. In 2015, Wands et al.[176] reported that CT bound to cell surface glycoproteins, the study by Wands et al. also gave weight to an epidemiological study which implicated variation among fucose containing glycoconjugates in human susceptibility to cholera to research conducted by Barua and Paguio in 1977 [179].

It has been shown that the gastrointestinal epithelium is exceptionally rich in glycoconjugates and expresses, in addition to GM1, glycolipids and glycoproteins that carry blood group antigens [180-182]. Consistent with this, individuals of a specific blood group, the O blood group, have been shown to have an increased risk of hospitalisation due to infection from *V. cholerae* O1 El Tor and O139 [58, 130, 132, 183, 184].
To explore this further, we must scrutinise the structure of the carbohydrate terminal residues of the blood group antigens which can be seen in Figure 3.8.

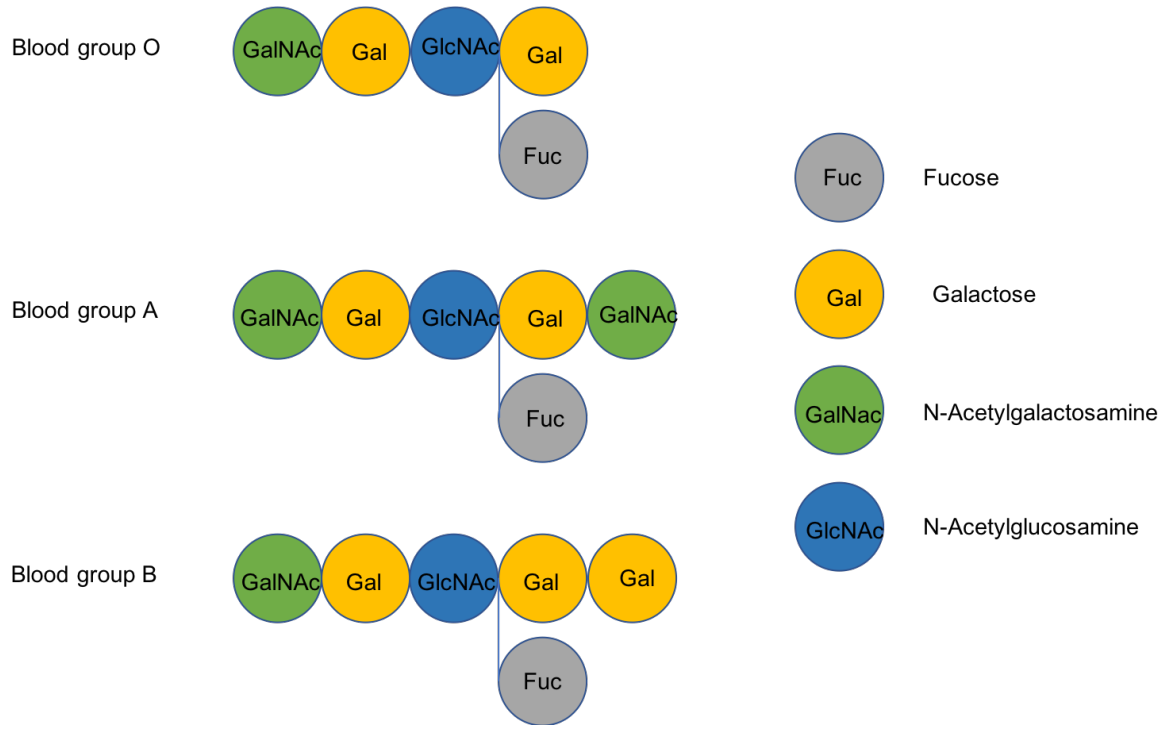


Figure 3.8. A schematic diagram of the carbohydrate residues that form the ABO blood group phenotype. Blood group O harbours a terminal fucose residue whereas blood groups A and B have additional sugar residues at their terminal ends, specifically a N- acetlygalactosamine for blood group A and a terminal galactose residue for blood group B.

The positioning of the terminal fucose residue belonging to blood group O individuals plays a key part in cholera toxin binding and ultimately cholera disease. Studies have shown that blood group O has the ability to bind CTB in two orientations, whereas the A determinant is only able to bind CTB in one orientation, increasing the affinity of CTB binding to blood group O individuals [185]. Interestingly, blood group O has been associated with a lower risk of colonisation in exposed household contacts [184] but once colonised, these individuals have a higher risk of more severe disease [130, 131, 186]. During the Latin American cholera outbreak (described in chapter 1 section 1.5), when cholera was introduced into Peru, it was discovered that individuals with blood group O were eight times more likely to be hospitalised with severe cholera [183].

3.4.3 Blood group of Kolf-2 donor

We have identified that the Kolf2 small intestinal organoids were differentiated from an individual with the AO blood group genotype and Kolf-2 cells ultimately having blood group A. It was important to identify the blood group of the donor so that we are aware of any impact that it might have on any results from this project. There are some previous studies that have examined the effect of CT upon blood group.

Kuhlmann et al. conducted a study in 2016 in which, enteroids were differentiated from primary biopsies of patients with blood group A and blood group O, treated with cholera toxin and levels of cAMP measured consequently. The study showed that although they could not identify blood group differences in binding of CT to the epithelial cells, the enteroids of blood group O had an increased response to CT and higher levels of cAMP production [187] compared with the blood group A enteroids. This study is relevant to this project, but ultimately, the models are different as described in the introduction chapter section 1.10.1, we still need to be aware of possible differences in response to CT in different blood groups.

3.4.4 Fucose residues in Kolf-2 organoids

Here, we explore the presence of fucose residues within our Kolf2 small intestinal organoids. Therefore, I immuno-stained the organoids with Ulex europaeus agglutinin 1 (UEA1) which is a lectin that preferentially binds to many glycoproteins and glycolipids containing α-linked fucose residues, such as ABO blood group glycoconjugates. UEA1 binds blood group O cells. UEA1 has been used in several studies to visualise mucus in human, mouse [188, 189], pig [190] and cell lines [191]. The results of this staining can be seen in Figure 3.9, with clear staining of fucose residues within the iHOs.

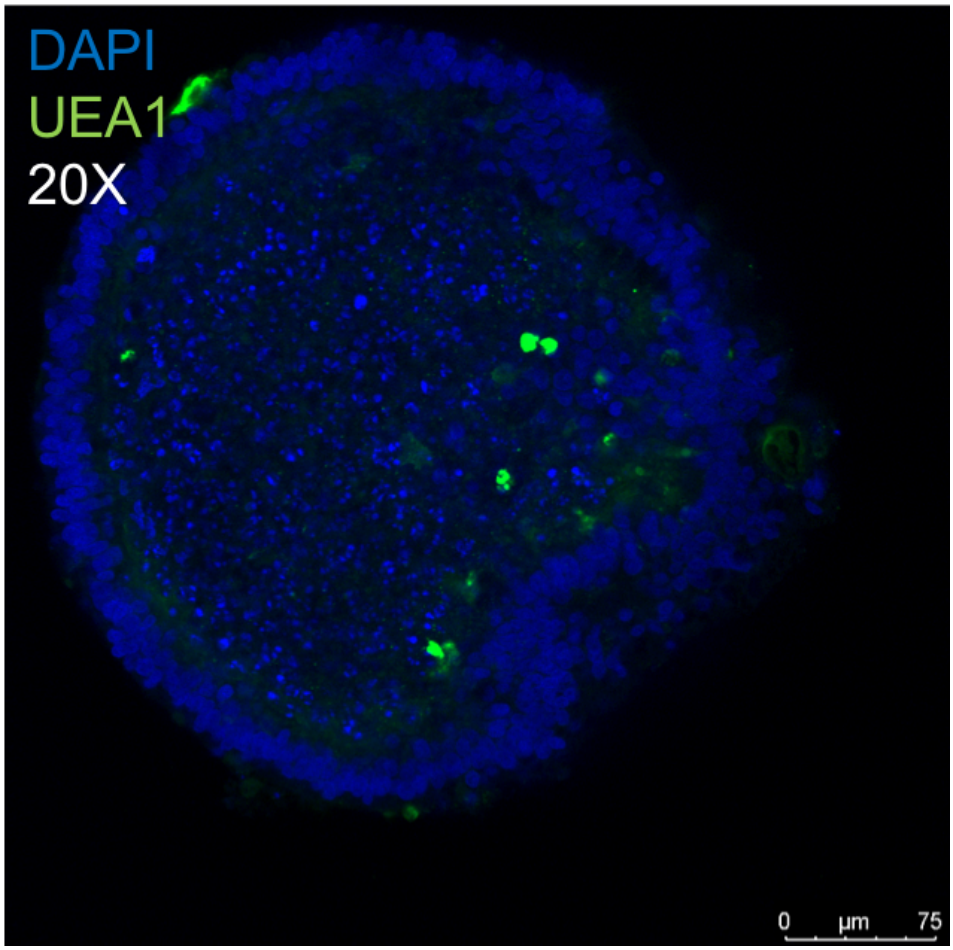


Figure 3.9. Immuno-staining followed by confocal imaging of Kolf-2 iHOs stained with UEA1 (green) to identify α-linked fucose residues and 4',6-diamidino-2-phenylindole (DAPI) to identify nuclei.

Following the identification of the fucose residues (seen in Figure 3.9), we wanted to examine whether these fucosylated residues did in fact bind to the B subunit of cholera toxin. We further stained the Kolf-2 iHOs with AlexaFlour 647 CTB and FitC UEA1 (used previously).

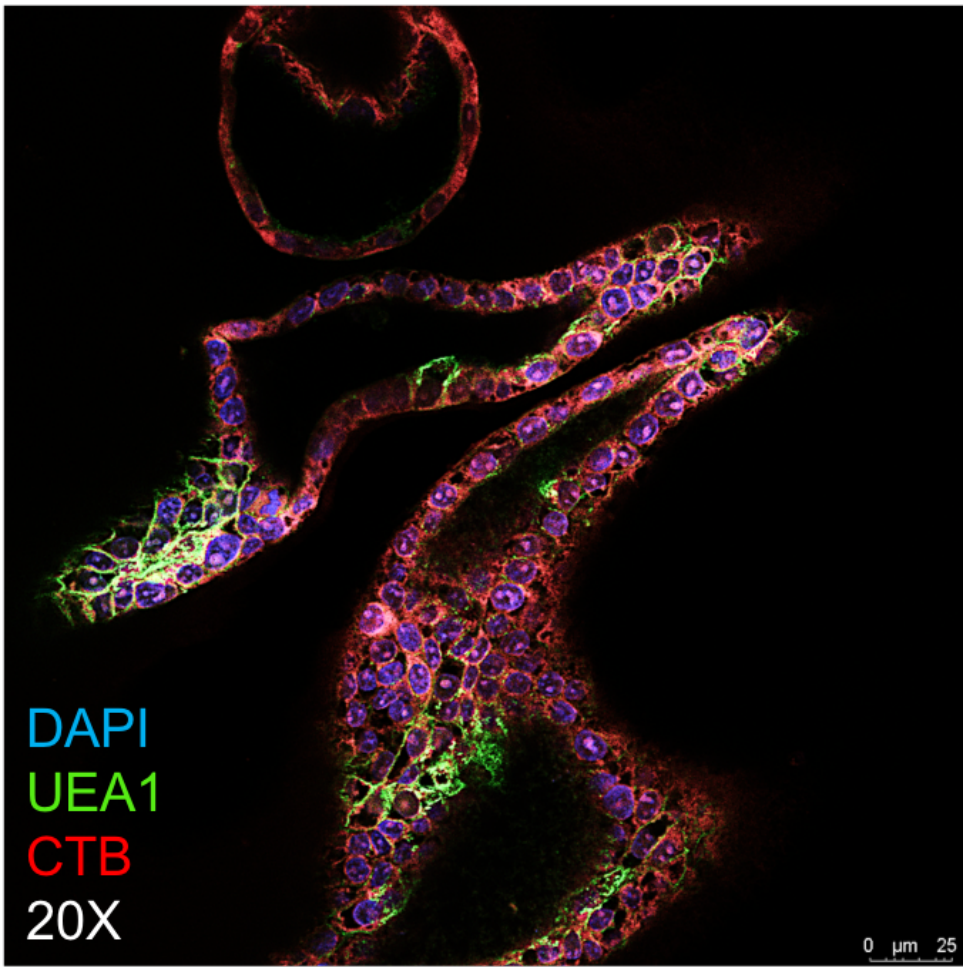


Figure 3.10. A confocal microscopy image of a Kolf2 small intestinal organoid stained with UEA1 to illustrate fucose residues (as previously used over the course of this project) Directly conjugated CTB AlexaFlour647 shown in red and DAPI to stain nuclei shown in blue. Although a slight co-staining for UEA1 and CTB can be seen, it is not explicit.

Figure 3.10 shows some overlap and co-staining in CTB UEA1 staining, indicating that although CTB might bind to fucose residues, it may not be the primary binding site.

3.5 Uncovering different morphologies in the organoids

During this project, as the cells were grown and passaged in culture over time, differences in the shape and morphology of the iHOs became apparent. IHOs that have only been passaged for a few weeks (~7 weeks) in culture appear to be very round and cystic, whereas iHOs that have been cultured (~32 weeks) acquire a budded and a seemingly different growth state as shown in Figure 3.11. I have investigated if these two maturation states hold different proportions of cell types and whether they yield different phenotypic responses to cholera toxin (this will be discussed in Chapter 4). Unfortunately, due to the mixed culture of the iHOs and the varying number of iHOs within the culture wells, it was not possible to calculate the proportion of budded and cystic iHOs at every stage of growth. It is important to note that during all stages there will be a proportion of mixed morphologies, however the bias is toward cystic at younger stages and budded as the organoids are passaged over time and the culture ages.

In addition, it was noticed that when iHOs were split more vigorously (disaggregated into very small clumps or by pipetting a greater number of times) the re-forming iHOs took on a more immature state, (similar to that of passage 7 in Figure 3.11 whereas when budded iHOs were split more gently, they continued to form budded iHOs after passage (similar to those seen in Figure 3.11 passage 32).

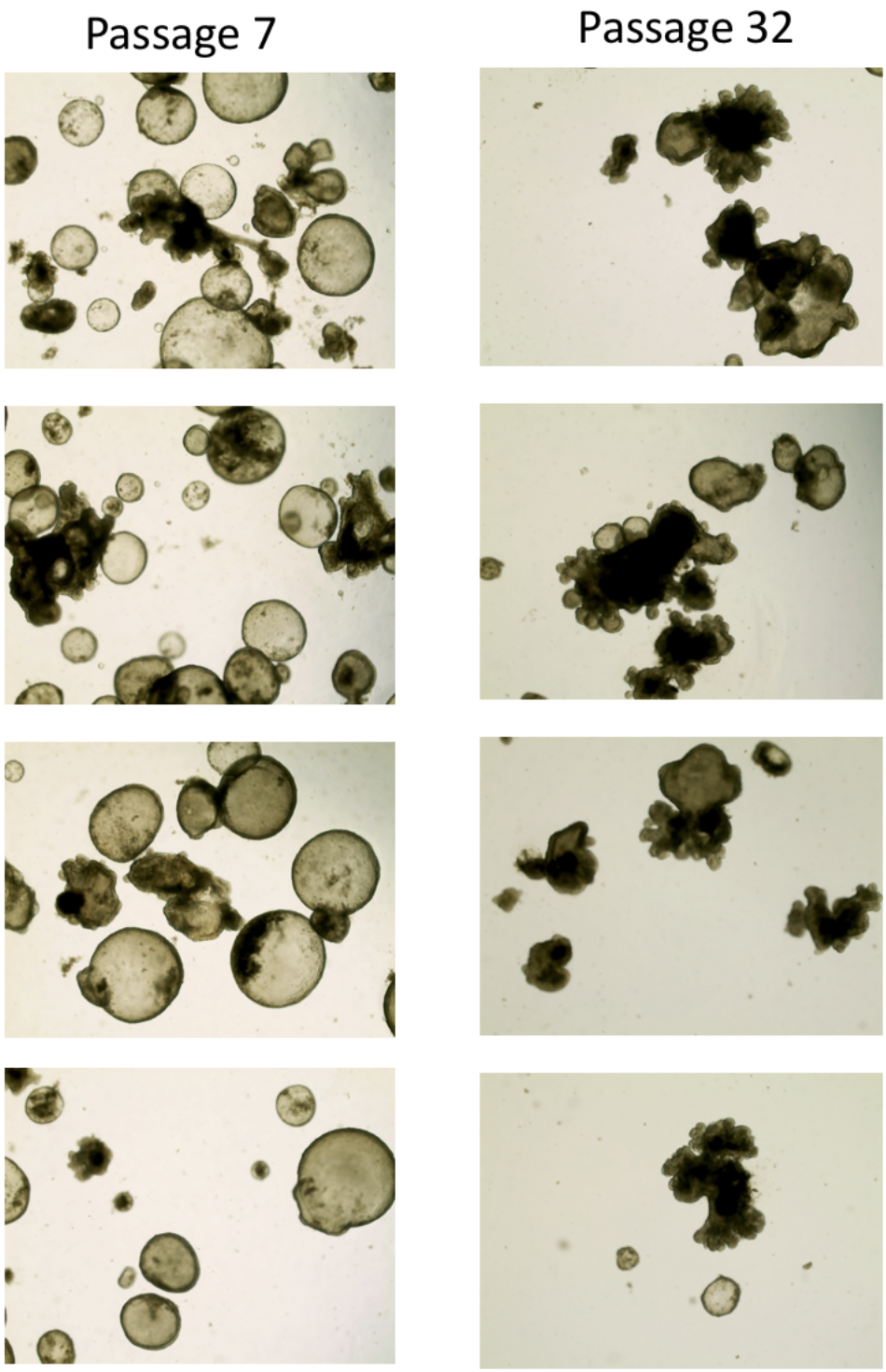


Figure 3.11. Light microscopy image to show the different morphologies witnessed over the course of the project associated with increasing passage number. Cystic or rounded organoids are common to lower passage numbers (see passage 7) whereas more budded and structured organoids are a feature present within older organoids (see passage 32).

3.5.1 Transcriptome analysis of budded vs cystic organoids

IHOs at passage 18 show both morphologies within the same culture. I wanted to investigate even at the same passage number (and therefore age of organoid), if the morphology of the organoid had any influence on gene expression. Organoids were manually extracted from the Matrigel dome and placed into RLT+ and processed for RNA seq analysis (method in chapter 2 section 2.6.2). Samples were processed for sequencing by the Research and Development team at Wellcome Sanger Institute and the resultant data QC'ed and formatted by Oliver Lorenz from the Pathogen Informatics team at Wellcome Sanger Institute. The transcriptome data was filtered per gene, with all genes showing greater than one log fold change and below one negative fold change with a p- value of less than 0.05. The full lists of differential genes can be found in Appendix A-Z.

Transcriptomic analysis comparing naïve non injected iHOs of the budded morphology compared with iHOs from the cystic morphology from the same passage (p18) shows that there are in fact some transcriptomic differences between the two morphologies in absence of CT (full gene lists in Appendix A-F) In total there were 116 genes significantly differentially expressed (following the filtering parameters described above), 102 genes were upregulated in budded iHOs compared with cystic and 14 genes were downregulated. Of the 102 genes that were upregulated, one of the genes with the highest fold change (8.1) was *SLC26A3* which is a chloride anion exchanger which is localised to the apical membrane of the columnar epithelium and on some goblet cells- it is instrumental in chloride re-uptake, helping to create the osmotic gradient for resorption of fluid from the lumen of the intestine [192]. This is of particular interest when considering the dynamics of CT and *V. cholerae* infection. The fact that this gene is up regulated in the budded morphology compared with the cystic morphology raises the question if this responsible for the lack of swelling witnessed in the iHOs of the cystic morphology discussed previously in this chapter. Although it is just speculation, future work could include creating a CRISPR knockout in the human cell line Kolf-2 in the *SLC26A3* gene to see if this does in fact impact the morphology or recover the swelling phenotype seen in the cystic iHOs. In addition to this, there are many genes involved in transport of small molecules upregulated in the budded morphology compared with the

cystic morphology- these include lipocalins and fatty acid binding proteins. Of the 14 downregulated genes in the budded compared with the cystic naïve iHOs, many are involved in embryonic development. Considering the upregulated genes in addition to the downregulated genes, it suggests that the budded iHOs are more mature in nature than their cystic counterparts. Of course, this should be considered for any future work, interestingly nothing has been documented in the literature highlighting or investigating the difference between the two morphologies.

From this data, I can conclude that the differences within the morphologies of the organoids in this mixed population are somewhat due to changes in gene expression that can be measured at a transcriptomic level. Therefore, I also propose to investigate the proteome and metabolome of the different morphologies of organoids in the future to uncover if there are any differences in the metabolic activity or protein expression of the two morphological states.

3.6 Discussion

The aim of this chapter of the thesis was to develop the iHO model for infection. The data produced here corresponds with previously accepted data in that the GM1 receptor is sparsely distributed within both the human GI tract [87, 176, 177] which is reflected within our small intestinal organoids (and can be seen using confocal imaging). We also raise the possibility of an additional cholera toxin binding receptor, which over the course of this project, an expanding number of research groups are also discovering.

We have observed and gained an insight into different maturation states of the iHOs over time and passage number, which to our knowledge is an unexplored area of the iPSC derived small intestinal iHO system. Transcriptomically, we do see a differential gene expression difference among these two morphological states (at the same passage number) which has been discussed in this chapter- within naïve organoids. The effect of CT on the transcriptome of the iHOs will be investigated in the next chapter of this thesis.

In this chapter CTB binding of the organoids is explored, the next chapter will investigate the effect of the whole toxin, the enzymatic A subunit in addition to the binding B subunit on the organoids.

Chapter 4:

Cholera Toxin and the Kolf-2 human intestinal organoids

Aim:

To explore the effects of cholera toxin on Kolf2 human intestinal organoids.

Chapter 3 explained how the Kolf2 human intestinal organoid (iHO) infection model was established and validated. In this chapter we will explore, visualise and quantify how Cholera Toxin B subunit (CTB; see chapter 1 section 1.6.1) binds to the small intestinal iHOs and how iHOs respond to the complete cholera holotoxin which comprises of the A and the B subunits (chapter 1 section 1.6.1). Here, I used techniques such as confocal and electron microscopy and measured cytokine production of both cholera toxin (CT) treated and naïve iHOs to better understand the effect of CT on the epithelium of small intestinal iHOs. Based on work presented in chapter 1 section 1.6.2, we know that GM1, the primary receptor for CT is only expressed on the apical surface of iHOs. We exposed both the apical and basolateral surface of the iHOs to CT, via surface administration or micro injection directly into the lumen of the iHO, in order to look for measurable differences in toxin binding and response.

4 Toxin Assays (whole toxin)

To determine whether it was possible to elicit a toxin-specific response by exposing the basolateral surface of the iHOs to CT, toxin was simply added to the standard growth medium. This practice has been used by others in the field on mouse primary and human derived iHOs to study the effects of different drug treatments. In relation to cholera, previous studies by Beekman and Clevers et al*.* have used this approach on mouse iHOs, that harbour a genetic mutation in the cystic fibrosis transmembrane receptor (CFTR), and human iHOs derived from biopsies of patients with cystic fibrosis, to look at forskolin and CT induced swelling. They have shown that treatment of iHOs with forskolin or CT leads to a rapid increase their cAMP (cyclic adenosine monophosphate) levels, which in turn results in the opening of the CFTR channel [193]. Interestingly, this study shows while both forskolin and CT induce CFTR mRNA expression and ultimately a swelling response within these primary derived iHOs, the timing of the response differs, with forskolin inducing an almost immediate response increasing to 150 % normalised area in less than 30 minutes whilst CT treated iHOs started to swell (increase in area) after 60 minutes, showing a 150 % increase compared to the normalised area in approximately 90 minutes post treatment [193]. Noting this difference, both treatment conditions reach the same point (250 % of the normalised area) after 120 minutes post treatment. So, although CT is slower at initiating an observable swelling phenotype, the size changes of the iHOs plateau with both treatments after 120 minutes.

In this chapter, I also explore the effects of CT over time to gain a longitudinal view of the response within our iPSC derived small intestinal iHO model. It is important to note that I use basolateral stimulation of CT in some experiments (mainly swelling responses) and microinjection of CT to elicit an apical response in others. I will make it clear which mode of administration took place when discussing the experimental design.

4.1 Whole toxin (CT) administration to the basolateral surface of Kolf2 iHOs

4.1.1 Response to CT (swelling)

It is important to note that the morphology in our cultured iHO populations changed with increasing passage number and therefore age (see chapter 3 section 3.5). the iHOs become biased towards iHOs that are increasingly budded and invaginated compared to the other observed morphology, cystic and round, which dominate in lower passage numbers. This experiment was conducted on early passage iHOs (between passage 8-15 and repeated at least 10 times over the course of this project with three replicate wells of iHOs at each experiment) with a bias in the population towards a cystic morphology (Figure 4.1), however there were a small proportion of budded iHOs present within the organoid culture.

Since I showed the known receptor for CT binding, GM1, is only expressed on the apical surface of the iHOs I added commercially available CT from Sigma (Cat. No. 227036 Inaba 569B) to the base growth media to expose the basolateral surface of the iHOs to CT. Base growth media alone was used as a negative control (see Figure 4.1 for a description of the experimental design). From Figure 4.2, it appears that the 'swelling response' of the iHOs to CT treatment is dependent on the concentration of CT applied, with swelling visually observed at all CT concentrations. Given the clear lack of GM1 on the apical surface of our iHOs, the swelling response in the presence of varying concentrations of purified commercial CT was perhaps unexpected. However, on broader consideration of the published literature this had been previously reported by Dekkers et al., using primary iHOs [194].

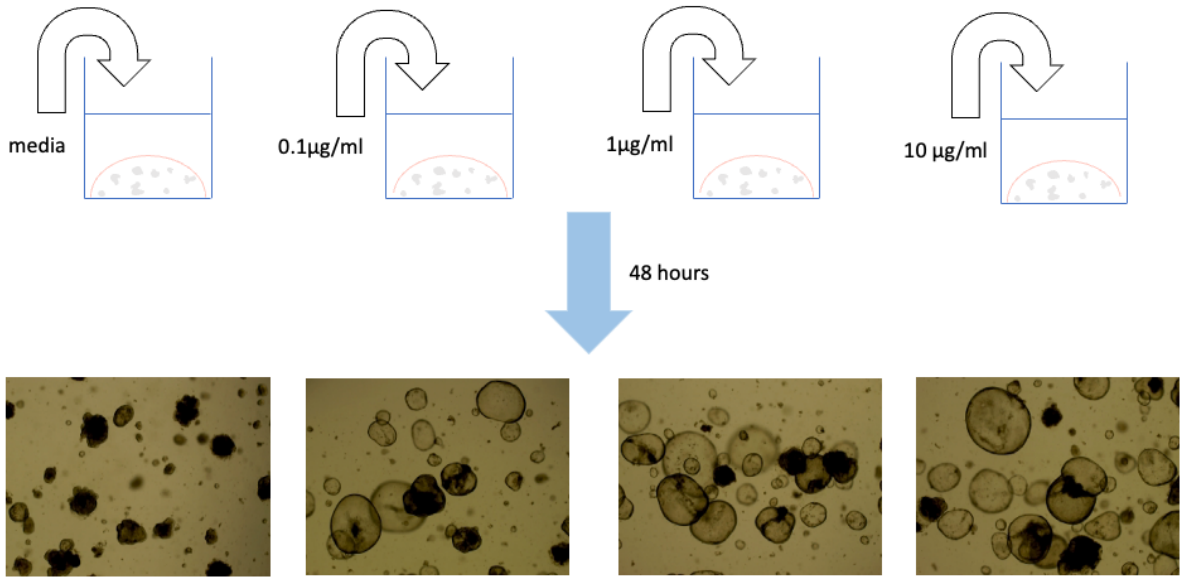


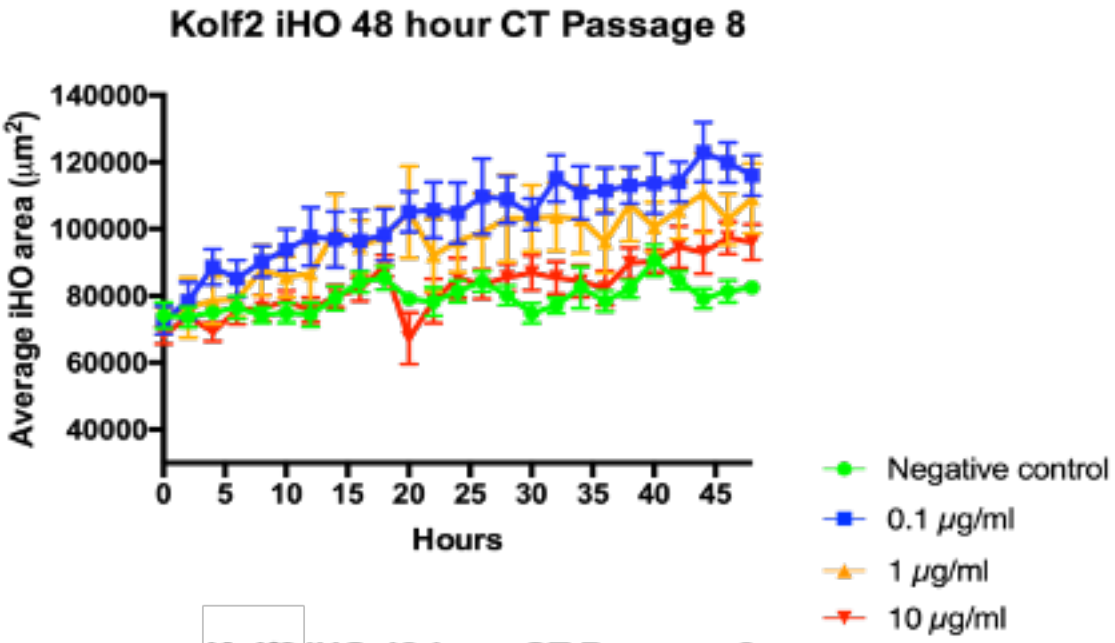
Figure 4.1. Experimental design for the exposure of purified CT to the basolateral surface of the iHOs. Varying concentrations (0.1µg/ml, 1µg/ml and 10µg/ml) of purified toxin were added to the media of the iHOs, in addition to a negative control of media alone, incubated for 48hours and visualised using EVOS XL Core.

Since basolateral administration of CT elicited a swelling response and it allows for all the iHOs in the culture dish to be exposed to the toxin it makes confocal imaging more efficient, and it is also less technically challenging than microinjecting individual iHOs. In addition, the basolateral stimulation has been used by many other research groups (as highlighted above). However, in section 4.2 I will consider microinjection as a mode of administration for CT. Until that point, all experiments have been conducted using basolateral administration of CT. All experiments below, for consistency, were performed using the experimental design illustrated in Figure 4.1 and described in detail in Sections 4.1.2 and 4.1.3 using 1 µg/ml of CT in the challenge and media only as a negative control.

To refine these qualitative observations, the experiment was repeated with the average iHO volume recorded over time using the IncuCyte Zoom with and without CT present. The IncuCyte Zoom is an automated imager which collects images at timepoints set by the user and includes downstream analysis software that enables calculations to made based upon

iHO size etc. Here, three individual wells of iHOs taken from passage 8 were exposed to between 0.1 and 10 µg of externally applied CT. The change in size of iHOs over 48-hours is shown in Figure 4.2 - using different concentrations of CT and a negative media only control (green line). From Figure 4.2 it is clear that all concentrations of CT induce an increase in area compared to the negative control (unpaired t-test comparing negative control to 0.1µg/ml of CT p-value <0.001).

a)



b)

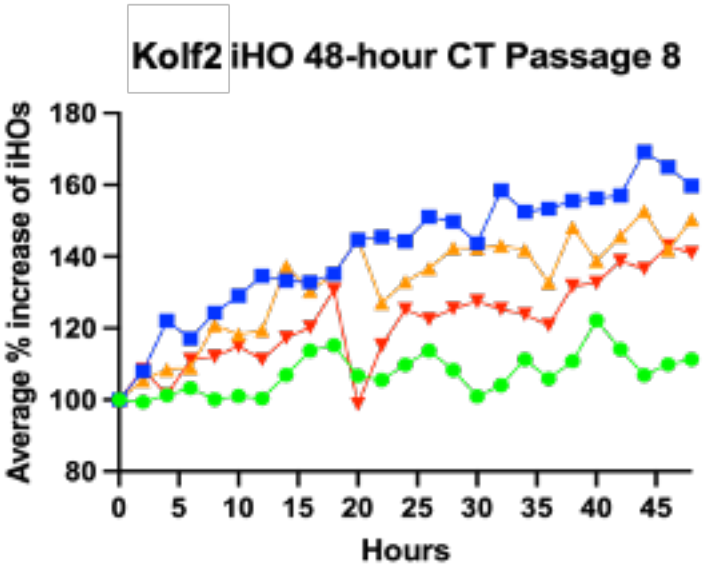


Figure 4.2. a) A Line graph showing an increase of average iHO area over time in response to selected concentrations of purified CT. Whole CT was added to the media of the iHOs (as described in Figure 1), the average area of the iHOs was measured using the IncuCyte ZOOM. Error bars are shown, displaying the standard error of the mean (SEM) due to measuring most of the iHOs within a well to allow for variation in iHO size and morphology. Line graph b) shows the % increase of the mean area of the iHOs over the 48-hour time frame of the experiment. Graphs were drawn using GraphPad Prism 9.

Interestingly from Figure 4.2 we can see an inverse relationship between swelling and CT concentration, with iHOs treated with more dilute CT (0.1 µg/ml) appearing to increase in size more than the iHOs that were treated with the higher concentration of CT (10 µg/ml). These experiments were repeated at least five times (using the same experimental design as Figure 4.1, with iHOs at different passages (between passage 8 and passage 12) with very similar results being reported at these earlier passage numbers. As noted above these iHOs were mostly of the cystic morphology.

Next, I wanted to understand if the differing iHO morphology had an impact on the 'swelling response' that we see in the younger-cystic iHOs. So, varying concentrations of CT were added to the basolateral surface of three individual wells of iHOs grown at passage 33 (see Figure 4.1 for experimental design) cells, see Figure 4.3 for the results.

a)

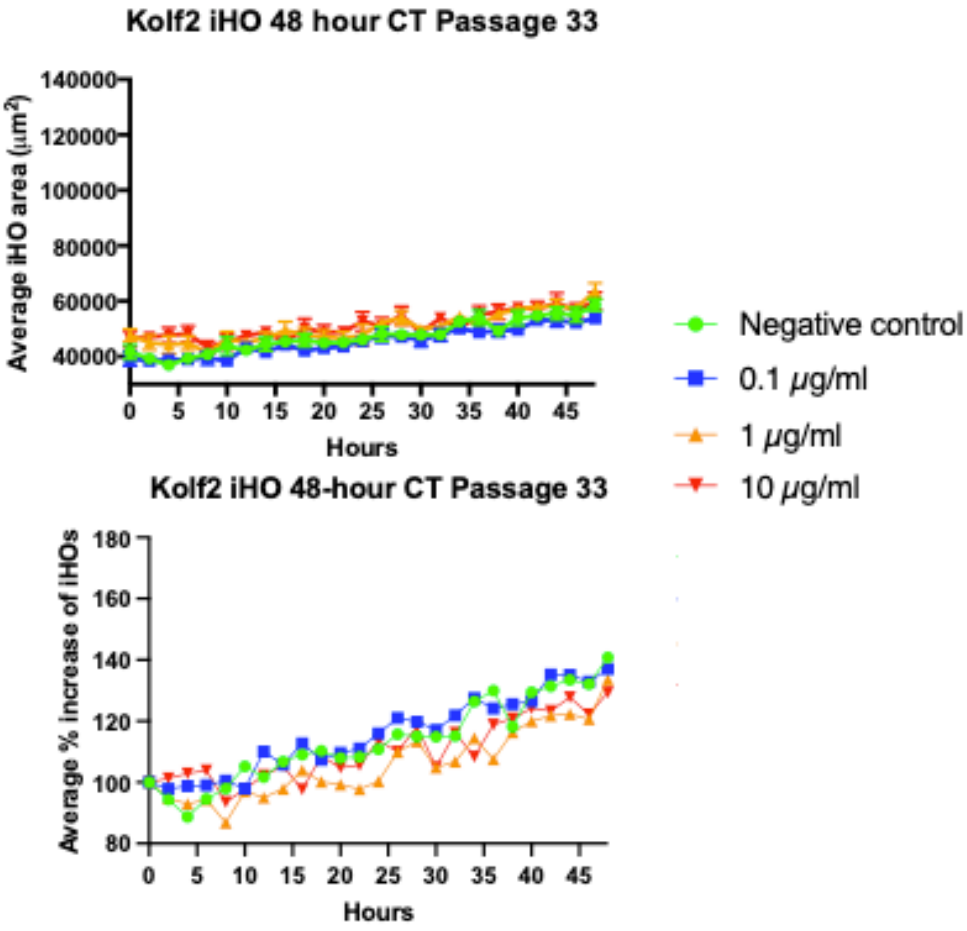


Figure 4.3. a) A line graph showing the average area of iHOs at passage 33 in response to varying concentrations of toxin compared with a negative control (green line). The varying concentrations of toxin were administered via the media therefore targeting the basolateral surface of the iHO. the average area of the iHOs was measured using the IncuCyte ZOOM. Error bars are shown, displaying the standard error of the mean (SEM) due to measuring most of the iHOs within a well to allow for variation in iHO size and morphology. Line graph b) shows the % increase of the mean area of the iHOs over the 48-hour time frame of the experiment. Graphs were drawn using GraphPad Prism 9.

Figure 4.3 reveals a markedly different response for passage 33 iHOs compared to the earlier passage (passage 8 as seen in Figure 4.2) iHOs shown in Figure 4.2. The major difference between these two experiments was the passage number and therefore the age of the iHOs and in turn the relative proportions of the two different iHO morphologies in the sample - with the younger passage having a higher proportion of cystic iHOs in comparison with passage 33 samples which have a higher proportion of budded iHOs (as shown in chapter 3 Figure 3.11). From Figure 4.3 that at passage 33, the iHOs mirror the control sample for

growth (shown by the green line and negative media only control) and show no statistically different response (using an unpaired two-tailed t-test comparing negative control to 0.1μg/ml of CT p-value 0.5386) and no swelling.

4.1.2 Understanding the impact of CT on iHOs: Toxin linked iHO pathology at the tight junction.

Whilst from the sections above it is clear that CT administered to the basolateral surface induces swelling in the iHOs. In order to gain a finer detailed view of the CT response, I first used transmission electron microscopy to visualise the iHOs at a cellular level. In doing so I wanted to understand whether the integrity of the cells or cell structures within the iHOs had been compromised following treatment of CT, for example if the cellular tight junctions showed any pathology, if the microvilli had been cleaved, or if the cells appear apoptotic. Transmission electron microscopy was performed by David Goulding at the Wellcome Sanger Institute.

Figure 4.4 shows electron microscopy images of naïve iHO compared with a basolateral CT treated iHOs. It can be seen at this resolution, that CT does indeed have an impact on the ultrastructure of the iHOs. The tight junctions of naïve iHOs seen in Figure 4.4a show the characteristic 'pinching' between the junctional complexes appearing 'tight' and near the adjoining cells. In comparison the 'pinching' between the tight junction complexes appears to be impaired and not as 'tight' or in as close proximity in the CT treated iHOs (Figure 4.4b). Simila images of these cells were taken at least five times for each condition.

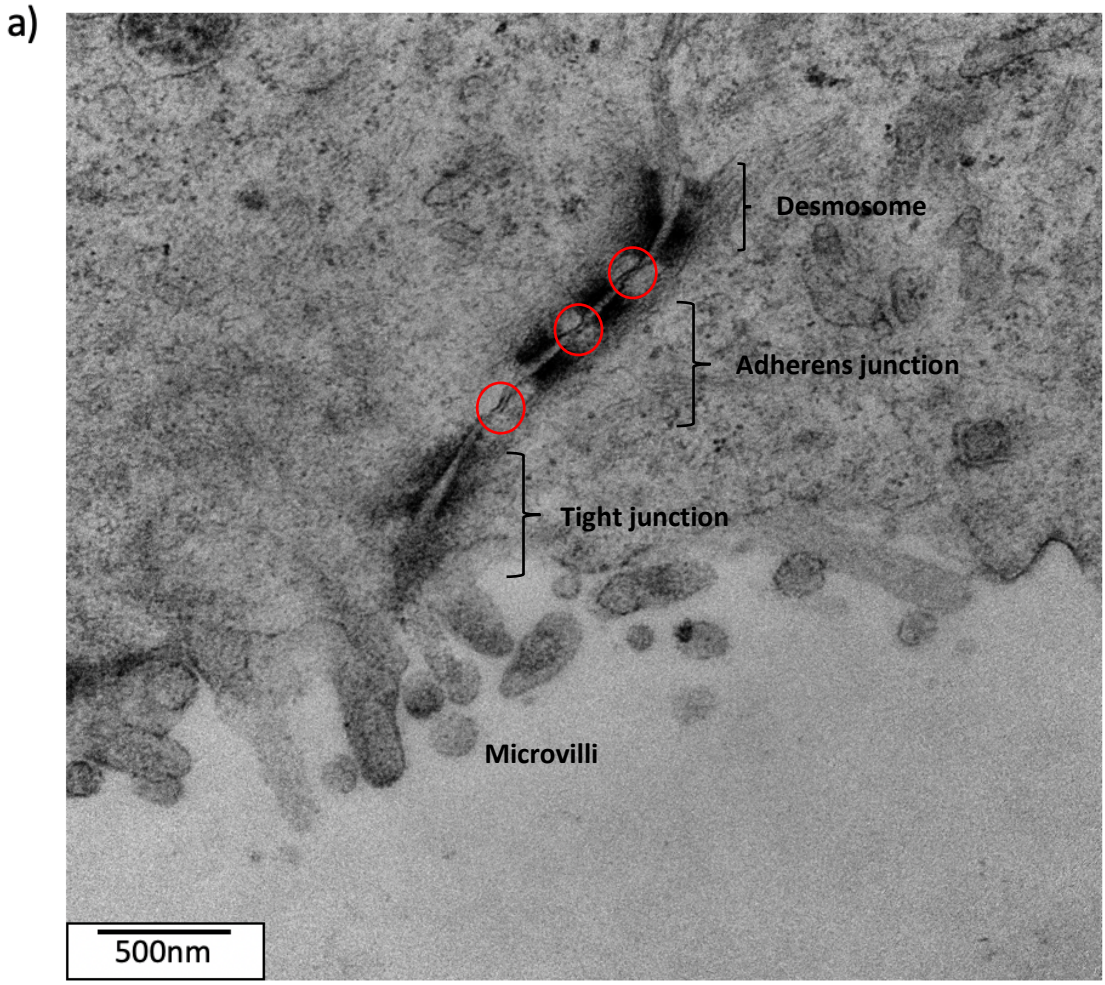


Figure 4.4a. Transmission electron microscopy images showing tight junction structures of a healthy naïve iHO. The 'pinches' (highlighted by the red circles) between the dark regions of the tight junction are very close in proximity to one another.

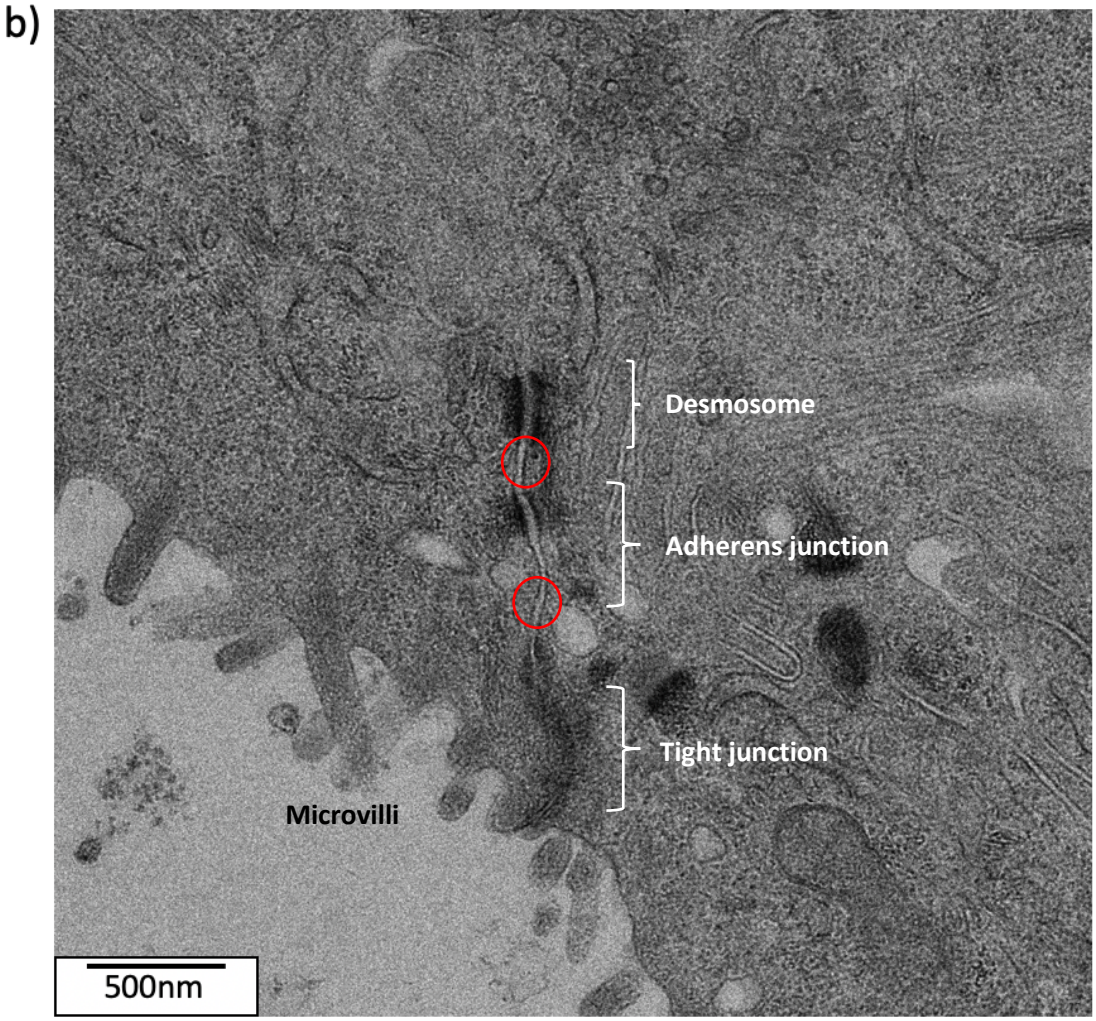


Figure 4.4b. Transmission electron microscopy images showing tight junction structures of an iHO that had CT administered to the basolateral surface for 48 hours. The 'pinches' (highlighted by the red circles) between the dark tight junctional strictures are not as close in proximity compared with the naïve control iHO. Suggesting a breakdown of the tight junctional complex.

4.1.3 Understanding the impact of CT on iHOs: degranulation of goblet cells.

Flakes of mucus are clearly seen in rice water stool produced from cholera patients. It has been reported for many years that CT is able to induce the hypersecretion of mucus. In 1988, using rabbit ileum, Leitch et al., showed this hypersecretion of mucus in response to CT occurred between 3-5 hours [195]. Later, in 1997, Fromm et al. explored this mucus secretion in the human colonic goblet cell line HT-29/B6 and reported the increase of mucus secretion for more than 20 hours [196]. This study also quantifies the intracellular mucin before (86%) and after (13%) CT treatment, with the extracellular mucin increased correspondingly from 14% to 87% total mucin after CT treatment [196].

It is the goblet cells in the intestine that contain the granules of mucus (see Chapter 3 section 3.1). Mucus is continually secreted by healthy cells of the intestinal epithelia into the intestinal lumen to protect and lubricate the intestine. During storage within the goblet cell granules, MUC2 is assembled on a ring-like platform which is formed by calcium dependent interactions between MUC2 N-termini [197]. Once the mucus has been released into the lumen of the intestine, the calcium is chelated and the mucin unfolds, expanding by > 1,000-fold in volume [198].

To look at the impact of CT on goblet cells and mucus secretion 1 µg/ml of CT was added to the media and therefore the basolateral surface of the iHOs. The CT treated iHOs were stained after 24 hours of exposure to CT for confocal imaging with *Ulex europaeus* agglutinin 1 (UEA1; for method see chapter 2 section 2.3.1). UEA1 is a lectin that binds specifically to many glycoproteins and glycolipids containing α-linked fucose residues, such as ABO blood group glycoconjugates. This lectin preferentially binds blood group O cells and has been used to determine secretor status of individuals. It has been established to be an excellent marker for human endothelial cells. Many other research groups have used UEA1 to stain mucus [199-203]. UEA1 has been used previously to stain α- linked fucose residues and ultimately staining for

mucus and goblet cells in addition to DAPI to stain the nuclei. In the negative control PBS was used in place of CT.

In Figure 4.5a, the confocal imaging for iHOs stained with UEA1 before and after CT treatment shows that in the naïve iHO the mucus appears to be both punctate, as expected by the granules held within goblet cells, and are diffuse, explained by mucus layer lining the lumen of the iHO. In comparison, Figure 4.5b shows that after CT treatment the mucus layer is now more dispersed with fewer concentrated spots of staining within cells indicating the degranulation of goblet cells and release of mucus into the lumen of the iHO.

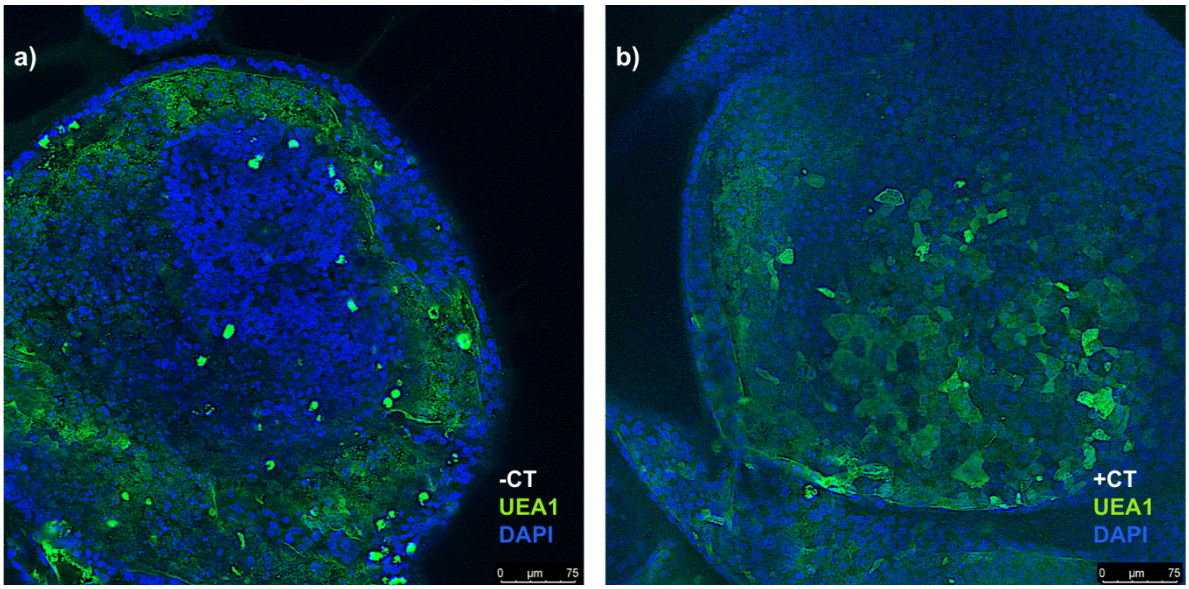


Figure 4.5. Immuno-staining of Kolf2 iHOs stained with Ulex europaeus agglutinin 1 (UEA1) lectin-which has been used by many research groups and is an accepted marker for staining mucus. Nuclei are stained with (DAPI) dilactate. (a) illustrates staining of a naïve iHO with concentrated UEA1 staining in comparison to (b) showing when CT is administered to the iHOs on the basolateral surface.

These results show that iHO model mirrors the findings seen using rabbit ileum and goblet cell lines, as well as the data from human cholera patients in terms of mucus secretion and degranulation of goblet cells. I show that treatment with CT does subsequently result in increased staining for UEA1 and likely an increase in mucus production within our model.

4.1.4 Understanding the impact of CT on iHOs: Host response to toxin.

To understand the host response to CT, I chose to look at the Interleukin-8 (IL-8) response because it had been previously shown to be an early marker of bacterial infection [204]. IL-8 is a proinflammatory cytokine that plays a key role in initiating the host cell defence mechanism towards bacterial pathogens acting by recruiting and activating neutrophils and T cells [205]. It has been shown that infection with *V. cholerae* can induce IL-8 in T84 cells [206]. T84 cells have been derived from a human colon carcinoma, they are able to grow in monolayers [207] and have been used for multiple bacterial infection studies including *V. cholerae* [206] and *S.* Typhimurium [208]. In addition, although cholera is traditionally considered to be a non-inflammatory disease, vaccine studies have shown that live attenuated *V. cholerae* strains can cause reactogenicity among volunteers and have raised levels of IL-8 [209], with colonised volunteers having increased levels of IL-8 in their stool [210]. Although we are focussed on the response of the host cells to CT and not the whole bacteria at this point, I thought it would be interesting to examine whether CT would illicit a similar immune response to that seen in human volunteers with whole bacteria within our iHOs. To do this (for consistency) the same experimental design detailed above was followed (see Figure 4.1). A media only control was used to measure the basal expression of cytokines by iHOs. Different concentrations of CT were added to the media to stimulate the basolateral surface and the supernatants were removed, filtered and then IL-8 concentrations in the supernatant were assayed on a Luminex Flex-Map 3D (for method see chapter 2 section 2.6.1). This was repeated using three biological replicates for each concentration, in which three individual wells containing many iHOs were used for the cytokine results at each timepoint with each concentration. From Figure 4.6 it is clear that there is an observable dose dependent increase in IL-8 production linked to toxin concentration.

Vaccine studies in human volunteers have shown that IL-8 response was dependant on inoculum size-demonstrated with dose-dependency [209]. Similarly, Figure 4.6 shows the IL-8 production at 48-hours post treatment by the iHOs increases in a dose dependant manner, albeit to purified toxin as opposed to whole bacteria. There is a significant difference between

the administration of media only (negative control) and the 10 μg/ml of toxin (p value of 0.0252 in an unpaired t test).

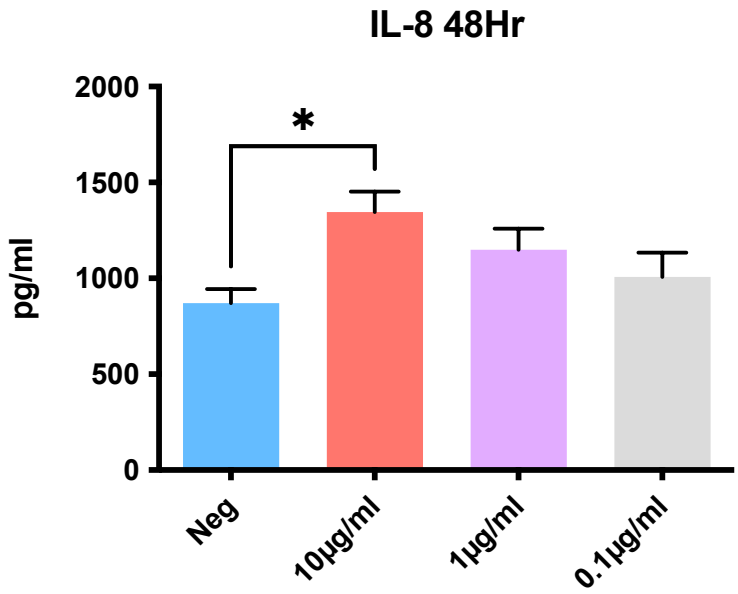


Figure 4.6. Interleukin 8 production measured from the iHO supernatant by Luminex at 48 hours post toxin exposure. Three biological replicates were used, the graph shows the mean with SEM plotted at each toxin concentration.

It should be noted that whilst the results of the above experiment are robust, I repeated this work using an alternative Human Cytokine Assay Kit that simultaneously measured the concentration of 48 different human cytokines (Bio-Rad Bio-Plex Pro™; Method in Chapter 2 section 2.6.1) with a wider range of CT concentrations (10 μg/ml, 1 μg/ml, 0.1 μg/ml, 0.01 μg/ml and 0.001 μg/ml) sampled at a wider range of timepoints (4, 24 and 48 hours post CT administration) shown in Figure 4.7. This experiment showed high variability and sometimes conflicting results compared to the single cytokine assay described above (Figure 4.6) The difference in levels of IL-8 produced form the cells should also be highlighted with the pg/ml in the smaller kit ranging from 800 pg/ml to 1548 pg/ml at 48 hours post toxin administration, whereas in the larger 48 plex Bio-Rad Bio-Plex Pro™ the IL-8 levels range from 3761 pg/ml to 11900 pg/ml at 48 hours post toxin administration. Due to interruptions in lab operations, linked to the UK Government Covid19 restrictions on working, I was not able to repeat these experiments. Using the larger 48-plex Bio-Rad Bio-Plex Pro™ kit, I did witness a significant difference only within one parameter (4hr media control vs 4hr 10 μg/ml p value of 0.0182 in

an unpaired t-test). However, there was no significant differences when conducting unpaired t-tests among any of the other parameters (including the media only control at 48 hours and the 10 μg/ml as seen in the smaller kit in Figure 4.6).

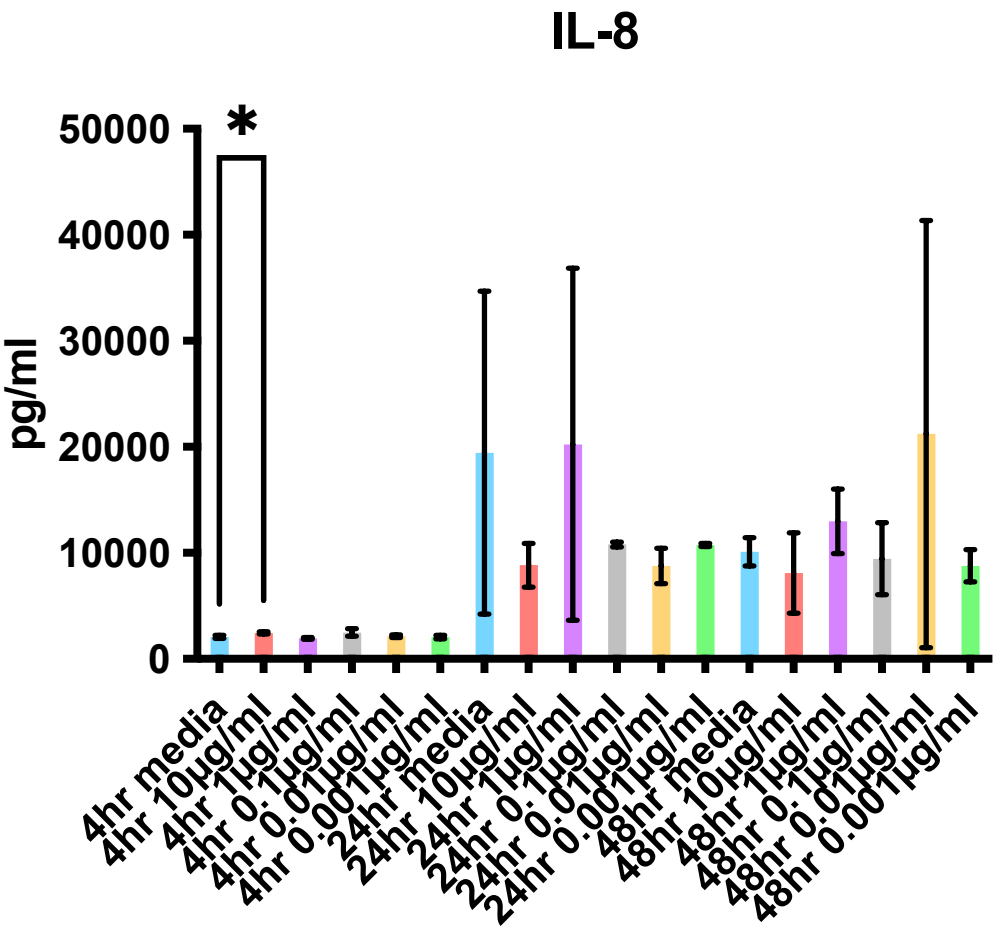


Figure 4.7 Interleukin 8 production measured from the iHO supernatants sampled at 4, 24 and 48 hours using a wider range of CT concentrations (10μg/ml, 1μg/ml, 0.1μg/ml, 0.01μg/ml and 0.001μg/ml) by Luminex using the Bio-Rad Bio-Plex Pro ™ 48 plex kit. Three biological replicates were used, and the graph shows the mean with SEM plotted at each timepoint with each toxin concentration.

This inconsistency between experimental approaches could be explained by a few factors, first systematic and technical challenges/variation from using the 48-cytokine assay whereby the 48-plex nature of the assay in might have an impact on the sensitivity of the kit, noting that the levels of IL-8 among the negative control (CT -ve naïve iHO) samples were highly variable. In addition, if repeated I would suggest mechanically disrupting the iHOs to release the intracellular cytokine levels and measuring the total cytokine levels the secreted levels present in the media. A study published in 2017 by Noel et al. showed that (in primary

organoids) IL-8 and IFN-γ were produced by the basolateral side of the organoid but released by the apical surface [211], which as we were measuring the cytokine levels in the supernatant would not be recorded. Disrupting the organoids in order to gain access to the apical surface would likely improve consistency between experiments. Ideally, the experiment would be repeated using both kits, the same supernatants from the same cells - with the same batch of CT, in order to make any comparisons between the two experiments. As stated above unfortunately, this was not possible.

4.2 Effect of microinjection of CT and differences within morphologies of iHOs

To determine the effect of toxin on the transcriptome of the iHOs, I wanted to compare the patterns of gene expression among naïve cells. I chose to microinject the iHOs for this experiment to access the apical surface of the iHO which would be the more 'natural' route of infection by *V. cholerae* and exposure to CT. This also allowed iHOs from the same passage microtiter plate to act as controls with the aim of improving consistency. To do this I microinjected toxin into the lumen, this would also allow for the injected iHOs to be identified and isolated from the culture dishes alongside the non-injected controls used for this experiment (summarised in Figure 4.7). Here I compared the response to CT between the budded and cystic iHOs, using trypan blue-only injected iHO as a negative injection control. The passage 18 iHOs were used to ensure that both cystic and budded iHO morphologies were present in the microtiter plate but the age variable would be removed. Injected iHOs were harvested after 4-hours, 24-hours and 48-hours. At this point the iHOs were extracted from the Matrigel matrix and individually placed into RLT+ buffer and the whole RNA extracted (Method 2.6.2). Samples were processed using the Research and Development pipeline at Wellcome Sanger Institute and the resultant data quality controlled and formatted and the RNAseq pipeline was run by Oliver Lorenz from the Pathogen Informatics team (for the method see chapter 2 section 2.6.2.). The transcriptome data was filtered per gene, with all genes showing greater than one log-fold change and below one negative log-fold change with a p-value < 0.05. The full lists of differential genes can be found Appendix A-Z. I interpreted the differential gene expression and ran subsequent pathway analysis and conducted literature searches relating to genes of interest.

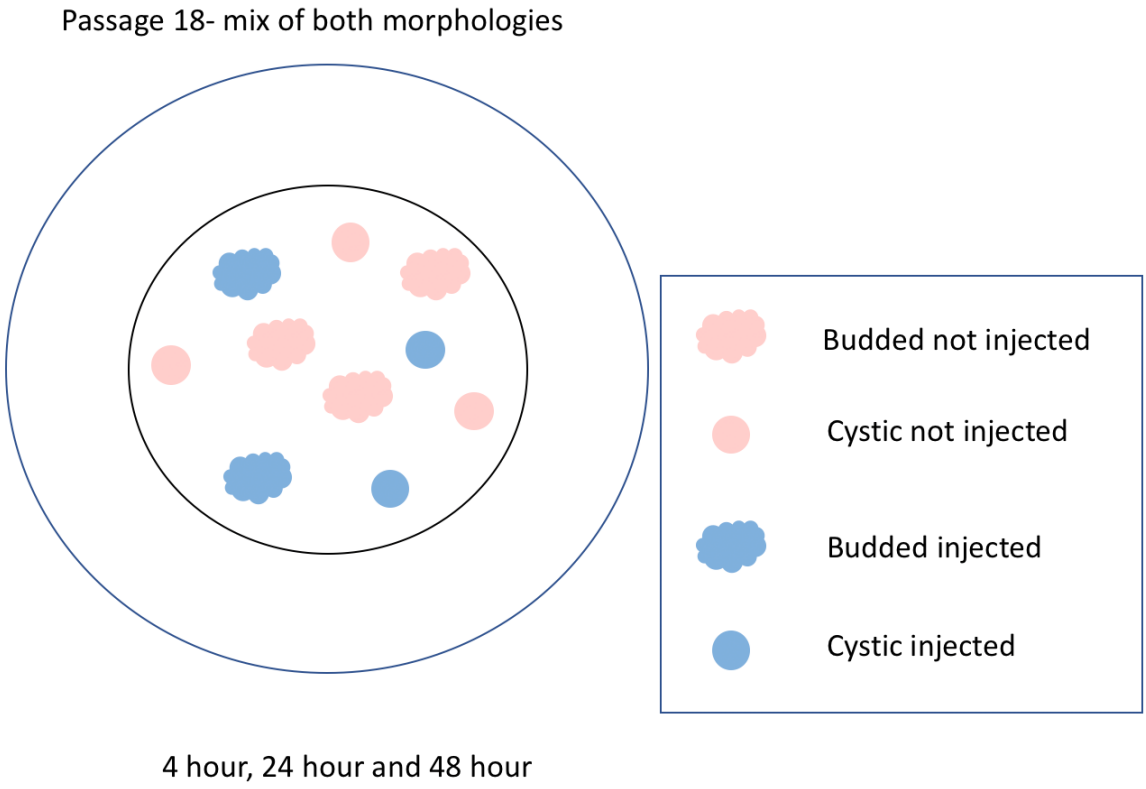


Figure 4.7. A schematic showing the experimental design used to explore differential gene expression within morphologically budded and cystic iHOs. The key indicates the morphologies and the microinjection state of iHOs investigated. Pink indicates not injected whereas blue show microinjected iHOs. Two microinjection dishes were used, one for the CT variable and one for the trypan blue variable. Here both budded and cystic iHOs were injected with CT. Control iHOs were injected trypan blue only. Time points in which iHOs were harvested were 4, 24, and 48 hours post injection with either trypan blue only, a mixture of CT with trypan blue or completely naïve and not injected.

The morphological transcriptomic differences of naïve iHOs have been discussed in chapter 3 section 3.5.1 of this thesis, here I will explore the effects of CT on the transcriptome of the iHOs looking at the morphological differences between the cystic and budded iHOs following treatment with CT.

Here, I compared the CT plus trypan blue microinjected iHOs with trypan blue injected controls (to allow for any stress response linked to the mechanical damage resulting the action of microinjection on the iHOs cells). Reads were mapped to the human reference genome hg38 using nfcore [145], rnaseq pipeline https://github.com/nf-core/rnaseq using "star_salmon" alignment method [146, 147]. Once the above filtering parameters (>1 log fold-change or <1 log fold-change with a p value >0.05) had been set, at the 4-hour timepoint, we

saw 559 genes differentially expressed in the cystic CT injected iHOs compared with the trypan blue only cystic injected iHO controls. 272 genes were upregulated and 287 were downregulated. I mapped the differentially expressed genes to pathways using Reactome [212]. The upregulated genes mapped to rRNA pathways including processing in the nucleus and cytosol and many RNA polymerase pathways (11 out of the top 25 most significant pathways). The downregulated genes mapped to pathways including response to heme deficiency and cytosolic unfolded proteins (the top pathway with a total of 29 genes- 11 of which were identified in our analysis) with one of the top five most down-regulated genes (*CHAC1*) being involved in this pathway.

When looking at the budded iHOs at the 4-hour timepoint (CT plus trypan blue compared with trypan blue alone), 366 genes were differentially expressed with 217 upregulated and 149 downregulated. Pathway analysis for the upregulated genes within the budded iHO population include genes involved in gap junction assembly and trafficking, gap junctions are responsible for the permitting the direct exchange or ions and small molecules between cells. Regulation and rRNA processing are present in 5 of the top 25 pathways for upregulated genes at the 4-hour timepoint. The downregulated pathway analysis for iHOs harvested at 4-hours included genes involved in the TNF pathway linked to inflammation, cell proliferation, survival or cell death [213-215]. Out of the top 25 most downregulated pathways, five of these pathways are involved in cellular responses to stress, this is also witnessed at later timepoints below.

At 24-hours post injection of CT plus trypan blue compared with trypan blue alone, we see 787 genes differentially expressed in the cystic morphology with 390 genes upregulated and 397 downregulated (following filtering parameters outlined above). Reactome pathway analysis of the upregulated genes include pathways involved in cellular responses to heat stress, stimuli and stress (3 of the top 5 pathways identified). Of the 397 down-regulated genes, this analysis revealed many pathways involved in cell cycle at various stages (again 3 of the top five pathways identified) suggesting that at the 24-hour timepoint within the cystic morphology, the iHOs are slowing their cell cycle and potentially using the energy on the stress response pathways instead- according to this experiment. Looking at the budded morphology, there are 289 genes differentially expressed after infection with CT plus trypan

blue compared with trypan blue alone 176 genes that were upregulated and 113 genes which were downregulated. Pathway analysis of the upregulated genes includes 15 of the top 25 pathways involving fibroblast growth factor receptor 2 (*FGFR2*) signalling, further literature searches of *FGFR2* yield studies of skin and cancers with one study showing that *FGFR1* and *FGFR2* are key regulators of keratinocyte migration *in vitro* and in wounded skin- this could indicate that the iHOs are responding similarly in the CT at 24-hours post injection. Pathway analysis of the 113 downregulated genes revealed pathways involved in necroptosis (five genes within the pathway, two of which were identified in our analysis, the p-value placed it second highest in significance). Necroptosis is an important type of programmed cell death which eventually leads to cell lysis, again indicating that the cells are not dying but using their resources to respond to CT.

When comparing the CT plus trypan blue against trypan blue alone at 48-hours post injection, in the cystic morphology I saw eight genes differentially expressed with one upregulated and seven gene downregulated. Unfortunately, it was not possible to conduct pathway analysis on these small numbers of genes, but a literature search shows the only upregulated gene (*ZBTB20* FC 6.36) is a zinc finger associated transcriptional regulator which plays a role in several processes such as neurogenesis, glucose homeostasis and post-natal growth. Of the seven downregulated genes, the gene with the greatest fold change (of -11.54) was *CH507.513H4* which is a novel gene with no known function. The remaining six genes include those that may play a role in transcriptional initiation (*EIF3CL*) and *ST14,* which has been shown to cleave and activate hepatocyte growth factor. However, none of these have any clear link to their role in CT transport or effect or their role in *V. cholerae* infection. There were 131 genes that were differentially expressed in the budded iHOs at the 48-hour timepoint. Of these, 116 genes were upregulated, and pathway analysis indicates involvement in O-linked glycosylation of mucins, in which three of the 116 genes are in this pathway- moreover, one of the genes with the highest fold change is Glucosaminyl (N-Acetyl) Transferase 4 (*GCNT4*). This link to glycosylation of mucus correlates with the degranulation of goblet cells seen phenotypically by confocal microscopy in the iHOs after administration of CT. Although pathway analysis was conducted on the 15 genes that were downregulated, it should not be over interpreted due to the small gene numbers. Three of the 15 downregulated genes are involved in cellular responses to stress- which is consistent with the

cellular response to stress pathways that have been seen at previous timepoints in this experiment. Interestingly, some genes can be seen most highly upregulated at all timepoints- these include *SIK1B* (salt inducible kinase 1B) which unfortunately has no published information relating to the gene function.

Of particular interest to this study was the expression of genes associated with tight junctional integrity, namely the claudin family and occludins to understand if the breakdown of the tight junctions seen by microscopic analysis was reflected in the transcriptome. At 48-hour post CT microinjection, *CLDN2* (claudin 2) had a -5.29-fold change and was in the top 10 most downregulated genes at this timepoint. Claudin 2 plays a key role in energy-efficient ion and water transport in the proximal tubules of the kidneys and in the gut [216]. Claudin 2 is typical of leaky, water-transporting epithelia [217] and induces channel formation within the tight junction and is mostly expressed in the proximal intestine [218]. At 4-hours post CT microinjection we witness a two-fold increase in the expression of *CLDN2* in the cystic iHOs- this is consistent of the swelling response that we can measure, in addition to Claudin 2 being upregulated in both Crohn's disease and ulcerative colitis [218], in which barrier function of the intestine is impaired. At 24-hours post CT microinjection, the expression of *CLDN2* drops to -2.96-fold change. We then see a further decrease of -5.29-fold change at 48-hour post CT microinjection in the budded iHOs with *CLDN2* being in the top 10 most downregulated genes at this timepoint.

4.3 Discussion

The aim of this chapter of the thesis was to explore the effects of whole cholera toxin on the Kolf2 human intestinal organoids. I have shown the release of mucus from goblet cells both using confocal microscopy and qualitatively by bulk RNAseq. I have also explored the impact that CT has on the ultrastructure of the iHOs, showing that CT does in fact lead to a breakdown of tight junctions in iHOs, this was shown via electron microscopy and by bulk RNAseq analysis. This apparent breakdown of the tight junctions upon treatment of CT could in fact be the reason that we see the swelling response of the iHOs after administration of CT. It is interesting to note that CT alone is capable of such disruption to the tight junction complex, particularly when considering the characterised zona occludens toxin (zot) produced by *V. cholerae* which has been shown to increase the mucosal permeability by disrupting the tight junctions [219]. I can only hypothesise at this stage that if CT alone causes such damage to the tight junctions of the organoids, once *V. cholerae* expressing both CT and zot are injected into the lumen of the organoids, this tight junction disruption would be increased. I have also identified the variation in the morphology of the iHOs that develops as a result of ageing and increased passage number of the iHOs, something that to my knowledge has not been recorded previously.

There are still many areas that have not been possible to study over the course of this thesis, for example, whether the number of GM1 receptors on the surface of the iHO become saturated after exposure to CT or whether there are in fact alternative receptors for CT. Comparing the rate of swelling after basolateral and apical administration of CT would also be interesting to explore- in addition to conducting z-stack or 3-dimensional measuring of the increase in iHO size as opposed to the 2-dimensional measurements available to me during this thesis, this would allow for the variation in the organoid morphology in that the whole area of the iHO would be measured.

In the future, I would repeat the RNAseq experiment and use a larger sample size- as opposed to 3 organoids per condition, I would also like to include a comparison between basolateral

administration and apical microinjection in which we would be able to compare swelling rates between the two modes of infection in addition to the comparison of the transcriptomic data. Unfortunately, this was beyond the scope of this PhD. I have touched upon the transcriptomic differences between the two morphologies witnessed in the iHO cultures, future work would also need to take place to characterise these differences in greater detail however I do highlight that there is a difference, at least at the transcriptomic level during this thesis- something that has not been recorded in the literature previously.

The next chapter of this thesis will focus on the bacterial strain selection for this project (chapter 5) and in chapter 6 will then examine the phenotype of these selected bacteria once they have been microinjected into the lumen of the iHO model to compare to the data presented in this chapter using purified commercial CT.

Chapter 5:

Comparative genomics of the *Vibrio cholerae* strains used in this project

Aim:

V. cholerae are a diverse bacterial species, widely dispersed on the planet occupying assorted niches. When investigating novel phenotypes in model systems, it is important to use bacterial strains that are generally representative of that species. Due to the complexity of operating the organoid model as well as the restrictions on working with Schedule 5 organisms it was not be feasible to investigate more than three isolates therefore selection of strains based on genome data was paramount in this study. The aim of this chapter was to understand, from the whole genome phylogeny of *V. cholerae,* if the isolates selected for this study were representative of important nodes within the species diversity to use in the infection studies. For this to be possible, I had to create high quality assemblies to enable a detailed comparison to take place. Here I performed a comprehensive comparative genomic analysis to determine the genetic similarities and differences between three selected isolate genomes. To understand any potential differences in behaviour in the organoid infection model, using several computational approaches I confirmed the presence of known virulence determinants, as well as identified and characterised novel DNA regions of interest. To understand how generalisable these observations were, I determined whether any of the novel DNA regions were specific to the isolate in question, specific to a whole lineage or present in other bacterial isolates distributed across the species phylogeny. The aim of accumulating this data was to determine how robustly we could link any differences in pathology to relative position within the phylogenetic tree.

5 Introduction

Work conducted previously by Domman et al., (2017) looking at isolate genomes from across South and Central America and Dorman et al., (2020) focusing on Argentina, highlighted that the role of the Seventh pandemic El Tor (7PET; see Chapter 1 section 1.2.2) as well as to non-7PET lineages in Latin America during in the 1990s. Using these data *V. cholerae* was divided into epidemic and endemic disease-causing lineages. Epidemic cholera is linked to two lineages both of which are characteristically of the O1-serogroup: classical and more recently 7PET [220] (see Chapter 1 section 1.2.2). Whereas endemic disease was linked to genetic lineages displaying a range of different serogroups including O1, but which do not appear capable of causing such expansive, widespread epidemic disease that is characterised by 7PET. These endemic lineages appear to be more diverse and localised, characteristic of the region, rather than part of globally radiating clonal expansions [69].

As discussed in Chapter 1 section 1.6, the principal virulence factors necessary for epidemic cholera include TCP (toxin coregulated pilus mediating bacterial attachment to epithelial cells), CTX phage, and the *ctxAB* genes, encoding the cholera toxin (CT) that it carries (responsible for production and delivery of the enterotoxin to the epithelial cells). However, whilst almost all *V. cholerae* causing epidemics carry these virulence genes, they are generally lacking from endemic lineages, although these *V. cholerae* lineages carry other genes that have been linked to virulence [83, 84].

Dorman et al. [221] showed that from genomic analysis of the 490 Argentine isolates responsible for the localised endemic disease, although they lack TCP and CTX phage, 21 of the 61 non-7PET isolates carry a Type Three Secretion system (T3SS). *V. cholerae* T3SSs have been well studied in animal infection models and have been shown to be necessary for intestinal colonisation in the absence of TCP and cause fatal diarrhoeal disease in infant rabbits [222]. Not only are T3SSs necessary for colonisation in bacterial isolates lacking TCP, but these isolates also produce a very different disease compared to the archetypal *V. cholerae* 7PET, strain N16961 in the infant rabbit model. *V. cholerae* strains with a T3SS

produce not only a more pronounced diarrhea but with a faster onset than that of *V. cholerae* O1 strains [222].

More broadly, T3SS are major virulence factors that have been identified in multiple Gram-negative enteric pathogens, including *Salmonella, Shigella, Escherichia, Yersinia, Pseudomonas, Vibrio* and *Aeromonas* species [95]. T3SS enable bacteria to inject bacterial proteins, or effectors into other bacteria or host cells, where they can manipulate the target cells function [96]. The first effector protein identified in *Vibrio cholerae,* named VopF, alters the actin cytoskeleton of the target cell [223]. VopF has been shown to influence the integrity of tight junctions in cultured mammalian monolayers and contributes to colonisation of the infant mouse intestine [97, 98, 224].

There are three known T3SS variants found within *V. cholerae*, the most common being T3SS-2α, which is present within the archetypal T3SS reference strain AM-19226 [95]. This system is most closely related to the *V. parahaemolyticus* T3SS2 which itself is associated with pandemic gastroenteritis caused by *V. parahaemolyticus* [224]. Furthermore, an ortholog of VopF has been identified in *V. parahaemolyticus* (VopL) [225]. The second T3SS, denoted T3SS-2β, is less common than T3SS-2α [226]. The third and the least common T3SS variant is a 'putative' T3SS and more genetically similar to T3SS from virulent *V. anguillarum* isolates [221, 227]. Interestingly, Dorman et al. identified all three versions of T3SS within endemic isolates collected within Argentina [221]. It's important to note that moreover, no 7PET lineage strain has been seen to harbour any T3SS, moreover T3SSs in this study were mutually exclusive with two distinct T3SSs never detected in the same genome at the same time [221]. In Chapter 6 of this thesis, I will explore the frequency of the T3SS variation among the isolates available to this project (all of those displayed in the phylogenetic tree in Figure 5.1).

As described below, I generated a whole genome phylogeny from 481 *Vibrio cholerae* isolates that were available to this study representing both endemic and epidemic lineages isolates collected from over 14 countries, comprised of both clinical and environmental samples. It was from these isolates and data I carefully selected three isolates for further phenotypic investigation. In this chapter the assemblies of the three selected isolates were improved. The complete characterisation of the isolate genomes aided the characterisation of the

similarities and differences between their genomes to determine if these were isolate specific or specific to broader phylogenetic groups. The overarching aim is to use this data to help interpret any differences in responses that may be seen using the organoid model: whether that be in the human host or the pathogen (see Chapter 6).

5.1 Specific Aims:

1. To generate high quality improved genome assemblies required to perform detailed comparative genomics.

2. To identify and catalogue regions of difference among the selected isolate genomes.

3. To assess whether the three selected strains were representative of the lineage in which they represent.

5.1.1 Phylogenetic analysis for live strains in the context of the species diversity

An extensive collection of 481 genomes was used to build the phylogenetic tree, which included a selection of live *V. cholerae* isolates that had been sequenced and were available to this study. In brief, DNA fragments of approximately 450 bp were produced from 0.5 μg gDNA and used for Illumina library creation for these isolates. Libraries were sequenced on the Illumina X10 platform in a 150 bp paired end run with a target coverage of 30 X and assembled using SPAdes v3.8.2 [228]. Kraken [229] was used to check for contamination and to confirm the genomes were *V. cholerae*. Assemblies were screened for appropriate length- knowing that the *V. cholerae* genome is 4.1 Mbp [230] any assemblies over 5 Mbp were ruled out of further analysis due to potential contamination. A list of all genomes used for the strain selection can be found in 'AppendixAA'.

Once the genomes had been quality controlled and checked for contamination, I visualised the genetic relationships between isolates by first determining their phylogenetic placement. Since the isolates in this collection should be representative of the species diversity initially, I inferred the phylogeny based on gene variation amongst core genes. To identify the core genome across the species Roary v3.12.0 [231] (Chapter 2 Methods sections 2.8.1) was used to generate a pangenome from the annotated sequences of the 481 bacterial isolate genomes analysed here. For this analysis, genes were considered to be core if they were present in 97% of genomes included and showed 100% DNA sequence identity. Using these cut-offs, in total 2736 genes were identified as core see Table 5.1 for a summary of pangenome analysis).
Roary was run using the options '-e –maft -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, the CD-HIT clustering cut off was 97%. Core-gene alignments were trimmed using trimAl v1.4.1 [153], SNP-sites v2.5.1 [154] was used to extract the variant sites from the trimmed alignment.

IQ-TREE v1.6.10 was used to create the maximum-likelihood phylogenetic tree [155] with the general time reversible (GTR) and ascertainment bias correction (ASC) models [156]. Five thousand approximate likelihood ratio tests were conducted to assess the robustness of the

computed phylogenies [152] and ultrafast bootstrap approximations took place for each tree calculation [157]. The phylogenetic tree was then visualised using Phandango [158]. The phylogenetic tree was rooted using *Vibrio metoecus* - the most closely related species, to provide an outgroup.

Core genes	97% <= strains <= 100%	2736
Soft core genes	95% <= strains <= 97%	40
Shell genes	15% <= strains <= 95%	884
Cloud genes	0% <= strains <= 15%	25790
Total genes	0% <= strains <= 100%)	29450

Table 5.1. A summary table illustrating the Roary pangenome output data. Core genes are defined as genes present within 97%- 100% of the genomes that were included in the phylogenetic tree, with a minimum 95% amino acid sequence similarity (based on blastp). Soft core genes are genes present within 95% -97% of the strains with a minimum 95% amino acid sequence similarity, shell genes are defined as genes within 15%- 95% of genomes in the tree with a minimum 95% amino acid similarity and cloud genes are defined as genes present within 0%-15% of genomes within the tree with a minimum 95% amino acid sequence similarity. The total genes is the total number of gene clusters found within all genomes that were used to build the phylogeny.

5.2 Strain selection for this project

Three live isolates were selected from the inferred phylogeny (see Figure 5.1). They were selected to represent the different disease potentials that belonged to three different lineages across the species. Importantly, I included isolates from the two archetypal pandemic lineages – the 7PET reference strain N16961 and a classical lineage strain NCTC10732 responsible for pandemics 1-6 [220]. Although both classical and 7PET isolates require CT and TCP to cause disease [8, 81] (as previously described in Chapter 1 section 1.6) there are key differences in gene content and gene expression. In brief, for classical isolates the regulation of CT and TCP is known to be influenced by pH, temperature and osmolarity [232]. Whereas the induction of CT and TCP gene expression in 7PET requires complex and specific growth conditions, including culturing bacteria at 37 °C static for 4 hours followed by shaking [233-235]. In addition, Beyhan et al. found that 524 genes (13.5% of the genome) were differentially expressed between classical, and El Tor biotype isolates [235]. These differences in expression included genes involved in chemotaxis, biofilm formation and transport of amino acids with El Tor isolates showing higher levels of expression in these key pathways - this has generally been taken to suggest that 7PET has a higher survival capacity in the marine environment. Interestingly, genes known to be modulators of *ctxAB* gene expression, that encode CT, with expressed five times higher levels of expression seen in classical strains [235]. More recently, El Tor isolates harbouring a mutation in the *ctxB* (*ctxB7* allele) have been shown to express more cholera toxin than wild type El Tor [236].

Details of the three strains used in this chapter are outlined in Table 5.2. Figure 5.1 illustrates the phylogeny of the species with key virulence factors (outlined in chapter 1 of this thesis) highlighted. The three chosen bacterial isolates are highlighted in blue within the phylogenetic tree in Figure 5.1 itself with a label indicating the 7PET lineage in which N16961 falls. NCTC10732 falls in the classical lineage which gave rise to the 7PET isolates This is consistent with previous work that showed that although existing at the same time, classical and 7PET isolates share a common ancestor [46]. The third isolate (VC0009 harbouring a T3SS-2α- which has been mentioned in chapter 1 and will be discussed further in chapter 6) falls into a distinct lineage which is more genetically diverse than both the classical and 7PET lineages.

Name of strain **Description**	**Place and year of isolation**	**Comparable Reference**
N16961 7PET O1, El Tor	Bangladesh, 1975	N16961- O1 El Tor
NCTC10732 Classical O1	Kasanli- India, 1952	O395- Classical
VC0009 T3SS, Non O1/O139	Argentina, 1993	AM-19226- O39 Type 3

Table 5.2. A list of the bacterial strains used over the course of this project, including place and year of isolation and the comparable reference strain used to align the hybrid assemblies.

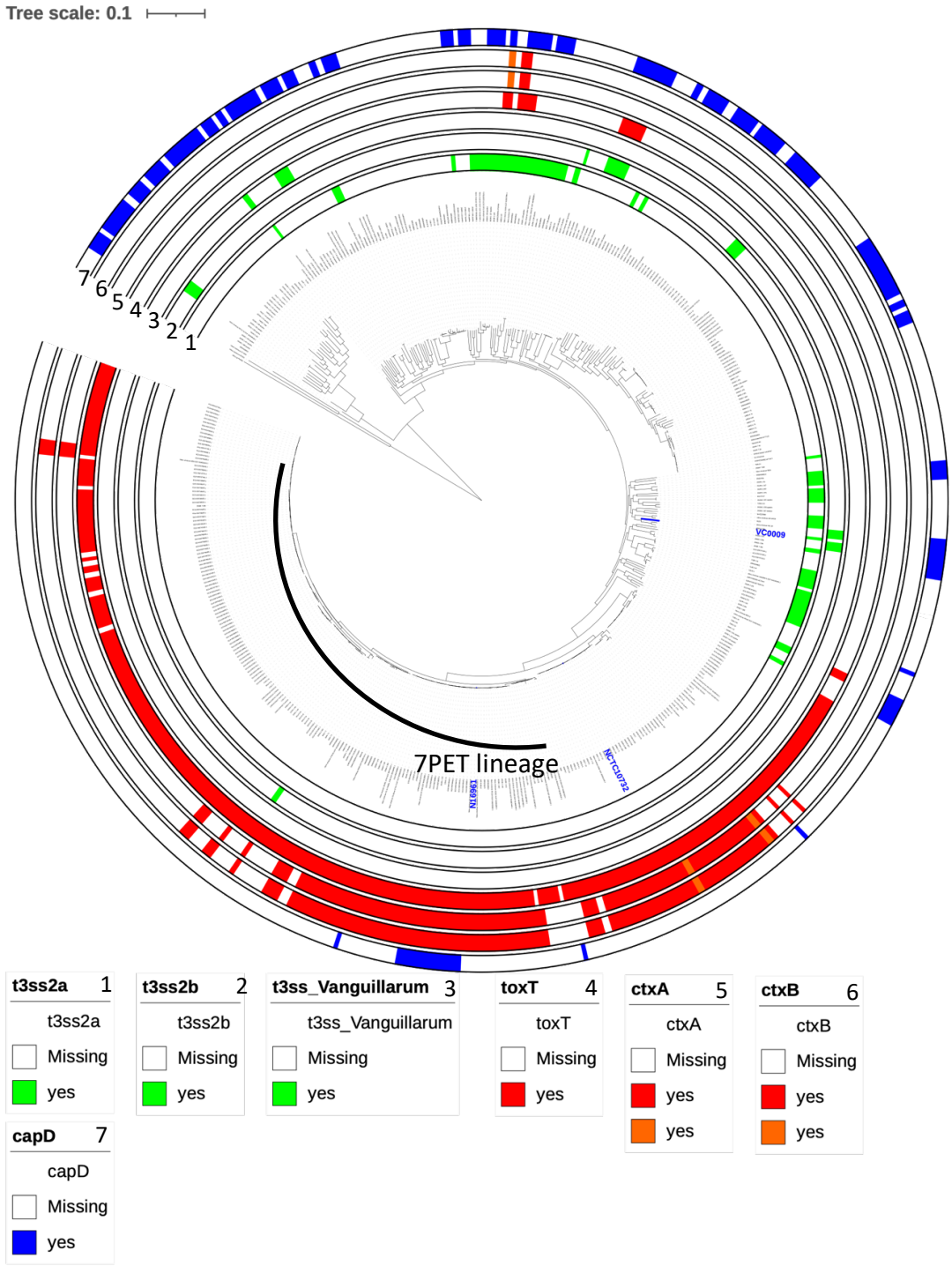


Figure 5.1. A phylogenetic tree of all 481 V. cholerae *strains sequence analysed in this study. This tree has been rooted against 3* Vibrio metoecus *genomes. Colours show presence and absence of key virulence determinants. The inner three rings (labelled 1-3) with green markings denote the presence of T3SS and its derivative, the middle three rings (labelled 4-6) with red markings show the presence of* toxT, ctxA *and* ctxB *and the outside ring (ring7) showing blue markings for the presence of* capD- *a genetic marker of capsule. The 3 bacterial strains selected for this project (N16961, NCTC10732 and VC0009) have been labelled blue within the phylogeny showing their position within the species phylogeny. The black semi-circle in the middle of the phylogeny illustrates where isolates belonging to the 7PET lineage reside. The phylogenetic tree was created using iTol v6.5.4 [237].*

5.3 Short read assemblies of selected strains

The 7PET reference strain N16961 selected for further investigation in iHOs here was originally sequenced in 2000 by Sanger sequencing from cosmid libraries and whole genome random shotgun sequencing [230] which generated a 'finished' genome with two circular chromosomes, one contig for each chromosome with a combined length of 4,033,460 base pairs [230], hence it did not require any improvement. A 'finished' genome is understood to have had any ambiguities or discrepancies resolved due to continuous segments of sequence being ordered and attached to one another to correct any errors [238]. In 2009, the Genomic Standards Consortium classified a finished genome as 'genome sequences with less than 1 error per 100,000 base pairs and where each replicon is assembled into a single contiguous sequence with a minimal number of possible exceptions commented in the submission report' [239].

Conversely the genomes for the other strains selected, classical biotype strain NCTC10732 and strain VC0009, were sequenced with the short read Illumina HiSeq X10 platform, only. Table 5.3 shows the assembly statistics for the short read assembled genomes produced by Illumina sequencing. It can be seen from Table 5.3 that the short-read assembly data showed that the assemblies for strains NCTC10732 and VC0009 had a relatively high number of contigs (46 for classical NCTC10732 and 37 for the type 3 strain VC0009) which we anticipated would include many artefactual rearrangements, gaps and errors. Since the aim of this analysis was to perform a detailed comparative analysis of all three genomes, I made the decision to improve their assemblies. Firstly, I attempted to improve the current assembly *in silico* by using SPAdes [228] as opposed to the Velvet assembler [240] due to the N50 for the SPAdes assembled Type 3 strain (VC0009) was much greater than that of the Velvet assembly. The number of total reads mapped for both strains was greater with the SPAdes assemblies also-indicating a higher quality of assembly. Then I re-sequenced NCTC10732 and VC0009 using a long read sequencing approach and creating a hybrid assembly by combining short-read and long-read data using methodologies detailed below in section 5.3.1. The specific steps undertaken are outlined below.

	Assembly Type	Total Length	No. Contigs	Avg. Contig Length	Largest Contig	N50	Total Raw Reads	Reads Mapped	Reads Unmapped	Reads Paired	Total Bases Mapped
NCTC10732 25455_8#270	Scaffold: SPAdes + Improvement	4045914	46	87954.65	590240	314978	1829924	1765609	64315	1714686	266606959
NCTC10732 25455_8#270	Scaffold: Velvet + Improvement	4035015	43	93837.56	590306	343450	1829924	1763504	66420	1711122	266289104
VC0009 27059_8#318	Scaffold: SPAdes + Improvement	4105229	37	110952.14	988109	611171	2247880	2155734	92146	2077872	325515834
VC0009 27059_8#318	Scaffold: Velvet + Improvement	4111322	45	91362.71	611187	398877	2247880	2152851	95029	2072662	325080501

Table 5.3. A table showing comparisons of SPAdes [228] and Velvet [240] assemblies for the Classical and VC0009 strains used over the course of this project.

5.3.1 Long read sequencing of selected strains

To improve the assemblies of the genomes selected here long-read genomic data was generated using either PacBio for classical NCTC10732 strain or Oxford Nanopore Technologies MinIon for strain VC0009. In brief, whole chromosomal DNA was extracted safely in the Schedule 5 laboratory using the Masterpure Complete DNA and RNA Purification kit (for a detailed description see Chapter 2 section 2.7.2). Once it was confirmed by culture that no live bacteria were present (by streaking the DNA preps on LB agar plates and incubated at 37°C overnight), the DNA was removed from the Schedule 5 laboratory for sequencing using the MinIon platform (see Chapter 2 section 2.7.3; Figure 5.2).

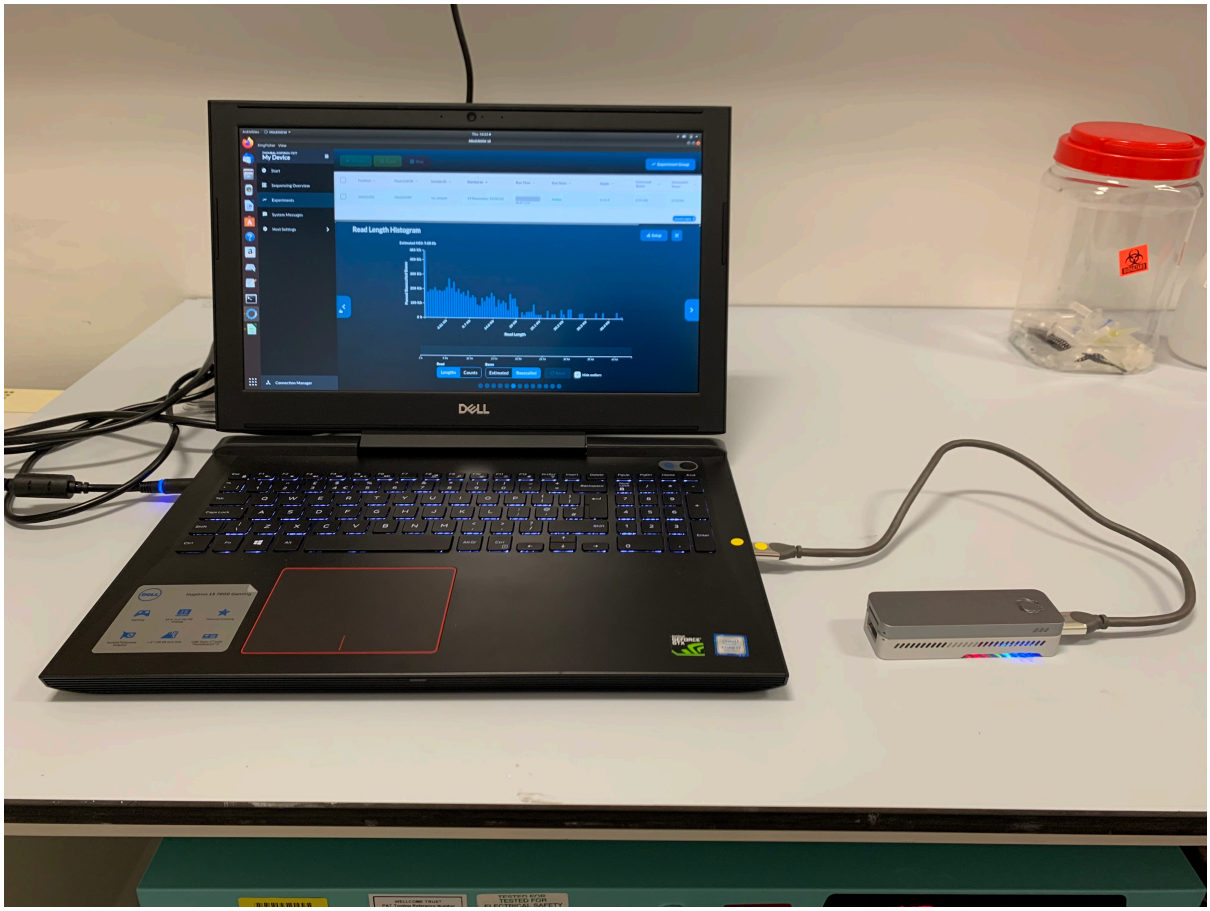


Figure 5.2 Oxford Nanopore Technologies MinIon flow cell connected to a laptop in the process of long read sequencing of VC0009.

Quality control checks comparing the short-read assemblies and the long-read assembly statistics once the long-read data had been generated (summarised in Table 5.3). In the case of NCTC10732, Mauve was used to identify 15 SNP differences between the long read and short read assemblies of NCTC10732. A low number of SNPs is expected during re-sequencing and confirmation, SNPs can be accumulated by long-term storage, so a small number of SNPs is expected. We also re-sequenced N16961 using Minion to confirm its sequence.

5.3.2 Hybrid assemblies of selected *V. cholerae* strains

In order to create these hybrid genomic assemblies, the short read data was combined with either PacBio or MinIon data using Unicycler v0.4.7 using the settings –mode normal -t12 (12 threads) [241]. Unicycler is an assembly pipeline for bacterial genomic data, it incorporates short, long or hybrid assembled genomes. Unicycler builds an initial assembly graph from the short reads using the *de novo* assembler SPAdes and then simplifies the graph using information from the long reads combined with the short reads. The SPAdes version used in this project was v3.8.2 [150].

Table 5.4 illustrates the improvements to the genome assemblies made by creating hybrid assemblies. It can be clearly seen that this hybrid sequencing approach greatly enhance the quality of the assemblies. The number of contigs was decreased (from 37 to 6) in VC0009 when the long-reads were combined with the short-reads in the construction of the hybrid assemblies. The reduction in number of contigs in addition to the increase in genome length suggest better genome coverage as well as fewer opportunities for error in the assemblies- the smaller the number of contigs, the fewer fragments of genome in need of reassembly. Although the number of contigs has not greatly decreased in the case of NCTC10732, the fact that the N50 in both genomes has more than doubled indicates that the hybrid assemblies are of better quality in comparison to the short-read assemblies. The N50 is the key indicator when it comes to genome quality. It is defined as the length of the shortest contig in the let of largest contigs that together constitute at least 50 % of the total assembly size.

a)

Genome	No. of Contigs	Total length	N50	Length of longest contig
Short-read NCTC10732	46	4,045,914	314,978	590,240
Hybrid NCTC10732	41	4,058,016	833,482	1,342,614

b)

Genome	No. of Contigs	Total length	N50	Length of longest contig
Short-read VC0009	37	4,105,229	611,171	988,109
Hybrid VC0009	6	4,165,744	2,854,030	2,854,030

Table 5.4. Comparisons of short-read and hybrid assembly statistics. 3a) shows short-read and hybrid assembly comparisons for the Classical strain NCTC10732. 3b) Differences in the short-read assemblies and the improved hybrid assemblies.

In order to visualise the hybrid assemblies, De Brujin graphs were created by uploading the .gfa (graphical fragment assembly) files created from the Unicycler output data. The .gfa files were then visualised using Bandage v0.8.0 [242]. Figure 5.3 illustrates the hybrid assemblies from Bandage.

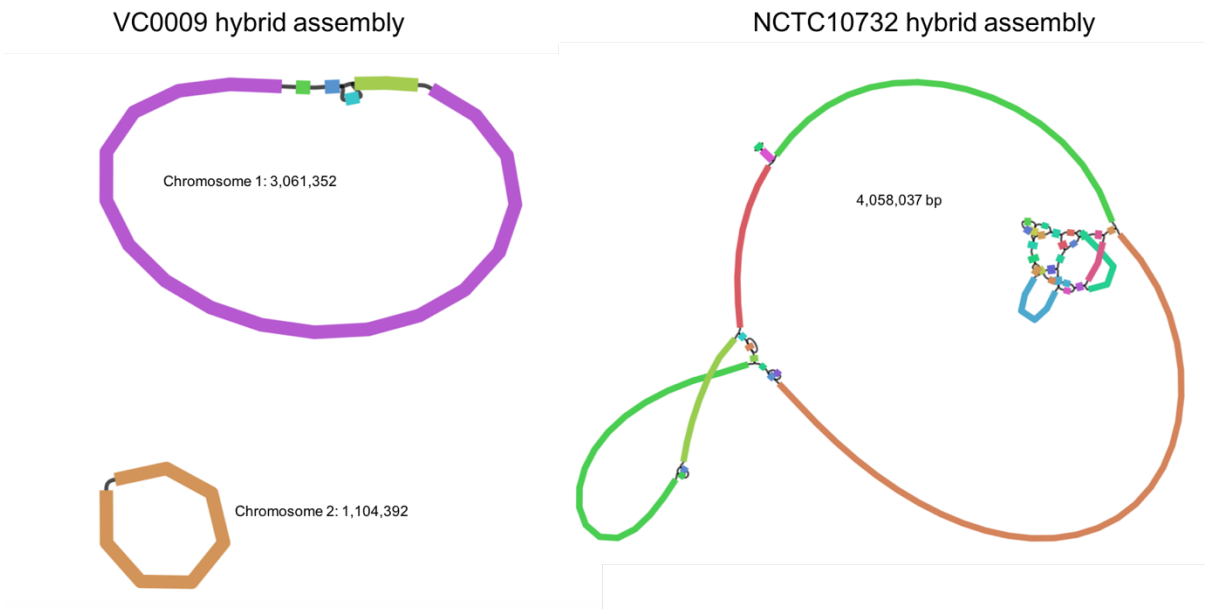


Figure 5.3. De Brujin graphs of the hybrid assemblies were visualised using Bandage v0.8.0 [242]. The coloured blocks represent contigs and the black lines show joins between the contigs. The hybrid assembly of VC0009 creates two circular chromosomes (which is what should be expected from a well resolved assembly - a contig for each of the chromosomes). The hybrid assembly for NCTC10732 is less well resolved, however all the genomic information needed is contained within the assembly.

5.3.3 Re-ordering of the assembled contigs of the selected strains

The assembled contigs from the classical strain NCTC10732 hybrid assembly were re-ordered against the reference strain N16961, based on manual inspection, however, two areas of the genome appeared to be inverted (as shown in Figure 5.4). As NCTC10732 is a classical strain, it was important to align with a relevant reference genome - particularly with the genome being very fragmented. The reference strain selected to align the classical strain NCTC10732 to was the classical reference strain O395.

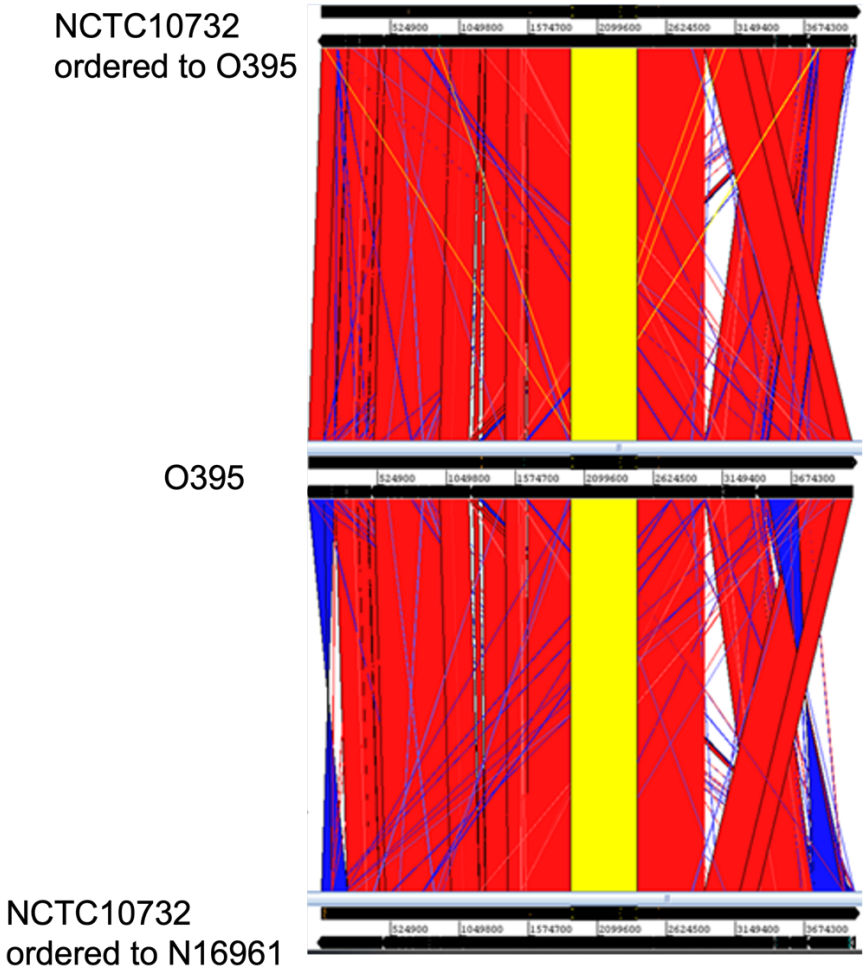


Figure 5.4. A synteny plot to show the ordering of NCTC10732 which was ordered to O395 (top) aligned to O395 (middle) and NCTC10732 ordered to N16961 (bottom) red and yellow bars indicate similarity within the genome and blue shows inversions, the two inversions seen when NCTC10732 was ordered to N16961 is resolved when NCTC10732 is ordered to O395.

In doing so the inversions that were seen from re-ordering the contigs against the N16961 genome disappeared. With O395 being a classical reference strain, it made sense that our classical strain should be re-ordered to it rather than N16961 as it is more genetically similar to O395 that to N16961- the positions of which can be seen in the species phylogeny (Figure 5.1). Figure 5.5 shows the re-ordering of NCTC10732 against O395. This re-ordering to O395 was used for all further analysis for the NCTC10732 genome.

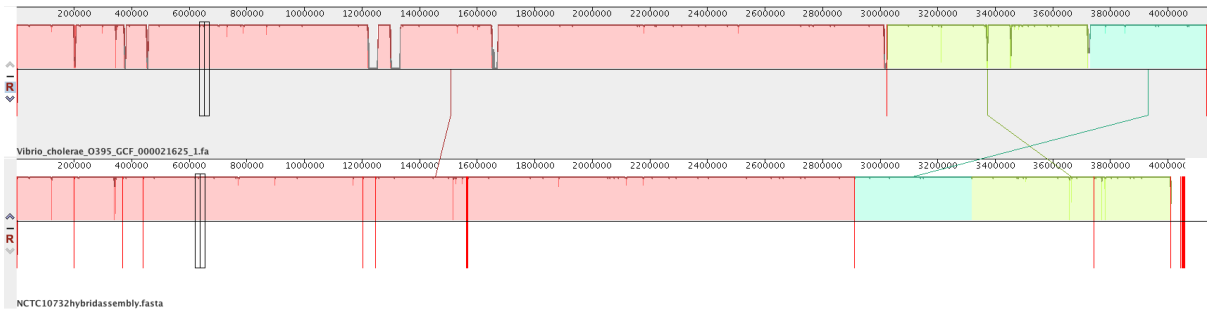


Figure 5.5. An example of the MAUVE software showing the original position of the contigs and the new re-ordered position within the hybrid assembly.

Once the contigs were re-ordered using MAUVE, these new hybrid assembled genomes were visualised and inspected in Artemis Figure 5.6.

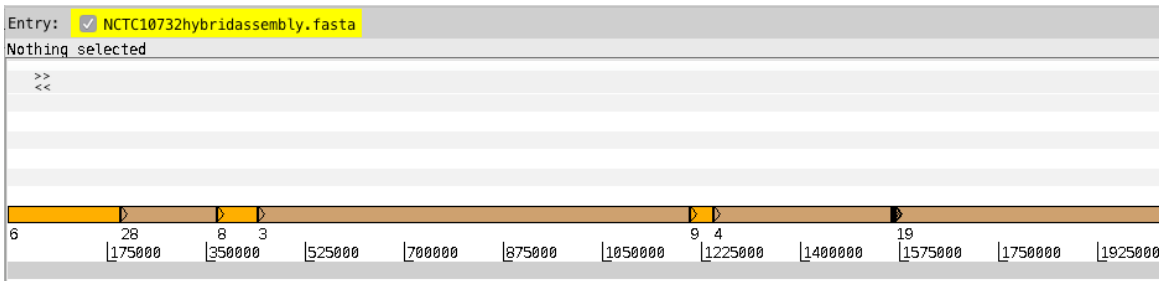


Figure 5.6. A screenshot from Artemis illustrating the re-ordered contigs of Classical strain NCTC10732 using MAUVE. The first contig in the sequence (reading the contig number in the orange bars from left to right) is now what was the 6th contig etc.

These assemblies were then re-annotated using Prokka v1.5[151, 152] (see Table 5.2) creating a new GFF3 file which records the position and functional annotation of all genes predicted to be present. The GFF3 data was also viewed in Artemis for manual inspection and curation and any refined annotations created in this process were used for all downstream analysis in other software, where necessary. Of note smaller contigs less than 200 bp were removed from the dataset because most on manual inspection were seen to be miss assemblies. I also wanted to manually confirm that no important genetic data would be lost by this cut-off: For example, it was found that contig number 5 in the hybrid assembly for classical strain NCTC10732 was tRNA. To enable me to look at what the contigs <200 bp contained, I created fasta files to the contig and then searched for that region in N16961 as it is a complete

genome Figure 5.7 shows an illustration of this process, a combination of the de Brujin graph created in Bandage alongside an Artemis screenshot showing the highlighted region of the reference genome in in fact tRNA.

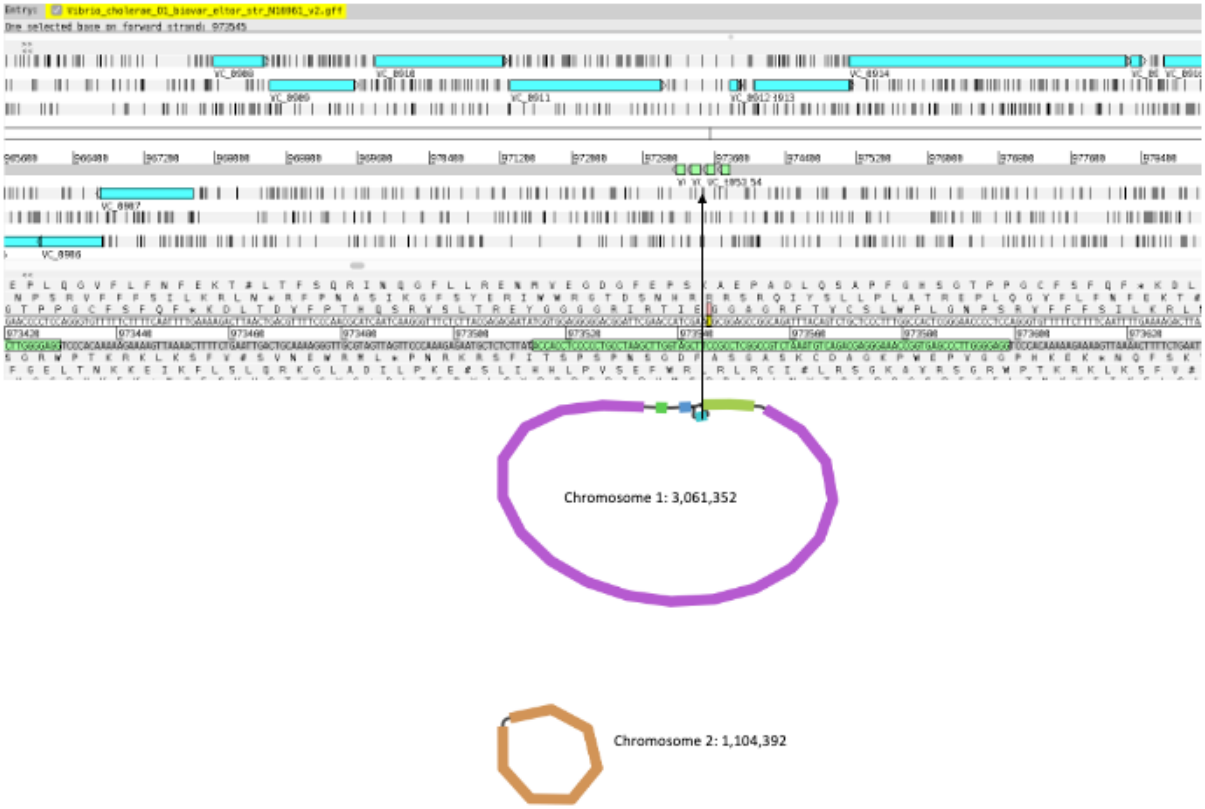


Figure 5.7. An illustration showing the process of identifying the contents of contigs smaller than 200 bp. In this case, contig 5 of NCTC10732 (shown with the arrow) has been blasted against N16961 for sequence similarity. It can be seen in the highlighted region of the genome that this region contains tRNA.

For detailed comparison of the remaining contigs between genomes nucleotide BLAST searches were used to create pairwise comparisons. These comparisons were visualised in ACT (Artemis Comparison Tool) [243]. ACT allows interactive visualisation of the genomes, which enables the user to zoom in and out of the genome comparison in order to identify regions, genes or coding sequences (CDSs) that are present or absent among the genomes of interest. Figure 5.8 shows an ACT screenshot illustrating all three genomes being compared against one another. It has been adjusted to show an overview of the whole genomes to highlight the major differences in contig order and orientation.

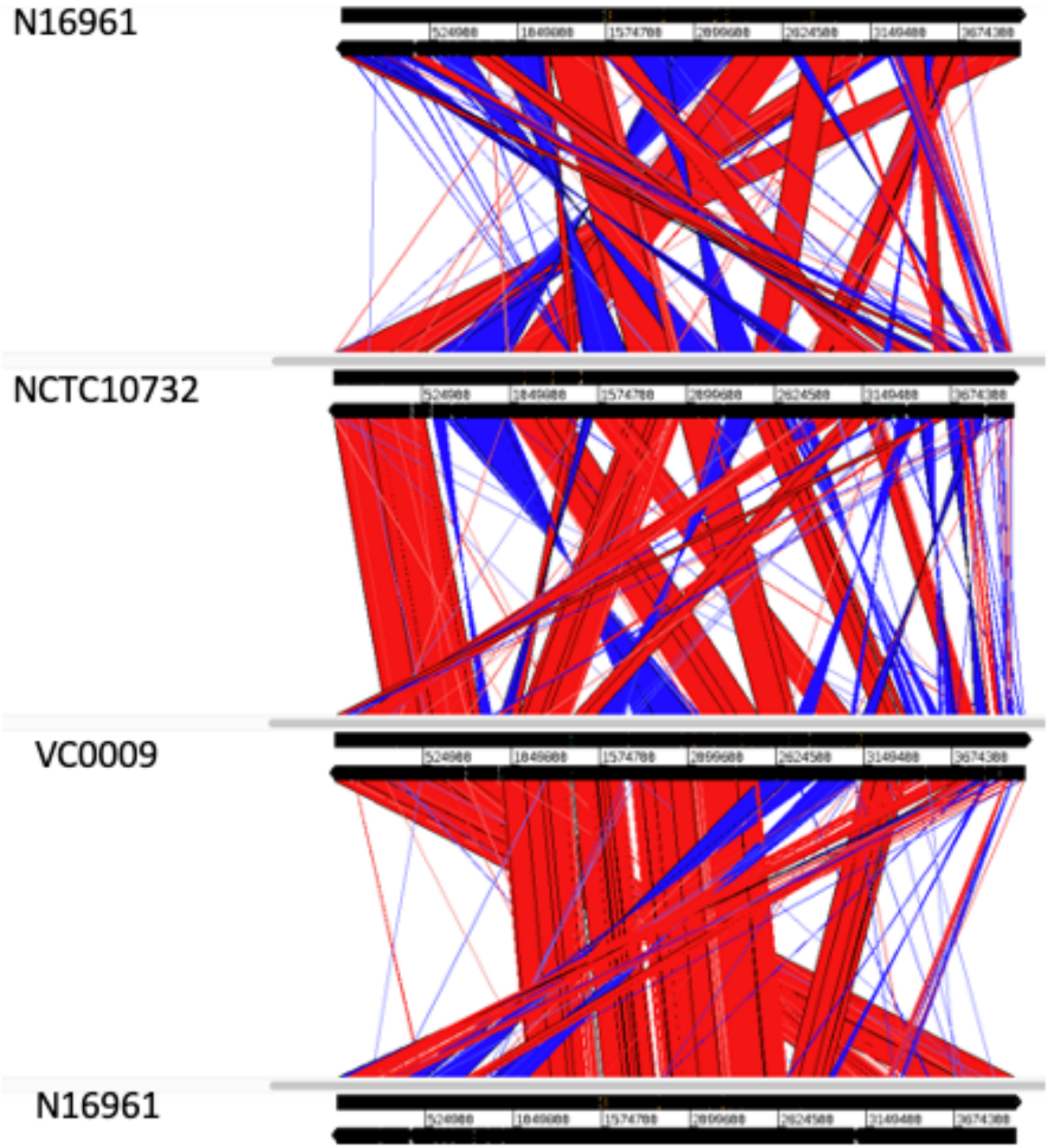


Figure 5.8. An ACT screenshot showing all three bacteria genomes compared to each other. The top row is N16961 with its comparison to NCTC10732, the one below is NCTC10732 with its comparison to VC0009 and the bottom comparison is VC0009 against N16961. The comparison was done this way in order to visualise each isolate compared with the other isolates of interest. The comparisons have been adjusted to show the whole of each genome and BLAST hits filtered to 999 to filter out similarities less than 999 bps. Red bars illustrate similarity amongst the genomes, blue shows inverted regions and white are regions that are absent.

5.4 Identifying genetic regions of difference among the three selected isolates

The three genomes (with chromosomes 1 and 2 concatenated) were visualised in GView (https://server.gview.ca) by uploading the .embl file onto the web-based tool. I wanted to plot the pangenome of all three strains and then look for areas of similarities and differences within the three genomes. Figure 5.9 shows the circular pan genome plot produced by GView, whereby the inner green circle represents the pangenome (all the three genomes combined) and the outer rings illustrate the genomes of interest. There are clear regions of difference- shown by 'gaps' in the chosen genomes. These gaps or regions of difference include known and novel regions.

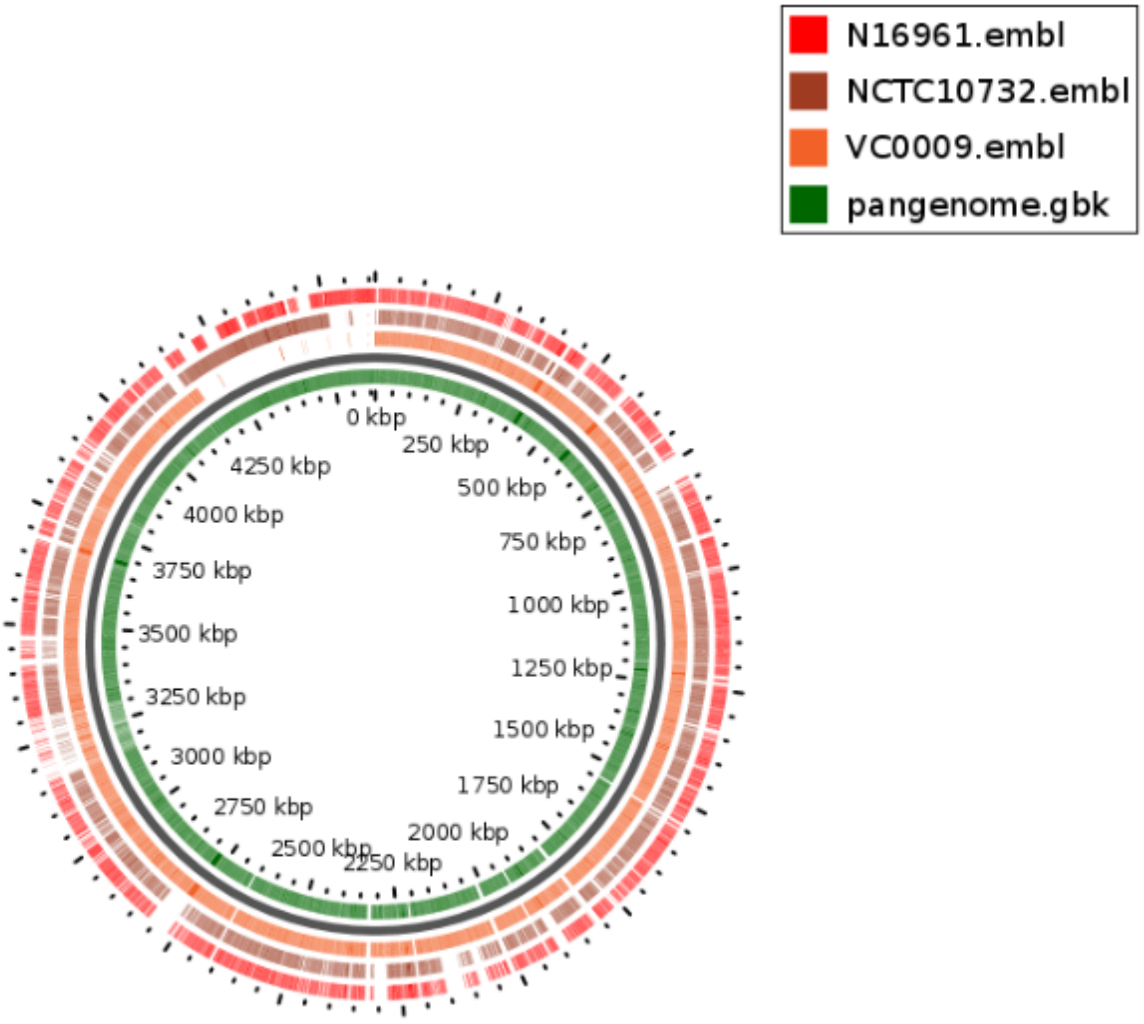


Figure 5.9. A circular genome comparison produced using GView illustrating presence and absence of genomic data among N16961, NCTC10732 and VC0009. Dashes confer presence and spaces indicate gene absence. The green circle at the centre, labelled pangenome, represents all genes within all the strains combined. V. cholerae chromosomes 1 and 2 have been concatenated for ease of display.

Figure 5.9 provided from GView provides an overview of the three genomes and shows key regions of variation amongst the three selected bacterial isolates. To understand what these regions encoded, I then manually inspected all the regions of differences between the genomes. A comprehensive table of these manually created regions of difference can be seen in Table 5.5.

Table 5.5. Illustrating Regions of Difference (RoDs) between the three genomes of interest. The table was developed from feature files manually made in ACT showing areas of the genomes that contained regions of difference based on the Prokka annotation. The co-ordinates of each genome are shown, and the presence/absence of each particular region can be seen within the table.

Genome	start	stop	Size of region (bp)	Number of genes in region	Average length of genes (bp)	GC % of region	Predicted genes in region	Published name/ Prokka annotation	R.O.D name	Present in N16961	Present in NCTC10732	Present in VC0009	Xsome 1 or 2	notes
N16961	175212	189441	14229	10	1422.9	39.24	VC_0175 - VC_0185	VSP-1		Y	N	N	1	
N16961	523303	531106	7803	8	975.375	39.67	VC_0490 - VC_0498	VSP-2		Y	N	N	1	
N16961	533104	550060	16956	14	1211.142857	40.05	VC_0502 - Vc_0516	VSP-2		Y	N	N	1	
N16961	1548970	1550563	1593	2	796.5	44.42	VC_1449, VC_1450	*rtxC* RTX toxin-activating lysine-acyltransferase RtxC		Y	N	N	1	
N16961	1564158	1564473	315	1	315	37.03	VC_1452	*rstC*		Y	N	N	1	just before CTX phage
N16961	3280308	3286896	6588	11	598.9090909	41.58	VC_A0307 - VC_A0317	putative hypothetical, putative *mutT,* putative lipoprotein *blc*		Y	N	N	2	
N16961	3288622	3291970	3348	8	418.5	42.73	VC_A0322 - VC_A0329	putative hypothetical, DNA damage inducible protein,		Y	N	N	2	
N16961	3293686	3296503	2817	5	563.4	37.65	VC_A0330 - VC_A0334	putative hypothetical		Y	N	N	2	
N16961	3298365	3300003	1638	3	546	42.53	VC_A0338 - VC_A0340	putative hypothetical		Y	N	partial	2	
N16961	3301644	3306549	4905	8	613.125	42.25	VC_A0344- VC_A0351	putative hypothetical, putative lipoprotein *blc*		Y	partial	N	2	
N16961	3321103	3323848	2745	6	457.5	43.7	VC_A0375 - VC_A0380	putative hypothetical		Y	Y	N	2	
N16961	3338473	3339148	675	1	675	34.91	VC_A0405 YrdB family protein	putative hypothetical		Y	N	N	2	
N16961	3340450	3340792	342	1	342	37.32	VC_A0408	putative hypothetical		Y	N	N	2	
N16961	3342071	3347966	5895	10	589.5	42.94	VC_A0411 - VC_A0420	putative hypothetical		Y	N	partial	2	
N16961	3373556	3375383	1827	3	609	42.61	VC_A0463 - VC_A0465	putative hypothetical		Y	N	N	2	
N16961	3524866	3527260	2394	1	2394	33.15	VC_A0628	SecA related protein		Y	N	N	2	
N16961	3633667	3638446	4779	3	1593	42.3	VC_A0728 - VC_A0730	putative hypothetical		Y	N	N	2	
N16961	86357	87752	1395	1	1395	46.56	VC_0089	*ccp*		Y	Y	N	1	haem cytochrome c peroxidase
N16961	232730	233729	999	1	999	32.3	VC_0228	*wavE*		Y	Y	N	1	lipopolysaccharide
N16961	239357	240410	1053	1	1053	44.4	VC_0235	putative LPS biosynthesis		Y	Y	N	1	
N16961	241473	242688	1215	1	1215	35.36	VC_0237	putative conserved		Y	Y	N	1	
N16961	246277	262234	15957	13	1227.461538	39.22	VC_0241 - VC_0253	*rfbG, rfbL, rfbM, rfbN*, IS1004 transposase related protein		Y	Y	N	1	
N16961	263562	268971	5409	6	901.5	40.46	VC_0258 - VC_0263	transferases, *rfbT, rfbV*		Y	Y	N	1	
N16961	274161	276975	2814	2	1407	41.07	VC_0269 - VC_0270			Y	Y	N	1	
N16961	837081	840375	3294	3	1098	47.53	VC_0781 - VC_0783	LysR family protein, putative conserved		Y	Y	N	1	
N16961	863281	870724	7443	5	1488.6	42.6	VC_0809 - VC_0813	VPI-1		Y	Y	N	1	
N16961	876756	908346	31590	25	1263.6	34.18	VC_0820 - VC_0844 tcp genes	VPI-1		Y	Y	N	1	

N16961	981853	983500	1647	2	823.5	44.9	VC_0919 - VC_0920	putative serine acetyltransferase related protein, putative exopolysaccharide biosynthesis protein EpsF		Y	Y	N	1	
N16961	1484761	1496074	11313	11	1028.454545	44.61	VC_1393 - VC_1404	putative *sugE*, methyl accepting chemotaxis protein, chemotaxis *cheA, cheY, cheR, cheB, cheW*		Y	Y	N	1	
N16961	1564192	1585198	21006	19	1105.578947	45.41	VC_1452 - VC_1478rts, zot, ctb	CTX phage		Y	Y	N	1	
N16961	1643146	1644172	1026	3	342	39.73	VC_1528 - VC_1530	putative hypothetical		Y	Y	N	1	
N16961	1673291	1675532	2241	2	1120.5	43.35	VC_1561 - VC_1562	LysR family protein, putative beta lactamase related protein		Y	Y	N	1	
N16961	1695345	1699071	3726	5	745.2	44.35	VC_1583 - VC_1587	putative superoxide dismutase, ankB, putative catalase protein, conserved hypothetical protein		Y	Y	N	1	
N16961	1882525	1887601	5076	5	1015.2	44.83	VC_1748 - VC_1752	putative hypothetical protein		Y	Y	N	1	
N16961	1897377	1922811	25434	13	1956.461538	42.21	VC_1760 - VC_1772	VP1-2		Y	Y	N	1	one gene after tRNA
N16961	2549349	2553426	4077	3	1359	33.47	VC_2385 - VC_2387	putative RNA directed DNA polymerase, putative hypothetical		Y	Y	N	1	
N16961	3954166	3955786	1620	2	810	50.76	VC_A1042- VC_A1043	putative product Ccm2, putative *tagE*		Y		N		
N16961	3016123	3017761	1638	2	819	47.47	VC_A0047 - VC_A0048	putative hypothetical		Y	Y	N	2	
N16961	3175456	3181954	6498	5	1299.6	37.07	VC_A0198 - VC_A0202	putative hypothetical, IS1004 transposase		Y	Y	N	2	
N16961	3231346	3234208	2862	4	715.5	45.9	VC_ A0250 - VC_A0253	putative hypothetical, antibiotic acetyltransferase protein		Y	Y	N	2	
N16961	3239709	3246000	6291	10	629.1	45.85	VC_A0258 - VC_A0267	putative hyppothetical, transcriptional regulator AraC/XylS family protein, putative multidrug resistance pprotein		Y	Y	N	2	
N16961	3253420	3253996	576	1	576	41.07	VC_ A0275	putative hypothetical, IS1004 transposase		Y	Y	N	2	
N16961	3261458	3267497	6039	6	1006.5	42.02	VC_A0281 - VC_A0286	putative integrase, putative integrase IS5, hypothetical protein		Y	Y	N	2	
N16961	3272133	3277110	4977	9	553	42.61	VC_A0292 - VC_A0300	putative hypothetical, putative chloramphenicol acetyltransferase protein		Y	partial	N	2	
N16961	3299502	3309591	10089	17	593.4705882	42.41	VC_A0340 - VC_A0356	putative hypothetical, putative lipoprotein Blc		Y	partial	N	2	
N16961	3310814	3318320	7506	10	750.6	39.43	VC_A0360 - VC_A0369	putative plasmid stabilisation element ParE, putative hypothetic		Y	Y	N	2	
N16961	3319630	3323824	4194	7	599.1428571	41.6	VC_A0373 - VC_A0380	putative hypothetical		Y	partial	N	2	
N16961	3325098	3328581	3483	6	580.5	41.62	VC_A0383 - VC_A0388	putative hypothetical, putative toxin resistance protein		Y	Y	N	2	
N16961	3330520	3331213	693	2	346.5	40.35	VC_A0391 - VC_A0392	putative killer protein		Y	Y	N	2	
N16961	3332777	3343802	11025	19	580.2631579	41.96	VC_ A0396 - VC_A0414	*blc3*, putative hypothetical		Y	partial	N	2	
N16961	3345095	3355715	10620	20	531	42.84	VC_A0417 - VC_A0436	*speG2*, putative hypothetical, putative conserved		Y	partial	N	2	
N16961	3357577	3366154	8577	13	659.7692308	39.34	VC_A0440 - VC_A0453	*blc4, yafN, mbIIM, relE*, putative haemagglutinin protein		Y	Y	N	2	

N16961	3367600	3381694	14094	24	587.25	43.04	VC_A0455 - VC_A0477	putative hypothetical, IS5, putative doc protein		Y	partial	N	2	
N16961	3383974	3396475	12501	23	543.5217391	42.59	VC_A0483 - VC_A0506	*relE2, aes, sspa2*, putative hypothetical, putative lipase, IS1004 transposase		Y	Y	N	2	
N16961	3506663	3507419	756	1	756	46.1	VC_A0608	*yjjG*		Y	Y	N	2	
N16961	3524772	3527337	2565	1	2565	33.36	VC_A0628	*secA2*		Y	N	N	2	
N16961	3661048	3662776	1728	2	864	45.75	VC_ A0755 - VC_A0756	transcrptional regulator LysR family protein		Y	Y	N	2	follows tRNA
N16961	3695563	3696940	1377	1	1377	39.55	VC_A0790	putative hypothetical		Y		N		
N16961	3698236	3700252	2016	3	672	40.41	VC_A0793 - VC_A0795	*pinR*, putative hypothetical, putative resolvase		Y	Y	N	2	
N16961	3793036	3799003	5967	7	852.4285714	40.78	VC_A0878 - VC_A0884	putative integrase, putative hypothetical protein		Y	Y	N	2	
N16961	3801528	3803238	1710	1	1710	43.83	VC_A0888	*luxR* trasncriptional regulator		Y	Y	N	2	
N16961	3906519	3909507	2988	3	996	48.24	VC_A0991 - VC_A0993	puative transcriptional regulator *lysR*		Y	Y	N	2	
N16961	3935848	3938062	2214	2	1107	46.24	VC_A1027- VC_A0128	putative maltose operon periplasmic protein, putative maltoporin		Y	Y	N	2	
NCTC10732	5887	21412	15525	14	1108.928571	40.22	NCTC10732_00005-00018	*hipA*- uncharacterised protein related to capsule biosynthesis enzymes, predicted ATPase, Dna helicase		N	Y	N	1	The hip proteins may be involved in cell division control and may interact with cell division genes or their products.
NCTC10732	88930	90307	1377	1	1377	46.44	NCTC10732_00082	ccp- di-haem cytochrome c peroxidase		Y	Y	N	1	
NCTC10732	260958	262038	1080	1	1080	32.1	NCTC10732_00252	lipopolysaccharide synthesis		Y	Y	N	1	
NCTC10732	267647	268696	1049	1	1049	44.48	NCTC10732_00259	lipopolysaccharide biosynthesis protein		Y	Y	N	1	
NCTC10732	274715	295802	21087	20	1054.35	39.54	NCTC10732_00265-00284	*rfba, alqC, qmd, arnB, RfbG. rfbI, fadK, adhE1, rfbV, capD, wcaJ1*		Y	Y	N	1	
NCTC10732	301316	303989	2673	2	1336.5	41.14	NCTC10732_00290-00291	*manA1*- phosphomanose isomerase type1, *mak*- fructokinase		Y	Y	N		
NCTC10732	326960	343511	16551	14	1182.214286	46.65	NCTC10732_00309-00324	*ints1*- phage integrase, *cas1, cse1, cse2, cse3, cse4, cas5e, cas2, cas3,*		N	Y	partial		
NCTC10732	841988	845246	3258	2	1629	47.5	NCTC10732_00797-00798	*leuo1*- leucine transcriptional activator, metallo-beta-lactamase superfamily		Y	Y	N		
NCTC10732	868182	875580	7398	5	1479.6	42.56	NCTC10732_00821-00825	type III restiction enzyme, *tehB*- tellurite resistance protein		Y	Y	N		
NCTC10732	881782	913264	31482	27	1166	34.17	NCTC10732_00831-00857	*tagA, tpx, mcpA1, tcpA, tcpD, epsE2, tcpF, tcpN1, pppA, mcpC1, acfC,tagE-1*		Y	Y	N		
NCTC10732	940736	941303	567	1	567	40.32	NCTC10732_00880	transposase IS200 like		Y	Y	N		
NCTC10732	986762	988409	1647	2	823.5	44.78	NCTC10732_00930-00931	bacterial transferase, putative exopolysaccharide biosynthesis protein		Y	Y	N		
NCTC10732	991028	991469	441	1	441	45.48	NCTC10732_00935	bacterial transferase		Y	Y	N		
NCTC10732	1489597	1500964	11367	11	1033.363636	44.67	NCTC10732_01382-01392	*suqE*- small multidrug resistance protein, *mcp3, cheY2, cheA1*- chemotaxis protein, *cheY-2, cheR1, cheB1, cheW1, mcpa3*		Y	Y	N		
NCTC10732	1518658	1522267	3609	3	1203	38.92	NCTC10732_01405-01407	hypothetical protein, lipase		Y	Y	N		
NCTC10732	1561199	1566581	5382	6	897	45.79	NCTC10732_01440-01445	*ctxB, ctxA, zot, ace*		partial	Y	N		phage protein missing in N16961- note on a separate contig

NCTC10732	1624486	1625503	1017	1	1017	39.88	NCTC10732_01495	bacterial extracellular solute-binding protein		Y	Y	N		
NCTC10732	1654627	1656859	2232	2	1116	43.3	NCTC10732_01525-01526	*lysR* substrate binding domain, metallo-bta-lactamase superfamily		Y	Y	N		
NCTC10732	1676568	1680312	3744	4	936	44.35	NCTC10732_01544-1547	superoxide dismutase, ankyrin repeats 3 copies, catalase precurser, inner membrane protein		Y	Y	N		
NCTC10732	1709131	1709900	769	2	384.5	41.04	NCTC10732_01578-01579	rRNa methyltransferase, IS200 like transposase		N	Y	Y		
NCTC10732	1802826	1803559	733	1	733	41.14	NCTC10732_01654	bacterial extracellular solute-binding protein		N	Y	Y		
NCTC10732	1863300	1868358	5058	4	1264.5	44.89	NCTC10732_01706-01709	*mcS2*, lipoprotein, surface antigen, mechanosensitive ion channel		Y	Y	N		
NCTC10732	1878185	1903538	25353	13	1950.230769	42.19	NCTC10732_01717-01729	*intA3*- prophage integrase, type VI sectretion system, resistin-like protein, *hsdR*- type I restriction enzyme, hypothetical protein, transcriptional regulator		Y	Y	N		
NCTC10732	1917532	1918270	738	1	738	42.22	NCTC10732_01744	phage tail tape measure fammily		Y	Y	N		
NCTC10732	1919604	1929522	9918	14	708.4285714	42.47	NCTC10732_01748-01761	Mu like prophage, mor transcriptional activator family, phage protein, putative secretion ATPase, integrase core domain, hypothetical proteins		Y	Y	N		
NCTC10732	2530220	2534306	4086	5	817.2	33.52	NCTC10732_02317-02321	RNA dependent Dna polymerase, cytochrome c biogenesis protein, predicted ATP-binding protein involved in virulence, hypothetical proteins		Y	Y	N		ncRNA at start or region
NCTC10732	2981924	2983958	2034	3	678	40.44	NCTC10732_02735-02737	*pinR*- putative DNA-invertase from lamboid prophage		Y	Y	N		
NCTC10732	3076781	3082703	5922	5	1184.4	40.72	NCTC10732_02811-02815	hypothetical proteins, bacillus haemplytic enterotoxin (HBL), non-hemolytic enterotixin lytic component L1		Y	Y	N		
NCTC10732	3085203	3086928	1725	2	862.5	43.8	NCTC10732_02818-02819	*hchA*- moleculer chaperpne Hsp31 and glyoxalase, *lasR* (luxR transcriptional regulator)/ chaperone *hchA* associated		Y	Y	N		
	3190286	3193251	2965	3	988.3333333	48.52	NCTC10732_02915-02517	serine deaminase transcriptional activator, glutaredoxin, NADH oxidase family		Y	Y	N		
NCTC10732	3219612	3221813	2201	1	2201	46.24	NCTC10732_02946	maltose operon periplasmic protein		Y	Y	N		
NCTC10732	3248073	3248594	521	1	521	51.53	NCTC10732_02972	*moeB*- rhodanese-like domain		Y	Y	N		
NCTC10732	3372191	3373817	1626	2	813	47.51	NCTC10732_03069-03070	*yiaV*- tripartite multidrug resistance system membrane fusion protein, yiaW- unknown function DUF3302		Y	Y	N		
NCTC10732	3456778	3458413	1635	1	1635	34.15	NCTC10732_03141	hypothetical protein		Y	Y	N		
NCTC10732	3531586	3572491	40905	47	870.3191489	48.27	NCTC10732_03205-03253	starts with *int, yhaV*, putatibe bacteriophage CI repressor, phage regulatory protein CII, bacteriophage protein, rRNA methylation protein, bacteriophage replication gene A protein, phage caspid scaffold protein, phage major caspid protein, phage small terminase subunit		partial	Y	N		ncRNA in middle of region

NCTC10732	3622257	3625116	2859	3	953	45.84	NCTC10732_03303-03305	*tam*- amylase, *cat*- bacterial transferase hexapeptide		Y	Y	N		
NCTC10732	3630614	3636851	6237	8	779.625	45.83	NCTC10732_03310-03317	*merR1*- mercuric resistance operon regulatory protein, *ripA*- transciptional activator, *ydhC2*- multidrug resistance protein, copper binding protein,		Y	Y	N		
NCTC10732	3644317	3644893	576	1	576	40.73	NCTC10732_03324	transposase IS200 like		Y	Y	N		
NCTC10732	3652425	3658536	6111	6	1018.5	42.1	NCTC10732_03330-03335	*intS2*- putative phage integrase, transposase, procyclic acidic repetitive protein, uropathogenic specific protein, hyppothetical proteins		Y	Y	N		
NCTC10732	3663176	3665714	2538	4	634.5	40.76	NCTC10732_03341-03344	plasmid stabilisation protein, prevent host death protein, hypothetical protein		Y	Y	N		
NCTC10732	3667669	3668326	657	2	328.5	39.97	NCTC10732_03345-03346	proteic killer supression protein *higB-1, higA-1* antidote protein		Y	Y	N		
NCTC10732	3669923	3675422	5499	7	785.5714286	42.38	NCTC10732_03348-03354	hypothetical proteins, bleomycin resistance protein,		Y	Y	N		
NCTC10732	3679366	3682327	2961	4	740.25	41.59	NCTC10732_03358-03361	*rhtB4*- lysine exporter protein, hypothetical proteins		Y	Y	N		
NCTC10732	3683566	3691702	8136	10	813.6	39.74	NCTC10732_03364-03373	*parD1*, hypothetical proteins, transposase IS3, resitiction endonuclease,		Y	Y	N		
NCTC10732	3694103	3712247	18144	30	604.8	41.55	NCTC10732_03377-03406	*blc-4, blc-1*- bacterial lipocalin like domain, *ycaC*- isochorismatase hydrolase, *relE2*- plasmid stabilisation protein, lipoprotein, hypothetical proteins,		partial	Y	N		
NCTC10732	3713463	3720159	6696	7	956.5714286	42.77	NCTC10732_03408-03414	uncharacterised protein conserved in bacteria, hypothetical proteins, unknown function,		Y	Y	N		
NCTC10732	3722033	3730592	8559	14	611.3571429	39.3	NCTC10732_03417-03430	*blc2*- bacterial lipocalin like domain, *relE, yafN, mboIIM*, hypothetical proteins, unknown function		Y	Y	N		
NCTC10732	3732057	3762351	30294	44	688.5	39.88	NCTC10732_03432-03475	*ypeA3, vat, yafN, relE*		partial	Y	N		
NCTC10732	3763436	3776637	13201	22	600.0454545	42.61	NCTC10732_03478-03499	lipase, stringent starvation protein A,		partial	Y	N		
NCTC10732	3777995	3781154	3159	6	526.5	41.17	NCTC10732_03503-03508	hypothetical proteins, site-specific recombinase *xerD*		Y	Y	N		
NCTC10732	3782452	3783802	1350	1	1350	39.6	NCTC10732_03511	site- specific recombinase *xerD*		Y	Y	N		
NCTC10732	3816602	3818303	1701	2	850.5	45.53	NCTC10732_03542-03543	cysL3- transcriptional regulator, glyoxalase-like domain, tRNAs at end		Y	Y	N		tRNA at end
NCTC10732	3840498	3843738	3240	1	3240	37.4	NCTC10732_03568	putative ATPase		partial	Y	N		blastp- Predicted ATPase,RecF/RecN/SMC N terminal domain. PFAM PHP- associated. DNA polymerase
NCTC10732	3949943	3952166	2223	3	741	45.5	NCTC10732_03662-03664	transposase Dde domain, hypothetical proteins		Y	Y	N		
NCTC10732	3969729	3970485	756	1	756	46.37	NCTC10732_03680	predicted hydrolase		Y	Y	N		
NCTC10732	4008927	4045458	36531	53	689.2641509	50.58	NCTC10732_03718-03770	DNA-binding transcriptional regulator Nlp, Mu-like prophage FluMu protein gp27, Protein of unknown function (DUF3486), Protein of unknown function (DUF2730), TraR/DksA C4-		N	Y	N		

								type zinc finger protein, Predicted lysozyme (DUF847), N-acetylmuramidase, Bacteriophage Mu transposase, transcriptional regulator, Mu-like prophage FluMu protein gp28, Mu-like prophage FluMu protein gp29, hypothetical proteins,						
NCTC10732	4045537	4047481	1944	4	486	48.84	NCTC10732_03772-03775	plasmid replication protein, hypothetical protein		Y	Y	N		
NCTC10732	4050721	4052845	2124	4	531	43.39	NCTC10732_03777-03780	replication initiation factor, phage RstB protein, hypothetical protein		Y	Y	N		on its own contig
NCTC10732	4054419	4056507	2088	4	522	44.18	NCTC10732_03782-03785	plasmid stabilisation system, toxin-antitoxin system, bleomycin resistance,		partial	Y	N		
NCTC10732	23875	31400	7525	5	1505	40.18	NCTC10732_00023-00027	*hin*- DNA integration/recombination/inversion protein, *mutL*, hypothetical protein, phage integrase family protein,		N	Y	Y		site-specific recombinase, phage integrase family protein, MutL protein,site-specific tyrosine recombinase XerC, hypothetical protein, hin DNA integration/recombination/inversion protein
VC0009	279094	282640	3546	4	886.5	44.52	type3_00252-00255	*yibH*, potassium ion channel, permease, multidrug transporter		N	N	Y		
VC0009	354450	355026	576	1	576	47.78	type3_00314	*sugE*, Small Multidrug Resistance protein		N	N	Y		
VC0009	420934	429268	8334	12	694.5	43.61	type3_00363-00374	*zot*, accessory enterotoxin protein, bacteriophage f237, hypothetical protein		N	N	Y		
VC0009	516277	517663	1386	1	1386	40.74	type3_00453	glycoporin		N	N	Y		
VC0009	616532	618827	2295	1	2295	41.77	type3_00531	Cholix toxin precursor, Exotoxin A binding protein		N	N	Y		
VC0009	739570	776317	36747	44	835.1590909	38.74	type3_00633-00676	tRNA before: *yopE, yscN1, invG, fliP, cadC, yscU, ssaV*		N	N	Y		
VC0009	780338	789113	8775	8	1096.875	34.05	type3_00678-00685	*tdh2*- thermostable direct hemolysin, Vibrio thermostable direct hemolysin		N	N	Y		
VC0009	806207	807206	999	1	999	25.9	type3_00703	cogP family protein		N	N	Y		
VC0009	915446	926138	10692	7	1527.428571	41.85	type3_00800-00806	*gsiA4, yejE1*, yej/b1 permease transporter, moxR like protein		N	N	Y		
VC0009	929195	929735	540	1	540	39.37	type3_00811	transposase IS200 like		N	N	Y		
VC0009	1415618	1417472	1854	1	1854	30.19	type3_01262	hypothetical protein		N	N	Y		
VC0009	1696879	1698472	1593	3	531	42.22	type3_01529-01531	hypothetical proteins		N	N	Y		ncRNA at end
VC0009	1755313	1762837	7524	6	1254	48.45	type3_01581-01586	*ydam3, ybbD, rbsC1, rbsA1*, transcriptional regulator		N	N	Y		ompR before
VC0009	1882923	1906008	23085	14	1648.928571	40.33	type3_01702-01715	antitoxin *higA*, putative *TnsA* endonuclease Transposon Tn7,		N	N	Y		
VC0009	2037157	2039047	1890	1	1890	44.88	type3_01847	chromosome segregation ATPase		N	N	Y		
VC0009	2066336	2086127	19791	19	1041.631579	45.71	type3_01869-01887	transposase one end, phage integrase the other. Hypothetical proteins. Predicted transciptional regulator		N	N	Y		
VC0009	2102137	2105188	3051	2	1525.5	39.19	type3_01906-01907	*oatA, vatD* putative lipopolysaccharide modification acyltransferase, putative lipopolysaccharide biosynthesis O-acetyl transferase *wbbJ*		N	N	Y		
VC0009	2111946	2114133	2187	2	1093.5	35.83	type3_01915-01916	*lex1*- LPS biosynthesis protein, 0-antigen ligase		N	N	Y		

VC0009	2117754	2155446	37692	29	1299.724138	39.01	type3_01920-01948	*wzzE, iolG, alqD, arnB, wcaJ1, capD, gfcB, gfcE, wzb, wecB1*. LPS biosynthesis		N	N	Y		
VC0009	2157291	2158047	756	2	378	43.02	type3_01953-01952	nucleotidyltransferase substrate binding protein		N	N	Y		
VC0009	2262084	2262570	486	1	486	38.6	type3_02050	transposase IS200 like		N	N	Y		
VC0009	2280718	2310526	29808	29	1027.862069	43.77	type3_02073-02101	phage integrase at end, potassium uptake protein, mating pair stabilisation protein		N	N	Y		
VC0009	2475935	2476988	1053	1	1053	44.59	type3_02260	transposase IS200 like. Followed by tRNAs		N	N	Y		
VC0009	2755486	2759239	3753	2	1876.5	42.83	type3_02514-02515	type 1 restriction enzyme EcoKI subunit		N	N	Y		
VC0009	2766817	2813473	46656	36	1296	45.45	type3_02526-02561	Flagellar biosynthesis proteins. integrase at end		N	N	Y		
VC0009	3007787	3010082	2295	2	1147.5	34.63	type3_02728-02729	hypothetical protein		N	N	Y		
VC0009	3037513	3040780	3267	4	816.75	46.79	type3_02751-02754	putative integral membrane protein ascorbate specific PTS system		N	N	Y		
VC0009	3086603	3090275	3672	1	3672	39.26	type3_02798	exonuclease V subunit alpha		N	N	Y		
VC0009	3145299	3145947	648	1	648	42.53	type3_02845	hypothetical protein		N	N	Y		
VC0009	3249045	3267918	18873	26	725.8846154	42.83	type3_02961-02986	proteic killer supression protein *higB, higA,* transposase IS30 family		N	N	Y		
VC0009	3279987	3301452	21465	24	894.375	42.17	type3_03000-03023	putative stabilisation protein, hypothetical protein, putative acetyltransferase		N	N	Y		
VC0009	3302621	3305087	2466	4	616.5	41.99	type3_03027-03030	putative O-methyltransferase, phosphotransferase enzyme family, hypothetical protein		N	N	Y		
VC0009	3377188	3378232	1044	1	1044	40.84	type3_03089	transposase IS200 like		N	N	Y		
VC0009	3415801	3430678	14877	13	1144.384615	45.28	type3_03126-03138	*amyM, malK, lpxD, cycB, malF1, malG, nplT, nahK, dmlR*		N	N	Y		
VC0009	3448298	3451727	3429	1	3429	38.2	type3_03154	hypothetical protein		N	N	Y		
VC0009	3466185	3476661	10476	8	1309.5	41.05	type3_03171-03178	hypothetical protein, type IV secretory pathway *Virb4*, transcriptional factor, response regulator containing a CheY-like reciever domain and a GGDEF domain.		N	N	Y		
VC0009	3568819	3571762	2943	2	1471.5	36.37	type3_03255-03256	hypothetical protein		N	N	Y		
VC0009	3676235	3678152	1917	2	958.5	47.55	type3_03359-03360	*yjlD*- pyridine nucleotide-disulphide oxidoreductase protein.		N	N	Y		
VC0009	3724195	3726571	2376	2	1188	39.71	type3_03388-03389	*nfdA*- cytosine deaminase and related metal-dependent hydrolase, hypothetical protein		N	N	Y		
VC0009	3729086	3730571	1485	2	742.5	51.21	type3_03392-03393	homocystein synthase, hypothetical protein		N	N	Y		
VC0009	3833747	3835907	2160	2	1080	40.81	type3_03487-03488	*AraC* famimly transcriptional regulator, carboxymuconolactone decarboxylase family protein		N	N	Y		
VC0009	3862362	3867195	4833	3	1611	42.53	type3_03518-03520	DNA-binding transcriptional activator XapR/ bacterial regulatory helix-turn-helix protein, amylase		N	N	Y		

VC0009	3868356	3874188	5832	3	1944	44.33	type3_03522-03524	putative glycogen debranching enzyme, ferric enterobactin esterase, maltose operon periplasmic protein,		N	N	Y		
VC0009	3899258	3905738	6480	8	810	47.08	type3_03548-03555	perdicted permease, marR antibiotic resistance protein, molybdopterin biosynthesis protein, putative LysR family transcriptional regulator, quinone oxidoreductase, thiroredoxin-like protein, Eam-like transporter family		N	N	Y		ncRNA around this region
VC0009	3966528	3967716	1188	1	1188	36.25	type3_03602	putative exopolysaccharide biosynthesis protein		N	Y	Y		
VC0009	4119722	4144670	24948	22	1134	41.02	type3_03741-03762	*TnsA* endonuclease N-terminal domain family, integrase, transposase, perdicted ATP-binding protain involved in virulence, putative restriction endonuclease, ATPase involved in DNA repair,		N	N	Y		ncRNA surround region
VC0009	2184564	2185842	1278	1	1278	43.55	type3_01976	hypothetical protein		N	Y	Y		
VC0009	3165321	3171990	6669	6	1111.5	40.19	type3_02862-02867	Bacterial lipocalin		N	N	Y		
VC0009	3173378	3174188	810	1	810	32.09	type3_02869	hypothetical protein		N	N	Y		
VC0009	3177689	3183548	5859	10	585.9	42.59	type3_02873-02882	hypothetical protein, MazG related protein, plasmid maintainance protein, post segregation antitoxin		N	N	Y		
VC0009	3184786	3198934	14148	14	1010.571429	41.01	type3_02883-02897	puative acyl-CoA N-acetyltransferase, phenazing biosynthesis		N	N	Y		
VC0009	3200251	3216694	16443	24	685.125	41.63	type3_02899-02922	*ccdB, relE2, ccdB, maa1, speG1, blc2-* Bacterial lipocalin		N	N	Y		
VC0009	3217948	3223375	5427	6	904.5	40.62	type3_02923-02928	bacterial protein of unknown function DUF925, hypothetical proteins		N	N	Y		
VC0009	3224441	3225098	657	1	657	38.15	type3_02931	conserved protein of unknown function DUF1349		N	N	Y		
VC0009	3227398	3236173	8775	14	626.7857143	41.91	type3_02934-02947	plasmid stabilisation protein, antitoxin stability system, death on curing protain Doc toxin RelE2 addiction molecule toxin		N	partial	Y		
VC0009	3237487	3242446	4959	5	991.8	41.57	type3_02950-02954	membrane protein, hypothetical protein, GNAT family protein		N	N	Y		
VC0009	3246809	3247234	425	1	425	36.62	type3_02959	hypothetical protein		N	N	Y		
VC0009	3269409	3270110	701	1	701	40.74	type3_02988	hypothetical protein		N	N	Y		
VC0009	3276286	3278410	2124	4	531	39.86	type3_02995-02998	uncharacterised protein conserved in bacteria DUF2059, transcriptional regulator		N	Y	Y		

Table 5.5 shows the regions of difference (RoDs) which were created from manual inspection of the three genomes using ACT comparisons and then cross referenced to the pan-genome output file to confirm that the approach undertaken did not miss any genes of importance. This analysis identified 1213 genes that were differentially distributed among the three bacterial strains of interest (summarised in Figure 5.10). Of the total 1213 genes that were contained within the regions of differences 420 of which had prokka annotations that were 'hypothetical', 11 had some integrase function, with an additional 3 genes that were classed as 'putative integrase'. 19 of the genes had an annotation associated with 'phage' whether that be phage protein or phage integrase. 15 genes were annotated with a T3SS tag, an additional two were putative T3SS related. Four of the genes of differences had an annotation linked to O antigen ligase.

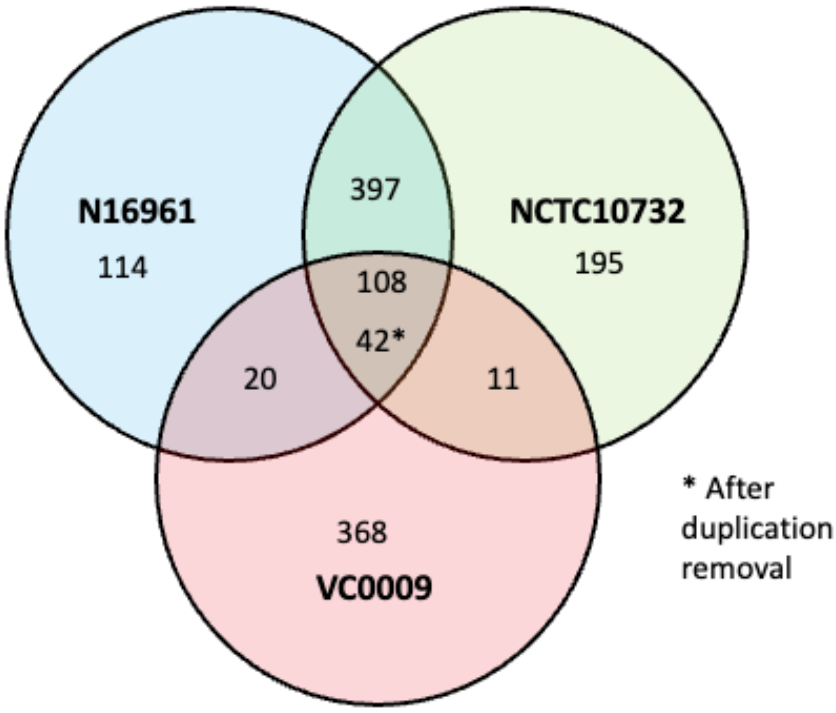


Figure 5.10. A Venn diagram showing the number of genes unique to the bacterial strains N16961 (114) coloured blue, NCTC10732 (195) coloured green and VC0009 (368) coloured red, The Venn diagram also shows the number of genes shared between two of the strains and the triangle in the middle illustrates 108 genes shared between all three isolates (42 of which are unique). Compiled using the Roary output.

It was surprising to find 108 genes that were classed as genes that were common to all three strains, given the approach I had taken. However, upon further inspection many of these genes were located in the integron area of the genome. The integron region of the genome in N16961 is located on chromosome 2 at the nucleotide positions of 309750 to 435418 consisting of 125,669 bp [244]. The integron comprises of genes that are flanked by highly conservative repetitive sequences of 123-126 bp which are called *V. cholerae* repeats (VCRs) [244-247]. From this analysis that transposases, such as IS1004 transposase is picked up many times within all of the genomes and accounts for five of these 108 genes in Table 5.6.

Next, I deleted the duplicate 'genes' and the 108 reduced to only 42 genes in common to all the isolates. Many of these duplicate genes are annotated as transposases which occur multiple times in all three of the genomes. The reason that I picked these up as differences in my manual synteny comparisons (blastn) whereas they appear to be the same in the pan genome analysis (blastp) is that they could have been on the edge of my regions of difference when I manually added the feature files in Artemis. Of the 42 genes that remain, the vast majority (27 of 42) of them are situated in the super-integron region of the chromosome and we can see from Table 5.6 that *higAB* and *parDE* orthologues within this super-integron region, these features (along with others) led Heidelberg to suggest that the smaller chromosome (chromosome 2) of *V. cholerae* was originally a mega-plasmid that had been captured by an ancestral *Vibrio* species [230].

It was not surprising that the classical strain NCTC10732 and the El Tor N16961 had a higher number of genes in common than either of the other two comparisons (NCTC10732 to VC0009 and VC0009 to N16961) as the classical and El Tor strains are more closely related to one another on the phylogenetic tree of the species see Figure 5.1. The classical and El Tor strains also have many virulence genes in common- for example the CTX Φ and *tcp* genes that are absent in VC0009.

Gene	Annotation	NCTC10732	VC0009	N16961
yopE	SpoVR family protein	NCTC10732_01828	type3_00766 type3_00634	AE003853.714
sugE	SugE protein	NCTC10732_01382	type3_00314	AE003852.1440
rlmF	23S rRNA mA1618 methyltransferase	NCTC10732_01580 NCTC10732_01578	type3_00501	AE003852.1653
intA_1	Phage integrase	NCTC10732_01716	type3_00632	AE003852.1794
VCJ_000727	Capsular polysaccharide biosynthesis protein	NCTC10732_00256	type3_01911	AE003852.246
kdtA	3-deoxy-D-manno-octulosonic-acid transferase	NCTC10732_00257	type3_01912	AE003852.247
VOA_002967	Mannosyltransferase OCH1-like enzyme	NCTC10732_00258	type3_01913	AE003852.248
xynD	Polysaccharide deacetylase	NCTC10732_00263 NCTC10732_00262	type3_01918	AE003852.253 AE003852.252
rfaD	ADP-L-glycero-D-manno-heptose-6-epimerase	NCTC10732_00264	type3_01919	AE003852.254
VCJ_000540	ISVch4 transposase, OrfA	NCTC10732_00021 NCTC10732_01746 NCTC10732_02734 NCTC10732_03501 NCTC10732_03510	type3_02563	AE003852.269 AE003852.1824 AE003853.364 AE003853.495 AE003853.779 AE003852.1515
VCJ_003149	Transposase OrfAB subunit B	NCTC10732_00020 NCTC10732_01745 NCTC10732_02733 NCTC10732_03502 NCTC10732_03509	type3_02562	AE003852.270 AE003852.1516 AE003853.496 AE003853.778
AAY52_08450	IS1004 transposase	NCTC10732_00880 NCTC10732_01579 NCTC10732_02132 NCTC10732_02500 NCTC10732_03259 NCTC10732_03324 NCTC10732_03489	type3_00811 type3_02050 type3_02260 type3_03089	AE003852.912 AE003853.202 AE003853.273 AE003853.481
VOA_001066	Rhodanese-like sulfurtransferase	NCTC10732_02972	type3_03550	AE003853.1036
VOA_001846	Hypothetical protein	NCTC10732_03315	type3_02845	AE003853.263
parE1_1	Plasmid stabilization system protein	NCTC10732_03341	type3_02921	AE003853.306 AE003853.376
group_1823	Prevent-host-death protein	NCTC10732_03342	type3_02922	AE003853.307 AE003853.377
blc	Outer membrane lipoprotein Blc	NCTC10732_03387 NCTC10732_03420 NCTC10732_03379	type3_02918 type3_02864	AE003853.312 AE003853.345 AE003853.397 AE003853.434
VCJ_000545	Glyoxalase family protein	NCTC10732_03499 NCTC10732_03353 NCTC10732_03784	type3_02872 type3_02971	AE003853.342 AE003853.494 AE003853.392 AE003853.465
AAY52_17260	Plasmid stabilization system protein	NCTC10732_03496	type3_02934 type3_02956 type3_02974	AE003853.343 AE003853.491

AAY52_17195	DNA-damage-inducible protein J	NCTC10732_03497	type3_02935 type3_02957 type3_02975	AE003853.344 AE003853.492
hutH	Histidine ammonia-lyase	NCTC10732_01203	type3_00132	AE003853.346 1246
parE1	Plasmid stabilization element ParE, putative	NCTC10732_03363	type3_02965	AE003853.353
higB-1	Proteic killer suppression protein	NCTC10732_03345	type3_02967	AE003853.382
higA-1	Antidote protein	NCTC10732_03346	type3_02968	AE003853.383
group_101	Hypothetical protein	NCTC10732_03349	type3_03030 type3_02871	AE003853.388
group_174	Hypothetical protein	NCTC10732_03350 NCTC10732_03402	type3_02995 type3_02910	AE003853.389
AAY52_16845	N-acetyltransferase GCN5	NCTC10732_03352 NCTC10732_03785	type3_02989	AE003853.391 AE003853.468
AAY52_04155	Putative acetyltransferase	NCTC10732_03354	type3_02964	AE003853.393
AAY52_04145	Lipoprotein	NCTC10732_03357 NCTC10732_03426 NCTC10732_03409 NCTC10732_03362	type3_02958	AE003853.405 AE003853.416
VCJ_003063	Glyoxalase family protein	NCTC10732_03378	type3_02870 type3_02937	AE003853.406 AE003853.333
group_78	Hypothetical protein	NCTC10732_03412	type3_02878 type3_02927 type3_02894	AE003853.411 AE003853.422
VOA_001482	Antitoxin of toxin-antitoxin stability system	NCTC10732_03391	type3_02969	AE003853.413
group_14	Predicted Fe-S protein	NCTC10732_03410	type3_02972 type3_02976	AE003853.417 AE003853.309
AAY52_16380	Lactoylglutathione lyase	NCTC10732_03413	type3_02985	AE003853.423
AAY52_17520	Hypothetical protein	NCTC10732_03430	type3_02962	AE003853.443
ypeA_3	N-acetyltransferase GCN5	NCTC10732_03440	type3_03020 type3_02946	AE003853.459
speG_6	N-acetyltransferase GCN5	NCTC10732_03444 NCTC10732_03419	type3_03010	AE003853.463 AE003853.433
AAY52_17165	Antitoxin of toxin-antitoxin stability system	NCTC10732_03403 NCTC10732_03483	type3_02939 type3_03004	AE003853.477
relE2	Death on curing protein, Doc toxin	NCTC10732_03404 NCTC10732_03484	type3_02940 type3_03005	AE003853.478
VOA_001493	Acetyltransferase (putative)	NCTC10732_03490 NCTC10732_03400 NCTC10732_03376 NCTC10732_03488	type3_02898 type3_02924	AE003853.482 AE003853.373
VCJ_003055	Autative acetyltransferase	NCTC10732_03415 NCTC10732_03498	type3_03008	AE003853.493 AE003853.427
yjjG_1	dUMP phosphatase	NCTC10732_03680	type3_03137	AE003853.595

Table 5.6. A table of the 42 genes that were identified as genes of difference using the Roary analysis. Many of these genes are repeated multiple times within the three genomes.

5.5 Discussion

To enable a detailed genomic comparison of the three chosen bacterial isolates, I needed to create improved high quality genome assemblies for two of the three strains selected. To do this, I combined the available short read assembly data with new long read data that was sequenced either by PacBio or Minion technologies. This allowed me to create hybrid assemblies which were much improved versions of the short read only assemblies that were available. The processes and reasoning behind this are provided in this chapter. I have catalogued the gross differences between the three bacterial isolates of interest in this chapter and in the following chapter I will explore the regions of difference between these high-quality genomes in finer detail and see if these regions are present across the species phylogeny in any patterns, be them lineage specific or random.

In the next chapter I will examine the three genomes in greater detail by manually assessing regions of difference. I will also inspect where the regions of difference that were catalogued in this chapter lie across the whole species phylogeny thus identifying whether these bacterial isolates are representative of the lineages in which they lie. Furthermore, a selected number of regions of difference will be discussed in more detail. I will also infect the small intestine organoid model with the three selected bacterial strains (N16961, NCTC10732 and VC0009) recording differences in colonisation and bacterial morphology among these isolates.

Chapter 6:

Cross species genomic analysis

6 Introduction

In Chapter 5 of this thesis, I investigated the gene presence and absence across the three *V. cholerae* strains of interest (N16961, NCTC10732 and VC0009). In this chapter, I place these differences in gene content in context of the whole species. In doing this I plan to be able to better interpret any observed differences in colonisation patterns or infectivity of these strains in the small intestinal organoid model system. These data were also used to determine if the bacterial isolates selected for this project were representative of the lineage in which they reside.

As described in Chapter 5 section 5.2.2, the three strains were selected to represent the different disease potentials that belonged to three different lineages across the *V. cholerae* species. Two of these strains are archetypes for their genetic lineages- N16961, the most well characterised laboratory strain representing 7PET *V. cholerae*, NCTC10732 which belongs to the classical lineage, the clinical predecessor to 7PET, responsible for cholera pandemics 1-6 [220]. The disease potential of the chosen strains has been discussed in the previous chapter of this thesis, in this chapter, I will discuss some genetic variation within these chosen strains and look at these in the context of the species phylogeny to identify if these differences are lineage specific or found within the strain alone. I will also identify key virulence determinants and look at these in the context of the species phylogeny. I will also infect the small intestinal organoid model with N16961, NCTC10732 and VC0009 to identify any differences in colonisation or pathology between these isolates.

6.1 Specific Aims:

4. To determine if the genetic differences noted correlate to the lineage(s) as a whole or are specific to the isolate that was selected.

5. To determine if there are observable differences in infection profile between N16961, NCTC10732 and VC0009 and whether we can link any differences identified to the genomic variation of the bacterial isolates.

6.2 Virulence determinants across the phylogeny

Using Roary [231], I generated a gene distribution matrix against the phylogeny of the species- as previously discussed in chapter 5 of this thesis. Using these data, firstly I plotted the key virulence factors associated with the selected bacterial isolates that were chosen for this project e.g., presence of cholera toxin (CT) and a type three secretion system (T3SS) and its derivative which have been previously described in chapter 1 of this thesis. Figure 5.1 shows the distribution of these key virulence factors.

As discussed in Chapter 5 Introduction, there are three derivatives of T3SSs: T3SS2α, T3SS2β and the putative T3SS that resembles that present in *V. anguillarum*. No isolate in our collection possessed two or more of T3SSs at once, consistent with the published literature [221]. It was curious to see that three of the isolates (two environmental isolates; 16853_4_65 and SRR4039819 and one clinical isolate- *Vibrio cholerae* V51) within the phylogeny not only had genes that encode a T3SS but also *toxT*; as a marker of VPI1 (Vibrio pathogenicity island 1 as termed because this genomic island is only present within epidemic strains due to the presence of the virulence factors that it encodes) in addition to the genes that encode CT, these three isolates are both located within the non-7PET lineage within the phylogenetic tree in Figure 5.1, marked with the asterisk. Upon further investigation, one of these three isolates (V51) has previously been reported to have a T3SS and encode cholera toxin [69], a similar gene pattern has been witnessed in strains belonging to serogroup O141, which encode a T3SS in addition to cholera toxin [248].

A greater number of genomes encode *toxT* than the genes for cholera toxin (*ctxA* and *ctxB*), this is not surprising as *toxT* is a transcriptional regulator whose expression is necessary for the two major virulence determinants in *V. cholerae*- TCP and CT. Therefore, a genome must possess *toxT* to not only colonise its host but to be able to encode the genes necessary for cholera toxin expression.

According to the Roary pan genome output, in my dataset, there was only one genome that possessed *ctxA* and *ctxB* in the absence of *toxT*, this finding is somewhat contradictory given the acquisition of the CTX phage is dependent on having the TCP genes because the TCP encodes the toxin co-regulated pilus (as discussed in Chapter 1 section 1.6). It is also

interesting to note the pattern of distribution of the *capD* gene across the species phylogeny (Figure 5.1). The gene *capD* is required for the biosynthesis of the capsular polysaccharide and ultimately responsible for capsule production. It is used as a marker of the capsule loci which is present across the phylogeny- but is more prevalent in the more diverse serogroups closer to the root of the phylogeny as opposed to within the clonal 7ET lineage- except for the O139 lineage. The distribution and significance of the capsule locus will be discussed further in section 6.3.4.

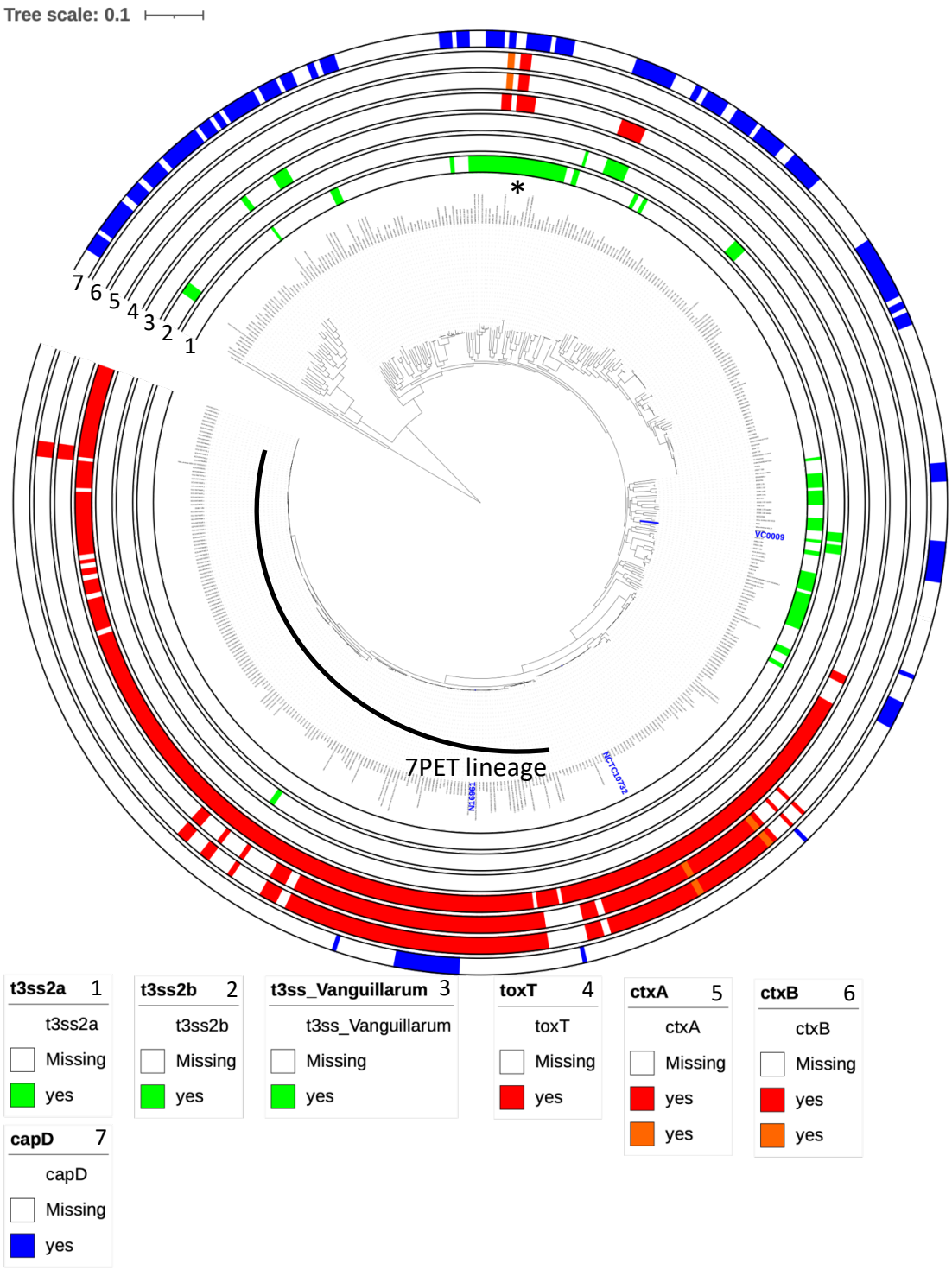


Figure 6.1 . A phylogenetic tree of all 481 V. cholerae *strains available for this project. This tree has been rooted against 3* Vibrio metoecus *genomes. colours show presence and absence of key virulence determinants. The inner three rings with green markings labelled 1-3, denote the presence of T3SS and its derivative, the middle three rings labelled 4-6 with red markings show the presence of* toxT, ctxA *and* ctxB *and the outside ring labelled 7, showing blue markings, for the presence of* capD- *a genetic marker of capsule. The bacterial strains for this project have been clearly labelled blue within the phylogeny showing their position within the species phylogeny. The black semi-circle in the middle of the phylogeny illustrates where isolates belonging to the 7PET lineage reside. The asterisk highlights the isolates discussed in section 6.2 that encode a T3SS but also toxT in addition to ctxA and ctxB. The phylogenetic tree was created using iTol v6.5.4 [237].*

6.3 Regions of difference

Next, in addition to the known virulence genes, and to be comprehensive, I plotted the distribution of the genes present in the regions of difference (RODs) I showed to differentiate the three selected *V. cholerae* strains in Chapter 5. Figure 6.2 was created using Phandango [158], the phylogenetic tree of the species can be seen down the left hand side of the image and the gene presence/absence matrix on the right (blue bars indicate gene presence). Figure 6.2 shows gene presence with the variable genes ordered from the highest in frequency (left-hand side) to the lowest (right-hand side). This view enabled me to see phylogenetically distinct genetic markers linked to lineages within the with phylogeny. Importantly, not only were the phylogenetic distribution of RODs among the three strains of interest important to record. It was also key to see if these regions were peculiar to that isolate alone or representative of a lineage(s) across the phylogenetic tree. Something that would not be possible if we were only looking at the RODs amongst the three isolates.

In order to illustrate how different RODs were distributed in the pan genome, I selected four regions that summarise the major patterns of distribution that I observed. These regions are marked a-c in Figure 6.2. The region marked ‘a’ is largely absent from 7PET but present in most other non-7PET lineages. Region marked ‘b’ in Figure 6.2 shows genes that are sporadically present in genomes that fall across different non-7PET lineages (albeit not all non-7PET lineages). The region marked ‘c and in Figure 6.2 highlights a region that stand out across the phylogeny as being rare and present in only one or two isolates in comparison to the 481 genomes included in the phylogeny.

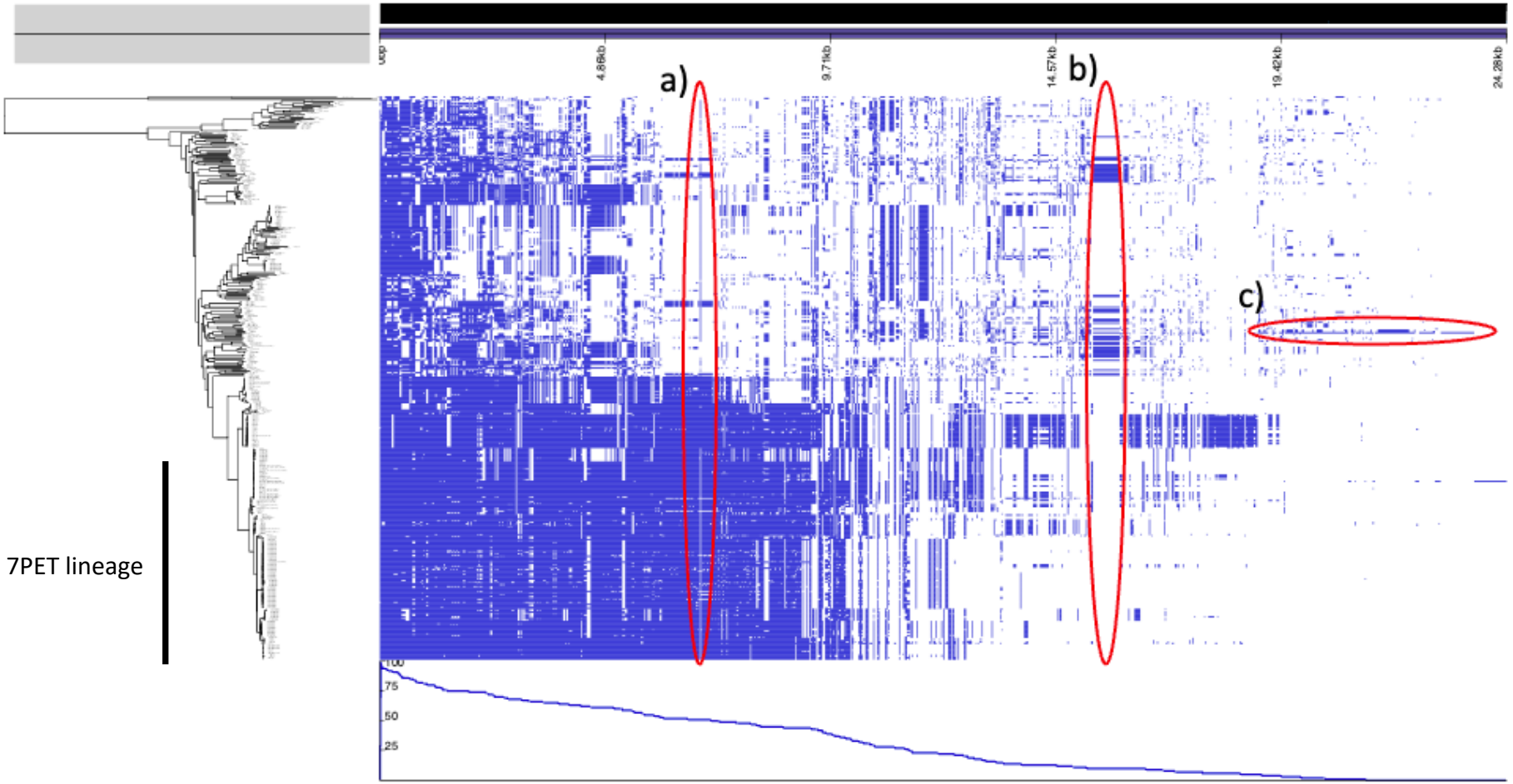


Figure 6.2. Regions of differences within the 3 bacterial isolates of interest plotted against the species phylogeny (tree on left) The blue blocks indicate gene presence and the white gaps illustrate gene absence. The graph along the bottom indicates what percentage of isolates have the gene present within their genome. Regions that will be discussed later in this chapter are labelled a, b, and c indicated by red circles. The 7PET lineage is marked with the black line to the left of the phylogeny.

6.3.1 Region labelled 'a' in Figure 6.2

The region labelled 'a' in Figure 6.2 was intriguing as it appears to have a distinct split in distribution down the phylogenetic tree, generally being absent in the 7PET lineage (as indicated with the black line to the left of the phylogeny in Figure 6.2), and present among isolates that are non-7PET. The pan genome analysis conducted using Roary [231] identified this gene to be *ycaC*. The gene *ycaC* is predicted to encode an isochorismatase hydrolase. Unfortunately, there is no published literature (for *Vibrio cholerae*) regarding its function. In *E. coli,* the protein product of *ycaC* was first identified in 1987 [249], however the specific function remains unknown [250]. Related prokaryotic proteins have functions in antibacterial resistance [251]. A previous study has shown that YcaC is exceptionally conserved across both bacterial (>75% conserved between both *Pseudomonas aeruginosa* and *E. coli*) and fungi (>50% sequence identity to *P. aeuriginosa*)- suggesting that whilst YcaC may not be part of an integral pathway, its function could confer a significant evolutionary advantage to microbial life [251]. Groftehauge et al. raise the possibility that if, like other isochorismate-like bacterial proteins, YcaC is involved in antimicrobial drug resistance suggesting that the lack of ubiquity and high level of conservation suggests that YcaC could have been acquired by horizontal gene transfer conferring a significant selective advantage for both bacteria and fungi.

In order to identify the genomic context of *ycaC*, I analysed the ACT comparisons of the three chosen bacterial isolates (N16961, NCTC10732 and VC0009) to look for the presence of *ycaC* within the three chosen bacterial isolates. *YcaC* is present in VC0009 and NCTC10732 but absent in the 7PET isolate N16961 (which is consistent with the presence of the gene within non-7PET genomes), the gene is also positioned in the integron region of the chromosome (as previously discussed in Chapter 5 section 5.4). It would be interesting to explore the presence and function of *ycaC* in relation to *V. cholerae* further to elucidate whether in fact there is/was a selective advantage to isolates encoding the gene prior to 7PET.

6.3.2 Type 3 Secretion System labelled 'b' on Figure 6.2

The region marked as 'b' in Figure 6.2 is the type 3 secretion system (T3SS) identified within one of the three selected *V. cholerae*, strain VC0009. Looking at the gene content of this region, identifies it as the T3SS-2α- as previously described in chapter 5 introduction and illustrated in Figure 6.3 which shows the synteny plot of the T3SS-2α region present in AM-19226 alongside that of VC0009. The schematic at the bottom of Figure 6.3 shows the characterised genes that are present within T3SS-2α systems.

From the 481 *Vibrio cholerae* isolates included in the species phylogeny 57 harboured T3SS-2α , including the archetypal AM-19226 strain, in which the T3SS was first identified. I wanted to explore the variation among the other alternative T3SSs and established that 17 of the 481 isolates encoded the T3SS-2β and 4 of the isolates contained the putative T3SS with genetic similarity to that first described in *V. anguillarum* (Figure 5.1). The presence and phylogenetic positioning of these derivatives of T3SS show that T3SSs are not generally present within the 7PET lineage of *V. cholerae* (illustrated by the internal black semi-circle on Figure 5.1.

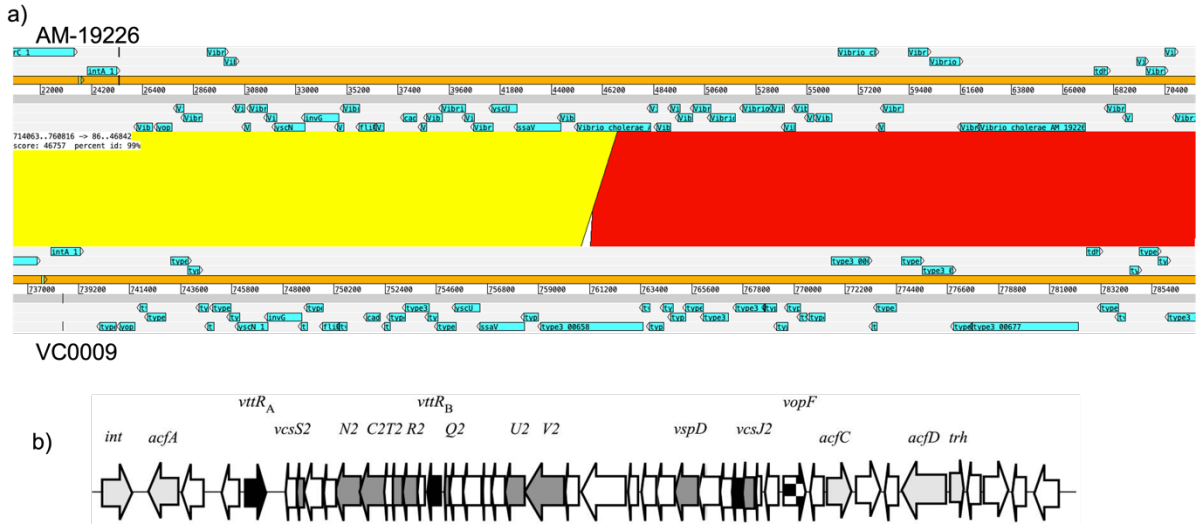


Figure 6.3. Schematics of T3SS-2α. a) shows an ACT synteny plot of AM-19226 (top) aligned with VC0009 (bottom) to show the presence of the T3SS-2α, red and yellow bars illustrate similarity between the isolates. b) illustrates the genes that encode the T3SS-2α reported in AM-19226 by Alam et al. [223].

6.3.3 Region labelled ‘c’ on Figure 6.2.

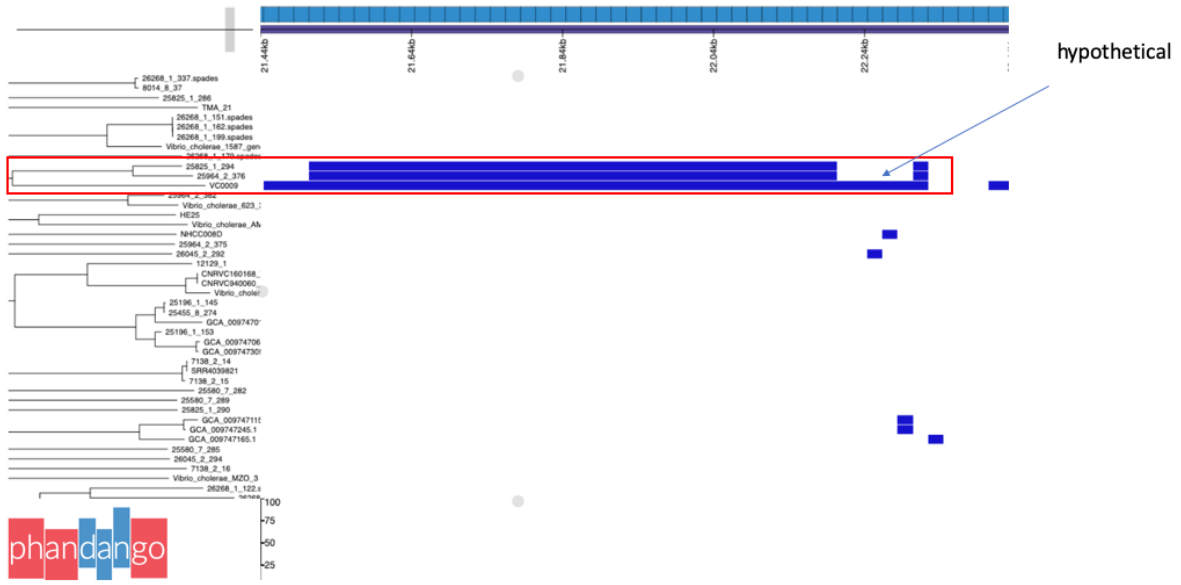


Figure 6.4. A Phandango image showing the phylogenetic tree of the species to the left and the region of interest that is present in VC0009, 25825_1_294 and 25964_2_376 to the right, highlighted with the blue bars indicating gene presence within that region. The red outline highlights the tree and the gene presence within this section of the phylogenetic tree. This is a zoomed in image of region ‘c’ marked in Figure 6.2.

Region ‘c’ is a region of over 44 Kb in size, situated between bases 2767304 and 2811611 in the VC0009 reference genome (Figure 6.2). Apart from VC0009, this locus was only detected in two other isolates that branched with VC0009: 25825_1_294 (H172860570) and 25964_2_376 (H172620738; Figure 6.4) which carried almost identical copies of region C, are non O1/O139 clinical isolates both taken in 2017, from Kenya and the Asian continent, respectively. Notably this region was seen to be absent in both the 7PET strain N16961 and classical strain NCTC10732 (Figure 6.2) and absent from the lineages in which they represent. Region ‘c’ consists of 33 predicted coding sequences (CDSs) with the average G/C content for this region being 45.54%, ranging from 34.85% to 49.4% (The overall average G/C content for VC0009 is 47.42%). The prokka annotations assigned to many of the CDSs within the region are flagellar biosynthesis or type III secretion system related.

To identify the site of insertion of this region, I inspected the comparison file comparing VC0009 to N16961 to identify where this region was inserted in relation to the well-studied N16961 genome, to understand if there had been any site-specific recombination. Interestingly, the region was situated at the same locus as the *tcp* genes (encoding the toxin co-regulated pilus which is necessary for colonisation of the host intestine in addition to being the receptor for infection by the CTXφ) are in N16961 (see the white CDSs in the top genome N16961 in Figure 6.5), this region is more commonly known as VPI-1. The insertion seen in the bottom (VC0009) genome in Figure 6.5 shows the presence of *ssrA* and integrase genes which have both been previously shown to be present at sites of integration. I investigated whether this region had integrated in the same position i.e., at the VPI-1 insertion site in the other two isolates (25825_1_294 and 25964_2_376) harbouring this region. Isolate 25825_1_294 did in fact show the same positioning as VC0009 (in place of VPI-1). However due to an incomplete assembly 25964_2_376, I was unable to conclude the exact insertion site for this isolate. The short read sequence of 25964_2_376 is composed of 62 contigs, of which, this genomic island sits on its own contig (contig number 19) highlighting the possibility of a mis-assembly in this sequence.

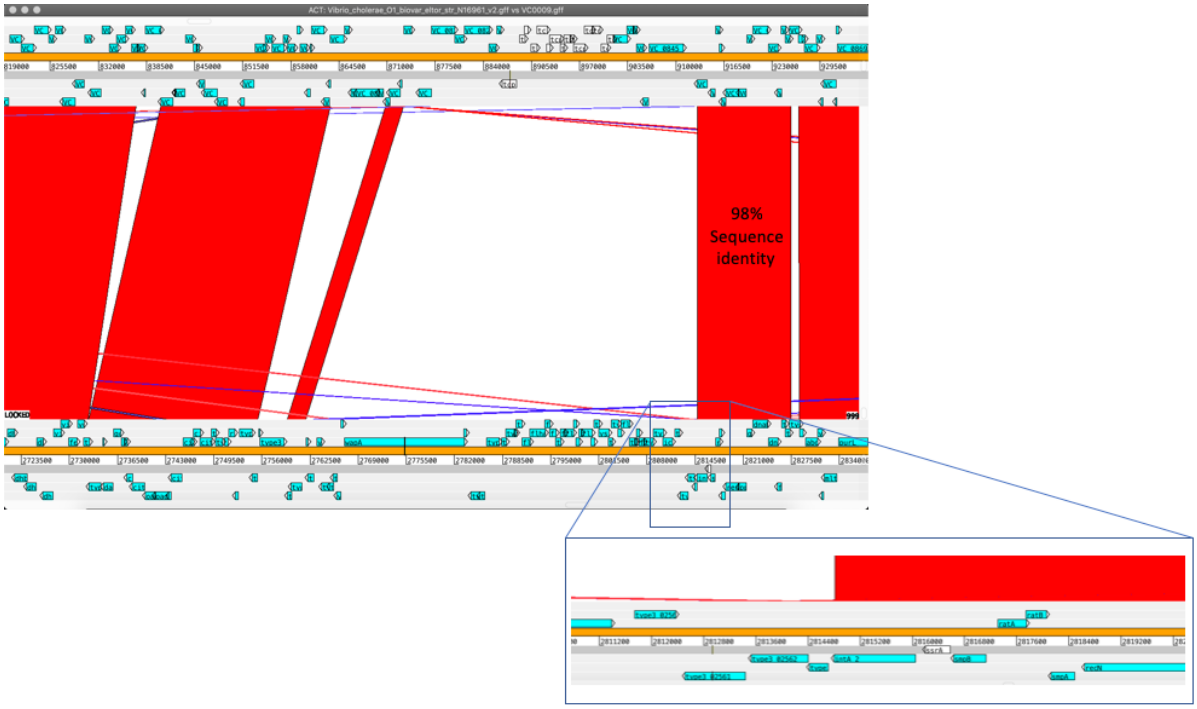


Figure 6.5. A synteny plot between N16961 and VC0009 showing presence of two different regions within the two genomes at the same position – indicating site specific recombination at this particular region. The insert on the VC0009 genome below shows the presence of ssrA and integrase genes which have both been previously shown to be present at sites of integration.

There have been limited studies showing that this locus is a site of genome flux, in 2016 Labbate et al. reported a genomic island in a non O1/O193 strain of *V. cholerae* isolated in Sydney, Australia which encode a CRISPR-Cas module and a T6SS. Interestingly, the region present within VC0009 did not harbour a CRISPR-Cas module [252-254]. As this locus has been associated with genome flux, I thought it would be beneficial to look for any known phage genes in the region using the web tool PHASTER [255, 256] which returned a questionable 'phage' sequence at the end of the region (see Figure 6.6) although this did not describe the complete region within the VC0009 genome.

Figure 6.6. Schematic produced from the PHASTER web tool of the 'questionable' phage present within genome VC0009. Present genes include hypothetical, transposases, integrases and phage like proteins. The region is flanked by two attachment sites (attL and attR) indicated by Att at the ends of the schematic.

Next, I wanted to check the positioning of this inserted region in a genome that did not contain VPI-1, I chose AM-19226 which is used as the reference strain for those isolates containing a T3SSs. The alignments show the insertion in VC0009 at the *ssRA* site (see Figure 6.7), but nothing related to this region within the AM-19226 genome.

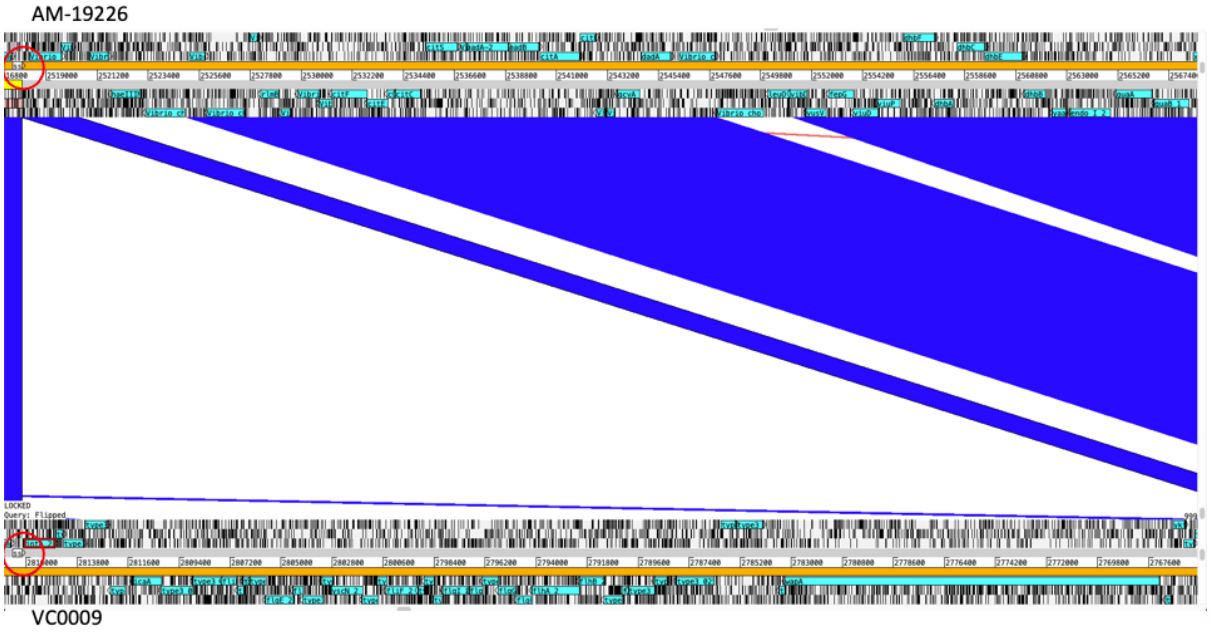


Figure 6.7. A Synteny plot showing ssRA *(red circle) present in both strain AM-19226 (top) and VC0009 (bottom). The genetic insertion is seen in VC0009 as the white area. Blue areas identify similarities between the two genomes.*

Following further analysis assisted by Avril Coghlan and Lowri Williams (Wellcome Sanger Institute) showed that this region included a basic set of proteins required for either an injectosome (T3SS) or flagellar (as they are both structurally similar). However, there are also many hypothetical proteins that would need to be characterised before we could conclude whether it was in fact a novel injectosome locus.

6.3.4 Capsule region

Exploration of the three genomes, both by blastn and Roary blastp analysis showed VC0009 possessed capsule genes, whereas the classical strain NCTC10732 and 7PET strain N16961 lacked the genes that encode capsule. It has been reported that over 85% of non-O1 *V. cholerae* isolates possess the genes to be able to produce capsule. Volunteer studies with non-O1 *V. cholerae* have shown that the presence of capsule appeared to mask certain critical surface antigens, resulting in a decreased host immune response [257].

Given its potential importance for virulence and because it differentiated the three selected isolates (N16961, NCTC10732 and VC0009), I wanted to investigate the nature of the VC0009 capsule in greater detail by generating high resolution images of the three bacterial strains after high pressure freezing followed by transmission electron microscopy of the bacterial isolates (this was conducted with the help of David Goulding at the Wellcome Sanger Institute). This high pressure freezing allows the ultrastructure of the bacteria to remain intact without the likelihood of ice crystals breaking down important structures such as organelles (see chapter 2 section 2.7.4). Figure 6.8 shows the transmission electron microscopy images following the high pressure freezing of the three selected bacterial isolates, a) shows the classical strain NCTC10732, b) shows the 7PET strain N16961 and c) shows the strain harbouring a T3SS, VC0009. The small uniform circles (shown within the red circle) surrounding VC0009 are characteristic of extracellular vesicles (EVs) and the darker shading surrounding the VC0009 bacteria (shown with the red rectangle) denotes presence of capsule and formation of a biofilm. Both features are absent from the NCTC10732 and N16961 bacterial images. Although genetically the capsule genes appeared present, we had hoped that we would be able to visualise the capsule with such high-resolution imaging, we did not know that there would be so many EVs present in the VC0009 strain. This highlights the benefit of being able to not only look at the genomes of the bacteria but also being able to have the opportunity to phenotype the bacteria to allow us to visualise directly the unknown properties of the bacteria.

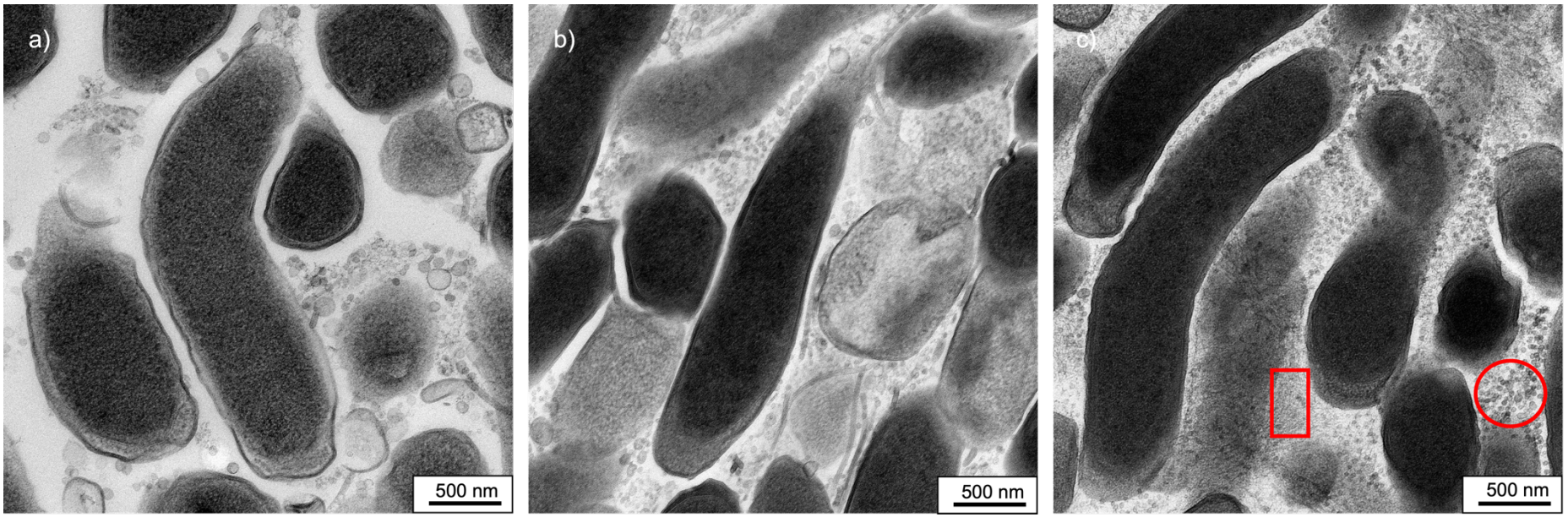


Figure 6.8. Transmission Electron Microscopy of the selected V. cholerae *bacterial isolates. A) illustrates the classical strain NCTC10732, b) shows the 7PET strain N16961 and c) shows the strain harbouring a T3SS, VC0009. There are many more smaller particles surrounding VC0009 (highlighted with the red circle), these particles are extracellular vesicles (EVs), characterised by their uniform size and shape. The darker shading surrounding VC0009 (shown within the red rectangle) denotes capsule and biofilm formation, which is absent from NCTC10732 and N16961.*

6.4 Based on the phylogeny and comparative genomic analysis- were the chosen isolates representative of their lineages?

By sequencing and comparing the selected isolates to themselves and to the genetic diversity held within genomes across the species I was able to answer the important question as to whether the three selected strains were representative of their respective lineages? From the phylogenetic tree (Figure 5.1) it is clear that 7PET strain N16961 is placed within wave one of the clonal 7PET lineage and has the same genetic characteristics as other 7PET isolates (as indicated by the phylogeny, N16961 is one of 175 7PET isolates within this phylogeny), N16961 has also been well characterised. The same was true of classical strain NCTC10732, being part of a clonal lineage with 27 classical isolates present in the phylogeny which also includes the classical reference strain O395 (used previously to align the hybrid sequence NCTC10732- see chapter 5 of this thesis for further details. NCTC10732 sits only three isolates away from O395.

The VC0009 isolate was selected because it originated as a clinical isolate, encoded a T3SS and was phylogenetically distinct from both the 7PET and classical lineages and so has scientific relevance.

However, from the phylogenetic analysis whilst VC0009 stands apart as a non-classical and non-7PET lineages strain located on its own branch within the phylogeny. More importantly VC0009 has many unique genetic differences that are not broadly represented across the phylogeny (some of which have been highlighted in this chapter). In order to definitively know it is representative, the experiments that have taken place over the course of this project would need to be repeated on a wider variety of isolates within the phylogeny. There are interesting genetic regions that I have highlighted in this chapter see section 6.3.3 that are present in other clinical isolates that were isolated in varying regions of the world that would be worth characterising in order to help answer if VC0009 is representative of its lineage. It is important to note that the sequencing and comparative genomics were performed in concurrently and so the nature of VC0009 only became apparent once the detailed comparative genomics took place. Despite this and given that this is a clinically derived strain all three isolates N16961, NCTC10732 and VC0009 were used in infection studied in iHOs to look for differences in early colonisation and pathology, as described below.

6.5 Infection of organoids with characterised bacterial isolates

Whilst it was crucial to characterise the genomic differences between the three isolates selected for this PhD, and to map these differences across the species phylogeny, it was also fundamentally important to identify any phenotypic differences within these isolates when they were microinjected into the lumen of our small intestinal organoid model.

Firstly, for the development of the organoid model, it was important to engineer fluorescently labelled bacteria that could then be used and tracked in organoid infection studies (see chapter 2 section). In brief, the selected strains (N16961, NCTC10732 and VC0009) were screened for any antibiotic resistance genes using ResFinder version 4.1 [258, 259]. This showed when screening for 25 different antibiotic classes containing 96 different antibiotics (data summarised in Table 6-1) that both the classical strain NCTC10732 and strain VC0009 possessed no known resistance genes. Conversely, 7PET strain N16961 showed the presence of a resistance gene predicted to confer resistance to chloramphenicol (*catB9*). This is consistent with the position of this strain in the 7PET tree and the acquisition of the *catB9* gene conferring chloramphenicol resistance in the 7PET lineage as has been previously reported [35]. Importantly, this genotype does not translate into phenotypic resistance, in essence being a silent chloramphenicol acetyl transferase mutant gene [260] which confers no phenotypic chloramphenicol resistance within the laboratory.

Table 6-1. A summary table of the antibiotic resistance of bacterial strains N16961, NCTC10732 and VC0009 as recorded by ResFinder version 4.1

Antibiotic	Antibiotic class	N16961	NCTC10732	VC0009
metronidazole	nitroimidazole	No	No	No
ethidium bromide	quaternary ammonium compound	No	No	No
chlorhexidine	quaternary ammonium compound	No	No	No
benzylkonium chloride	quaternary ammonium compound	No	No	No
cetylpyridinium chloride	quaternary ammonium compound	No	No	No
temperature	heat	No	No	No
clindamycin	lincosamide	No	No	No
lincomycin	lincosamide	No	No	No
fosfomycin	fosfomycin	No	No	No
formaldehyde	aldehyde	No	No	No
tylosin	macrolide	No	No	No
erythromycin	macrolide	No	No	No
telithromycin	macrolide	No	No	No
carbomycin	macrolide	No	No	No
azithromycin	macrolide	No	No	No
oleandomycin	macrolide	No	No	No
spiramycin	macrolide	No	No	No
chloramphenicol	amphenicol	Resistant catB9_AF462019	No	No
florfenicol	amphenicol	No	No	No
sulfamethoxazole	folate pathway antagonist	No	No	No

trimethoprim	folate pathway antagonist	No	No	No
fusidic acid	steroid antibacterial	No	No	No
unknown beta-lactam	beta-lactam	No	No	No
ampicillin	beta-lactam	No	No	No
amoxicillin	beta-lactam	No	No	No
ticarcillin	beta-lactam	No	No	No
meropenem	beta-lactam	No	No	No
piperacillin+clavulanic acid	beta-lactam	No	No	No
cefotaxime	beta-lactam	No	No	No
cefotaxime+clavulanic acid	beta-lactam	No	No	No
cephalotin	beta-lactam	No	No	No
piperacillin+tazobactam	beta-lactam	No	No	No
ticarcillin+clavulanic acid	beta-lactam	No	No	No
cefixime	beta-lactam	No	No	No
aztreonam	beta-lactam	No	No	No
ampicillin+clavulanic acid	beta-lactam	No	No	No
ceftazidime+avibactam	beta-lactam	No	No	No
piperacillin	beta-lactam	No	No	No
penicillin	beta-lactam	No	No	No
cefoxitin	beta-lactam	No	No	No
ceftriaxone	beta-lactam	No	No	No
temocillin	beta-lactam	No	No	No
cephalothin	beta-lactam	No	No	No

ertapenem	beta-lactam	No	No	No
ceftazidime	beta-lactam	No	No	No
imipenem	beta-lactam	No	No	No
cefepime	beta-lactam	No	No	No
amoxicillin+clavulanic acid	beta-lactam	No	No	No
tiamulin	pleuromutilin	No	No	No
quinupristin	streptogramin b	No	No	No
virginiamycin s	streptogramin b	No	No	No
pristinamycin ia	streptogramin b	No	No	No
mupirocin	pseudomonic acid	No	No	No
nalidixic acid	quinolone	No	No	No
unknown quinolone	quinolone	No	No	No
ciprofloxacin	quinolone	No	No	No
fluoroquinolone	quinolone	No	No	No
hydrogen peroxide	peroxide	No	No	No
ceftiofur	under_development	No	No	No
spectinomycin	aminocyclitol	No	No	No
rifampicin	rifamycin	No	No	No
linezolid	oxazolidinone	No	No	No
teicoplanin	glycopeptide	No	No	No
vancomycin	glycopeptide	No	No	No
tetracycline	tetracycline	No	No	No
tigecycline	tetracycline	No	No	No
minocycline	tetracycline	No	No	No
doxycycline	tetracycline	No	No	No
colistin	polymyxin	No	No	No
pristinamycin iia	streptogramin a	No	No	No
dalfopristin	streptogramin a	No	No	No
virginiamycin m	streptogramin a	No	No	No

quinupristin+dalfopristin	streptogramin a	No	No	No
gentamicin	aminoglycoside	No	No	No
kasugamycin	aminoglycoside	No	No	No
arbekacin	aminoglycoside	No	No	No
amikacin	aminoglycoside	No	No	No
kanamycin	aminoglycoside	No	No	No
ribostamycin	aminoglycoside	No	No	No
fortimicin	aminoglycoside	No	No	No
tobramycin	aminoglycoside	No	No	No
hygromycin	aminoglycoside	No	No	No
astromicin	aminoglycoside	No	No	No
sisomicin	aminoglycoside	No	No	No
apramycin	aminoglycoside	No	No	No
butirosin	aminoglycoside	No	No	No
lividomycin	aminoglycoside	No	No	No
butiromycin	aminoglycoside	No	No	No
unknown aminoglycoside	aminoglycoside	No	No	No
paromomycin	aminoglycoside	No	No	No
neomycin	aminoglycoside	No	No	No
streptomycin	aminoglycoside	No	No	No
dibekacin	aminoglycoside	No	No	No
netilmicin	aminoglycoside	No	No	No
isepamicin	aminoglycoside	No	No	No
bleomycin	aminoglycoside	No	No	No

Hence, these data showed all three strains were sensitive to ampicillin. To create fluorescently labelled derivative strains, N16961, NCTC10732 and VC0009 were transfected with the pDiGc plasmid (plasmid map available in chapter 2 section 2.7.1). Following transfection, the bacteria were selected by growing on LB plates containing ampicillin, the

resultant bacteria should express the green fluorescence protein (GFP). As the GFP on the pDiGc plasmid is under an ampicillin promoter, so it was pertinent to check that the bacterial isolates were not already ampicillin resistant.

For the iHO infection studies bacterial cultures were prepared by streaking LB plates (containing ampicillin) with the chosen bacterial isolate and incubating over night at 37 °C. The following day, the bacteria were extracted from the LB plates using a loop and placed into a universal containing PBS, the density of the suspension was then measured on a spectrophotometer to a desired optical density of 0.2 at 600 nm. 1000 μl of the bacterial suspension was then centrifugated at 16,000 x g for 5 minutes in order to produce a bacterial pellet. The pellet was then resuspended in 25 μl of PBS and 25 μl of trypan blue (in order to visually identify which organoids have been microinjected) the bacterial inoculum was then microinjected into the lumen of the small intestinal organoids and incubated for the desired period of time. It was important to quantify the number of bacterial cells that were being injected into the lumen of the organoids. To do this, once the organoids had been microinjected, an Eppendorf containing 100 μl of PBS over the end of the microinjection needle was held and pressed 'inject' on the injector system, in the same fashion as I would when injecting the organoids. I did this three times per bacterial strain, every time I conduced an experiment. I then plated this on ampicillin plates to quantify the number of bacteria that had been injected to the lumen of the organoids- labelled as 'inject on Figure 6.9 and Figure 6.10.

As each of the three strains of interest had been transformed with pDiGc, the bacteria constitutively expressed GFP and were visualised on a fluorescence microscope (EVOS FL2) using the green channel. This enabled me to not only see the organoids that were injected with the naked eye due to the trypan blue (that was used to identify microinjected organoids with the naked eye- as previously used in the CT experiments in Chapter 3 and 4 of this thesis) marking the organoids, but also to visualise the green, fluorescent bacteria within the lumen of the organoids once they had been injected.

I also wanted to understand what was happening to the bacteria once they were inside the organoid lumen. To do this, I manually isolated injected organoids from the Matrigel and

pipetted the injected organoid into 100 µl of PBS multiple times to disrupt the integrity of the organoid- subsequently disaggregating the organoid, then pipetting the suspension of disaggregated organoid on to LB to enable me to count the bacterial colonies and quantify the number of live bacteria in the lumen of the organoids at any given timepoint. First, I conducted this experiment using passage 16 organoids (which mostly conform to the cystic phenotype previously discussed), the results can be seen in Figure 6.9. All three strains were injected at comparable levels, however, over time, fewer colonies of VC0009- approximately 10-fold less (black triangles) were recovered from the organoids compared with both the classical NCTC10732 (red circles) and the 7PET N16961 (green squares). There also seems to be a plateau in colony forming units (CFUs) at T2 for the classical strain NCTC10732 and the 7PET strain N16961. Further timepoints need to be taken to assess whether this is indeed the case or not.

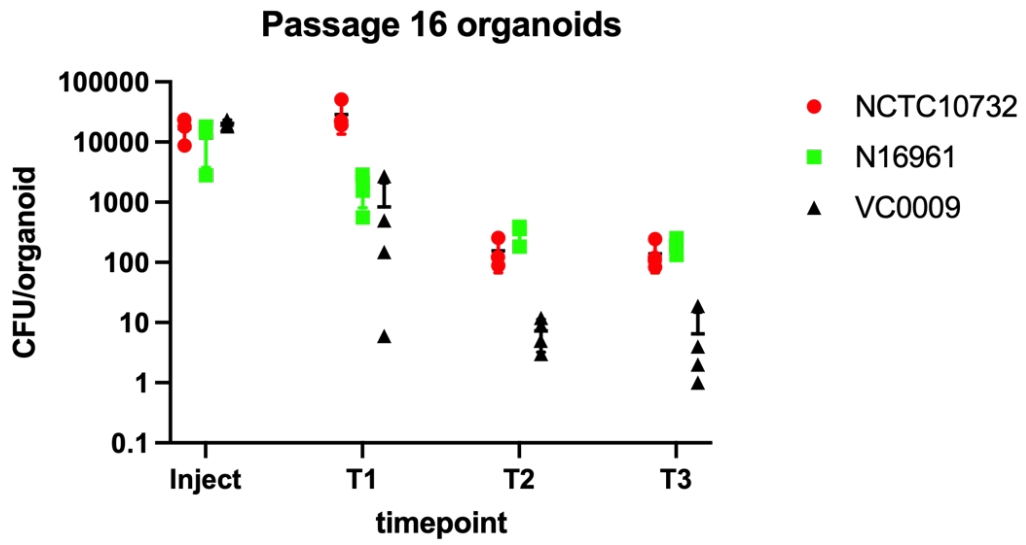


Figure 6.9. A graph illustrating the number of CFUs per organoid at a given time. The red circles show data from the classical strain NCTC10732, the green squares show data from the 7PET strain N16961 and the black triangles show data from the strain harbouring a T3SS, VC0009. The Y axis depicts CFUs per organoid and the X axis shows the timepoint at which the organoids were isolated and plated. The timepoints T1-T5 show hourly timepoints collected over the course of the project- T1 is one hour post injection etc.

Next, based on the observations in chapters 3 and 4 of this thesis on the two different morphologies present in the differentiated organoids, I wanted to understand if the passage number affected the results of the infection studies. The experiment was conducted using passage 17 organoids, derived from the same cell batch, and sampling for longer. The experimental protocol was identical to that that had been conducted for the passage 16 organoids. Figure 6.10 shows the resultant data for this replicate experiment (at passage 17). As had been seen before (Figure 6.9) all bacterial strains were injected into the lumen of the organoids at a comparable inoculum (approximately 10,000 CFU per organoid).

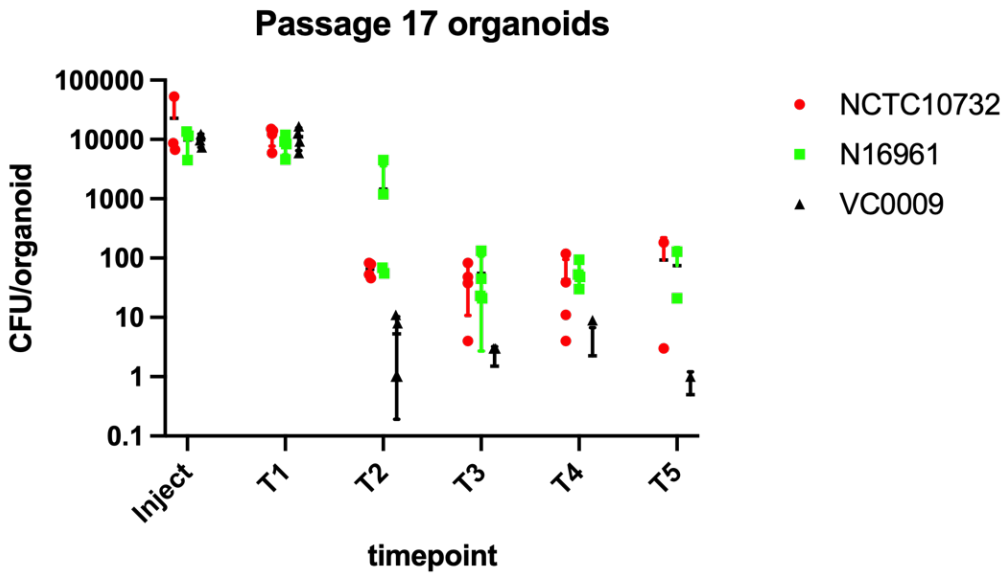


Figure 6.10. A graph showing passage 17 organoids injected with the three chosen bacterial strains. Red circles show data for the classical strain NCTC10732, green squares show data for the 7PET strain N16961 and black triangles illustrate CFU counts for the strain harbouring a T3SS, VC0009. The timepoints T1-T5 show hourly timepoints collected over the course of the project- T1 is one hour post injection etc.

The additional timepoints, T4 and T5 allowed me to differentiate the dynamics of VC0009, the classical and 7PET strains more clearly. No live colonies of VC0009 grew after T5, whereas for both NCTC10732 and N16961 viable bacteria were see at the T5 timepoint.

Infected organoids were fixed and immunostained (see chapter 2 section 2.3.1). As the bacteria were already visible in the green channel, I had the opportunity to stain with

something that could be visualised in the red and blue channels. As we have previously seen an association with CT and UEA1 (chapter 4 of this thesis) I took this opportunity to use UEA1 rhodamine for the bacterial injected organoids (as the bacteria constitutively expressed GFP so showed green via microscopy, the rhodamine and therefore red of the UEA1 was clear to visualise. I then stained the nuclei with DAPI as previously used and imaged the organoids on the confocal microscope (Leica SP8). Figure 6.11 shows the images captured with the confocal microscope.

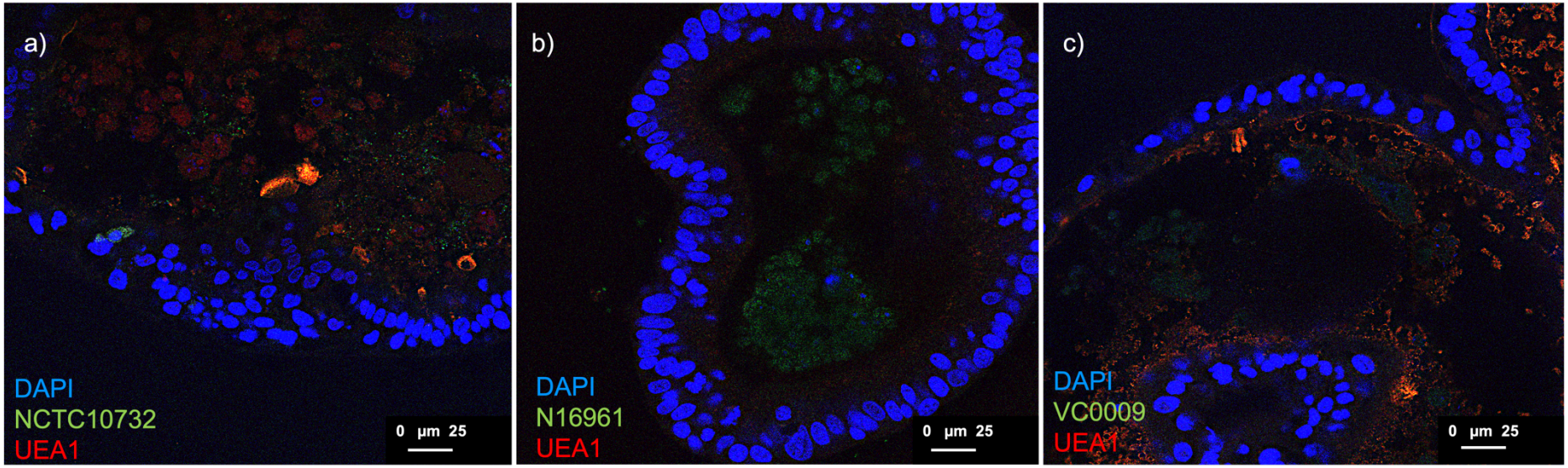


Figure 6.11. Confocal microscopy images of GFP labelled bacterial strains with GFP shown in green, UEA1 labelled with rhodamine, shown in red and nuclei stained with DAPI shown in blue. A) shows the localisation of the classical strain NCTC10732 (green) with the red labelled UEA1, b) shows the 7PET strain N16961 (green) with red labelled UEA1 and c) shows the strain harbouring a T3SS, VC0009 (green) with red labelled UEA1. All organoids have been stained with DAPI to show the presence of nuclei. Any co-localisation between the green bacterial strains and UEA1 appear orange in colour.

Figure 6.11a) shows the presence of the classical strain NCTC10732 in green and UEA1 (as previously used over the course of this project- see Chapters 3 and 4) in red with nuclei stained with DAPI in blue. There is co-localisation of the green NCTC10732 bacteria with the red UEA1 as shown by orange staining. Figure 6.11b shows the 7PET strain N16961 in green and UEA1 in red with nuclei stained with DAPI, there is no co-localisation between the bacteria and the UEA1 in this strain that can be visualised with the staining methods used. Figure 6.11c shows the strain harbouring a T3SS, VC0009 in green and UEA1 stained with red, and the organoid nuclei stained in blue with DAPI. There appears to be a co-localisation with the bacteria and UEA1 in the VC0009 strain in a similar fashion to that seen in the classical strain NCTC10732. This staining pattern was surprising as I had expected the pandemic strains to show more of a similarity to one another than the non-pandemic strain. The association of VC0009 and UEA1 was also surprising as UEA1 has been shown to become more disperse after administration of cholera toxin, we know that VC0009 lacks the ability to encode cholera toxin, whereas N16961 is capable of encoding cholera toxin but shows no association between the bacteria and UEA1. It is important to note when looking at the organoids once they were fixed (after 24 hours) in the phase channel of the EVOS FL2 auto microscope. Figure 6.12 shows these phase images. It can be seen clearly that the darker organoids in Figure 6.12b and Figure 6.12c are those that have been microinjected with bacteria (this was confirmed by switching the phase to the green channel prior to taking the photograph in order to visualise the green fluorescent bacteria). Those organoids that were injected in Figure 6.12a do not appear darker, but they have increased considerably in volume compared to those that were not injected.

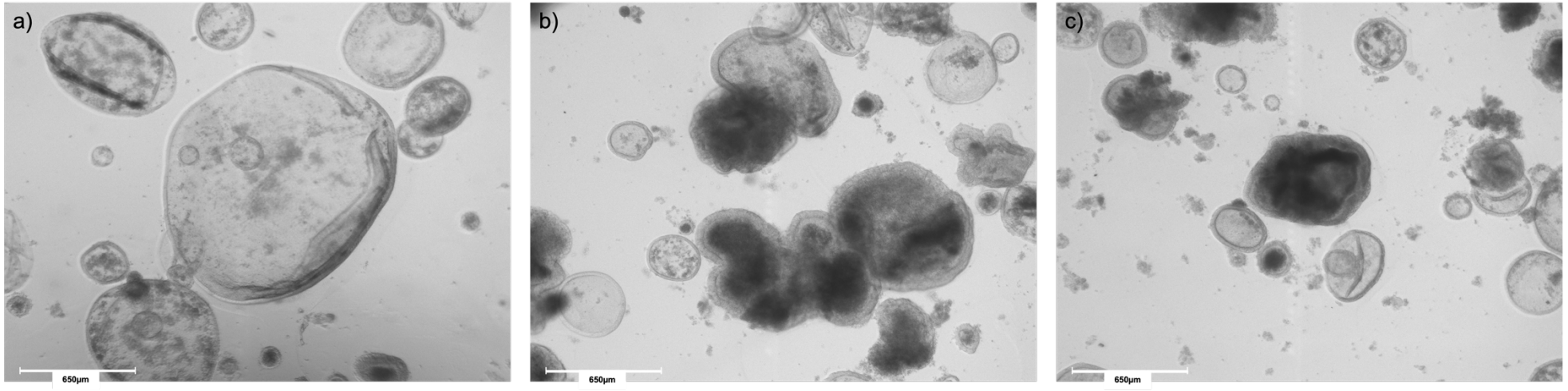


Figure 6.12. Phase images of small intestinal organoids microinjected with a) Classical strain NCTC10732, b) 7PET strain N16961 and c) T3SS harbouring VC0009. It can be seen more clearly in (b) and (c) which organoids had been injected with bacterium due to the dark shading. organoids that were injected with the classical strain NCTC10732 in (a) in a have increased dramatically in size compared with those injected in (b) and (c). Taken using EVOS FL auto microscope.

The swelling of the organoids injected with classical strain NCTC10732 was not surprising as I have previously discussed the ability of classical strains of *V. cholerae* to express toxin constitutively whereas 7PET strains need a complex set of growth conditions to enable toxin expression. The swelling that is seen in organoids injected with NCTC10732 closely resembles that of the cholera toxin only treated organoids within chapter 4 of this thesis.

6.6 Discussion

This chapter of the thesis aimed to catalogue the genetic regions of difference in addition to any gross phenotypic differences witnessed between the three bacterial strains upon microinjection into the organoid lumen. I have seen both genetic differences that we can phenotypically visualise e.g., the presence of capsule in strain VC0009 as well as phenotypic differences that cannot be explained using the bacterial genomes at this point.

The selection of VC0009 may not have been the most representative selection for a strain harbouring a T3SS. However, as the genetic characterisation and the phenotyping were done concurrently, making the decision to change strain once a lot of time was invested in the isolate would have been difficult. I was also limited by the live isolates that were available to the study. Being a Schedule 5 organism, shipping of live *V. cholerae* isolates is an added complexity. However, VC0009 is still a clinical strain, holds scientific relevancy and it was important to characterise any phenotypic differences between the three chosen isolates. Further characterisation would need to take place on the regions of differences found in this project to ascertain whether the genome encodes an additional injectosome or flagellar genes that may be of interest to the wider scientific community.

The reduction in CFUs in VC0009 compared with N16961 and NCTC10732 (as seen in Figure 6.9 and Figure 6.10) was intriguing as we know from previous work in this chapter that VC0009 possesses a capsule, and that a capsule should form a protective layer around the bacteria, however we are recovering fewer CFUs after injection into the organoid. The presence of capsule is unusual for *V. cholerae* strains other than those in the O139 serogroup, however literature relating to the serogroup O139 suggests that there could be a link to the pathogenicity of O139 *V. cholerae* due to the presence of capsule [261], and that synthesis of the capsular material is required for efficient colonisation of the suckling mouse small intestine [262]. Although we do see colonisation by VC0009 within the organoid model, it is at a lower level than that of N16961 and NCTC10732.

Further work would be needed to characterise the differences found amongst these three bacterial isolates upon injection into the organoid lumen, including possible host response (as was conducted with cholera toxin treated organoids). I believe that it would also be feasible

to use a wider selection of *V. cholerae* bacterial strains to look for phenotypic differences in colonisation etc. within the small intestinal organoid system. Additionally, the selected isolates could be tested in further surrogate/model systems such as wax moth larvae, *C. elegans* and zebra fish used for other bacterial pathogens. Using a more comprehensive selection of bacterial isolates may also show any clear lineage specific colonisation patterns or differences as opposed to strain specific differences within this small intestinal organoid model.

Chapter 7:

Future Works

Summary:

This thesis aimed to explore the use of an induced human intestinal organoid model of cholera. Here, we used both the live bacterium and the commercially available purified cholera toxin to understand host transcriptomic responses to the purified cholera toxin and differences in colonisation patterns across diverse *V. cholerae* isolates. I have also explored strain-specific differences in bacterial colonisation and pathology using this model. In doing so, *V. cholerae* isolates representing three different lineages were directly injected into the iPSC derived small intestinal organoid model to understand differences in pathology and colonisation. Comparative genomics was conducted to catalogue the genetic differences that distinguish the three isolates to better understand observed differences in pathology.

7 Introduction

The iPSC derived small intestinal organoid model has revealed important insights into the pathology of several pathogens, including here with *V. cholerae.* Despite this, there are constrains and limitations with this model, one of which being the exhibition of dual morphologies. This dual morphology phenomenon has, to our knowledge, not been previously mentioned by any other research group. These organoids represent a closed model system and so this only allowed for a short window of infection. Therefore, only early colonisation patterns could be studied as opposed to monitoring longer term infection. Given this there are the key questions that can still be answered with this model:

7.1 The nature of the 'hyper infectious' state of *V. cholerae*

Previous studies have shown that *V. cholerae* (along with other enteric bacteria) are able to occupy a hyper infectious state after infection and upon being shed from human hosts. This has also been shown to be true when using animal models to 'prime' the bacteria [263]. It would be of interest to this model to investigate whether this 'priming' of the bacteria through an animal model would reduce the load of bacteria needed to gain an infection in the organoids.

7.2 Interactions with the host using host mutant cell lines

The use of induced small intestinal organoids has opened a realm of possibilities for the ability to create host mutations using technologies such as CRISPR-Cas9.

This investigation specifically would have benefited greatly from the creation of both heterozygous and homozygous mutations in the CFTR, which result in cystic fibrosis allele carriage and cystic fibrosis respectively. The ability to create isogenic controls reduces the impact of host variation and focuses on the specific mutation created and provides for robust genetic controls. CRISPR-Cas would also facilitate making complemented cell lines in which the mutation is corrected. An interesting mutation to make would be in the blood group loci, from this study. We have shown that the Kolf-2 donor is blood group A, it would be of interest to this study and to the wider scientific community if we could engineer Kolf-2 cells that possess blood group O.

7.3 Larger scale infection phenotyping of bacterial isolates

This project has provided a model that result in measurable outputs in a human cell-based system, not only reducing the number of animals that would be used to study cholera in the future but also providing a screen to better understand isolate-specific differences that have not been reported in human cell technologies. This project has proven that it is possible to microinject *V. cholerae* into the lumen of the induced intestinal organoids and to measure phenotypic response by electron microscopy to investigate cellular damage and any isolate specific differences in colonisation patterns.

In summary, this thesis has shown that we are able to measure host response to cholera toxin within the iPSC derived small intestinal organoid model while identifying and acknowledging the differences in organoid morphology on a transcriptomic level. This study has also shown that it is possible to highlight differences between bacterial isolates not only genetically but also providing the opportunity to link these genetic distinctions to phenotypic differences in terms of bacterial colonisation within the organoid and gross differences between the bacterial isolates using electron microscopy.

The organoid model fits with the Global Task Force on Cholera Control (as discussed in chapter 1, section 1.3) under the surveillance and reporting arm by being able to understand that cholera as a disease has two very different clinical outcomes dependant on whether the disease was caused by pandemic isolates or non-pandemic isolates. Using the organoid model, we can model non-pandemic cholera to try to understand why and how it causes disease and how this disease is different to pandemic cholera and to potentially identify virulence markers of non-pandemic bacterial isolates. The model also fits under the leadership and coordination arm of the strategy by being able to quickly take relevant bacterial isolates that are causing disease and study them in a human derived model of disease, being able to see if the bacterial isolates of interest cause similar disease to pandemic cholera and being able to inform others of the findings to influence how the outbreak is controlled.

List of Appendices

Appendix name	Excel file name
Appendix A	4_hour_CT_budded_non-injected_v_4_hour_CT_cystic_non_injected_degs
Appendix B	4_hour_TB_budded_non-injected_v_4_hour_TB_cystic_non_injected_degs
Appendix C	24_hour_CT_budded_non-injected_v_24_hour_CT_cystic_non_injected_degs
Appendix D	24_hour_TB_budded_non-injected_v_24_hour_TB_cystic_non_injected_degs
Appendix E	48_hour_CT_budded_non-injected_v_48_hour_CT_cystic_non_injected_degs
Appendix F	48_hour_TB_budded_non-injected_v_48_hour_TB_cystic_non_injected_degs
Appendix G	4_hour_CT_budded_injected_v_4_hour_CT_budded_non_injected_degs
Appendix H	4_hour_CT_budded_injected_v_4_hour_Trypan_blue_budded_injected_degs
Appendix I	4_hour_CT_cystic_injected_v_4_hour_CT_non_injected_degs
Appendix J	4_hour_CT_cystic_injected_v_4_hour_Trypan_blue_cystic_injected_degs
Appendix K	4_hour_Trypan_blue_budded_injected_v_4_hour_Trypan_blue_budded_non_injected_degs
Appendix L	4_hour_Trypan_blue_cystic_injected_v_4_hour_Trypan_blue_cystic_non_injected_degs
Appendix M	24_hour_CT_budded_injected_v_24_hour_CT_budded_non_injected_degs
Appendix N	24_hour_CT_budded_injected_v_24_hour_Trypan_blue_budded_injected_degs
Appendix O	24_hour_CT_cystic_injected_v_24_hour_CT_cystic_non_injected_degs
Appendix P	24_hour_CT_cystic_injected_v_24_hour_Trypan_blue_cystic_injected_degs
Appendix Q	24_hour_Trypan_blue_budded_injected_v_24_hour_Trypan_blue_budded_non_injected_degs
Appendix R	24_hour_Trypan_blue_cystic_injected_v_24_hour_Trypan_blue_cystic_non_injected_degs

Appendix S	48_hour_CT_budded_injected_v_48_hour_CT_budded_non_injected_degs
Appendix T	48_hour_CT_budded_injected_v_48_hour_TB_injected_all_genes
Appendix U	48_hour_CT_budded_injected_v_48_hour_TB_budded_injected_degs
Appendix V	48_hour_CT_cystic_injected_v_48_hour_CT_cystic_non_injected_degs
Appendix W	48_hour_CT_cystic_injected_v_48_hour_TB_cystic_injected_degs
Appendix X	48_hour_CT_cystic_injected_v_48_hour_TB_injected_degs
Appendix Y	48_hour_Trypan_blue_budded_injected_v_48_hour_Trypan_blue_budded_non_injected_degs
Appendix Z	48_hour_Trypan_blue_cystic_injected_v_48_hour_Trypan_blue_cystic_non_injected_degs
Appendix AA	Isolates used to create the phylogeny with Accession numbers, year and geographical location of isolation

1. Clark, A., et al., *Estimating global, regional and national rotavirus deaths in children aged < 5 years: Current approaches, new analyses and proposed improvements.* Plos One, 2017. **12**(9).
2. Ali, M., et al., *Updated global burden of cholera in endemic countries.* PLoS Negl Trop Dis, 2015. **9**(6): p. e0003832.
3. Harris, J.B., et al., *Cholera.* Lancet, 2012. **379**(9835): p. 2466-2476.
4. Snow, J., *On the Mode of Communication of Cholera.* Edinb Med J, 1856. **1**(7): p. 668-670.
5. Bynum, W., *In retrospect: On the Mode of Communication of Cholera.* Nature, 2013. **495**(7440): p. 169-170.
6. *Microscopic Observations and Pathological Deductions on Asiatic Cholera.* The British and Foreign Medico-Chirurgical Review, 1855. **16**(31): p. 144-145.
7. Pollitzer, R., S. Swaroop, and W. Burrows, *Cholera.* Monogr Ser World Health Organ, 1959. **58**(43): p. 1001-19.
8. Kaper, J.B., J.G. Morris, Jr., and M.M. Levine, *Cholera.* Clin Microbiol Rev, 1995. **8**(1): p. 48-86.
9. Vezzulli, L., et al., *Environmental reservoirs of Vibrio cholerae and their role in cholera.* Environ Microbiol Rep, 2010. **2**(1): p. 27-33.
10. Colwell, R.R. and A. Huq, *Environmental reservoir of Vibrio cholerae. The causative agent of cholera.* Ann N Y Acad Sci, 1994. **740**: p. 44-54.
11. Martinelli Filho, J.E., et al., *Vibrio cholerae O1 detection in estuarine and coastal zooplankton.* Journal of Plankton Research, 2011. **33**(1): p. 51-62.
12. Colwell, R.R., et al., *Reduction of cholera in Bangladeshi villages by simple filtration.* Proc Natl Acad Sci U S A, 2003. **100**(3): p. 1051-5.
13. Shimada, T., et al., *Extended serotyping scheme forVibrio cholerae.* Current Microbiology, 1994. **28**(3): p. 175-178.
14. Gardner, A.D. and K.V. Venkatraman, *The Antigens of the Cholera Group of Vibrios.* J Hyg (Lond), 1935. **35**(2): p. 262-82.
15. Cabral, J.P., *Water microbiology. Bacterial pathogens and water.* Int J Environ Res Public Health, 2010. **7**(10): p. 3657-703.
16. Stroeher, U.H., et al., *Serotype Conversion in Vibrio-Cholerae O1.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(7): p. 2566-2570.
17. Ito, T., et al., *Mutations in the rfbT gene are responsible for the Ogawa to inaba serotype conversion in Vibrio cholerae O1.* Microbiol Immunol, 1993. **37**(4): p. 281-8.
18. Alam, M.T., et al., *Major Shift of Toxigenic V. cholerae O1 from Ogawa to Inaba Serotype Isolated from Clinical and Environmental Samples in Haiti.* PLoS Negl Trop Dis, 2016. **10**(10): p. e0005045.
19. Khan, A.I., et al., *Comparison of clinical features and immunological parameters of patients with dehydrating diarrhoea infected with Inaba or Ogawa serotypes of Vibrio cholerae O1.* Scand J Infect Dis, 2010. **42**(1): p. 48-56.
20. Longini, I.M., Jr., et al., *Epidemic and endemic cholera trends over a 33-year period in Bangladesh.* J Infect Dis, 2002. **186**(2): p. 246-51.
21. Redmond, J.W., M.J. Korsch, and G.D. Jackson, *Immunochemical studies to the O-antigens of vibrio cholerae. Partial characterization of an acid-labile antigenic determinant.* Aust J Exp Biol Med Sci, 1973. **51**(2): p. 229-35.

22. Sakazaki, R. and K. Tamura, *Somatic antigen variation in Vibrio cholerae.* Jpn J Med Sci Biol, 1971. **24**(2): p. 93-100.
23. Wachsmuth, K., P.A. Blake, and Ø. Olsvik, *Vibrio cholerae and cholera : molecular to global perspectives*. 1994, Washington, D.C.: ASM Press. xi, 465 p.
24. Organization, W.H., *Manual for laboratory investigations of acute enteric infections: programme for control of diarrhoeal diseases*. 1987: World Health Organization.
25. Colwell, R.R. and W.M. Spira, *The Ecology of Vibrio cholerae*, in *Cholera*, D. Barua and W.B. Greenough, Editors. 1992, Springer US: Boston, MA. p. 107-127.
26. Hasan, N.A., et al., *Nontoxigenic Vibrio cholerae non-O1/O139 isolate from a case of human gastroenteritis in the U.S. Gulf Coast.* J Clin Microbiol, 2015. **53**(1): p. 9-14.
27. Jenson, D., V. Szabo, and F.H.I.H.H.L.S.R.T. Duke, *Cholera in Haiti and other Caribbean regions, 19th century.* Emerg Infect Dis, 2011. **17**(11): p. 2130-5.
28. JM, K., *The U.N.'s Cholera Admission and What Comes Next*, in *The New York Times*. 2016.
29. CDC. *Cholera- general information*. 2022 13/04/2022 [cited 2022 26/09/2022]; Available from: https://www.cdc.gov/cholera/general/index.html.
30. Safa, A., G.B. Nair, and R.Y. Kong, *Evolution of new variants of Vibrio cholerae O1.* Trends Microbiol, 2010. **18**(1): p. 46-54.
31. Devault, A.M., et al., *Second-pandemic strain of Vibrio cholerae from the Philadelphia cholera outbreak of 1849.* N Engl J Med, 2014. **370**(4): p. 334-40.
32. Colwell, R.R., *Global climate and infectious disease: the cholera paradigm.* Science, 1996. **274**(5295): p. 2025-31.
33. Sack, D.A., et al., *Cholera.* Lancet, 2004. **363**(9404): p. 223-33.
34. Hu, D., et al., *Origins of the current seventh cholera pandemic.* Proc Natl Acad Sci U S A, 2016. **113**(48): p. E7730-E7739.
35. Weill, F.X., et al., *Genomic history of the seventh pandemic of cholera in Africa.* Science, 2017. **358**(6364): p. 785-789.
36. D, B., *History of Cholera.* Current Topics in Infectious Disease, 1992.
37. Gaffga, N.H., R.V. Tauxe, and E.D. Mintz, *Cholera: a new homeland in Africa?* Am J Trop Med Hyg, 2007. **77**(4): p. 705-13.
38. Vugia, D.J., et al., *Epidemic cholera in Trujillo, Peru 1992: utility of a clinical case definition and shift in Vibrio cholerae O1 serotype.* Am J Trop Med Hyg, 1994. **50**(5): p. 566-9.
39. Dziejman, M., et al., *Comparative genomic analysis of Vibrio cholerae: genes that correlate with cholera endemic and pandemic disease.* Proc Natl Acad Sci U S A, 2002. **99**(3): p. 1556-61.
40. Cvjetanovic, B. and D. Barua, *The seventh pandemic of cholera.* Nature, 1972. **239**(5368): p. 137-8.
41. Yoon, S.S. and J.J. Mekalanos, *2,3-butanediol synthesis and the emergence of the Vibrio cholerae El Tor biotype.* Infect Immun, 2006. **74**(12): p. 6547-56.
42. Kovacikova, G., W. Lin, and K. Skorupski, *Dual regulation of genes involved in acetoin biosynthesis and motility/biofilm formation by the virulence activator AphA and the acetate-responsive LysR-type regulator AlsR in Vibrio cholerae.* Mol Microbiol, 2005. **57**(2): p. 420-33.
43. Read, W.a.P., S, *Distribution of V. cholerae and El Tor type starins in certain rural areas in India. Indian Journal of Medical Reasearch*, 1941(29): p. 403-405.

44. Panja, G. and S.K. Ghosh, *Isolation of cholera vibrios from Hooghly river water at Calcutta.* Indian J Med Res, 1947. **35**(1): p. 1.
45. Azurin, J.C., et al., *A long-term carrier of cholera: cholera Dolores.* Bull World Health Organ, 1967. **37**(5): p. 745-9.
46. Mutreja, A., et al., *Evidence for several waves of global transmission in the seventh cholera pandemic.* Nature, 2011. **477**(7365): p. 462-5.
47. Alam, M., et al., *Diagnostic limitations to accurate diagnosis of cholera.* J Clin Microbiol, 2010. **48**(11): p. 3918-22.
48. Siriphap, A., et al., *Characterization and Genetic Variation of Vibrio cholerae Isolated from Clinical and Environmental Sources in Thailand.* PLoS One, 2017. **12**(1): p. e0169324.
49. Bhattacharya, M.K., et al., *Outbreak of Vibrio cholerae non-O1 in India and Bangladesh.* Lancet, 1993. **341**(8856): p. 1346-7.
50. Nair, G.B., et al., *Spread of Vibrio cholerae O139 Bengal in India.* J Infect Dis, 1994. **169**(5): p. 1029-34.
51. Dorman, M.J., et al., *High quality reference genomes for toxigenic and non-toxigenic Vibrio cholerae serogroup O139.* Sci Rep, 2019. **9**(1): p. 5865.
52. Aaron, J.E., et al., *Parathyroid gland hormones in the skeletal development of the ovine foetus: the effect of parathyroidectomy with calcium and phosphate infusion.* Bone Miner, 1992. **16**(2): p. 121-9.
53. Finkelstein, R.A., *Cholera, Vibrio cholerae O1 and O139, and Other Pathogenic Vibrios*, in *Medical Microbiology*, th and S. Baron, Editors. 1996: Galveston (TX).
54. Swerdlow, D.L. and A.A. Ries, *Vibrio cholerae non-O1--the eighth pandemic?* Lancet, 1993. **342**(8868): p. 382-3.
55. Stroeher, U.H., et al., *Genetic rearrangements in the rfb regions of Vibrio cholerae O1 and O139.* Proc Natl Acad Sci U S A, 1995. **92**(22): p. 10374-8.
56. Bik, E.M., et al., *Genesis of the novel epidemic Vibrio cholerae O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis.* EMBO J, 1995. **14**(2): p. 209-16.
57. Bik, E.M., R.D. Gouw, and F.R. Mooi, *DNA fingerprinting of Vibrio cholerae strains with a novel insertion sequence element: a tool to identify epidemic strains.* J Clin Microbiol, 1996. **34**(6): p. 1453-61.
58. Faruque, S.M., et al., *Molecular analysis of rRNA and cholera toxin genes carried by the new epidemic strain of toxigenic Vibrio cholerae O139 synonym Bengal.* J Clin Microbiol, 1994. **32**(4): p. 1050-3.
59. Weill, F.X., et al., *Genomic insights into the 2016-2017 cholera epidemic in Yemen.* Nature, 2019. **565**(7738): p. 230-233.
60. Singh, S., C.C. Kerndt, and D. Davis, *Ringer's Lactate*, in *StatPearls*. 2022: Treasure Island (FL).
61. Organization, W.H. *WHO position paper on Oral Rehydration Salts to reduce mortality from cholera*. 2008 [cited 2022 14th February]; Available from: https://www.who.int/cholera/technical/ORSRecommendationsForUseAtHomeDec2008.pdf.
62. Leibovici-Weissman, Y., et al., *Antimicrobial drugs for treating cholera.* Cochrane Database Syst Rev, 2014(6): p. CD008625.
63. Weil, A.A., et al., *Clinical outcomes in household contacts of patients with cholera in Bangladesh.* Clin Infect Dis, 2009. **49**(10): p. 1473-9.

64. Echevarria, J., et al., *Efficacy and tolerability of ciprofloxacin prophylaxis in adult household contacts of patients with cholera.* Clin Infect Dis, 1995. **20**(6): p. 1480-4.
65. Towner, K.J., et al., *Resistance to antimicrobial agents of Vibrio cholerae E1 Tor strains isolated during the fourth cholera epidemic in the United Republic of Tanzania.* Bull World Health Organ, 1980. **58**(5): p. 747-51.
66. Weber, J.T., et al., *Epidemic cholera in Ecuador: multidrug-resistance and transmission by water and seafood.* Epidemiol Infect, 1994. **112**(1): p. 1-11.
67. Rebaudet, S., et al., *Cholera in coastal Africa: a systematic review of its heterogeneous environmental determinants.* J Infect Dis, 2013. **208 Suppl 1**: p. S98-106.
68. Shah, M.A., et al., *Genomic epidemiology of Vibrio cholerae O1 associated with floods, Pakistan, 2010.* Emerg Infect Dis, 2014. **20**(1): p. 13-20.
69. Domman, D., et al., *Integrated view of Vibrio cholerae in the Americas.* Science, 2017. **358**(6364): p. 789-793.
70. J. Kumate, J.S., G. Gutiérrez,, *Cholera epidemiology in Latin America and perspectives for eradication,.* Bulletin de l'Institut Pasteur,, 1998. **96**(4): p. 217-226.
71. Cash, R.A., et al., *Response of man to infection with Vibrio cholerae. I. Clinical, serologic, and bacteriologic responses to a known inoculum.* J Infect Dis, 1974. **129**(1): p. 45-52.
72. Ritchie, J.M. and M.K. Waldor, *Vibrio cholerae interactions with the gastrointestinal tract: lessons from animal studies.* Curr Top Microbiol Immunol, 2009. **337**: p. 37-59.
73. Waldor, M.K., et al., *Regulation, replication, and integration functions of the Vibrio cholerae CTXphi are encoded by region RS2.* Mol Microbiol, 1997. **24**(5): p. 917-26.
74. Waldor, M.K. and J.J. Mekalanos, *Lysogenic conversion by a filamentous phage encoding cholera toxin.* Science, 1996. **272**(5270): p. 1910-4.
75. Pant, A., B. Das, and R.K. Bhadra, *CTX phage of Vibrio cholerae: Genomics and applications.* Vaccine, 2019.
76. Herrington, D.A., et al., *Toxin, toxin-coregulated pili, and the toxR regulon are essential for Vibrio cholerae pathogenesis in humans.* J Exp Med, 1988. **168**(4): p. 1487-92.
77. Karaolis, D.K., et al., *A bacteriophage encoding a pathogenicity island, a type-IV pilus and a phage receptor in cholera bacteria.* Nature, 1999. **399**(6734): p. 375-9.
78. Karaolis, D.K., et al., *A Vibrio cholerae pathogenicity island associated with epidemic and pandemic strains.* Proc Natl Acad Sci U S A, 1998. **95**(6): p. 3134-9.
79. Karaolis, D.K., R. Lan, and P.R. Reeves, *Molecular evolution of the seventh-pandemic clone of Vibrio cholerae and its relationship to other pandemic and epidemic V. cholerae isolates.* J Bacteriol, 1994. **176**(20): p. 6199-206.
80. Krebs, S.J. and R.K. Taylor, *Protection and attachment of Vibrio cholerae mediated by the toxin-coregulated pilus in the infant mouse model.* J Bacteriol, 2011. **193**(19): p. 5260-70.
81. Taylor, R.K., et al., *Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin.* Proc Natl Acad Sci U S A, 1987. **84**(9): p. 2833-7.
82. Lee, C.A., *Vibrio cholerae TCP: a trifunctional virulence factor?* Trends Microbiol, 1999. **7**(10): p. 391-2; discussion 393.

83. Faruque, S.M., et al., *Induction of the lysogenic phage encoding cholera toxin in naturally occurring strains of toxigenic Vibrio cholerae O1 and O139.* Infect Immun, 1998. **66**(8): p. 3752-7.
84. Said, B., et al., *Detection and differentiation of the gene for toxin co-regulated pili (tcpA) in Vibrio cholerae non-O1 using the polymerase chain reaction.* FEMS Microbiol Lett, 1995. **125**(2-3): p. 205-9.
85. Koch, R., *An Address on Cholera and its Bacillus.* Br Med J, 1884. **2**(1236): p. 453-9.
86. Mekalanos, J.J., et al., *Cholera toxin genes: nucleotide sequence, deletion analysis and vaccine development.* Nature, 1983. **306**(5943): p. 551-7.
87. Holmgren, J., et al., *Interaction of cholera toxin and membrane GM1 ganglioside of small intestine.* Proc Natl Acad Sci U S A, 1975. **72**(7): p. 2520-4.
88. Clemens, J., et al., *New-generation vaccines against cholera.* Nat Rev Gastroenterol Hepatol, 2011. **8**(12): p. 701-10.
89. Viswanathan, V.K., K. Hodges, and G. Hecht, *Enteric infection meets intestinal function: how bacterial pathogens cause diarrhoea.* Nat Rev Microbiol, 2009. **7**(2): p. 110-9.
90. Sears, C.L. and J.B. Kaper, *Enteric bacterial toxins: mechanisms of action and linkage to intestinal secretion.* Microbiol Rev, 1996. **60**(1): p. 167-215.
91. Serrano, A., et al., *Holotoxin disassembly by protein disulfide isomerase is less efficient for Escherichia coli heat-labile enterotoxin than cholera toxin.* Sci Rep, 2022. **12**(1): p. 34.
92. Eidels, L., R.L. Proia, and D.A. Hart, *Membrane receptors for bacterial toxins.* Microbiol Rev, 1983. **47**(4): p. 596-620.
93. van Heyningen, S., *A conjugate of the A1 peptide of chloera toxin and the lectin of Wisteria floribunda that activates the adenylate cyclase of intact cells.* FEBS Letters, 1983. **164**(1): p. 132-134.
94. Merritt, E.A., et al., *Crystal structure of cholera toxin B-pentamer bound to receptor GM1 pentasaccharide.* Protein Sci, 1994. **3**(2): p. 166-75.
95. Dziejman, M., et al., *Genomic characterization of non-O1, non-O139 Vibrio cholerae reveals genes for a type III secretion system.* Proc Natl Acad Sci U S A, 2005. **102**(9): p. 3465-70.
96. Coburn, B., I. Sekirov, and B.B. Finlay, *Type III secretion systems and disease.* Clin Microbiol Rev, 2007. **20**(4): p. 535-49.
97. Tam, V.C., et al., *A type III secretion system in Vibrio cholerae translocates a formin/spire hybrid-like actin nucleator to promote intestinal colonization.* Cell Host Microbe, 2007. **1**(2): p. 95-107.
98. Tam, V.C., et al., *Functional analysis of VopF activity required for colonization in Vibrio cholerae.* mBio, 2010. **1**(5).
99. Okada, N., et al., *Identification and characterization of a novel type III secretion system in trh-positive Vibrio parahaemolyticus strain TH3996 reveal genetic lineage and diversity of pathogenic machinery beyond the species level.* Infect Immun, 2009. **77**(2): p. 904-13.
100. Marlovits, T.C. and C.E. Stebbins, *Type III secretion systems shape up as they ship out.* Curr Opin Microbiol, 2010. **13**(1): p. 47-52.
101. Moraes, T.F., T. Spreter, and N.C. Strynadka, *Piecing together the type III injectisome of bacterial pathogens.* Curr Opin Struct Biol, 2008. **18**(2): p. 258-66.

102. Mueller, C.A., P. Broz, and G.R. Cornelis, *The type III secretion system tip complex and translocon.* Mol Microbiol, 2008. **68**(5): p. 1085-95.
103. Worrall, L.J., E. Lameignere, and N.C. Strynadka, *Structural overview of the bacterial injectisome.* Curr Opin Microbiol, 2011. **14**(1): p. 3-8.
104. Blocker, A., K. Komoriya, and S. Aizawa, *Type III secretion systems and bacterial flagella: insights into their function from structural similarities.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3027-30.
105. Domenighini, M., et al., *Identification of errors among database sequence entries and comparison of correct amino acid sequences for the heat-labile enterotoxins of Escherichia coli and Vibrio cholerae.* Mol Microbiol, 1995. **15**(6): p. 1165-7.
106. Critchley, D.R., J.L. Magnani, and P.H. Fishman, *Interaction of cholera toxin with rat intestinal brush border membranes. Relative roles of gangliosides and galactoproteins as toxin receptors.* J Biol Chem, 1981. **256**(16): p. 8724-31.
107. Orlandi, P.A., D.R. Critchley, and P.H. Fishman, *The heat-labile enterotoxin of Escherichia coli binds to polylactosaminoglycan-containing receptors in CaCo-2 human intestinal epithelial cells.* Biochemistry, 1994. **33**(43): p. 12886-95.
108. Yamada, K.M., et al., *Exogenous gangliosides enhance the interaction of fibronectin with ganglioside-deficient cells.* Exp Cell Res, 1983. **143**(2): p. 295-302.
109. Butterton, J.R., et al., *Development of a germfree mouse model of Vibrio cholerae infection.* Infect Immun, 1996. **64**(10): p. 4373-7.
110. Nygren, E., et al., *Establishment of an adult mouse model for direct evaluation of the efficacy of vaccines against Vibrio cholerae.* Infect Immun, 2009. **77**(8): p. 3475-84.
111. Angelichio, M.J., et al., *Vibrio cholerae intestinal population dynamics in the suckling mouse model of infection.* Infect Immun, 1999. **67**(8): p. 3733-9.
112. Banwell, J.G., et al., *Intestinal fluid and electrolyte transport in human cholera.* J Clin Invest, 1970. **49**(1): p. 183-95.
113. Hughes, K.J., et al., *Isolation and characterization of the Vibrio cholerae acfA gene, required for efficient intestinal colonization.* Gene, 1995. **156**(1): p. 59-61.
114. Franzon, V.L., A. Barker, and P.A. Manning, *Nucleotide sequence encoding the mannose-fucose-resistant hemagglutinin of Vibrio cholerae O1 and construction of a mutant.* Infect Immun, 1993. **61**(7): p. 3032-7.
115. Baselski, V.S., S. Upchurch, and C.D. Parker, *Isolation and phenotypic characterization of virulence-deficient mutants of Vibrio cholerae.* Infect Immun, 1978. **22**(1): p. 181-8.
116. Chiang, S.L. and J.J. Mekalanos, *Use of signature-tagged transposon mutagenesis to identify Vibrio cholerae genes critical for colonization.* Mol Microbiol, 1998. **27**(4): p. 797-805.
117. Camilli, A. and J.J. Mekalanos, *Use of recombinase gene fusions to identify Vibrio cholerae genes induced during infection.* Mol Microbiol, 1995. **18**(4): p. 671-83.
118. Henderson, D.P. and S.M. Payne, *Vibrio cholerae iron transport systems: roles of heme and siderophore iron transport in virulence and identification of a gene associated with multiple iron transport systems.* Infect Immun, 1994. **62**(11): p. 5120-5.
119. Bishop, A.L., B. Patimalla, and A. Camilli, *Vibrio cholerae-induced inflammation in the neonatal mouse cholera model.* Infect Immun, 2014. **82**(6): p. 2434-47.
120. Millet, Y.A., et al., *Insights into Vibrio cholerae intestinal colonization from monitoring fluorescently labeled bacteria.* PLoS Pathog, 2014. **10**(10): p. e1004405.

121. Sawasvirojwong, S., et al., *An Adult Mouse Model of -induced Diarrhea for Studying Pathogenesis and Potential Therapy of Cholera.* PLoS Negl Trop Dis, 2013. **7**(6): p. e2293.
122. Olivier, V., N.H. Salzman, and K.J. Satchell, *Prolonged colonization of mice by Vibrio cholerae El Tor O1 depends on accessory toxins.* Infect Immun, 2007. **75**(10): p. 5043-51.
123. Metchnikoff, E., *Recherches sur le cholera et les vibrions. Receptivite des jeunes lapins pour le cholera intestinal.* Ann. Inst. Pasteur (Paris), 1894. **8**: p. 557.
124. De, S.N. and D.N. Chatterje, *An experimental study of the mechanism of action of Vibriod cholerae on the intestinal mucous membrane.* J Pathol Bacteriol, 1953. **66**(2): p. 559-62.
125. Spira, W.M., R.B. Sack, and J.L. Froehlich, *Simple adult rabbit model for Vibrio cholerae and enterotoxigenic Escherichia coli diarrhea.* Infect Immun, 1981. **32**(2): p. 739-47.
126. Schrank, G.D. and W.F. Verwey, *Distribution of cholera organisms in experimental Vibrio cholerae infections: proposed mechanisms of pathogenesis and antibacterial immunity.* Infect Immun, 1976. **13**(1): p. 195-203.
127. Ritchie, J.M., et al., *Back to the future: studying cholera pathogenesis using infant rabbits.* MBio, 2010. **1**(1).
128. Cohen, M.B., et al., *Validation and characterization of a human volunteer challenge model for cholera by using frozen bacteria of the new Vibrio cholerae epidemic serotype, O139.* Infect Immun, 1999. **67**(12): p. 6346-9.
129. Pitisuttithum, P., et al., *A human volunteer challenge model using frozen bacteria of the new epidemic serotype, V. cholerae O139 in Thai volunteers.* Vaccine, 2001. **20**(5-6): p. 920-5.
130. Glass, R.I., et al., *Predisposition for cholera of individuals with O blood group. Possible evolutionary significance.* Am J Epidemiol, 1985. **121**(6): p. 791-6.
131. Harris, J.B. and R.C. LaRocque, *Cholera and ABO Blood Group: Understanding an Ancient Association.* Am J Trop Med Hyg, 2016. **95**(2): p. 263-264.
132. Harris, J.B., et al., *Blood group, immunity, and risk of infection with Vibrio cholerae in an area of endemicity.* Infect Immun, 2005. **73**(11): p. 7422-7.
133. Levine, M.M., et al., *New knowledge on pathogenesis of bacterial enteric infections as applied to vaccine development.* Microbiol Rev, 1983. **47**(4): p. 510-50.
134. Dictionary, C.E., *Collins English Dictionary*. 2014. p. 1920.
135. Wataya, T., K. Muguruma, and Y. Sasai, *[Human pluripotent stem cell and neural differentiation].* Brain Nerve, 2008. **60**(10): p. 1165-72.
136. Sato, T., et al., *Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche.* Nature, 2009. **459**(7244): p. 262-5.
137. Spence, J.R., et al., *Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro.* Nature, 2011. **470**(7332): p. 105-9.
138. Kretzschmar, K. and H. Clevers, *Organoids: Modeling Development and the Stem Cell Niche in a Dish.* Dev Cell, 2016. **38**(6): p. 590-600.
139. Hannan, N.R., et al., *Generation of multipotent foregut stem cells from human pluripotent stem cells.* Stem Cell Reports, 2013. **1**(4): p. 293-306.
140. Lees, E.A., et al., *Using Human Induced Pluripotent Stem Cell-derived Intestinal Organoids to Study and Modify Epithelial Cell Protection Against*

Salmonella and Other Pathogens.* Journal of Visualized Experiments, 2019(147).

141. Leha, A., et al., *A high-content platform to characterise human induced pluripotent stem cell lines.* Methods, 2016. **96**: p. 85-96.
142. Forbester, J.L., et al., *Interaction of Salmonella enterica Serovar Typhimurium with Intestinal Organoids Derived from Human Induced Pluripotent Stem Cells.* Infect Immun, 2015. **83**(7): p. 2926-34.
143. Leslie, J.L., et al., *Persistence and toxin production by Clostridium difficile within human intestinal organoids result in disruption of epithelial paracellular barrier function.* Infect Immun, 2015. **83**(1): p. 138-45.
144. Bartfeld, S., et al., *In vitro expansion of human gastric epithelial stem cells and their responses to bacterial infection.* Gastroenterology, 2015. **148**(1): p. 126-136 e6.
145. Ewels, P.A., et al., *The nf-core framework for community-curated bioinformatics pipelines.* Nat Biotechnol, 2020. **38**(3): p. 276-278.
146. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.
147. Patro, R., et al., *Salmon provides fast and bias-aware quantification of transcript expression.* Nat Methods, 2017. **14**(4): p. 417-419.
148. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
149. Ulgen, E., O. Ozisik, and O.U. Sezerman, *pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks.* Front Genet, 2019. **10**: p. 858.
150. Bankevich, A., et al., *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.* J Comput Biol, 2012. **19**(5): p. 455-77.
151. Page, A.J., et al., *Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data.* Microb Genom, 2016. **2**(8): p. e000083.
152. Seemann, T., *Prokka: rapid prokaryotic genome annotation.* Bioinformatics, 2014. **30**(14): p. 2068-9.
153. Capella-Gutierrez, S., J.M. Silla-Martinez, and T. Gabaldon, *trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses.* Bioinformatics, 2009. **25**(15): p. 1972-3.
154. Page, A.J., et al., *SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments.* Microb Genom, 2016. **2**(4): p. e000056.
155. Nguyen, L.T., et al., *IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies.* Mol Biol Evol, 2015. **32**(1): p. 268-74.
156. Lewis, P.O., *A likelihood approach to estimating phylogeny from discrete morphological character data.* Syst Biol, 2001. **50**(6): p. 913-25.
157. Hoang, D.T., et al., *UFBoot2: Improving the Ultrafast Bootstrap Approximation.* Mol Biol Evol, 2018. **35**(2): p. 518-522.
158. Hadfield, J., et al., *Phandango: an interactive viewer for bacterial population genomics.* Bioinformatics, 2018. **34**(2): p. 292-293.
159. Sawasvirojwong, S., et al., *An Adult Mouse Model of Vibrio cholerae-induced Diarrhea for Studying Pathogenesis and Potential Therapy of Cholera.* PLoS Negl Trop Dis, 2013. **7**(6): p. e2293.

160. Ferreira, L.P., V.M. Gaspar, and J.F. Mano, *Design of spherically structured 3D in vitro tumor models -Advances and prospects.* Acta Biomater, 2018. **75**: p. 11-34.
161. Costa, E.C., et al., *3D tumor spheroids: an overview on the tools and techniques used for their analysis.* Biotechnol Adv, 2016. **34**(8): p. 1427-1441.
162. Langhans, S.A., *Three-Dimensional in Vitro Cell Culture Models in Drug Discovery and Drug Repositioning.* Front Pharmacol, 2018. **9**: p. 6.
163. Pontes Soares, C., et al., *2D and 3D-organized cardiac cells shows differences in cellular morphology, adhesion junctions, presence of myofibrils and protein expression.* PLoS One, 2012. **7**(5): p. e38147.
164. Ravi, M., et al., *3D cell culture systems: advantages and applications.* J Cell Physiol, 2015. **230**(1): p. 16-26.
165. Lang, L., et al., *Simultaneously inactivating Src and AKT by saracatinib/capivasertib co-delivery nanoparticles to improve the efficacy of anti-Src therapy in head and neck squamous cell carcinoma.* J Hematol Oncol, 2019. **12**(1): p. 132.
166. Landy, J., et al., *Tight junctions in inflammatory bowel diseases and inflammatory bowel disease associated colorectal cancer.* World J Gastroenterol, 2016. **22**(11): p. 3117-26.
167. Turner, J.R., *Intestinal mucosal barrier function in health and disease.* Nat Rev Immunol, 2009. **9**(11): p. 799-809.
168. Edelblum, K.L. and J.R. Turner, *The tight junction in inflammatory disease: communication breakdown.* Curr Opin Pharmacol, 2009. **9**(6): p. 715-20.
169. Shen, L., C.R. Weber, and J.R. Turner, *The tight junction protein complex undergoes rapid and continuous molecular remodeling at steady state.* J Cell Biol, 2008. **181**(4): p. 683-95.
170. Berkes, J., et al., *Intestinal epithelial responses to enteric pathogens: effects on the tight junction barrier, ion transport, and inflammation.* Gut, 2003. **52**(3): p. 439-51.
171. Schneeberger, E.E. and R.D. Lynch, *Structure, function, and regulation of cellular tight junctions.* Am J Physiol, 1992. **262**(6 Pt 1): p. L647-61.
172. Lee, B., K.M. Moon, and C.Y. Kim, *Tight Junction in the Intestinal Epithelium: Its Association with Diseases and Regulation by Phytochemicals.* J Immunol Res, 2018. **2018**: p. 2645465.
173. Rao, R., *Occludin phosphorylation in regulation of epithelial tight junctions.* Ann N Y Acad Sci, 2009. **1165**: p. 62-8.
174. Lees, E.A., et al., *Using Human Induced Pluripotent Stem Cell-derived Intestinal Organoids to Study and Modify Epithelial Cell Protection Against Salmonella and Other Pathogens.* Jove-Journal of Visualized Experiments, 2019(147).
175. Sipione, S., et al., *Gangliosides in the Brain: Physiology, Pathophysiology and Therapeutic Applications.* Front Neurosci, 2020. **14**: p. 572965.
176. Wands, A.M., et al., *Fucosylated Molecules Competitively Interfere with Cholera Toxin Binding to Host Cells.* ACS Infect Dis, 2018. **4**(5): p. 758-770.
177. Cervin, J., et al., *GM1 ganglioside-independent intoxication by Cholera toxin.* PLoS Pathog, 2018. **14**(2): p. e1006862.
178. Wands, A.M., et al., *Fucosylation and protein glycosylation create functional receptors for cholera toxin.* Elife, 2015. **4**: p. e09545.
179. Barua, D. and A.S. Paguio, *ABO blood groups and cholera.* Annals of Human Biology, 1977. **4**(5): p. 489-492.

180. Finne, J., et al., *Novel polyfucosylated N-linked glycopeptides with blood group A, H, X, and Y determinants from human small intestinal epithelial cells.* J Biol Chem, 1989. **264**(10): p. 5720-35.
181. Breimer, M.E., *Tissue specificity of glycosphingolipids as expressed in pancreas and small intestine of blood group A and B human individuals.* Arch Biochem Biophys, 1984. **228**(1): p. 71-85.
182. Bjork, S., et al., *Structures of blood group glycosphingolipids of human small intestine. A relation between the expression of fucolipids of epithelial cells and the ABO, Le and Se phenotype of the donor.* J Biol Chem, 1987. **262**(14): p. 6758-65.
183. Swerdlow, D.L., et al., *Severe life-threatening cholera associated with blood group O in Peru: implications for the Latin American epidemic.* J Infect Dis, 1994. **170**(2): p. 468-72.
184. Harris, J.B., et al., *Susceptibility to Vibrio cholerae Infection in a Cohort of Household Contacts of Patients with Cholera in Bangladesh.* PLOS Neglected Tropical Diseases, 2008. **2**(4): p. e221.
185. Heggelund, J.E., et al., *High-Resolution Crystal Structures Elucidate the Molecular Basis of Cholera Blood Group Dependence.* PLoS Pathog, 2016. **12**(4): p. e1005567.
186. Levine, M.M., et al., *Genetic susceptibility to cholera.* Ann Hum Biol, 1979. **6**(4): p. 369-74.
187. Kuhlmann, F.M., et al., *Blood Group O-Dependent Cellular Responses to Cholera Toxin: Parallel Clinical and Epidemiological Links to Severe Cholera.* Am J Trop Med Hyg, 2016. **95**(2): p. 440-3.
188. Cornick, S., et al., *VAMP8-mediated MUC2 mucin exocytosis from colonic goblet cells maintains innate intestinal homeostasis.* Nat Commun, 2019. **10**(1): p. 4306.
189. Fakih, D., et al., *Normal murine respiratory tract has its mucus concentrated in clouds based on the Muc5b mucin.* Am J Physiol Lung Cell Mol Physiol, 2020. **318**(6): p. L1270-L1279.
190. Ermund, A., et al., *The mucus bundles responsible for airway cleaning are retained in cystic fibrosis and by cholinergic stimulation.* Eur Respir J, 2018. **52**(2).
191. Cornick, S., F. Moreau, and K. Chadee, *Entamoeba histolytica Cysteine Proteinase 5 Evokes Mucin Exocytosis from Colonic Goblet Cells via alphavbeta3 Integrin.* PLoS Pathog, 2016. **12**(4): p. e1005579.
192. Singla, A., et al., *LPA stimulates intestinal DRA gene transcription via LPA2 receptor, PI3K/AKT, and c-Fos-dependent pathway.* Am J Physiol Gastrointest Liver Physiol, 2012. **302**(6): p. G618-27.
193. Dekkers, J.F., et al., *A functional CFTR assay using primary cystic fibrosis intestinal organoids.* Nat Med, 2013. **19**(7): p. 939-45.
194. Dekkers, J.F., et al., *Characterizing responses to CFTR-modulating drugs using rectal organoids derived from subjects with cystic fibrosis.* Sci Transl Med, 2016. **8**(344): p. 344ra84.
195. Leitch, G.J., *Cholera enterotoxin-induced mucus secretion and increase in the mucus blanket of the rabbit ileum in vivo.* Infect Immun, 1988. **56**(11): p. 2871-5.
196. Epple, H.J., et al., *Differential stimulation of intestinal mucin secretion by cholera toxin and carbachol.* Pflugers Arch, 1997. **433**(5): p. 638-47.
197. Ambort, D., et al., *Calcium and pH-dependent packing and release of the gel-forming MUC2 mucin.* Proc Natl Acad Sci U S A, 2012. **109**(15): p. 5645-50.

198. Johansson, M.E., H. Sjovall, and G.C. Hansson, *The gastrointestinal mucus system in health and disease.* Nat Rev Gastroenterol Hepatol, 2013. **10**(6): p. 352-61.
199. Ermund, A., et al., *Mucus threads from surface goblet cells clear particles from the airways.* Respir Res, 2021. **22**(1): p. 303.
200. Gouyer, V., F. Gottrand, and J.L. Desseyn, *The extraordinarily complex but highly structured organization of intestinal mucus-gel unveiled in multicolor images.* PLoS One, 2011. **6**(4): p. e18761.
201. Nystrom, E.E.L., et al., *An intercrypt subpopulation of goblet cells is essential for colonic mucus barrier function.* Science, 2021. **372**(6539).
202. Sharpen, J.D.A., et al., *Transglutaminase 3 crosslinks the secreted gel-forming mucus component Mucin-2 and stabilizes the colonic mucus layer.* Nat Commun, 2022. **13**(1): p. 45.
203. Gustafsson, J.K., et al., *Intestinal goblet cells sample and deliver lumenal antigens by regulated endocytic uptake and transcytosis.* Elife, 2021. **10**.
204. Hirao, Y., et al., *Interleukin-8—An Early Marker for Bacterial Infection.* Laboratory Medicine, 2000. **31**(1): p. 39-44.
205. Oppenheim, J.J., et al., *Properties of the novel proinflammatory supergene "intercrine" cytokine family.* Annu Rev Immunol, 1991. **9**: p. 617-48.
206. Zhou, X., et al., *Induction of interleukin-8 in T84 cells by Vibrio cholerae.* Infect Immun, 2004. **72**(1): p. 389-97.
207. Murakami, H. and H. Masui, *Hormonal control of human colon carcinoma cell growth in serum-free medium.* Proc Natl Acad Sci U S A, 1980. **77**(6): p. 3464-8.
208. McCormick, B.A., et al., *Salmonella typhimurium attachment to human intestinal epithelial monolayers: transcellular signalling to subepithelial neutrophils.* J Cell Biol, 1993. **123**(4): p. 895-907.
209. Rodriguez, B.L., et al., *Differential interleukin-8 response of intestinal epithelial cell line to reactogenic and nonreactogenic candidate vaccine strains of Vibrio cholerae.* Infect Immun, 2001. **69**(1): p. 613-6.
210. Silva, T.M., et al., *New evidence for an inflammatory component in diarrhea caused by selected new, live attenuated cholera vaccines and by El Tor and Q139 Vibrio cholerae.* Infect Immun, 1996. **64**(6): p. 2362-4.
211. Noel, G., et al., *A primary human macrophage-enteroid co-culture model to investigate mucosal gut physiology and host-pathogen interactions.* Scientific Reports, 2017. **7**(1): p. 45270.
212. Gillespie, M., et al., *The reactome pathway knowledgebase 2022.* Nucleic Acids Res, 2022. **50**(D1): p. D687-D692.
213. Czermak, B.J., et al., *In vitro and in vivo dependency of chemokine generation on C5a and TNF-alpha.* J Immunol, 1999. **162**(4): p. 2321-5.
214. Micheau, O. and J. Tschopp, *Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes.* Cell, 2003. **114**(2): p. 181-90.
215. Widera, D., et al., *Nuclear Factor-kappaB controls the reaggregation of 3D neurosphere cultures in vitro.* Eur Cell Mater, 2006. **11**: p. 76-84; discussion 85.
216. Venugopal, S., S. Anwer, and K. Szaszi, *Claudin-2: Roles beyond Permeability Functions.* Int J Mol Sci, 2019. **20**(22).
217. Rosenthal, R., et al., *Claudin-2, a component of the tight junction, forms a paracellular water channel.* J Cell Sci, 2010. **123**(Pt 11): p. 1913-21.

218. Luettig, J., et al., *Claudin-2 as a mediator of leaky gut barrier during intestinal inflammation.* Tissue Barriers, 2015. **3**(1-2): p. e977176.
219. Di Pierro, M., et al., *Zonula occludens toxin structure-function analysis. Identification of the fragment biologically active on tight junctions and of the zonulin receptor binding domain.* J Biol Chem, 2001. **276**(22): p. 19160-5.
220. Faruque, S.M., M.J. Albert, and J.J. Mekalanos, *Epidemiology, genetics, and ecology of toxigenic Vibrio cholerae.* Microbiol Mol Biol Rev, 1998. **62**(4): p. 1301-14.
221. Dorman, M.J., et al., *Genomics of the Argentinian cholera epidemic elucidate the contrasting dynamics of epidemic and endemic Vibrio cholerae.* Nat Commun, 2020. **11**(1): p. 4918.
222. Shin, O.S., et al., *Type III secretion is essential for the rapidly fatal diarrheal disease caused by non-O1, non-O139 Vibrio cholerae.* mBio, 2011. **2**(3): p. e00106-11.
223. Alam, A., et al., *Identification of Vibrio cholerae type III secretion system effector proteins.* Infect Immun, 2011. **79**(4): p. 1728-40.
224. Chaand, M., et al., *Type 3 Secretion System Island Encoded Proteins Required for Colonization by Non-O1/non-O139 Serogroup Vibrio cholerae.* Infect Immun, 2015. **83**(7): p. 2862-2869.
225. Liverman, A.D., et al., *Arp2/3-independent assembly of actin by Vibrio type III effector VopL.* Proc Natl Acad Sci U S A, 2007. **104**(43): p. 17117-22.
226. Carpenter, M.R., et al., *CRISPR-Cas and Contact-Dependent Secretion Systems Present on Excisable Pathogenicity Islands with Conserved Recombination Modules.* J Bacteriol, 2017. **199**(10).
227. Castillo, D., et al., *Comparative Genome Analyses of Vibrio anguillarum Strains Reveal a Link with Pathogenicity Traits.* mSystems, 2017. **2**(1).
228. Prjibelski, A., et al., *Using SPAdes De Novo Assembler.* Curr Protoc Bioinformatics, 2020. **70**(1): p. e102.
229. Wood, D.E. and S.L. Salzberg, *Kraken: ultrafast metagenomic sequence classification using exact alignments.* Genome Biol, 2014. **15**(3): p. R46.
230. Heidelberg, J.F., et al., *DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.* Nature, 2000. **406**(6795): p. 477-83.
231. Page, A.J., et al., *Roary: rapid large-scale prokaryote pan genome analysis.* Bioinformatics, 2015. **31**(22): p. 3691-3.
232. Gardel, C.L. and J.J. Mekalanos, *Regulation of cholera toxin by temperature, pH, and osmolarity.* Methods Enzymol, 1994. **235**: p. 517-26.
233. Iwanaga, M., et al., *Culture conditions for stimulating cholera toxin production by Vibrio cholerae O1 El Tor.* Microbiol Immunol, 1986. **30**(11): p. 1075-83.
234. Miller, V.L. and J.J. Mekalanos, *A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in Vibrio cholerae requires toxR.* J Bacteriol, 1988. **170**(6): p. 2575-83.
235. Beyhan, S., et al., *Differences in gene expression between the classical and El Tor biotypes of Vibrio cholerae O1.* Infect Immun, 2006. **74**(6): p. 3633-42.
236. Naha, A., et al., *Deciphering the possible role of ctxB7 allele on higher production of cholera toxin by Haitian variant Vibrio cholerae O1.* PLoS Negl Trop Dis, 2020. **14**(4): p. e0008128.
237. Letunic, I. and P. Bork, *Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation.* Nucleic Acids Res, 2021. **49**(W1): p. W293-W296.

238. Mardis, E., et al., *What is finished, and why does it matter.* Genome Res, 2002. **12**(5): p. 669-71.
239. Chain, P.S., et al., *Genomics. Genome project standards in a new era of sequencing.* Science, 2009. **326**(5950): p. 236-7.
240. Zerbino, D.R. and E. Birney, *Velvet: algorithms for de novo short read assembly using de Bruijn graphs.* Genome Res, 2008. **18**(5): p. 821-9.
241. Wick, R.R., et al., *Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads.* PLoS Comput Biol, 2017. **13**(6): p. e1005595.
242. Wick, R.R., et al., *Bandage: interactive visualization of de novo genome assemblies.* Bioinformatics, 2015. **31**(20): p. 3350-2.
243. Carver, T.J., et al., *ACT: the Artemis Comparison Tool.* Bioinformatics, 2005. **21**(16): p. 3422-3.
244. Ramirez, D.A., H.A. Saka, and R. Nores, *Detection of Vibrio cholerae aDNA in human burials from the fifth cholera pandemic in Argentina (1886-1887 AD).* Int J Paleopathol, 2021. **32**: p. 74-79.
245. Clark, C.A., et al., *The Vibrio cholerae O1 chromosomal integron.* Microbiology (Reading), 2000. **146 (Pt 10)**: p. 2605-2612.
246. Petroni, A., et al., *CARB-9, a carbenicillinase encoded in the VCR region of Vibrio cholerae non-O1, non-O139 belongs to a family of cassette-encoded beta-lactamases.* Antimicrob Agents Chemother, 2004. **48**(10): p. 4042-6.
247. Rowe-Magnus, D.A., A.M. Guerout, and D. Mazel, *Super-integrons.* Res Microbiol, 1999. **150**(9-10): p. 641-51.
248. Hounmanou, Y.M.G., et al., *Genomic and Phenotypic Insights for Toxigenic Clinical Vibrio cholerae O141.* Emerg Infect Dis, 2022. **28**(3): p. 617-624.
249. Bilous, P.T., et al., *Nucleotide sequence of the dmsABC operon encoding the anaerobic dimethylsulphoxide reductase of Escherichia coli.* Mol Microbiol, 1988. **2**(6): p. 785-95.
250. Colovos, C., D. Cascio, and T.O. Yeates, *The 1.8 A crystal structure of the ycaC gene product from Escherichia coli reveals an octameric hydrolase of unknown specificity.* Structure, 1998. **6**(10): p. 1329-37.
251. Groftehauge, M.K., et al., *Crystal Structure of a Hidden Protein, YcaC, a Putative Cysteine Hydrolase from Pseudomonas aeruginosa, with and without an Acrylamide Adduct.* Int J Mol Sci, 2015. **16**(7): p. 15971-84.
252. Grissa, I., G. Vergnaud, and C. Pourcel, *CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W52-7.
253. Abby, S.S., et al., *MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems.* PLoS One, 2014. **9**(10): p. e110726.
254. Couvin, D., et al., *CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins.* Nucleic Acids Res, 2018. **46**(W1): p. W246-W251.
255. Arndt, D., et al., *PHASTER: a better, faster version of the PHAST phage search tool.* Nucleic Acids Res, 2016. **44**(W1): p. W16-21.
256. Zhou, Y., et al., *PHAST: a fast phage search tool.* Nucleic Acids Res, 2011. **39**(Web Server issue): p. W347-52.
257. Johnson, J.A., et al., *Vibrio cholerae O139 synonym bengal is closely related to Vibrio cholerae El Tor but has important differences.* Infect Immun, 1994. **62**(5): p. 2108-10.

258. Bortolaia, V., et al., *ResFinder 4.0 for predictions of phenotypes from genotypes.* J Antimicrob Chemother, 2020. **75**(12): p. 3491-3500.
259. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**: p. 421.
260. Rowe-Magnus, D.A., A.M. Guerout, and D. Mazel, *Bacterial resistance evolution by recruitment of super-integron gene cassettes.* Mol Microbiol, 2002. **43**(6): p. 1657-69.
261. Sengupta, D.K., M. Boesman-Finkelstein, and R.A. Finkelstein, *Antibody against the capsule of Vibrio cholerae O139 protects against experimental challenge.* Infect Immun, 1996. **64**(1): p. 343-5.
262. Waldor, M.K., R. Colwell, and J.J. Mekalanos, *The Vibrio cholerae O139 serogroup antigen includes an O-antigen capsule and lipopolysaccharide virulence determinants.* Proc Natl Acad Sci U S A, 1994. **91**(24): p. 11388-92.
263. Alam, A., et al., *Hyperinfectivity of human-passaged Vibrio cholerae can be modeled by growth in the infant mouse.* Infect Immun, 2005. **73**(10): p. 6674-9.