The Molecular Characterisation of *Shigella* spp. from Papua New Guinea and Pacific Island Nations

Elisheba Malau

BMedLabSc (James Cook University), BSc (Hons) (University of Melbourne)

This thesis is submitted in total fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

School of Health and Life Sciences

Federation University Australia

Northways Road

Churchill Victoria, 3841

Australia

Submitted

November 2019

Statement of Authorship

Except where explicit reference is made in the text of the thesis, this thesis contains no material published elsewhere or extracted in whole or in part from a thesis by which I have qualified for or been awarded another degree or diploma. No other person's work has been relied upon or used without due acknowledgement in the main text and bibliography of the thesis.

Signed:_____	Signed:_____
Dated: 28/11/2019	Dated: 28/11/2019
NAME	NAME
Elisheba MALAU	Andrew GREENHILL
Candidate	Principal Supervisor

Statement of Ethics

Relevant research reported in this thesis received approval from the Papua New Guinea Medical Research Advisory Committee (MRAC no. 16.43) and the Federation University Australia Ethics Review Committee (Human ethics no. A17-074).

E Malau

November 2019

Abstract

In Papua New Guinea (PNG) and other low-middle income countries (LMICs) in Oceania, diarrhoea remains a leading cause of hospitalisation and death in children <5 years old; and is an important cause of illness in older children and adults. *Shigella* is one of four leading causes of diarrhoea globally and is likely a major cause of diarrhoea in Oceanic LMICs, particularly PNG. A survey was conducted among parents of children with diarrhoea in the PNG highlands, demonstrating limited access to, and application of, sanitation and hygiene; likely contributing to the transmission of *Shigella* and other pathogens in PNG. To better understand *Shigella* in Oceanic LMICs, multiple analyses of isolates from within PNG (n=38), far-northeastern Australia, and from travellers returning to Australia from Oceanic LMICs with shigellosis (n=34) were conducted. Antimicrobial susceptibility (by disc diffusion), detection of virulence genes by polymerase chain reaction (PCR) were conducted on all isolates, and whole genome sequencing (WGS) conducted on 63 isolates. *Shigella* spp. were commonly resistant to two or more classes of first-line antibiotics, with resistance more common in post-2010 relative to pre-2010 isolates. WGS was used to verify PCR detection of virulence genes, determine whether resistance could be predicted genetically, and conduct phylogenetic analysis of *Shigella* spp. in Oceania. WGS surpassed PCR in the detection of virulence genes, but correlated poorly with phenotypic antimicrobial resistance. Phylogenetic analysis revealed the intra- and inter-country relatedness. Three phylogenetic groups of *S. flexneri* co-exist in Oceanic LMICs (and far-northeastern Australia), the result of multiple incursions. Two lineages of *S. sonnei* were detected, one circulating in PNG and New Caledonia, and the other in various other countries including Fiji; with no geographical overlap of the two *S. sonnei* lineages. Incursions of *Shigella* into the Oceanic LMICs occur regularly, and are

likely to occur again. This study provides evidence of the need for, and potential approach to, expanded surveillance of *Shigella* in the region.

Table of Contents

List of Tables

List of Figures

Acknowledgements

Firstly, I would like to thank my supervisors Dr Andrew Greenhill, Assoc Prof Jenny Mosse and Dr Paul Horwood for their guidance and support throughout this PhD project.

This research would not have been possible if not for the help, support and collaborations, with the staff from institutions such as the PNGIMR, the MDU Public Health Laboratory at the Doherty Institute in Melbourne and the Public Health Microbiology Forensic and Scientific Services from the Queensland Department of Health in Brisbane, for the provision of *Shigella* isolates from the Pacific region; and the Sanger Institute in the UK for doing WGS of my *Shigella* isolates as well as, Assoc Prof Kathryn Holt and Dr Jane Hawkey from the Monash University, in helping me understand the bioinformatic tools used to analyse the WGS data for my *Shigella* isolates from Sanger. A special thank you also to Dr William Pomat and Dr Rebecca Ford, whose help was essential towards the acquiring and processing of samples at PNGIMR, Goroka; Mary Valcanis at MDU, Dr Amy Jennison at Queensland Health; and Prof Gordon Dougan at Sanger Institute.

I would also like to thank the academic and general staff at Federation University Australia, for all their help and support during the undertaking of this project. To my PhD crew, Hannah, Jacqui, Sami and Tinni who started this PhD journey with me, thanks for the Chinese, chocolates, home cooked meals, stories and laughs, which kept me healthy and happy.

To my family; this word in itself is quite broad as culturally, in true island ‘pasin’ (i.e. spirit/way), apart from my own immediate family, my extended family in Goroka played

a paramount role also towards the success and completion of this study. Without the help of people like Sarah Javati, Nicole Kotale, Mition Yoannes, Florence Lawrence, Alfred Alawaki, Tobias Maure and many others, my field work and processing of samples would not have been possible. A special thank you also to my little sister Marcella for her help with making my phylogenetic tree diagrams look legible. Thank you all (both close and extended) for your love, patience and support throughout this journey. I definitely could not have accomplished what I had set out to do these last few years were it not for your support. To my parents, thank you for your words of wisdom, guidance and encouragement. Thank you for always believing in me and for giving me the confidence to find that inner strength that pushed me out of my comfort zone and into positions of high achievements.

To conclude I wish to dedicate this thesis in memory of my dearest uncle, the late Dr Peter Wai'in. You were the first of our family to achieve a PhD degree and though God called you home much too soon, you remain forever in our hearts and minds, and your stories and wisdom continue to guide and motivate us, your children, to live lives filled with meaning, love and kindness. Thank you!

The undertaking of this PhD would not have been possible without the support and grant of the Australian Government Research Training Program (RTP) stipend and fee-offset Scholarship through Federation University Australia.

List of Publications

The following publications have arisen from this thesis. The accepted manuscripts are provided in Appendix 1.

Malau, E., Mosse, J., Horwood, P.F., Greenhill, A.R. (2016). Shigellosis: A truly neglected disease in Papua New Guinea. PNG Medical journal. *59*(3-4), 147-154.

Malau, E., Ford, R., Valcanis, M., Jennison, A.V., Mosse, J., Bean, D., Yoannes, M., Pomat, W., Horwood, PF., Greenhill, AR. (2018). Antimicrobial sensitivity trends and virulence genes in *Shigella* spp. from the Oceania region. Infection, Genetics and Evolution. *64*, 52-56

Abbreviations

°C	Degrees Celsius
µg	micrograms
µl	microliters
µmol	micromoles
AB	Applied Biosystems
AMP	Ampicillin
AMR	Antimicrobial resistance
ARG-ANNOT	Antibiotic Resistance Gene Annotation
C	Chloramphenicol
CDC	Center for Disease Control and Prevention
CIP	Ciprofloxacin
CLSI	Clinical Laboratory Standards Institute
CRO	Ceftriaxone
DALY	Disability Adjusted Life Years
DF	Degrees of Freedom
DNA	Deoxyribonucleic acid
EHP	Eastern Highlands Province
EIEC	Enteroinvasive *E.coli*
EUCAST	European Committee on Antimicrobial Susceptibility Testing
GEMS	Global Enteric Multi-centre Study
HIV	Human Immunodeficiency Virus
HPD	Highest posterior density
IHME	The Institute for Health Metrics and Evaluation

L	Lineage
LMIC	Low-middle income countries
MDR	Multi-drug resistance
MDU	Microbiology Diagnostic Unit
MGE	Mobile genetic elements
MRCA	Most recent common ancestor
min	minutes
NA	Nalidixic acid
ng	Nanogram
NNDSS	National Notifiable Diseases Surveillance System
PAI	Pathogenicity islands
PCR	Polymerase chain reaction
PGs	Phylogenetic groups
PNG	Papua New Guinea
PNGIMR	Papua New Guinea Institute of Medical Research
sec	seconds
SNPs	Single Nucleotide Polymorphisms
sp.	Species (referring to a single unidentified species)
spp.	Species (referring to two or more species)
SXT	Trimethoprim sulfamethoxazole/Co-trimoxazole
TET	Tetracycline
T3SS	Type III Secretion System
UNICEF	United Nations Children's Fund
USA	United States of America
UV	Ultra Violet

VFDB	Virulence factors database
WASH	Water, Sanitation and Hygiene
WGS	Whole Genome Sequencing
WHO	World Health Organization

Chapter 1 - General Introduction

1.1 Background

Diarrhoea and common infectious diseases are, collectively, among the leading contributors to morbidity and mortality globally, second only to cardiovascular diseases. In contrast to most non-communicable diseases such as cardiovascular disease, the burden of diarrhoeal diseases is greatest in the young, particularly those under 5 years of age (Roser & Ritchie, 2019). Though data obtained over the last 10 – 15 years indicates a decrease in the burden of diarrhoeal diseases globally, diarrhoea still remains a leading cause of morbidity and mortality in low and middle-income countries (Fischer Walker, Perin, Aryee, Boschi-Pinto, & Black, 2012; Kotloff *et al*., 2013). Sustained efforts are required to continue the trend towards decreasing burden of diarrhoeal diseases.

There is a plethora of aetiological agents that can cause diarrhoea, including numerous physiological conditions. Identifying and correctly diagnosing the aetiological agent is essential to understanding the true impact of diarrhoeal disease, with a view to improving public health management and reducing the overall burden of diarrhoeal disease. This is particularly important in low-income settings, where the aetiology is most likely to be infectious agents, and where the burden of diarrhoeal disease is greatest (Fischer Walker *et al*., 2012; L. Liu *et al*., 2012).

The Oceania region is economically and socially diverse, consisting of high-income countries with advanced economies (e.g. Australia and New Zealand), and countries that have only recently transitioned from low income to lower-middle income, such as Papua New Guinea and Solomon Islands (Figure 1.1) (International Monetary Fund, 2018). In

brief, the Oceania region consists of low-middle income countries (LMICs) and high-income countries. However, the distinction is not perfectly clear, with health indicators in parts of Australia such as indigeous communities comparable to health indicators in some Oceanic LMICs. Despite recent economic growth in the Oceania region, many people still have a low disposable income and thus reside in resource-limited settings, with limited access to basic hygiene and health services. As such, the burden of diarrhoeal disease continues to remain high in Pacific Island countries (Hoy, Roth, Viney, Souares, & Lopez, 2014).

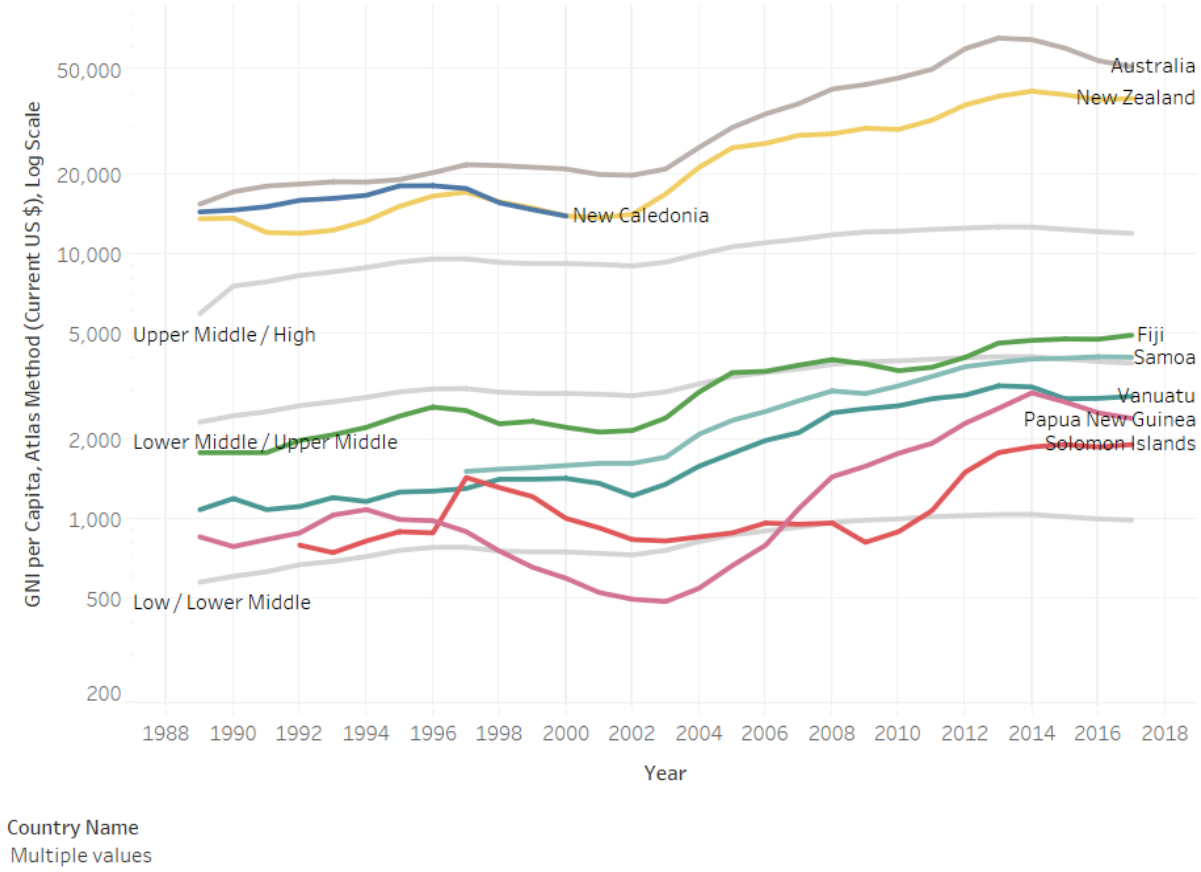


Figure 1.1 The economy of selected countries in the Oceania region based on gross national income per capita. Relatively few countries in the region are upper middle or high income countries (The World Bank, 2019).

1.2 Reason for the Study

It is likely that an integrated approach to combat diarrhoeal diseases, similar to the approach advocated for in the fight against malaria in recent decades (Shiff, 2002), would have the greatest impact. Part of such an approach is an understanding of circulating pathogens that cause diarrhoea, and logically microbiological studies should first focus on pathogens responsibly for a high burden of diarrhoeal disease. Studies in the last decade, analysing global data on enteric pathogens, reveal *Shigella* to be an important causative agent of diarrhoea (Fischer Walker *et al*., 2012; Kotloff *et al*., 2013). Current evidence suggests that shigellosis is one of the leading causes of diarrhoea in Papua New Guinea (PNG) (Benny *et al*., 2014; Greenhill *et al*., 2014; Soli *et al*., 2014). Although data are lacking in other LMIC Oceanic nations, based on global findings (Kotloff *et al*., 2013), it is likely that shigellosis is an important disease in those countries as well.

Efforts to reduce the burden of *Shigella* infection appear to lag behind control of many other infectious agents. This study seeks to improve understanding of *Shigella* in LMICs in Oceania, with a major focus on strains from PNG; the most populated LMIC in the region and one where the burden of diarrhoea and other infectious disease remains high (PNG National Department of Health, 2010; World Health Organization, 2013). To date, very little research has been conducted that seeks to increase our knowledge of *Shigella* circulating in Oceanic LMICs. To facilitate better control, prevention and treatment of shigellosis, there needs to be a greater understanding of circulating strains of *Shigella*.

One reason for the current lack of understanding of *Shigella* (and various other bacterial pathogens) circulating in Oceanic LMICs is that microbial culture and antibiotic sensitivity tests are not routinely conducted in countries in the region. Many countries

may have culture facilities at a relatively small number of hospitals, and may or may not conduct culture for diarrhoeal disease. When isolates are obtained they are seldom retained for further characterisation. These issues became clear in the PNG context during the cholera outbreak of 2009-11 (Greenhill *et al*., 2012). While bacterial culture has distinct advantages for disease diagnosis, in high-income settings culture is now complemented with molecular biology techniques. Such approaches are seldom used in the Oceanic LMICs, but consideration should be given to their in-country implementation.

1.3 Research Questions

With this background in mind, this study aims to answer the following research questions:

1. Is there evidence of increasing antimicrobial resistance in contemporary strains of *Shigella* relative to historical strains in Oceanic LMICs, and if so, what are the main mechanisms of resistance?
2. What are the virulence traits of *Shigella* circulating in PNG and other LMICs in Oceania?
3. Is there evidence of high levels of intra-and/or inter-country relatedness between *Shigella* isolates in Oceania?
4. Could molecular methods and/or characterisation using polymerase chain reaction (PCR) and whole genome sequencing (WGS) be used to detect *Shigella* infections in low-income, low-resource settings, such as the LMICs in Oceania?

By understanding the epidemiology of *Shigella* in the Oceania (with a focus on LMICS in the region), the data generated from this study will form the basis for more targeted studies. This study, in combination with future studies, should seek to assist local

governments in improving patient outcomes and lead to the implementation of better public health measures to control the spread of *Shigella*.

Chapter 2 - Literature Review

2.1 Overview of *Shigella* and Shigellosis

Shigella is a Gram negative, non-motile, non-spore forming, facultative anaerobic bacillus belonging to the Family Enterobacteriaceae. *Shigella* is closely related to *Escherichia coli*, an organism commonly present in the gut of many mammals and birds (Lan & Reeves, 2002; Niyogi, 2005). Most strains of *E. coli* are non-pathogenic, but some strains cause diseases of the gastrointestinal tract and other organ systems. *Shigella* species have a genome akin to that of *E. coli*, but with the deletion of some segments of the genome and the presence of mobile genetic elements such as insertion sequences and a virulence plasmid. There are similarities in the (genetic determinants of) pathogenesis of *Shigella* and some of the diarrhoeaogenic strains of *E. coli*, in particular the enteroinvasive *E. coli* (EIEC) (Lan & Reeves, 2002).

The *Shigella* genus consists of four species: *S. dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei* (Niyogi, 2005). Based on genetic analysis, these four species of *Shigella* could be more accurately considered different varieties of a single species. However, the four species remain due to historical, epidemiological and clinical considerations. Each species can also be further classified into different biochemical and serological groups. *S. dysenteriae* in serogroup A has 13 serotypes; *S. flexneri* in serogroup B has 15 serotypes*; S. boydii* in serogroup C has18 serotypes and *S. sonnei* in serogroup D has only one serotype (Niyogi, 2005).

Humans are the only natural host and therefore main reservoir for *Shigella*, however the organism can survive on inanimate surfaces for months (Islam *et al*., 2001; Kramer,

Schwebke, & Kampf, 2006) and will survive in water and foods. Reflecting this, transmission is primarily through the faecal-oral route from person-to-person contact and also through the consumption of contaminated food and water supplies. Transmission is facilitated through inadequate basic hygiene practices.

Shigella is highly infectious; ingestion of as few as 10 – 100 bacteria can cause infection (DuPont, Levine, Hornick, & Formal, 1989; Niyogi, 2005). Incubation time depends on the species serotype causing infection; typically 1-3 days, but can range from 12 hours to 7 days (Department of Health and Human Services, 2011; Heymann, 2014). Infection with *Shigella* is often characterised by blood and/or mucus in stools, which is clinically referred to as dysentery. However, these dysenteric symptoms are not always present in infected people, with the disease commonly resulting in diarrhoea without dysentery. A mild case of shigellosis may result in fever, fatigue, malaise, anorexia and watery, non-bloody diarrhoea. These symptoms are similar to those caused by other enteric pathogens, and can therefore be difficult to distinguish clinically (Niyogi, 2005). In severe cases of shigellosis, severe abdominal cramps, dysentery and sometimes seizures occur (Niyogi, 2005; Zaidi & Estrada-Garcia, 2014). Other signs and symptoms may include dehydration, intestinal complications, and decreased immune function and nutritional status leading to co-infection (Niyogi, 2005). Shigellosis can result in death. There are various mechanisms of pathogenesis that are encoded by multiple virulence genes distributed across multiple loci on the *Shigella* chromosome and the virulence plasmid.

2.2 *Shigella* Pathogenesis and Virulence

2.2.1 Pathogenesis

The various virulence genes present in *Shigella* enable the bacteria to cause infection due to its ability to invade, colonise and cause inflammation in the colon and epithelium of the intestine (Zaidi & Estrada-Garcia, 2014). The mechanisms associated with pathogenesis have been previously described by Mattock and Blocker (2017). A summary of the process is described below, and the process of *Shigella* spp. invasion and pathogenesis is illustrated in Figure 2.1.

Shigella accomplishes infection with the aid of the body's host defence via microfold cells (M cells) situated along the surface of the intestinal epithelium (Figure 2.1, stage 1). The M cells are responsible for sampling particles coming from the gut lumen and deliver them into the mucosal lymphoid tissue where immune responses are initiated. Once inside the epithelial layer, the bacteria are released into an intraepithelial pocket where they encounter resident macrophages that engulf and prepare to degrade them (Figure 2.1, stage 2). To prevent degradation, the bacteria induce pyroptosis, an inflammatory form of programmed cell death. This results in the activation of inflammatory cytokines in the macrophage, followed by the macrophage plasma-membrane rupturing and releasing its proinflammatory intracellular contents extracellularly (Figure 2.1, stage 6). As the macrophage undergoes cell death it produces cytokines, signalling other pro-inflammatory cells such as polymorphonuclear cells (i.e. neutrophils and natural killer-cells) to the site of infection (Figure 2.1, stage 4). This causes massive inflammation, which results in the destruction of the epithelial lining of the intestine. The ability to produce toxins such as the Shiga toxin (produced by *S. dysenteriae* 1) and *Shigella* enterotoxins 1 and 2 (ShET-1 & ShET-2) further aids invasiveness (Bergsbaken, Fink, &

Cookson, 2009; Mattock & Blocker, 2017; Schroeder & Hilbi, 2008; Zaidi & Estrada-Garcia, 2014).

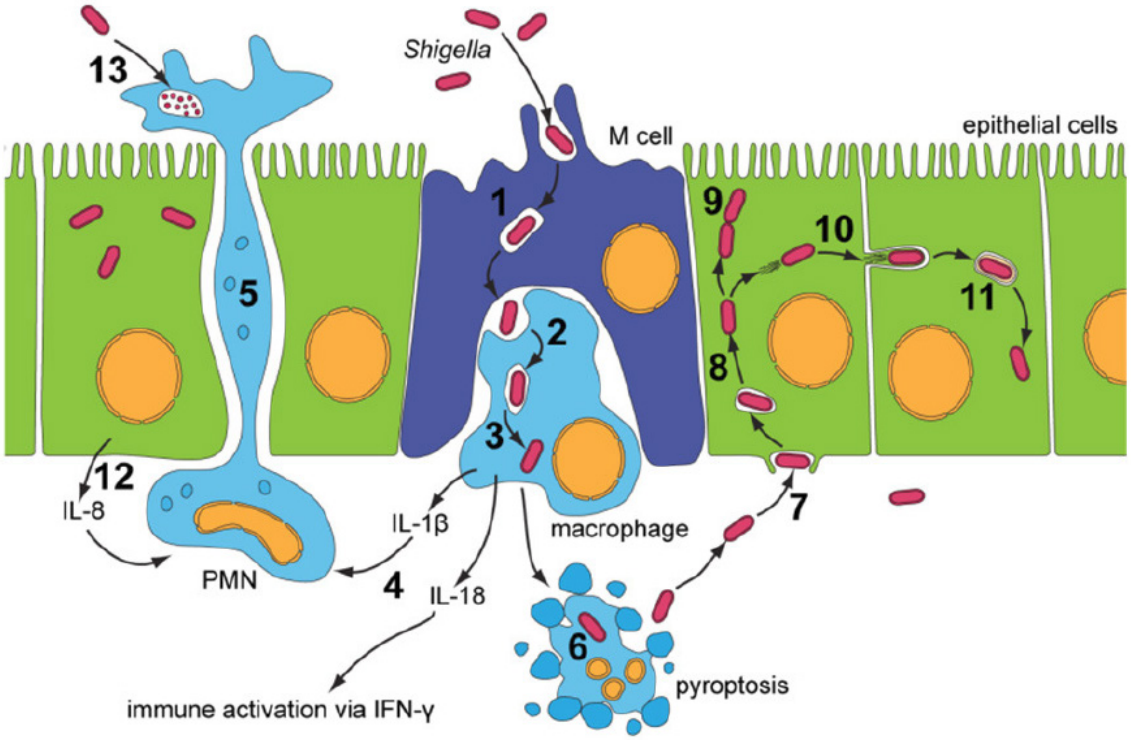


Figure 2.1 The key stages of *Shigella* infection. *Shigella* enters colonic epithelium via M-cell membrane ruffling (1), and is then endocytosed by resident macrophages (2). Pyroptotic/caspase 1-dependent cell death of macrophage is induced after Shigella escape/evade the phagocytic vacuole (3 and 6) and pro-inflammatory cytokines are released (4). Escaped *Shigella* are then taken up via the macropinocytic process at the basolateral membrane of the epithelial cells (7). GTPases are stimulated, triggering actin polymerization allowing *Shigella* mobility inside the epithelial cells to infect and spread to adjacent epithelial cells (8-11). The epithelial barrier is destroyed by pro-inflammatory cells (12) then the bacteria travel up from the basolateral to apical colonic (5) epithelium which weakens the junction and allows for further invasion. Polymorphonuclear leukocytes eventually remove *Shigella* from the colonic membrane (13). Adapted from Mattock and Blocker (2017).

2.2.2 *Shigella* virulence genes

The pathogenesis of *Shigella* spp. is directly related to virulence mechanisms. *Shigella* spp. possess a number of virulence genes that are present on its chromosome and/or virulence plasmid; these genes, and indeed the virulence plasmid, is absent from *E. coli* (other than EIEC, which shares some virulence genes). Products of the virulence genes in *Shigella* aid in the adhesion, invasion, intracellular replication and cell-to-cell spread of *Shigella* (Marteyn, Gazi, & Sansonetti, 2012; Mattock & Blocker, 2017). The virulence genes recognised as being important in *Shigella* are summarised in Table 2.1.Virulence genes important for establishing initial infection are confined to a 30 kb region in the bacteria's virulence plasmid. This region contains the *mxi-spa* locus which codes for the Type III secretion system (T3SS), and the *ipa* and *ipg* genes which are crucial for the invasion and initiation of *Shigella* infection (Yang *et al*., 2005). Establishing infection requires a complex and coordinated regulation of all virulence genes. This is achieved by the activity of two transcriptional activators, virF and virB (Broach, Egan, Wing, Payne, & Murphy, 2012). VirF is important in the activation of virG/icsA which facilitates the intra- and inter-cellular spread of *Shigella* and virB regulates the gene components involved in the T3SS (Broach *et al*., 2012). There are also distinct regions located within the *Shigella* chromosome known as pathogenicity islands (PAI), that contain unstable transferable elements found in a variety of combinations in different *Shigella* species and subtypes (Yang *et al*., 2005). The composition of these gene clusters, when combined with other plasmid virulence factors, determines disease presentation and severity of symptoms (Mattock & Blocker, 2017).

Various *Shigella* virulence genes are located on PAIs, which are divided into five main groups: SHI-1, SHI-2, SHI-3, SHI-O and Stx-phage P27. The two most important groups

in *Shigella* pathogenesis are SHI-1 and Stx-phage P27. SHI-1 contains the genes *sigA*, *pic*, *set1A* and *set1B*. The *set1A* and *set1B* genes encode for the *Shigella* enterotoxin 1 (ShET1) protein; the *pic* gene encodes for intestinal colonisation and *sigA* gene encodes for a putative enterotoxin (Mattock & Blocker, 2017). Stx-phage P27 contains the gene *stxAB*, which encodes the Shiga toxin (Yang *et al*., 2005).

Effector genes are also important in the virulence of *Shigella*. They (effector genes) are encoded on the virulence plasmid, and the resulting effector proteins are secreted by the T3SS (Mattock & Blocker, 2017; Veenendaal *et al*., 2007). Important effector virulence genes include *icsB*, *ipaABCD*, *ipaH7.8*, *ipaH9.8*, *ospD3* and *virA*. The *icsB*, *virA, ipaB* and *ipaC* genes encode for effectors that aid in the lysis of the host cell membrane after the intra- and inter-cellular spread of *Shigella* in the epithelial cells; products of the *icsB* and *virA* genes also inhibit autophagy. *IpaABCD* aids in the macropinocytic uptake of *Shigella* into the colonic epithelial cells; more specifically *ipaD* and *ipaB* activate the T3SS, *ipaC* promotes actin polymerisation and induction of effector translocation and *ipaA* promotes actin depolymerisation of the actin filaments (Mattock & Blocker, 2017). *IpaH7.8* and *ipaH9.8* are part of the *ipaH* gene family and both are found on the virulence plasmid. The former is noted to promote macrophage pyroptosis and the latter inhibits NFκB activation, therefore reducing cellular inflammation in response to *Shigella* infection (Mattock & Blocker, 2017). *Shigella* enterotoxin 2 (ShET2) is encoded by the *ospD3/sen* gene and is thought to mediate early fluid secretion in the jejunum of the small intestine, producing watery diarrhoea as a result (Mattock & Blocker, 2017).

1 Table 2.1 An overview of important *Shigella* virulence genes.

Genes	Proteins	Plasmid or Chromosomal genes	Bacteria species	Function	Reference
set1a	*Shigella* enterotoxin 1	Chromosome	All *S. flexneri* 2a isolates	Active toxin of ShET-1	Vargas *et al*. (1999); da Cruz *et al*. (2014)
set1b	*Shigella* enterotoxin 1	Chromosome	All *S. flexneri* 2a isolates	Active toxin of ShET-1	Vargas *et al*. (1999); da Cruz *et al*. (2014)
ipaH	Invasion plasmid antigen	Plasmid & chromosome	All *Shigella*	Essential for host cell invasion & intracellular survival	Vargas, Gascon, Jimenez De Anta, and Vila (1999)
ial	Invasion-associated locus	Plasmid - 120-140 MDa	*S. flexneri*, *S. sonnei* & *S. dysenteriae*	Codes for epithelial cell penetration	Ghosh, Pazhani, Niyogi, Nataro, and Ramamurthy (2014)
invE	Transcriptional activator	Plasmid - 230-kb	*S. flexneri* & EIEC	Helps facilitate the intracellular spread of *Shigella* during infection & promotes the expression of multiple genes associated with virulence	Broach, Egan, Wing, Payne, and Murphy (2012)
ipaBCD	Effector proteins (Invasion plasmid antigens)	Plasmid	*S. flexneri* & *S. dysenteriae*	Key virulence factors in *Shigella* spp. Have effector functions essential for host cell invasion & intracellular survival; & also control the secretion & translocation of other effector proteins. These proteins also help with the polymerization & depolymerization of actin, facilitating bacterial invasion of the host cell.	Schroeder and Hilbi (2008); Lluque *et al*. (2015)
sen/osp3	*Shigella* enterotoxin 2	Plasmid - 140 MDa	Most, but not all *Shigella* & also EIEC	Active toxin of ShET-2 - associated with invasion	Vargas *et al*. (1999); da Cruz *et al*. (2014)
virF	Virulence regulon transcriptional activator	Plasmid	Predominantly in *S. flexneri* & in some *S. dysenteri*ae & *S. sonnei*	Primary regulator of plasmid-encoded virulence genes. Activates the transcription of icsA (VirG) & VirB, which is an activator of the ipaABCD virulence regulon. Also helps facilitate the intracellular spread of *Shigella* during infection.	Broach *et al*. (2012)
virG/icsA	Virulence facilitator	Plasmid	Predominantly in *S. flexneri*	Facilitates the intra- and intercellular spread of *Shigella* through actin-based motility.	Broach *et al*. (2012)

2

2.3 Epidemiology

2.3.1 Global burden of diarrhoeal diseases and shigellosis

Diarrhoeal diseases remain one of the greatest contributors to morbidity and mortality globally, particularly among children <5 years of age. The burden of diarrhoeal disease is greatest in the world's most vulnerable populations, particularly those residing in low and low-middle income countries. The importance of diarrhoea is sometimes overlooked, but it ranks behind only pneumonia as an infectious cause of morbidity and mortality in children and adults; causing more all-age illness than malaria, tuberculosis or HIV (Murray *et al*., 2012; C. L. Walker *et al*., 2013). Interventions are required to lower the burden of diarrhoea.

One of the main challenges addressing the high burden of diarrhoeal diseases is the myriad of aetiological agents that cause gastrointestinal infections. It is difficult to ascertain which pathogens are the greatest contributors to the burden of diarrhoeal disease, and thus which pathogens warrant targeted intervention. Studies have shown that *Shigella* is an important cause of diarrhoea and/or determined the burden of *Shigella* in developed countries such as the United States of America (USA) (Fischer Walker *et al*., 2012; Kotloff *et al*., 1999; McCrickard, Crim, Kim, & Bowen, 2018) and in some low-income settings in Africa (Fischer Walker *et al*., 2012; Kotloff *et al*., 1999; Platts-Mills *et al*., 2018) and Asia (Kotloff *et al*., 1999; Li *et al*., 2015; H. Liu *et al*., 2018). However, comparative data that enable a true indication of the burden of *Shigella* across multiple settings have been lacking.

The Global Enteric Multi-center Study (GEMS) was conducted to overcome this lack of data. The 3 year study sought to determine factors contributing to paediatric diarrhoeal

disease in sub-Saharan Africa and South Asia. It found rotavirus, *Cryptosporidium*, enterotoxigenic *E.coli* and *Shigella* to be the main pathogens responsible for moderate-severe diarrhoea in children (Kotloff *et al*., 2013). Thus, GEMS provided comprehensive, laboratory confirmed evidence that demonstrated *Shigella* to be a major contributor to diarrhoeal disease in children in low-income countries. *Shigella* was identified as one of the four leading pathogens responsible for causing moderate to severe diarrhoea in children. Moreover, unlike other leading causes of diarrhoea such as rotavirus, *Cryptosporidium* and *E.coli*, a child's chances of being infected with *Shigella* increased with age within the first 5 years of life (Kotloff *et al*., 2013). Although older children and adults are generally less susceptible to moderate to severe diarrhoeal illness, shigellosis does occur in these populations as well. Indeed, shigellosis is the second leading cause of diarrhoeal mortality amongst all ages globally (Lozano *et al*., 2012). Thus, *Shigella* is an important cause of moderate to severe gastrointestinal illness from infancy through to adulthood. Evidence of the importance of *Shigella* in high-burden settings provides impetus and justification for further research on this pathogen.

Despite improvements in our knowledge of the important role *Shigella* plays in the aetiology of diarrhoea in low and low-middle income countries, very little is known about the burden of *Shigella* in Oceanic LMICs. Outbreaks of shigellosis and transmission of *Shigella* have been reported in Fiji, New Caledonia, Samoa, Solomon Islands, Marshall Islands, PNG and the Torres Straits in far northern Australia (Germani *et al*., 1994; Greenhill *et al*., 2014; Gunzburg, Gracey, Burke, & Chang, 1992; Howard *et al*., 2000; Korff, 2015; Morahan, 1968; NNDSS Annual Report Writing Group, 2013; Painter *et al*., 2015; Rosewell *et al*., 2010; Schuurkamp, Bulungol, & Kereu, 1990; Soli *et al*., 2014; Storch, Gunn, Martin, Pollard, & Sinclair, 1980; Watson, 2006). Based on surveillance

data, shigellosis is likely to be one of the leading causes of diarrhoea in PNG (Benny *et al*., 2014; Greenhill *et al*., 2014; Soli *et al*., 2014). Despite cases and outbreaks being reported via the respective local media outlets and online sources in Pacific island nations, no further study or surveillance data has been published. However, based on global findings (Kotloff *et al*., 2013) and other piecemeal data derived from studies conducted in Oceania from the World Health Organization (WHO) and other health monitoring organisations (Germani *et al*., 1994; IHME, 2013a, 2013b, 2013c, 2013d; Painter *et al*., 2015; SPEHIS, 1999a, 1999b; Storch *et al*., 1980; Watson, 2006; World Health Organization, 2013), it is likely that shigellosis is an important contributor to disease in other Oceanic LMICs as well.

While the greatest burden of diarrhoeal illness occurs in low-income settings, where sanitation and hygiene is lacking, diarrhoeal diseases can also affect people in high-income settings. In countries such as the USA and European countries, *Shigella* is often found amongst children attending day-care centres, migrant workers, people living in correctional facilities, men who have sex with men, travelers visiting developing countries, and in communities struck by natural disasters or experiencing conflict (Niyogi, 2005). An improved understanding of the pathogen could have a globally positive impact.

Despite the challenges of determining the cause of diarrhoeal diseases, it seems reasonable to conclude that *Shigella* is one of the most important causes of diarrhoea (Fischer Walker *et al*., 2012; Kotloff *et al*., 2013). However, compared to many other agents of infectious disease, *Shigella* is largely neglected as a cause of diarrhoeal illness, particularly in developing countries, and remains an important public health problem (Kotloff *et al*., 1999; Niyogi, 2005). Indeed, *Shigella* has been recognised as a major

contributor to the burden of diarrhoeal disease for approximately 2 decades (Kotloff *et al*., 1999; Murray *et al*., 2012) although, until recently, there was a lack of strong aetiological data to support the epidemiological modelling. With those data now available, further research should be encouraged to combat shigellosis.

2.3.2 Global distribution of *Shigella* species

The distribution of each species of *Shigella* varies according to socio-economic circumstances. *S. sonnei* is the most common of the four species to be found in high-income settings (Niyogi, 2005; Zaidi & Estrada-Garcia, 2014). In low-income settings, *S. flexneri* is the most common (Kotloff *et al*., 1999; Niyogi, 2005), with *S. dysenteriae* and *S. boydii* also present. Moreover, it appears that as countries transition to middle-income status, the epidemiology of circulating strains also transitions. This supposition is based, in part, on the findings of von Seidlein *et al*. (2006): *Shigella* surveillance across six Asian countries over 4 years found *S. flexneri* to predominate in Bangladesh, Pakistan, Indonesia, China and Vietnam, while *S. sonnei* accounted for most cases in Thailand. It was postulated that this difference in distribution may correspond to Thailand's rapid progression towards being a middle-income country (von Seidlein *et al*., 2006). The findings of Vinh *et al*. (2009) seem to support this theory, with a shift in species distribution from *S. flexneri* to *S. sonnei* in southern Vietnam (Vinh *et al*., 2009; von Seidlein *et al*., 2006). A similar phenomenon was observed as early as the 1960s in countries such as Israel, the USA and parts of Europe. Circulating *Shigella* species transitioned from the once predominant *S. dysenteriae* to *S. flexneri* and *S. sonnei*; with the latter (*S. sonnei*) gradually becoming the predominant species in developed countries (Ashkenazi *et al*., 1993).

2.3.3 Molecular epidemiology of *Shigella*

S. flexneri is the most important *Shigella* species in terms of burden of disease. Seven phylogenetic groups (PGs) have been proposed by Connor *et al.* (2015). PGs 1, 2, 4 and 6 descended from an older lineage of *S. flexneri*, with an estimated most recent common ancestor (MRCA) dated to between the years 1341 to 1659. The most recent lineages of *S. flexneri* are PG 5 and 3, having an estimated MRCA dating to 1822 (PG5) and 1848 (PG3). Phylogenetically distinct lineages are observed to co-exist, with long-lived strains circulating within a country, alongside strains of recent incursion (Connor *et al*., 2015).

Four phylogenetic lineages (L) of *S. sonnei* have been identified. Three of these lineages include isolates spanning from the 1940s through to 2000s that are still in circulation; and the fourth lineage is derived from a single isolate from France (Holt *et al*., 2012). The progenitor for *S. sonnei* was estimated to exist <500 years ago, making it a more recent species of *Shigella* compared to *S. flexneri* (Connor *et al*., 2015; Holt *et al*., 2012). The earliest isolates in LI and LII have an estimated MRCA as recent as the 19th century, and all LIII isolates have their MRCA in the 20th century. The LIV isolates are ancestrally older, with the MRCA existing in the 17th century; that being the single isolate originating from France (Holt *et al.*, 2012). Data from Holt and colleagues (2013) indicated that unlike European isolates, isolates from non-European countries form tight, shallow-rooted phylogenetic clusters suggesting contemporary dispersal, (Holt *et al*., 2013). This suggests that the global spread of *S. sonnei* infections originated from small numbers of clones in Europe, and that its global distribution in part reflects the movement of Europeans over the past ~500 years.

2.3.4 *Shigella* and shigellosis in PNG

According to PNG's current National Health Plan for 2011-2020, diarrhoeal diseases are one of the five most frequently recorded illnesses in the country, with approximately 40 people per 1,000 presenting to hospital outpatients and/or health centres annually with diarrhoeal related illness (PNG National Department of Health, 2010). Over 2 people per 1,000 are admitted for diarrhoeal related illnesses. Age specific data on the burden of diarrhoea are not provided in the current health plan, but the burden is likely highest in children. Data published by the WHO (World Health Organization, 2013) states that 9% of deaths occurring in children <5 years of age in 2012 were diarrhoea related. However, this could be an underestimation of the true burden of diarrhoeal disease in PNG because of poorly documented outpatient records. Indeed, other WHO published data state that diarrhoeal illness accounted for 14% of deaths in children aged 1 month to 5 years of age in 2013, with no appreciable decrease in the burden over the past 14 years (World Health Organization, 2013). Moreover, the impact of diarrhoeal diseases on the general health of children, and therefore their susceptibility to other causes of death, may not be fully accounted for in these estimates.

Despite recognition that diarrhoea is an important cause of morbidity and mortality in PNG, particularly among children, data on the aetiology of diarrhoeal disease in PNG are lacking. The most robust data comes from a case-control study by Howard and colleagues (Howard *et al*., 2000), although the data are now outdated with recruitment completed more than one-quarter of a century ago. The study found *Shigella* to be one of the most important aetiologies for children hospitalised with diarrhoea in Goroka, Eastern Highlands Province (EHP), being detected in 13% of cases. Recent surveillance at the PNG Institute of Medical Research (PNGIMR), conducted in the same setting and a

similar cohort (children hospitalised with diarrhoea in Goroka, EHP), found *Shigella* spp. and rotavirus to be the most commonly detected pathogens, with *Shigella* detected in 26.6% of children (Soli *et al*., 2014). In another recent study, focusing on both children and adults presenting with diarrhoea to hospital outpatients or an urban health clinic, *Shigella* was isolated from 22% of all study participants (Greenhill *et al*., 2014). On the basis of the aforementioned studies, it appears that *Shigella* remains an important pathogen in the Eastern Highlands of PNG.

Other studies provide evidence of the nationwide distribution of *Shigella*. The first published reports of *Shigella* infection in PNG (which was then the territory of Papua and New Guinea under Australian administration) were by Curtis (1964) and Morahan (1968). Curtis's study observed antimicrobial sensitivity on 70 isolates of *Shigella*, with 86% of them being *S. flexneri* and the remaining 14%, *S. sonnei*. Resistance to antibiotics was uncommon: of 70 isolates nine (13%) were resistant to streptomycin, 3 (4%) to chloramphenicol and 6 (9%) to tetracycline, with all isolates demonstrating resistance being *S. flexneri* (Curtis, 1964).

Morahan's study occurred over a 33 month period from March 1965 to November 1967. Faecal specimens were collected from the Wewak Hospital and other hospitals in the Sepik districts. A total of 907 stool samples were obtained from 848 symptomatic persons, of which 90 were *Shigella* positive. *S. flexneri* was the most common species isolated (and was responsible for two fatal cases of shigellosis), followed by *S. sonnei* and then *S. boydii* (Morahan, 1968).

In early 1987, Schuurkamp *et al.* (1990) monitored food handlers in PNG working at the Ok Tedi Mines in Western Province. The mining company re-enforced their pre-medical health checks on all potential food handlers following two cases of typhoid fever imported from the Highlands. An initial screen of 155 food handlers and 85 non-food handlers resulted in the isolation of *Shigella* spp. from 2.6% of food handlers and 3.5% of non-food handlers; all were asymptomatic carriers. Interestingly, *S. boydii* and *S. sonnei* were responsible for the majority of *Shigella* infections experienced on site (Schuurkamp et al., 1990). A parallel survey of 160 food handlers, including those from private fast food establishments, detected one *S. boydii* infection in a local fast food shop staff member (Schuurkamp *et al.*, 1990). The two species detected in this study are infrequently detected in PNG or other countries in the region, as *S. boydii* is predominantly confined to the Indian subcontinent and *S. sonnei* to more developed countries such as the USA and European countries (Niyogi, 2005).

In 2013 there was an outbreak of shigellosis (diarrhoea and dysentery) in a settlement camp near Bulolo, Morobe Province. Six samples were sent for laboratory analysis, with four confirmed positive for *Shigella*. There were an estimated 1200 cases of suspected shigellosis (with five fatalities), making it one of the largest outbreaks of shigellosis ever reported in the scientific literature (Benny *et al*., 2014).

Access to safe water, sanitation and hygiene (WASH) plays a crucial part in efforts to lower the burden of all causes of diarrhoea, and of shigellosis in particular, especially in resource poor settings like PNG. Challenges in the uptake of such measures exist (Horwood, Barrington, & Greenhill, 2013; Phuanukoonnon, Namosha, Kua, Siba, &

Greenhill, 2013), but the impact on diarrhoeal illness, and overall health would be significant (Horwood *et al*., 2013) if these key factors are addressed accordingly.

The study conducted on food handlers by Schuurkamp *et al.* (1990) demonstrates that asymptomatic carriage of *Shigella* occurs in PNG: a finding supported by recent published data (Horwood *et al*., 2017). Indeed, asymptomatic carriage has also been reported amongst refugees in South Sudan present in refugee camps and in European refugee processing centres, with approximately 14% of the 88 participants from a range of ages (i.e. 1 – 18+ years) testing positive for carriage (Bliss *et al*., 2018). Another study by Phantouamath *et al*. (2005) conducted in Laos also observed *Shigella* carriage in 6% non-diarrhoea presenting participants. The role of asymptomatic carriage in the persistence and transmission of shigellosis may be worthy of further investigation, as to date it has not been considered an important contributor to the epidemiology of the infection.

2.3.5 Burden of *Shigella* in the Pacific region

There are scarce data pertaining to the burden of diarrhoeal disease in other LMICs in Oceania. While there is evidence to suggest the burden of diarrhoeal disease has decreased significantly in many Oceanic LMICs over the past ~20 years (IHME, 2013a, 2013b, 2013c, 2013d); the comprehensive Global Burden of Disease study ranked diarrheal diseases as the second leading contributor to disability adjusted life years (DALYs) in the Pacific Islands for the years 1990 – 2010 (Hoy *et al*., 2014). Other sources also suggest a high burden of diarrhoeal disease in parts of the South Pacific (SPEHIS, 1999a, 1999b). Despite some improvements in gastrointestinal health in Oceanic LMICs, there remains a considerable burden of diarrhoeal disease in this region.

A concern for the future may be climate change, which could limit access and storage of fresh, uncontaminated water in small Oceanic island nations due to rising seawater levels and flash flooding, (Singh *et al*., 2001). Thus, despite improvements in recent decades, it may be difficult to sustain these improvements in LMICs in Oceania.

Limited data exists on the burden of shigellosis in some areas of Oceania; this is likely a reflection of poor ongoing surveillance. Despite limitations in the data, there is evidence that shigellosis has been an important cause of diarrhoeal illness over previous decades, and across various countries within the region. For instance, in June-July of 1977 an outbreak of shigellosis occurred in the Marshall Islands (Storch *et al*., 1980). A total of 147 faecal samples from patients with diarrhoea was analysed. *S. flexneri* type 1 was isolated from 32 of the stool samples, *S. flexneri* type 1b from 8 of the samples, and *S. flexneri* type 4a from one sample.

In New Caledonia, a 2 year study of endemic enteric pathogens associated with acute diarrhoea was conducted in 1990-1991. *Shigella* was found to be a predominant enteric pathogen causing diarrhoeal illness in rural areas of New Caledonia. Interestingly in this setting, of the 62 *Shigella* strains isolated, most were *S. sonnei*, with *S. flexneri* less frequent (Germani *et al*., 1994).

In a more recent study from American Samoa in May-June 2014, an outbreak of diarrhoeal illness caused by *S. flexneri* was reported (Painter *et al*., 2015). Of six stool samples sent to the Centers for Disease Control and Prevention (CDC) for analysis, four were positive for *Shigella* (Painter *et al*., 2015).

Although Australia is a high-income country (Figure 1.1), there is an uneven distribution in wealth and living conditions. In particular, indigenous Australians are considerably more at risk of infectious diseases (Gibney, Cheng, Hall, & Leder, 2017). A study by Gunzburg *et al*. (1992) looked at the epidemiology and microbiology of diarrhoea in young Aboriginal children in the Kimberley region of Western Australia and observed *E. coli*, *Salmonella* and *Shigella* as being the predominant bacterial enteric pathogens responsible for causing diarrheal disease among children in this region. Thus, it appears that shigellosis is likely an important diarrhoeal pathogen amongst Aboriginal and Torres Strait Islanders. Limited data exist on the epidemiology of diarrhoeal disease in the Torres Straits. However, there are strong geographical and cultural ties between the Torres Straits to PNG. Cross border transmission of infectious agents, for example tuberculosis (Bainomugisa *et al*., 2019), is known to occur and could well apply to diarrhoeal pathogens.

2.4 Diagnosis of Shigellosis

Diagnosis of shigellosis in low-resource settings is predominantly based on clinical presentation. For example, in PNG, patients presenting at the local clinic/hospital with a high fever and acute-severe watery diarrhoea with the presence of blood in their stools are usually suspected of having a bacterial infection. But confirmed diagnosis is strongly recommended to determine the exact cause of infection where resources are available, as diarrhoeal illnesses can be caused by a plethora of infectious agents (PNG National Department of Health, 2012; PNG Paediatric Society, 2016; Poka & Duke, 2013). Classically, diagnosis is based on bacterial culture and confirmation by serotyping with poly-O antisera, followed by screening isolates for AMR (where resources are available to determine cause of infection). However, these resources are often not available in low-

resource settings, so patients are treated (where treatment is available) based on the presentation of symptoms. Thus, unfortunately, the underlying cause of infection is not identified which than affects the outcome of the patients treatment regimen (PNG National Department of Health, 2012; PNG Paediatric Society, 2016; Poka & Duke, 2013).

Due to the lack of definitive diagnosis of the aetiology of diarrhoea there is a risk that antibiotics are prescribed more often than required; thus increasing the risk of antimicrobial resistance (AMR). There is a need for accurate, cost-effective diagnostic methods for diarrhoea that yield fast and easy to interpret results in low-income, low-resource settings.

2.5 Treatment of Shigellosis and Development of Antimicrobial Resistance

2.5.1 Global treatment and resistance

Current treatment guidelines for shigellosis provided by the WHO in 2005 (yet to be updated), recommend persons presenting with bloody diarrhoea to be prescribed specific antimicrobials effective towards *Shigella* (World Health Organization, 2005). The WHO guidelines in 2005 recommended treatment of shigellosis by prescribing ciprofloxacin as a first-line treatment followed by azithromycin as a second-line treatment for adults; and pivmecillinam and ceftriaxone to treat multi-drug resistant *Shigella* (Williams & Berkley, 2018). Unfortunately resistance to antimicrobials ampicillin, co-trimoxazole, nalidixic acid and tetracyclines has increased globally, and ciprofloxacin resistance *Shigella* has been reported in Asia in recent years (Chung The *et al*., 2016; Kim *et al*., 2015).

2.5.2 Treatment and resistance in PNG and Oceania

Irrespective of the cause, hydration is central to the treatment of diarrhoea and is especially important in children. With this in mind, treating diarrhoea based on severity of illness is routinely conducted in PNG (PNG National Department of Health, 2012; PNG Paediatric Society, 2016) and can ensure adequate health outcomes in the absence of aetiological data. In children with mild diarrhoea, with no signs of dehydration, parents are advised to give extra fluids; oral rehydration solution is recommended for children with moderate cases of diarrhoea and signs of dehydration; intravenous fluids are recommended for patients with severe diarrhoea and severe dehydration. Continued breastfeeding of babies is strongly advised to reduce the severity of diarrhoea and complications of malnutrition, which can be experienced in infants and young children; and zinc supplementation is also recommended (Duke, 2011; PNG Paediatric Society, 2016; Poka & Duke, 2013; Vince, 1995).

The optimal treatment and management of diarrhoeal diseases requires an understanding of the cause of the infectious agent responsible for the disease, which unfortunately is not currently obtainable in most Oceanic LMICs. This can lead to the overuse of antibiotics in the treatment of diarrhoea, which has been reported in PNG (Poka & Duke, 2013; Vince, 1995). The use of antibiotics may be warranted if there is evidence of shigellosis (evidence usually in the form of dysentery); however, there are other (sometimes non-bacterial) causes of dysentery, for which administered antibiotics will have limited or no impact. Nonetheless, ciprofloxacin is among the most commonly recommended antibiotics for both children and adults with dysentery (PNG National Department of Health, 2012; PNG Paediatric Society, 2016; Poka & Duke, 2013; Vince, 1995). Previously co-trimoxazole (or nalidixic acid or ampicillin) was recommended (Vince,

1995), but with the rise of AMR reported globally and also observed in recent studies in PNG, the recommendations for the antibiotic treatment of shigellosis have been updated (PNG National Department of Health, 2012; PNG Paediatric Society, 2016). Two recent studies have reiterated the need for the change in antimicrobial treatment regimes. Rosewell *et al.* (2010) reported that *Shigella* was commonly resistant to amoxicillin, chloramphenicol and co-trimoxazole; these findings are consistent with the more recent study conducted by Greenhill *et al*. (2014).

Resistance to frequently used antibiotics is common among *Shigella* isolates globally and in PNG. In 2009, an epidemic caused by multi-drug resistant *S. flexneri* was reported in four provinces of PNG. This led to a review of antimicrobial susceptibility of *Shigella* spp. isolated at the Port Moresby General Hospital between 2000 and 2009. Of 3,419 faecal samples cultured, approximately 4% were positive for *Shigella*, with *S. flexneri* isolates showing antimicrobial resistance to amoxicillin (98%), chloramphenicol (64%) and co-trimoxazole (86%) (Rosewell *et al*., 2010). Unfortunately, there are little data from other parts of the country pertaining to antimicrobial resistance for comparative purposes; though the recent study in Goroka, Eastern Highlands Province (EHP), yielded similar antimicrobial resistance patterns in *Shigella* (Greenhill *et al*., 2014). Notwithstanding the shortcomings addressed above (with other causes of dysentery), given that *Shigella* is the most common bacterial cause of dysentery and circulating strains are commonly resistant to other antibiotics (Greenhill *et al*., 2014; Rosewell *et al*., 2010), the current practice of recommending ciprofloxacin for treatment of dysentery in PNG is appropriate.

It is important to note that shigellosis does not always result in dysenteric stools. Indeed, in a recent surveillance study of children hospitalised with acute watery diarrhoea,

children with dysentery were excluded from the surveillance; yet *Shigella* was detected in 26.6% of children. If antibiotics are required during the treatment of diarrhoea with signs of severe dehydration, the standard treatment manual for children recommends chloramphenicol as a key component of treatment (PNG Paediatric Society, 2016). Recent data (Greenhill *et al*., 2014; Soli *et al*., 2014) suggests that the most common bacterial cause of diarrhoea (in the absence of dysentery) is *Shigella*, where chloramphenicol treatment is likely to be ineffective.

Other studies in Oceanic LMICs also point towards *Shigella* being a cause of diarrhoea and observe a rise in resistance to chloramphenicol. Approximately 20 years ago, multi-drug resistant *Shigella* was reported in Fiji. The Fiji Ministry of Health recorded 68 cases of shigellosis in 1996, 173 cases in 1997 and 334 cases in 1998 (Watson, 2006). Drug resistance to chloramphenicol was reported for 82% of the cases (Watson, 2006). In the Marshall Islands, multi-drug resistant (MDR) bacteria were commonly detected. While some isolates (n=18) were resistant only to tetracycline; one isolate was resistant to tetracycline, ampicillin and carbenicillin; two isolates were resistance to tetracycline, chloramphenicol, ampicillin and carbenicillin; and five isolates were resistant to all antibiotics tested (Storch *et al.*, 1980). Given the role of *Shigella* in non-dysenteric diarrhoea and the high rates of resistance to chloramphenicol in *Shigella* (Greenhill *et al*., 2014; Rosewell *et al*., 2010; Watson, 2006), treatment with chloramphenicol may be of little benefit to the patient. This situation may warrant the change in standard treatment protocols in the region, and potentially other high-burden, low-income settings; with ciprofloxacin likely to be a more effective treatment option.

2.6 Conclusions

Despite limited data existing on the burden of *Shigella* in the region, extrapolation from global data, various reports of *Shigella* outbreaks, and diarrhoeal disease surveillance suggest that *Shigella* is likely an important contributor to diarrhoea and dysentery in Oceanic LMICs. *S. flexneri* and *S. sonnei* appear to be the most common species of *Shigella* circulating in the region.

As with any pathogen, virulence genes are essential for the pathogenesis of *Shigella*. Without the presence of such genes, initiation and establishment of infection by *Shigella* in the host would not be possible. Examination of virulence genes could provide further insight into the strains of *Shigella* circulating, and may provide a proxy for virulence of circulating strains.

There is evidence of the development of AMR in *Shigella* in Oceanic LMICs. AMR is now commonplace to first-line treatments such as amoxicillin, chloramphenicol and co-trimoxazole in *Shigella*.

Ongoing and improved surveillance of *Shigella* in PNG and other Oceanic LMICs could improve our understanding of this important pathogen in the region. In doing so, consideration might be given to both traditional and genomic approaches to strain profiling and AMR determination.

Chapter 3 - Survey to Investigate Factors Associated with Diarrhoeal Illness in Eastern Highlands, Papua New Guinea

3.1 Introduction

Collectively, access to clean drinking water and appropriate disposal of faecal waste remains a major determinant of global health outcomes. This has led to the development of progams focusing on water, sanitation and hygiene (WASH) as an important intervention to improve health and social factors. Access to WASH is generally considered to impact positively on health, educational outcomes and gender equity. There have been improvements in the provision of WASH globally over the past ~20 years; with the recent WHO/UNICEF Joint Monitoring Program for Water Supply, Sanitation and Hygiene (WHO/UNICEF, 2019) stating that 1.8 billion people have gained access to improved water since 2000, and 2.1 billion people have gained access to basic sanitation services. However, the same report states that there remain areas of the world where adequate WASH is not available. Approximately 10% of the global population lack basic water services, and almost as many people (~8% of the world's population) practice open defecation. Particularly pertinent is the inequality in access to improved WASH, with rural populations poorly served (WHO/UNICEF, 2019).

There is global recognition that lack of access to safe water and inadequate sanitation and hygiene is inextricably linked to diarrhoeal disease. The WHO suggests that "lack of access to safe, clean drinking-water and basic sanitation, as well as poor hygiene cause nearly 90% of all deaths from diarrhoea" (World Health Organization, 2011), with most of those deaths occurring in children in low- and middle-income countries. Given the

global situation, it is likely that one of the main contributing factors to diarrhoeal diseases in PNG and elsewhere in Oceanic LMICs is inadequate WASH.

The 'inadequate WASH' in PNG and some other LMICs in Oceania is likely borne out of various contributing factors, including lack of appropriate infrastructure, inadequate maintenance of existing infrastructure, lack of perceived need to utilize such utilities, and a lack of appreciation of the importance of personal sanitation and hygiene. However, these are assumptions largely based on findings in international settings outside of Oceania. Factors contributing to, and outcomes of, inadequate WASH in PNG and more broadly Oceanic LMICs have been addressed in the scientific literature (Barrington *et al*., 2016; Fleming *et al*., 2019; Horwood *et al*., 2013; Horwood & Greenhill, 2012; Jenkins, 1995; Passey, 1995); including an excellent regional-focused review by Barrington (2016). Moreover, data suggesting poor standards and access to WASH are commonly stated by relevant organisations who monitor WASH and other public health indicators (e.g. JMP, https://washdata.org/data). Studies focusing on first hand, lived experiences of people living in Oceanic LMICs have been conducted (Kodish *et al*., 2019; Phuanukoonnon *et al*., 2013; Psutka *et al*., 2013), but are relatively rare in the scientific literature. However, it is these populations that are at risk of illness as a direct result of inadequate WASH, and their insights are valuable, as recognized by Jenkins (1995).

This study therefore aimed to gain an understanding of participants' awareness of, and behaviour towards, WASH.

The global importance of shigellosis in the aetiology of diarrhoea (Kotloff *et al*., 2013), and evidence of shigellosis in PNG, including in children admitted to hospital for acute

watery diarrhoea (Soli *et al*., 2014), led to an exploratory study seeking to gain an insight into factors associated with diarrhoeal illness in PNG.

3.2 Methods

3.2.1 Study population

A questionnaire (see Section 3.2.2) was administered to parents of children presenting to Eastern Highlands Provincial Hospital (EHPH) with symptoms of diarrhoea. Most of those children participated in routine surveillance of rotavirus, a program supported by the National Department of Health and WHO in PNG. The rotavirus surveillance program seeks to obtain samples from children hospitalized with watery diarrhoea; however, at the time this study was being conducted patients presenting to the paediatrics outpatients of EHPH with watery diarrhoea were also invited to participate. Voluntary informed consent was obtained from a parent/guardian of the hospitalized child; allowing for the questionnaire to be conducted, and for further diagnostic tests to be conducted on stool samples collected as part of the rotavirus surveillance. The questionnaire was administered in 2017.

Ethics approval was sought to conduct a questionnaire (Section 3.2.2) and conduct additional diagnostic testing (Section 3.3.4) through the PNGIMR Institutional Review Board, the PNG Medical Research Advisory Committee (MRAC no. 16.43), and Federation University Australia Human Ethics (A17-074).

3.2.2 Questionnaire

A study nurse from the PNGIMR delivered the questionnaire to parents of patients presenting to the outpatients, or hospitalized (in-patient) with clinical presentation of

diarrhoea at the EHPH, PNG. Questions were asked based on demographics (i.e. where the person(s) lived, housing, education, etc.); their access to water and sanitation facilities; and their personal hygiene practices. A copy of the questionnaire is provided in Appendix 2. Additional questions were asked pertaining to treatment; however, the responses to those questions were not analysed in this study.

3.2.3 Data analysis

Simple descriptive statistics were applied to the resulting data. To further investigate health seeking behaviour, approximate travel time of participants from their place of residence to Goroka town centre was calculated. Distance from Goroka, topography and local road conditions were considered when conducting the calculations. Chi-square analysis was conducted to determine whether correlations existed between travel time to EHPH and selected variables (as sought in the questionnaire).

3.2.4 Diagnosis of diarrhoeal illness

Faecal samples were collected from participants as part of the rotavirus surveillance program being conducted in collaboration with WHO PNG, The PNG National Department of Health and PNGIMR. Written permission (in addition to ethical approval, as stated in Section 3.2.1) was granted to the study team by WHO PNG to use rotavirus samples to better understand factors contributing to diarrhoea in PNG.

Where feasible, an additional sample was collected to enable bacterial culture to be conducted; however, the ongoing rotavirus surveillance took precedence. Bacterial culture was conducted based on widely accepted methods that target the detection of *Salmonella* and *Shigella*. In brief, culture was conducted using xylose lysine desoxycholate (XLD)

agar and MacConkey agar, with enrichment in selenite broth. Presumptive colonies were confirmed using biochemical tests and PCR. These methods were as outlined in Greenhill *et al*. (2014).

3.3 Results

3.3.1 Participants

There were 84 respondents to the questionnaire, who were parents/guardians of participants of the aforementioned rotavirus surveillance. Some questions in the questionnaire (Appendix 2.1) pertained directly to the parent/guardian, and other questions to the child. During the application of the questionnaire, respondants were asked to respond to Section 4 (on sanitation and hygiene) from their own perspective (i.e. that of the parent/guardian), not on behalf of the sick child.

Of the 84 rotavirus surveillance participants (i.e. the sick children), 46 were male and 38 female (ratio 1.2:1), with no significant difference). The average age of participants (where recorded, n=78) was 2.12 years, median 1.03 years, and the age range of all participants was 1 week old to 14.6 years old. The age of the respondents (parent/guardian of sick children) was not recorded.

Children aged 0-1 years old (infants) made up 48% (37/77) of the participants, of which 22% (8/37) were positive for *Shigella*. All *Shigella* positive participants presented at the health centre with decreased appetite and fever. Six of the eight *Shigella* positive infants had experienced vomiting and all (8/8) were dehydrated. Stool samples were reported to be ‘very watery’ and had a foul stench.

Children aged 1-5 years old made up 42% (32/77) of the participants, of which 25% (8/32) were positive for *Shigella*. Positive participants experienced vomiting, presented with fevers and some had decreased appetites and were dehydrated, with two experiencing stomach pains. Most stool samples (7/8) were 'very watery'and were also reported to have a foul stench, while 1/8 was reported to have blood and mucus/pus in stool samples.

Children aged >5 years made up 11% (8/77) of the participants, of which none were positive for *Shigella*. All stool samples (8/8) were 'very watery' and were reported to have a foul stench, while 1/8 contained blood in the stools and 1/8 had very dark (black) stools.

3.3.2 Demographics

Location data were recorded for 83 of the 84 rotavirus surveillance participants, with the majority of participants coming from the highlands region of PNG (77/83, 93%), particularly EHP (69/77, 90% of highland participants); the province in which the study was conducted. However, a small number of participants (6/83; 7%) resided in coastal, island or low-land regions of PNG at the time of their involvement in the study (Figure 3.1).

Approximately 41% of participants resided <30 min from the EHPH, 13% resided an estimated 30 – 60 min away, 31% resided 60 – 120 min away, and 14% resided >120 min away. Those that most likely travelled by plane (and/or boat for part of the journey) to get to Goroka were all considered to live >120 min away.

Seventy-nine respondents provided an answer for the number of people with whom they share a house. Approximately 9% of respondents shared a house with up to two other people, with 91% sharing a house with up to five other people. Interestingly, 80% of respondents provided more than one answer, which suggests that the number of occupants varies (see Section 3.4 for further explanation).

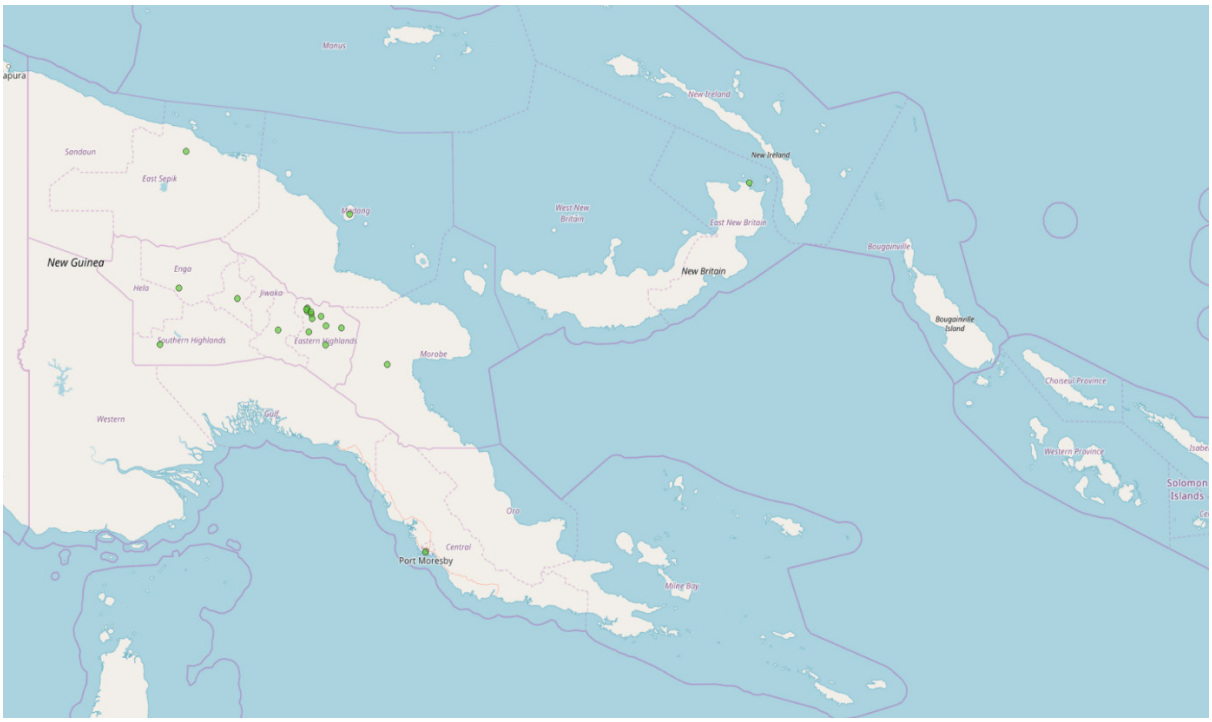


Figure 3.1 Place of residence of participants who answered the questionnaire. While most participants were from EHP, participants also came from other provinces, particularly other highland provinces.

3.3.3 Symptoms and disease diagnosis

Most participants (48/84, 57%) were admitted to hospital 2 – 6 days after parent-reported onset of symptoms of diarrhoea. There was no significant association between distance from the EHPH and time to admission after onset of symptoms.

The two most common symptoms, as reported by the parent/guardian of the child participant, were very watery diarrhoea and foul smelling faeces: 77 of the 82 (94%) respondents stated that the child participant had both these symptoms. Of the 77 children with both watery diarrhoea and foul smelling faeces, 64 (83%) reported only those two characteristics associated with faeces. At least one other symptom (in addition to watery diarrhoea and foul smelling faeces) was reported by 13 respondents. The most common additional symptom was blood in stools (7/84, 8% of all participants).

Symptoms other than characteristics of child faeces were also reported by parents/guardians. Of the 82 respondents to that section of the questionnaire, 80 (98%) reported their child to be suffering decreased appetite, and 80 (98%) children reported the child to be suffering dehydration (two children suffered decreased appetite in absence of dehydration, two children suffered dehydration in the absence of decreased appetite, and the remaining 78 children suffered both dehydration and reduced appetite). Other common symptoms were fever (63/82, 77%) and vomiting (51/82, 62%).

Of 84 participants, faecal samples from three participants were found to be positive for *Shigella*, and 15 positive for rotavirus. No correlations existed between symptoms reported and the detection of either pathogen.

3.3.4 WASH

3.3.4.1 Water

Analysis of the questions about aspects of WASH revealed that most respondents use a variety of water sources. Almost all respondents (93%) listed two or more sources of water. A summary of the responses for access and source of water are provided in Table

3.1. Two respondents stated that water was available either within the house or close to the house, but did not select the option of water being a long distance from the dwelling (indicative that they typically obtain their water from close proximity to their dwelling). In comparison the vast majority of respondents (73/81; 90%) stated that they travel a long distance to get water for at least some of their water needs.

Chi-square analysis was not conducted on any WASH data, as the majority of respondents answered with the same response (thus no differences would be determined).

Table 3.1 Summary statistics of responses to questions pertaining to access to water for household use.

Water Source and Distance	Proportion (%)
Multiple sources of water used (80)	93
Distance to water (81)	
In-house or close only	2
Long distance (at least sometimes)	90
Water sources used (80)	
Piped water	86
Flowing water*	83
Tank water	30
Well water	5

*Flowing water was from a spring, stream or river.
Number in parenthesis indicates number of responses to that question, out of a possible total of 84. Proportions are given using the number in parenthesis as the denominator.

3.3.4.2 Toilets

Similar to the use of multiple water sources, respondents also stated that they used multiple forms of toilets to dispose of faecal waste (Table 3.2). A small proportion of respondents (12%) used a flushing toilet, with half of those using only a flushing toilet. The most common form of toilet used was a pit toilet, with all respondents except those that listed only a flushing toilet using a pit toilet. A pit toilet was the only form of toilet

listed by over one-third of respondents, and a pit toilet or bush disposal accounting for an additional one-third of participants.

Table 3.2 Summary statistics of responses to questions pertaining to access to toilets.

Toilet Access and Use	Proportion (%)
Toilet (81)	
Only uses flushing toilet	6
Has access to flushing toilet	12
Pit toilet only	37
Pit toilet or bush toilet	33
Pit toilet and other toilet	23

Number in parenthesis as for Table 3.1.

3.3.4.3 Hygiene

Handwashing is an infrequent occurrence in this setting. Only 1% of respondents stated that they usually wash their hands after defecation, or before food preparation. The use of soap was similarly infrequent (Table 3.3).

Table 3.3 Summary statistics of responses to questions pertaining to hand washing.

Hand Washing	Proportion (%)	
	Usually	Sometimes
Handwashing after toilet (n=81)	1	99
Handwashing before food prep (n=77)	1	99
Use soap (n=80)	1	99
Soap in house (n=80)	1	99

Number in parenthesis as for Table 3.1.

Of 77 respondents, 12 (16%) stated that they share their house with pigs.

Study participants were asked to rate their own personal hygiene. Many participants provided more than one answer, thus during analysis the worst-case ranking was

considered. Of 77 respondents, one respondent (1%) rated their hygiene as excellent and 46 (60%) as good. The remaining respondents (~38%) stated that at best their personal hygiene could be better.

3.4 Discussion

This study demonstrates that access to WASH remains an important challenge in regional and rural PNG. The study design did not allow for risk ratios to be determined, and the data did not lend itself to correlations being detected. Nonetheless, the study provides preliminary data to support the notion that poor sanitation and hygiene (in parts of PNG, and by extension potentially other MICs in Oceania) contribute to the transmission of gastrointestinal pathogens and thus diarrhoeal disease.

Interestingly, the data obtained in this study suggest that access to WASH may be lower in EHP than expected based on commonly reported data for PNG (e.g. https://washdata.org/ and Figures 3.2 and 3.3). It is likely that data obtained by organisations such as WHO and UNICEF use different approaches to those used in this study, so direct comparisons should be interpreted with caution. Nonetheless, the data obtained in the current study are indicative of poor access to sanitation and hygiene measures, with pit toilets and bush toilets commonly used; and evidence of only sporadic hand washing and use of soap. Similarly, a variety of water sources are commonly used in this study setting, much of which is at risk of faecal contamination. Further investigation into access to safe water and sanitation is warranted in this setting.

Figure 3.2 Service levels for drinking water in urban and rural areas of PNG, Solomon Islands and Vanuatu for years 2000 and 2015. Most drinking water in rural PNG is surface water or from other unimproved sources. Data from WHO/UNICEF database for sanitation services in PNG, Solomon Islands and Vanuatu, https://data.unicef.org/topic/water-and-sanitation/drinking-water/

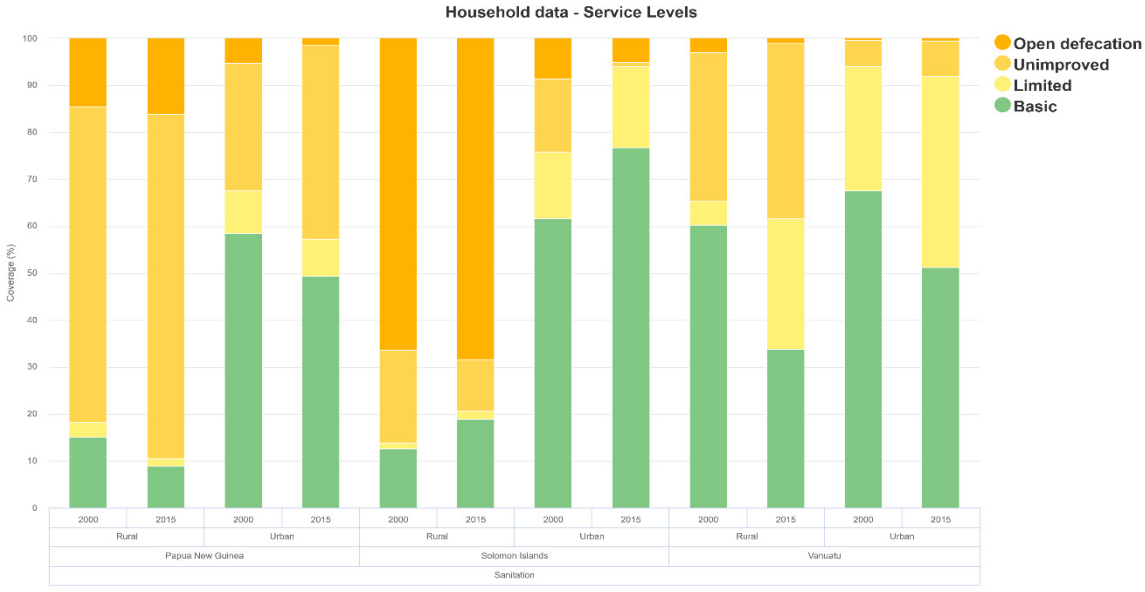


Figure 3.3 Basic access to sanitation services in urban and rural areas of PNG, Solomon Islands and Vanuatu for years 2000 and 2015. Open defication and unimproved sanitation remains common in rural PNG where most of the population live. Data from WHO/UNICEF database for sanitation services in PNG, Solomon Islands and Vanuatu, https://data.unicef.org/topic/water-and-sanitation/drinking-water/

Assuming this study reflects the true WASH situation, it is interesting to note that access to WASH is lower in study participants than commonly stated national standards. It could be that study participants have lower standard of, and access to, WASH than the broader population; and this low level leads to illness in the family, particularly diarrhoea.

This study is indicative of low levels of sanitation and hygiene in EHP of PNG, but it did not seek to provide accurate or precise data. The answer options in the questionnaire survey were intentionally broad, to prevent the need for a high number of options for the answer. Moreover, many of the questions were also open to interpretation by the respondent; and it was common for the respondent to provide multiple answers when the questionnaire was designed for only one response (for most questions). This led to the interpretation that when more than one answer was provided, it was indicative of multiple correct answers (e.g. a family can foreseeably use more than one type of toilet, or access water from more than one source). Less easily interpreted was that the majority of participants provided more than one response for the number of people with which they share a house; perhaps suggesting that the number of people within a dwelling changes.

It is difficult to ascertain what the outcomes of this study means in relation to other LMICs in Oceania. PNG is constantly reported as having among the worst metrics for WASH in the Oceania region (https://washdata.org/). Thus it is unlikely that other countries in Oceania, on the whole, have such limited access to WASH. However, as stated in Section 3.1, there are inequalities in access to WASH, and rural areas are commonly poorly served (WHO/UNICEF, 2019). Other rural areas in PNG and some countries within Oceania may have poorer access to WASH than the currently available data indicate.

This study was unable to determine correlations between symptoms and cause of disease; likely attributable to two factors. First and foremost, it is difficult to differentiate aetiologies of diarrhoea based on signs and symptoms. Consequently, there was considerable overlap in the symptoms (as reported by parents) for the majority of sick children. Secondly, a confirmed cause was determined for relatively few cases of diarrhoea in this study (18/84; 21%). In previous studies using rotavirus samples from EHP in PNG, rotavirus was detected in 26 – 31% of samples (Horwood *et al*., 2012; Soli *et al*., 2014), compared to 18% in this study. More telling, in a previous study *Shigella* was detected in 27% of samples collected for rotavirus surveillance (Soli *et al*., 2014); although in that study detection was conducted by PCR on archived samples, as opposed to culture used in the current study. In establishing this study it was expected that the detection of rotavirus and *Shigella* (and thus passive detection of *Salmonella* as well) would result in an aetiology being detected in ~50% of cases.

Given the relatively recent study looking at a broader range of aetiologies (Soli *et al*., 2014), it was not considered a good use of resources to repeat a broad-scoped aetiological survey. Although conducted more than 30 years ago, a study by Howard et al, also noted similar results to the aforementioned recent studies. The most commonly isolated pathogens from children aged 0-11 years in Goroka, were rotavirus (23%), *Shigella* spp. (13%), *Campylobacter* spp. (12%), *Cryptosporidium parvum* (10%) and enteropathogenic *Escherichia coli* (8%) (Howard et al, 2000).

The inability to differentiate the cause of diarrhoea based on symptoms is a reminder of the need for improved diagnosis and pathogen detection in people with diarrhoea, particularly if it impacts on optimal treatment.

Despite the lack of correlation between symptoms and aetiology in this study, the collection of data on parent/guardian reported symptoms provides some insights. One point of interest is that for two of the three culture confirmed cases of shigellosis, the parent/guardian suspected typhoid fever. This suggests that the parent/guardian recognized that the child had a serious illness. Typhoid fever is commonly clinically diagnosed in PNG (in the absence of adequate laboratory diagnosis): it might be that in some cases other gastrointestinal bacterial pathogens such as *Shigella* are the cause of illness.

Interestingly, despite the link between inadequate WASH and diarrhoea, insufficient studies have been conducted to date. A meta-analysis conducted by Cairncross et al. (2010) using studies conducted anywhere in the world, concluded that the risk of diarrhoea was reduced by 48% by handwashing with soap, 17% by access to improved water, and 36% by improved public sanitation (appropriate disposal of excreta). However, the authors also noted that most of the evidence was of 'poor quality', and that more studies are needed. A review of the available literature reveals that the situation in Oceanic LMICs mirrors that of other low- and middle- income countries globally. There appears to be a catch-22 situation where due to the desperate need to improve WASH in numerous Oceanic LMICs, the focus is on timely and cost-efficient rollout. This, to some degree, may preclude the integration of research such as case-control studies, which

would help determine what interventions work in a given setting, but add time and financial costs.

This questionnaire was exploratory in nature, and there is no intention to over-interpret a reasonably rudimentary questionnaire administered primarily by clinical staff under the direction of a laboratory scientist. Nonetheless, members of the broader study team do have experience in preparing, analyzing and interpreting qualitative data. Given the lack of data currently available, this work makes a valuable contribution to our understanding of the challenges in WASH in regional PNG. The outcomes of the study demonstrate a need for further investigation into access to WASH alongside the rollout of improved WASH.

Chapter 4 - Virulence Genes and Antimicrobial Resistance of *Shigella* from Oceania with a Focus on Papua New Guinean Isolates

4.1 Introduction

Shigella is an important cause of moderate to severe diarrhoea in both children and adults (Kotloff *et al*., 1999; Lozano *et al*., 2012). Analysis of global data from the year 2010 by Pires *et al*. (2015) reported 188 million cases of *Shigella*, with mortality estimated at just over 64,000. Despite limited data on the current burden of *Shigella* in Oceanic LMICs, data obtained in the late 1980s showed *Shigella* was an important cause of diarrheoa in children in PNG (Howard *et al*., 2000). Recent surveillance data and an outbreak investigation in PNG also suggest that this pathogen remains an important cause of diarrhoea (Section 2.3.3). Relative to PNG, less is known about the circulating strains and burden of *Shigella* in the broader Oceania region; however, it is likely to be an important contributor to infectious gastrointestinal illness in Oceanic LMICs.

Shigella has a number of virulence genes that aids its pathogenesis. It is able to cross the epithelial barrier of the colon, evade the host's immune system, and invade cells; causing inflammation and damage to the epithelial lining (Sansonetti, 2001). This pathogenicity is attributed to the presence of virulence genes on both the main bacterial chromosome and the virulence plasmid (Sansonetti, 2001). Virulence genes, such as *ial*, *ipaBCD* and *ipaH* are commonly associated with the ability of the *Shigella* to colonise and invade intestinal cells (Vargas *et al*., 1999). V*irF* and *invE* genes, located on the virulence plasmid, are primary regulators of transcription activation for the remaining plasmid encoded virulence genes. These plasmid encoded virulence genes make up the main components of the *Shigella* T3SS (Mitobe, Morita-Ishihara, Ishihama, & Watanabe, 2009). Genes

set1A and *set1B* are chromosomal genes that encode ShET-1 (*Shigella* enterotoxin 1), while ShET-2 (*Shigella* enterotoxin 2) is encoded by the *sen/ospD3* gene which is present on the virulence plasmid (Faherty *et al*., 2016; Vargas *et al*., 1999). A detailed overview of the function of virulence genes of *Shigella* is provided in Section 2.2.2 and Table 2.1.

Due to the potential severity of illness, PNG and international guidelines state that antibiotics can be considered for administration to patients with shigellosis (PNG National Department of Health, 2012; PNG Paediatric Society, 2016; Williams & Berkley, 2016; Williams & Berkley, 2018). Globally, many *Shigella* isolates exhibit AMR (Sack, Lyke, McLaughlin, & Suwanvanichkij, 2001; Williams & Berkley, 2018). In Asia and Africa resistance to antibiotics that only relatively recently have become widespread in their application (e.g. ciprofloxacin, nalidixic acid and cephalosporins) is now commonly detected (Gu *et al*., 2012). Although limited data exists on the status of AMR in Oceanic LMICs, there is evidence of increasing antimicrobial resistance to first-line treatment options such as ampicillin, chloramphenicol, tetracycline and co-trimoxazole in *Shigella* isolates (Greenhill *et al*., 2014; Rosewell *et al*., 2010; Storch *et al*., 1980; Watson, 2006).

A multitude of variables contribute to patient outcome in any infectious disease; only some of these variables are intrinsically associated with the pathogen. In shigellosis, two *Shigella* associated variables that have a major impact on disease outcome are the genetic determinants of virulence, and the susceptibility of the pathogen to antibiotics. Therefore, this study was conducted to detect virulence genes and phenotypic AMR of isolates from PNG and other LMICs in the Oceania region; including whether virulence determinants had changed over time, and whether isolates are becoming increasingly resistant to antibiotics.

4.2 Materials and Methods

4.2.1 *Shigella* isolates

Archived isolates of *Shigella* spp. were used in this study. Most were obtained from the PNGIMR, with additional isolates from reference laboratories in Australia. An overview of the isolates and their origin is provided in Section 4.3.1. Isolates had previously been confirmed as *Shigella* using traditional phenotypic tests (all isolates) and poly-O antisera (most isolates).

Approval was granted by the PNG Medical Research Advisory Committee for *Shigella* isolates from PNG to be sent to Australia for phenotypic and genotypic characterisation (MRAC no. 16.43). Approval to characterise all isolates was also covered by Federation University Australia Human Ethics (A17-074).

Upon receipt at Federation University Australia, all isolates were stored in cryovials containing beads (Mast Group Ltd) at -80°C. Isolates were cultured on nutrient agar through overnight incubation at 37°C. If there was a need to confirm purity of culture, subculture was conducted using MacConkey agar (incubated overnight at 37°C) prior to subsequent DNA extraction and antibiotic sensitivity testing.

4.2.2 Detection of *Shigella* virulence genes

DNA was extracted from selected isolated colonies using the FavorPrep Tissue Genomic DNA Extraction Mini Kit (Favorgen, Taiwan) as instructed by the manufacturer.

A real-time PCR assay targeting the *ipaH* gene was used to confirm each isolate as *Shigella* (Lin, Cheng, & Van, 2010). Real-time PCR was performed on an Applied

Biosystems (AB) StepOnePlus real-time PCR system (ThermoFisher Scientific, Australia). A total reaction volume of 20 µl for one sample consisted of 2X AB Taqman Universal PCR master mix (ThermoFisher Scientific, Australia) at a final concentration of 1X, 0.8 µl of each forward and reverse primers (10µM), 0.2 µl of probe (10µmol/µl), 6.2 µl of nuclease free water and 2 µl of DNA template. Cycling conditions were: pre-PCR hold at 50℃ for 2 min; initial denaturation at 95℃ for 20 secs; and then 40 cycles at 95℃ for 10 sec and 60℃ for 20 sec.

Conventional PCR was used to confirm the presence/absence of known *Shigella* virulence genes: *ipaBCD*, *ipaH*, *ial*, *virF*, *invE*, *set1A*/*set1B*, and *sen/ospD3*. Refer to Table 2.1 for details of the role of the genes targeted and their location on the *Shigella* genome. The genes were selected based on their importance in *Shigella* virulence, and to enable comparisons with other published studies (Faruque *et al*., 2002; Gomez-Duarte, Bai, & Newell, 2009; Lin *et al*., 2010; Muller *et al*., 2007; Talukder *et al*., 2007; Tornieporth *et al*., 1995).

For conventional PCR (Table 4.1), a total reaction volume of 14µl for one sample comprised 2X GoTaq Green master mix (Promega Corporation) at a final concentration of 1X, 0.4 µl of each forward and reverse primers (10µM), 4.2µl of nuclease free water and 2 µl of DNA template. Amplification was conducted using an AB Veriti 96-well thermal cycler and PCR conditions were: initial denaturation 94°C for 3 min; 25 cycles of denaturation at 94°C for 30 sec, annealing temperature specific for each primer set (between 54 and 63°C; Table 4.1) and elongation at 72°C for 90 sec; followed by a final elongation phase at 72℃ for 10 min. The amplicons were visualised by gel electrophoresis on a 2% agarose gel with SYBR safe gel stain and bands were observed

under UV light using a Dolphin-Doc Plus gel documentation image system (Wealtec Corp, USA).

Table 4.1 List of primers and probe used in real-time PCR and conventional PCR for confirmation of identification and for virulence gene profiling.

	Primer/probe	Sequence 5'-3'	Annealing temp. °C	Amplicon size (bp)	Reference
Real-time PCR: *Shigella* spp. (*ipaH*)	Shig_F	ACCATGCTCGCAGAGAAACT	60	181	Lin *et al*. (2010)
	Shig_R	TACGCTTCAGTACAGCATGC			
	Shig_P	HEX-TGGCGTGTCGGGAGTGACAGC-BHQ1			
Standard PCR: Virulence genes	ipaH_F1	GCTGGAAAAACTCAGTGCCT	56	424	Tornieporth *et al.* (1995)
	ipaH_R1	CCAGTCCGTAAATTCATTCT			
	ial_F	CTGGATGGTATGGTGAGG	58	320	Talukder *et al*. (2007)
	ial_R	GGAGGCCAACAATTATTTCC			
	invE_F	CGATAGATGGCGAGAAATTATATCCCG	57	766	Muller *et al*. (2007)
	invE_R	CGATCAAGAATCCCTAACAGAAGAATCAC			
	ipaBCD_F	GCTATAGCAGTGACATG	55	500	Faruque *et al*. (2002)
	ipaBCD_R	ACGAGTTCGAAGCACTC			
	sen/ospD3_F	ATGTGCCTGCTATTATTTAT	54	799	Talukder *et al*. (2007)
	sen/ospD3_R	CATAATAATAAGCGGTCAGC			
	set1A_F	TCACGCTACCATCAAAGA	57	309	Talukder *et al*. (2007)
	set1A_R	TATCCCCCTTTGGTGGTA			
	set1B_F	GTGAACCTGCTGCCGATATC	57	147	Talukder *et al*. (2007)
	set1B_R	ATTAGTGGATAAAAATGACG			
	virF_F	TCAGGCAATGAAACTTTGAC	58	618	Gomez-Duarte *et al.* (2009)
	virF_R	TGGGCTTGATATTCCGATAAGTC			

4.2.3 Antimicrobial resistance testing

The Kirby-Bauer disc diffusion method was used to test for antimicrobial resistance in all isolates, following the Clinical Laboratory Standards Institute (CLSI) performance standards for antimicrobial susceptibility testing (CLSI, 2015). Antibiotics tested were ampicillin 10µg (AMP), ceftriaxone 30µg (CRO), chloramphenicol 30µg (C), ciprofloxacin 5µg (CIP), nalidixic acid 30µg (NA), tetracycline 30µg (TET) and co-trimoxazole (trimethoprim-sulfamethoxazole) 25µg (SXT). *E.coli* ATCC 25922 was used as a control, as per the CLSI guidelines.

4.2.4 Statistical analysis

Throughout this thesis, *Shigella* spp. isolated up until and including 2009 are referred to as pre-2010 isolates. *Shigella* isolated during or after 2010 are referred to as post-2010 isolates, thus representing isolates circulating in the current decade and mostly obtained in recent surveillance and outbreak-response work conducted in PNG (Benny *et al*., 2014; Greenhill *et al*., 2014).

Differences in antibiotic resistance and virulence profiles were analysed by comparing pre-2010 *Shigella* isolates to post-2010 isolates. Differences in AMR profiles and virulence gene profiles were also sought between *S. flexneri* and *S. sonnei*. To conduct the analyses, a Chi-square test of independence was performed using an online, interactive calculation tool by Preacher (2001) (http://quantpsy.org/calc.htm). Where 0 values were detected, a Fisher's exact test was conducted, also using an online, interactive calculation tool by Preacher (2001) (http://quantpsy.org/fisher/fisher.htm).

4.3 Results

4.3.1 Overview of isolates

Seventy-two *Shigella* isolates from PNG (n=60) and neighbouring nations in Oceania (n=12) were available for analysis. The sample collection comprised 53 *S. flexneri* isolates (from 1985-2014), 16 *S. sonnei* isolates (from 1999-2015) and 3 *S. dysenteriae* isolates (from 1985 and 2010). Given the small number of *S. dysenteriae* isolates, these isolates were not included in this study, thus results are presented for 69 *Shigella* isolates.

Of the PNG isolates, 38 were obtained from within PNG and 22 isolates were from travellers returning to Australia from PNG. Thirty PNG isolates were from a study of adult and paediatric patients presenting to the Lopi Urban Clinic and the Goroka General Hospital with diarrhoea, previously conducted by our research team (Greenhill *et al*. 2014). One PNG isolate was from a shigellosis outbreak in September 2013, from Wau in the Morobe Province (Benny *et al*. 2014). Seven PNG isolates were archived samples from a previous case-control study of children admitted to the paediatric ward at the Goroka Hospital with diarrhoea during 1985 – 1990 (Howard *et al*. 2000).

Isolates from travellers returning to Australia from PNG (n=22) and neighbouring Pacific Island nations (n=12), who presented at their local health centre/ hospital with diarrhoea, were obtained from the Microbiology Diagnostic Unit (MDU) located in the Doherty Institute in Melbourne (n=10) and from Queensland Health (n=24). The 12 non-PNG isolates were from travellers returning to Australia from Pacific Island nations (Fiji, Vanuatu, Samoa, Solomon Islands), including three isolates from people who lived in or had spent time in the Torres Strait Islands (situated in far-northern Queensland, Australia).

4.3.2 Presence/absence of *Shigella* spp. virulence genes

The 69 isolates were confirmed as *Shigella* by real-time PCR detection of the *ipaH* gene prior to further analyses. Conventional PCR targeted the *ipaH* gene, with results corroborating the real-time PCR results; and an additional eight genes.

Isolates of each species were grouped according to their virulence gene profiles (presence/absence of genes); producing four groups of *S. flexneri* and two groups of *S. sonnei* (Tables 4.2 and 4.3). The *ial*, *invE*, *ipaBCD*, *sen/ospD3* and *virF* genes were detected in 61% of all isolates. The *set1A*/*set1B* genes were present in approximately 43% of *S. flexneri* isolates, but were absent from *S. sonnei*. The *sen/ospD3* gene was present in 79% of *S. flexneri* isolates and only one *S. sonnei* isolate. A higher proportion of *Shigella* isolated during or since 2010 harboured targeted virulence genes relative to pre-2010 *Shigella* isolates (irrespective of species), as shown in Figure 4.1. There were statistically significant differences in the proportion of virulence genes present pre- and post-2010 for all non-*ipaH* virulence genes when tested individually; *ial*, *invE*, *ipaBCD*, *sen/ospD3*, *virF,* Chi-square value = 4.56; DF = 1; p = 0.03; and *set1A/set1B*, Chi-square value = 11.59; DF =1; p = 0.0007.

Table 4.2 Virulence profiles based on presence/absence of virulence genes in *S. flexneri*.

Species	*ipaH*	*sen/ospD3*	*set1A/set1B*	*invE*	*ial*	*ipaBCD*	*virF*	Year of Isolation	No. Isolates	Origin
S. flexneri	+	-	-	-	-	-	-	1985	3	PNG
								1985	1	PNG
								1992	1	PNG
								2000	1	PNG
								2004	1	PNG
								2010	3	PNG
S. flexneri	+	-	+	-	-	-	-	2009	1	PNG
S. flexneri	+	+	-	+	+	+	+	1985	1	PNG
								1994	1	PNG
								2002	1	PNG
								2006	1	PNG
								2006	1	PNG
								2007	1	PNG
								2007	1	Fiji
								2010	4	PNG
								2011	1	PNG
								2012	1	Samoa
S. flexneri	+	+	+	+	+	+	+	2009	1	Torres Strait
								2010	1	Torres Strait
								2010	20	PNG
								2011	2	PNG
								2013	1	Torres Strait
								2013	3	PNG
								2014	1	PNG

Table 4.3 Virulence profiles based on presence/absence of virulence genes in *S. sonnei*.

Species	*ipaH*	*sen/ospD3*	*set1A/set1B*	*invE*	*ial*	*ipaBCD*	*virF*	Year of Isolation	No. Isolates	Origin
S. sonnei	+	-	-	-	-	-	-	1999	1	PNG
								2007	1	Solomon Islands
								2008	1	PNG
								2011	2	PNG
								2012	3	Fiji
								2013	1	Samoa
								2013	1	Vanuatu
								2014	3	PNG
								2015	1	PNG
								2015	1	Vanuatu
S. sonnei	+	+	-	+	+	+	+	2010	1	PNG

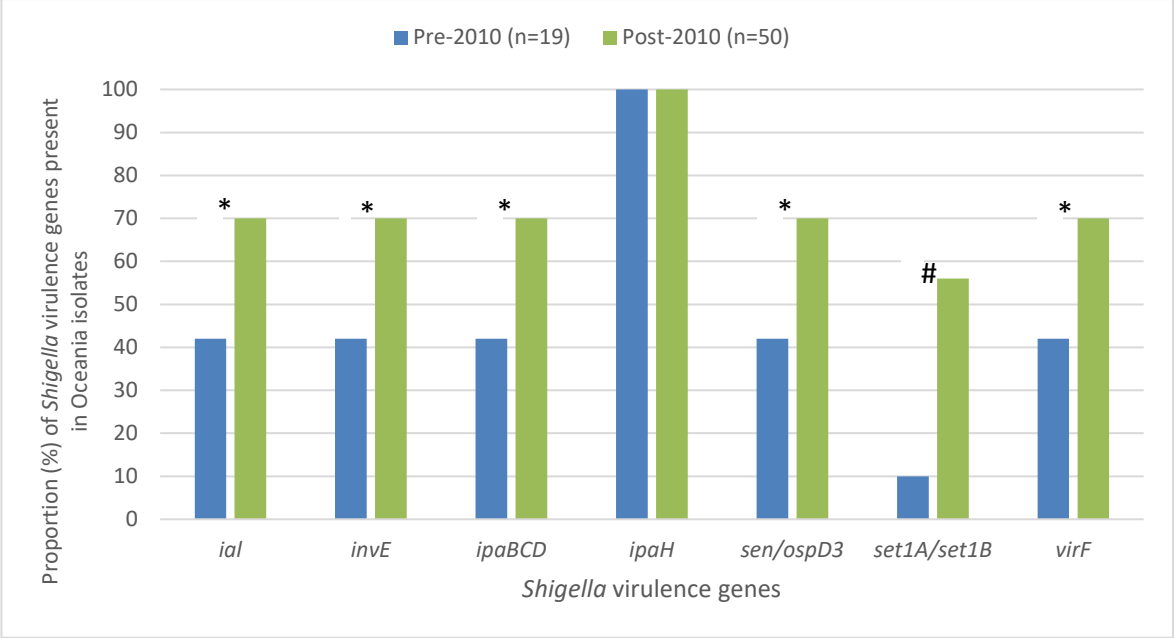


Figure 4.1 The prevalence of virulence genes detected in *Shigella* isolates circulating pre-2010 (i.e. 1994 - 2009) and post – 2010 (i.e. 2010 – 2015) in Oceania. * represents p=0.03; # represents p<0.01.

4.3.3 Antimicrobial resistance

4.3.3.1 Proportion of isolates resistant to antibiotics

S. flexneri (n=53) was commonly resistant to the antimicrobials ampicillin (77%, n=41), tetracycline (74%, n=39), chloramphenicol (60%, n=32) and co-trimoxazole (49%, n=26). Amongst the *S. sonnei* (n=16) isolates, resistance to co-trimoxazole (75%, n=12) and ampicillin (56%, n=9) was common, followed by tetracycline (19%, n=3) and nalidixic acid (6%, n=1) (Table 4.4). The proportion of *S. flexneri* and *S. sonnei* isolates resistant to individual antimicrobials differed, with resistance more common in *S. flexneri* for chloramphenicol (Fisher's exact test value, p <0.001) and tetracycline (Chi-square value = 15.5.2; DF = 1; p <0.01); but resistance to co-trimoxazole was more common in *S. sonnei* (Chi-square value = 3.34; DF = 1; p = 0.07).

In comparing the prevalence of isolates resistant to specific antibioitics pre- and post-2010 there was a significant difference for ampicillin (Chi-square value = 8.28, DF=1, p = 0.004) and co-trimoxazole (Chi-square value = 8.76, DF = 1, p = 0.003); with resistance more common in post-2010 isolates (Figure 4.2).

Table 4.4 The number of *S. flexneri* and *S. sonnei* isolates from Oceania resistant to selected antibiotics (proportion as percentage in brackets).

	S. flexneri n=53	*S. sonnei* n=16	Total n=69
AMP	41 (77)	9 (56)	50 (72)
C	32 (60)	0	32 (46)
CIP	0	0	0
CRO	0	0	0
NA	0	1 (6)	1 (1)
TET	39 (74)	3 (19)	42 (61)
SXT	26 (49)	12 (75)	38 (55)

AMP – Ampicillin, C – Chloramphenicol, CIP – Ciprofloxacin, CRO – Ceftriaxone, NA – Nalidixic acid, TET – Tetracycline, SXT – Trimethoprim-sulfamethoxazole or Co-trimoxazole

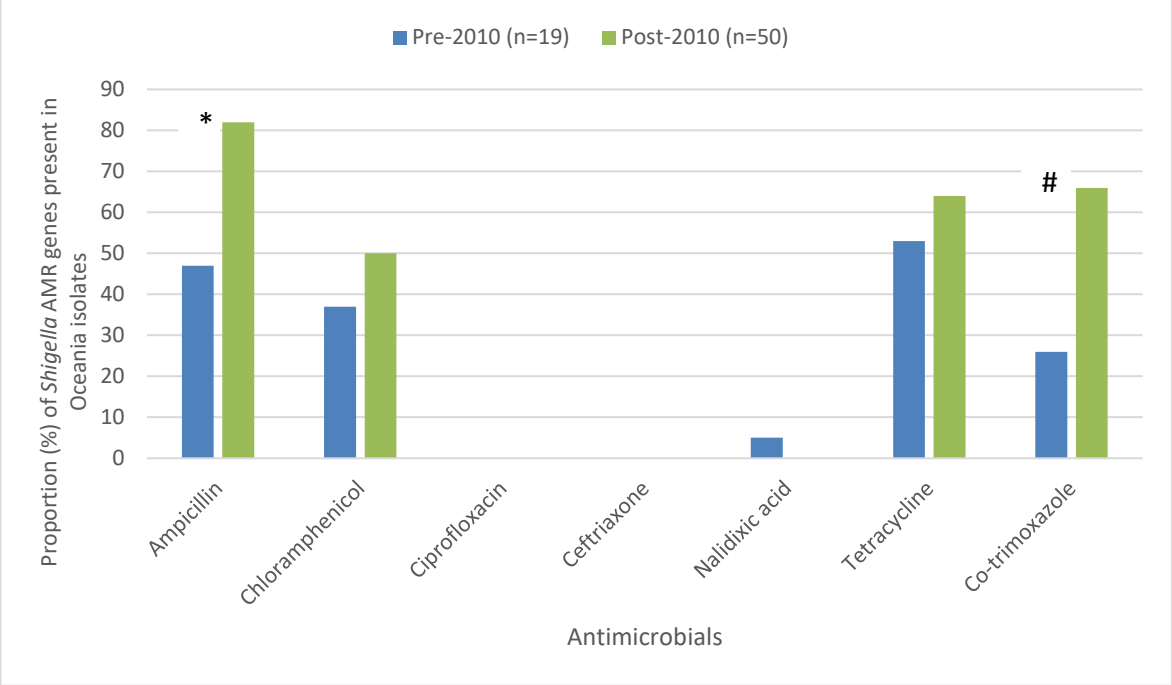


Figure 4.2 *Shigella* isolates resistance to antimicrobials pre- 2010 (19/69 isolates) and post-2010 (50/69 isolates). * represents p=0.004; # represents p=0.003.

4.3.3.2 Multi-drug resistance

In this study all *Shigella* isolates resistant to at least two antimicrobials were considered MDR. Of the 69 isolates analysed, 50 were MDR. The most common MDR profile was resistance to AMP/C/TET/SXT in 25% (17/69) of isolates, all of which were *S. flexneri*. The second most common MDR profile was AMP/C/TET in 17% (12/69) of *S. flexneri* isolates. AMP/SXT and AMP/TET/SXT was observed in 12% (8/69) and 9% (6/69) of isolates, respectively; both resistance patterns were observed in *S. flexneri* and *S. sonnei*. Other combinations were observed less frequently (4% of isolates or less) as shown in Table 4.5 and Figure 4.3.

There was a trend towards increasing prevalence of AMR. *Shigella* isolates post-2010 were more commonly resistant to one or more antibiotics than were isolates from pre-

2010 (Figure 4.4). Approximately 50% of isolates displayed MDR pre-2010 and approximately 80% of isolates displayed MDR post-2010.

Table 4.5 Multi-drug resistance profiles of *S. flexneri* and *S. sonnei* isolates from Oceania.

AMR	***S. flexneri***	***S. sonnei***	**Total (%)**
AMP/C/TET/SXT	17 (32)	0	17 (25)
AMP/C/TET	12 (23)	0	12 (17)
AMP/SXT	1 (2)	7 (44)	8 (12)
AMP/TET/SXT	4 (8)	2 (13)	6 (9)
AMP/TET	3 (6)	0	3 (4)
C/TET/SXT	2 (4)	0	2 (3)
AMP/C/SXT	1 (2)	0	1 (1)
NA/TET/SXT	0	1 (6)	1 (1)
AMP	3 (6)	0	3 (4)
TET	1 (2)	0	1 (1)
SXT	1 (2)	2 (13)	3 (4)
Susceptible	8 (15)	4 (25)	12 (17)
Total	53	16	69

*Refer to Table 4.4 for abbreviations of antibiotics

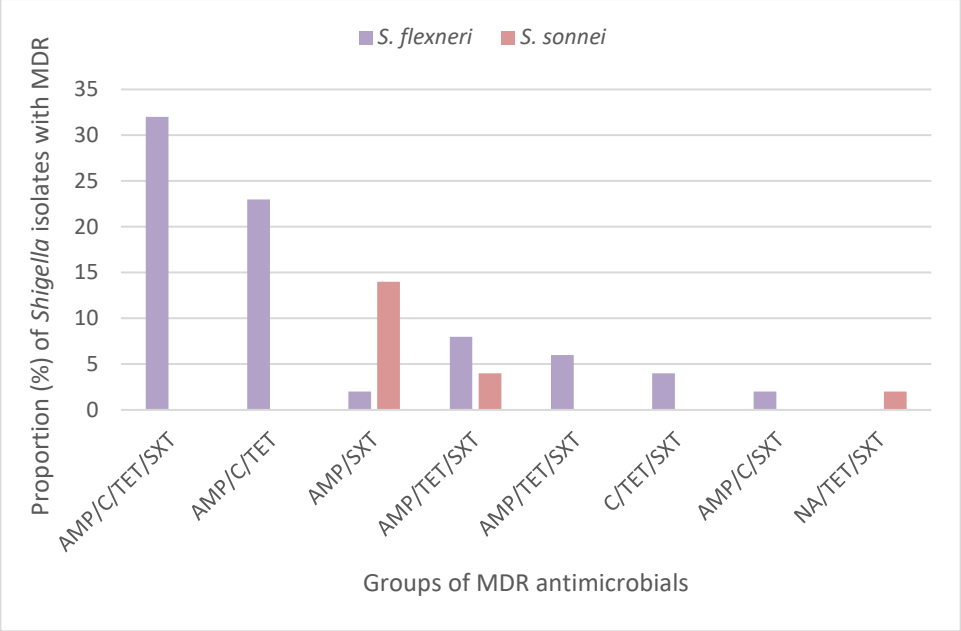


Figure 4.3 Resistance to two or more antibiotics was observed in 50 *Shigella* isolates. The proportion of those 50 isolates with various resistance profiles is illustrated.

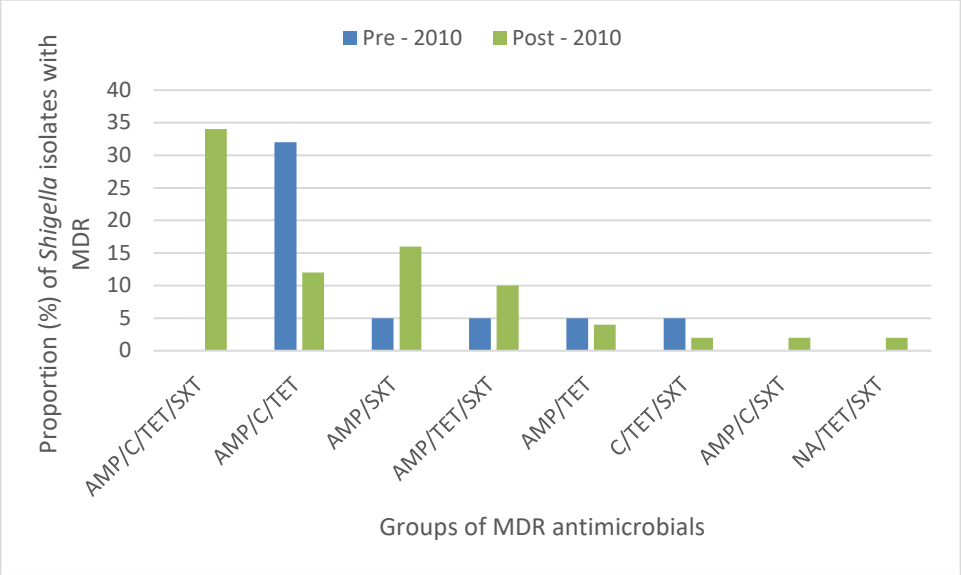


Figure 4.4 Proportion of *Shigella* isolates harbouring MDR to combinations of antibiotics pre-2010 (19/69 isolates) and post-2010 (50/69 isolates).

4.4 Discussion

This study provides basic characterisation of *Shigella* strains from Oceania, by detecting the presence/absence of *Shigella* virulence genes and phenotypic antimicrobial resistance patterns. The study provides preliminary evidence to suggest that AMR in *Shigella* isolates from PNG and other Oceanic LMICs is increasing. While there is no evidence of direct correlation between presence of virulence genes and pathogenesis, this study has detected more virulence genes in recent isolates than historical isolates. The significance of more virulence genes present in recent isolates is difficult to ascertain, particularly given the potential for plasmid associated virulence genes to be lost in storage.

Isolates of each species were grouped according to their virulence gene profiles (presence/absence of genes), producing four groups of *S. flexneri* and two groups of *S. sonnei* (Tables 4.2 and 4.3). Results suggest post-2010 (i.e. 2010-2015) *Shigella* isolates harbour more virulence genes compared to pre-2010 isolates (Figure 4.1) with *S. flexneri* isolates presenting with more virulence genes compared to *S. sonnei* isolates.

Prior to this study there were no published studies investigating virulence genes in *Shigella* from the LMICs in the Oceania region and relatively few studies have considered isolates from Asia. *Shigella* enterotoxin genes are primarily responsible for the clinical manifestation of acute-watery diarrhoea and dysentery in shigellosis (Yaghoubi *et al*., 2017). In this study, genes encoding ShET1 (*set1A* and *set1B*; particularly *set1A*) were present in *S. flexneri* isolates collected from 2009 onwards (57% of *S. flexneri* isolates) but not detected in *S. sonnei,* as was also observed in previous studies by Vargas *et al*. (1999) and more recently by Yaghoubi *et al.* (2017). All isolates positive for *set1A/set1B*

were from PNG or from the Torres Strait Islands; the latter being a region of northern Australia with close geographical and cultural links to PNG.

The ShET2 genes (*sen/ospD3*) were present in 79% of *S. flexneri* isolates, a detection rate higher than two comparable studies conducted in South America (Casabonne, Gonzalez, Aquili, & Balague, 2016; da Cruz et al., 2014) but lower than studies involving *S. flexneri* isolates from north-eastern Brazil (Medeiros *et al*., 2018). One PNG isolate of *S. sonnei* harboured *sen/ospD3*, which has been reported previously in other isolates (Casabonne *et al*., 2016; da Cruz *et al*., 2014); and indeed appears common in parts of Brazil (Medeiros *et al*., 2018) and China (Fan *et al*., 2017; Qu *et al*., 2014). Isolates positive for *sen/ospD3* were from PNG, the Torres Strait Islands, Fiji and Samoa.

S. flexneri isolates with different virulence gene profiles were isolated as recently as 2012-14 (Table 4.2), suggesting co-circulation of strains. A study of global isolates of *S. flexneri* using whole genome sequencing revealed old and new strains of *S. flexneri* to persist alongside one another and continue causing infection (Connor *et al*., 2015). Based on virulence gene profiling conducted in this study, it appears that different strains of *S. flexneri* coexist in PNG.

Shigella has a large unstable plasmid that encodes several virulence genes, including *ipaH*, *ipaBCD*, *virF*, *invE* and *ial*; the *ipaH* gene is also chromosomally encoded, so is considered a stable target gene. The *ipaH* gene was present in all isolates, while 62% of isolates contained all four *ipaBCD*, *virF*, *invE* and *ial* genes (Tables 4.2 and 4.3). The virulence genes *(ipaBCD*, *virF*, *invE* and *ial)* all contribute to facilitating and maintaining the spread and pathogenesis of *Shigella* during infection (Broach *et al*., 2012; Lluque *et*

al., 2015; Schroeder & Hilbi, 2008). The *virF* gene helps to facilitate the intracellular spread of *Shigella* during infection, while *invE* promotes the expression of multiple genes associated with virulence. Both genes are found on the bacteria's invasion plasmid (Broach *et al*., 2012) and the results of this study demonstrate that both genes are simultaneously present or absent in a particular isolate, as typically observed in *Shigella* isolates (Section 2.2, Table 2.1). This association is also observed with the other plasmid invasion genes, *ial* and *ipaBCD*, (Lin *et al*., 2010; Qu *et al*., 2014; Vargas *et al*., 1999). Our results suggest that there is a strong association between all of these genes, as they were invariably present/absent as a set (Table 4.2 and 4.3). In regards to the stability of these invasion plasmid virulence genes, some studies have shown that the *ial* gene in particular is prone to deletion (Zhang *et al*., 2013).

Based on this study's statistical analysis, the proportion of virulence genes present has increased from 2010 onwards, compared to previous years. Some virulence genes are located on the virulence plasmid and because of this can be lost by the isolate during prolonged storage (Yavzori, Cohen, & Orr, 2002). Interestingly, regardless of which country the isolates originated from, similar virulence profiles were observed in PNG, Solomon Islands, the Torres Strait, Fiji, Vanuatu and Samoa, regardless of species.

An increase in the prevalence of antimicrobial drug resistance to older antimicrobials, namely ampicillin, tetracycline, chloramphenicol and co-trimoxazole (Figure 4.3), was observed in post-2010 isolates compared to pre-2010 isolates. This coincides with similar findings observed in Fiji, where there is evidence of increasing resistance to ampicillin and chloramphenicol over time, especially in *S. flexneri* isolates (Watson, 2006). Also, the

majority of Oceania *S. sonnei* isolates in this study were resistant to co-trimoxazole and ampicillin.

The trend of increasing AMR in *Shigella* observed in this study is reflected in other low-income settings (Anandan *et al*., 2017; Bhattacharya *et al*., 2015; Bhattacharya *et al*., 2014; Dhital *et al*., 2017; Niyogi, 2007; Poramathikul *et al*., 2016). In Mozambique, *Shigella* isolates are commonly resistant to the same antibiotics as was observed in this study (Mandomando *et al*., 2009), namely ampicillin, tetracycline, chloramphenicol and co-trimoxazole. In the south of India, there was resistance to ampicillin, co-trimoxazole and nalidixic acid in *S. flexneri* isolates and resistance to co-trimoxazole and nalidixic acid in *S. sonnei* isolates for the years 2014 – 2015 which had also been observed in previous years (Anandan *et al*., 2017; Mamatha, Pusapati, & Rituparna, 2007). Clearly, antibiotic resistant *Shigella* is a global problem.

With increasing resistance comes MDR; this study shows *S. sonnei* being collectively resistant to antimicrobials, ampicillin, co-trimoxazole and tetracycline; compared to *S. flexneri* which was resistant to ampicillin, chloramphenicol, tetracycline and co-trimoxazole collectively. The MDR problem is greater in *S. flexneri* than it is in *S. sonnei* in Oceania isolates. Other studies have also shown that *S. sonnei* has low rates of MDR compared to *S. flexneri* (Shiferaw *et al*., 2012). In this study MDR, defined as resistance to at least two classes of antimicrobials, was detected in 72% (n=50) of isolates. Using a more stringent definition (resistance to at least three classes of antimicrobials), MDR was detected in 57% (n=39) of isolates (Table 4.5); a rate also similar to that observed in Mozambique (Mandomando *et al*., 2009).

This study provides evidence of increasing AMR in *Shigella* in PNG, but it is difficult to draw wider conclusions due to the low number of *Shigella* isolates prior to 2010 from countries in Oceania other than PNG. However, other published data from previous studies in Oceanic LMICs, noted in Section 2.6, suggests the presence of AMR among isolates in the region as early as the 1970s. The over prescription of antimicrobials to treat diseases without proper diagnosis may contribute to this increase in AMR in the region (Poka & Duke, 2013; Storch *et al*., 1980; Vince, 1995; Watson, 2006).

Fortunately, resistance to ciprofloxacin or ceftriaxone was not observed in any *Shigella* isolates in this study; however, one *S. sonnei* isolate was resistant to nalidixic acid. Ciprofloxacin resistance in *S. flexneri* strains and nalidixic acid resistance in *S. sonnei* is increasingly reported in Asia and Africa (Gu *et al*., 2012). Given the close proximity of Asia to Oceania, the possibility of AMR strains being introduced into the Oceania region is a cause for concern and warrants monitoring.

This study provides initial data on the distribution of *Shigella* virulence genes in PNG and neighbouring Oceanic LMICs and provides insight into the AMR trends. It is notable that many of the recently isolated *S. flexneri* harbour more virulence genes than most pre-2010 isolates. Similarly, there appears to be increasing resistance to commonly used antibiotics. Further genotyping is required to determine whether these represent new strains; or they are the same strains which have acquired genetic elements associated with virulence and antibiotic resistance (Malau *et al*., 2018).

The findings of this research suggest an emergence of new strains; however, it is difficult to differentiate strains based on the presence/absence of virulence genes and AMR

profiles. Phenotypic AMR testing remains the mainstay of bacterial AMR surveillance globally. In Oceanic LMICs the current limitation is the lack of hospital laboratories that routinely conduct microbial culture and sensitivity testing. It is imperative that we continue to monitor the antimicrobial resistance trends of *Shigella* in the region, to manage clinical cases and prevent morbidity and mortality amongst vulnerable populations in the region, and provide epidemiological data. In addition, complementary techniques for both AMR surveillance and strain differentiation may further improve our understanding of *Shigella* in high-burden settings.

Chapter 5 - Detection of Virulence and Antimicrobial Resistance Genes in *Shigella* Isolates in PNG and Oceania Using Whole Genome Sequencing

5.1 Introduction

Knowledge of *Shigella* in Oceanic LMICs is lacking, due to limited diagnostic and medical research capacity and facilities in the region and in low-income settings globally (Tickell *et al*., 2017).

Detection of AMR is commonly conducted via conventional (phenotypic) means of culture followed by sensitivity testing using the Kirby-Bauer disc diffusion method. Guidelines such as CLSI or EUCAST are commonly followed, although other systems also exist. Ideally, sensitivity testing should be routinely conducted as part of culture based diagnosis of bacterial infections, as the results can influence treatment. In reality AMR testing is only sporadically conducted in a small number of diagnostic laboratories in PNG; with culture for gastrointestinal pathogens particularly limited (Greenhill *et al*., 2012). This situation is likely to be mirrored in other Pacific island countries and territories, where resources are limited. However, with rising AMR in many bacterial pathogens globally, expanded monitoring is imperative.

In the previous chapter, PCR was used to detect the presence/absence of virulence genes. This approach required multiple PCR reactions to be conducted, making it time consuming. It is also prone to amplification failure and is necessarily targeted at selected gene targets. Here characterisation of *Shigella* from PNG and other Oceanic LMICs is

conducted by Whole Genome Sequencing (WGS); enabling detection of virulence genes and genes/mutations associated with AMR.

5.2 Materials and Methods

5.2.1 *Shigella* isolates for sequence analysis

The 69 isolates of *Shigella* characterised in the previous chapter and three *S. dysenteriae* were complemented with an additional six *Shigella* isolates. The additional six isolates were obtained in 2017 from clinical (n=3) and asymptomatic (n=3) persons in PNG; a total of 78 isolates were prepared and sent for sequencing.

Genomic material (25µl) from isolates was prepared using the same extraction methods as described in section 4.2.2. Extractions with DNA concentrations of 25-60 ng/µl were submitted to the Wellcome Sanger Institute in England for sequencing. Sequencing was conducted on an Illumina HiSeq genome analyser.

5.2.2 Detection of virulence genes using WGS

Genomic data acquired from sequencing were screened for virulence genes, using the Virulence Factors Database (commonly referred to as VFDB) (Chen *et al*., 2005) and srst2 v0.2.0 (standing for short read sequence typing) for bacterial pathogens (Inouye *et al*., 2014), all using default parameters. The main virulence genes *ipaH*, *ipaBCD*, *set1A/set1B*, *sen/ospD3* and *virF* were selected to enable comparison of WGS data to conventional PCR data (Chapter 4). The function of each target gene is addressed in Section 2.2.2 and Section 4.1.

5.2.3 Detection of AMR genes using WGS

Genomic data acquired from sequencing results were screened for AMR genes using the ARG-ANNOT (Antibiotic Resistance Gene-ANNOTation) database (Gupta *et al*., 2014). Screening for AMR genes for aminoglycosides, beta-lactams, chloramphenicol, ciprofloxacin, tetracycline and trimethoprim was conducted because of their predominant usage in PNG and other Oceanic LMICs. Fisher's exact and Chi-square test was used for statistical data analysis, as outlined in Section 4.2.4.

5.3 Results

5.3.1 Sequence outputs

From a total of 63 successfully sequenced *Shigella* isolates (47 for *S. flexneri* and 16 for *S. sonnei*), 57 yielded phenotypic or specific gene target results (Chapter 4) and WGS data. This enabled comparison of the presence/absence of virulence genes as determined by standard PCR and WGS methods; and comparison of phenotypic AMR results and WGS data.

5.3.2 Detection of *Shigella* virulence genes via conventional PCR and WGS

Overall, concordance by PCR and WGS was 100% for *ipaH*, 98% for set 1B, 84% for *set1A* and 75 – 77% for the remaining virulence genes (Figure 5.1).

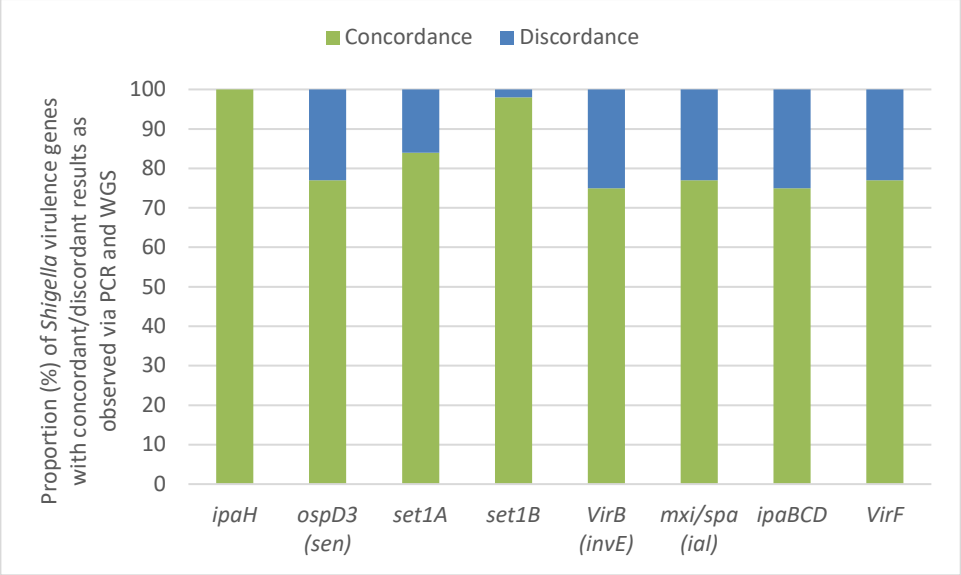


Figure 5.1 Proportion of *Shigella* isolates (n=57) for which the detection of virulence genes by conventional PCR and WGS was concordant (green) and discordant (blue).

Table 5.1 shows the results for all isolates included in analyses, with non-concordant results in blue. Visual inspection of Table 5.1 reveals that discordance occurred in specific isolates; i.e. for a single isolate, typically multiple genes were not detected by PCR but were detected by WGS. Of the 16 isolates where discordance between detection methods was observed, only one isolate was discordant for a single virulence gene, one isolate discordant for two virulence genes, and one isolate discordant for three virulence genes. For all 13 other isolates where discordance was observed, five or more virulence genes were absent by PCR but present by WGS.

All 16 *Shigella* isolates depicted in Table 5.1 with discordant PCR and WGS results had the primers for virulence gene detection re-aligned with their respective reference genomes using Geneious; i.e. AF386526 for *S. flexneri* and HE616529 and HE616528 for *S.sonnei*. The 'test with save primers' function in Geneious was used to double check their discordant results.

From the re-alignments on Geneious, it was observed that the reverse primer for the *invE* virulence gene commonly had a mismatch with the last base on the 3' end. The 3' end is the most important place for the mismatches, therefore this may have contributed to the false negative results for this virulent gene primer set.

For *S. sonnei* there were numerous mis-matches for the re-alignments to the reference genomes. Further visual representation of these primer alignments for the virulence genes that aligned successfully to their respective primers are portrayed in the appendices section. Also some of the *S. sonnei* isolates had a mismatch of one base in the middle of the *ipaBCD* forward primer. This is not likely to affect the performance of the primer as it was in the middle of the primer.

The *set1A* forward and reverse primers and one *set1B* primer were re-aligned with the reference genome AE005674 for *S. flexneri*. For *S. sonnei*, the *set1A* primers were aligned to the reference genome HE616528 showing some positive alignments, but only for one primer set; mainly the forward primer.

Table 5.1 Concordance of detection of virulence genes in *Shigella* isolates (n=57) when detected by conventional PCR and WGS. WGS data listed first, then detection of virulence gene via PCR (i.e. +/- indicates virulence gene detected by WGS, but not observed by PCR). Discordant results are highlighted in blue.

Isolate	*ipaH*	*ospD3 (sen)*	*set1A*	*set1B*	*VirB (invE)*	*mxi/spa (ial)*	*ipaBCD*	*VirF*	Year
S. flexneri #32_PNG	+/+	+/+	-/-	-/-	+/+	+/+	+/+	+/+	1990
S. flexneri #20_PNG	+/+	+/+	-/-	-/-	+/+	+/+	+/+	+/+	2006
S. flexneri #48_PNG	+/+	+/+	+/-	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #26_PNG	+/+	+/+	-/-	-/-	+/+	+/+	+/+	+/+	2007
S. flexneri #47_PNG	+/+	+/+	-/-	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #50_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #62_PNG	+/+	+/-	-/-	-/-	+/-	+/-	+/-	+/-	2010
S. flexneri #49_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #34_PNG	+/+	+/-	+/-	-/-	+/-	+/-	+/-	+/-	1990
S. flexneri #23_PNG	+/+	-/-	-/-	-/-	-/-	-/-	-/-	-/-	2004
S. flexneri #33_PNG	+/+	+/-	-/-	-/-	+/-	+/-	+/-	+/-	1990
S. flexneri #35_PNG	+/+	-/-	-/-	-/-	-/-	-/-	-/-	-/-	1990
S. flexneri #21_FJ	+/+	+/+	-/-	-/-	+/+	+/+	+/+	+/+	2007
S. flexneri #24_PNG	+/+	+/+	-/-	-/-	+/+	+/+	+/+	+/+	1994
S. flexneri #1_TS	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2009
S. flexneri #19_PNG	+/+	+/+	-/-	-/-	+/+	+/+	+/+	+/+	2006
S. flexneri #4_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2011
S. flexneri #25_PNG	+/+	-/-	+/+	-/-	-/-	-/-	-/-	-/-	2009
S. flexneri #51_PNG	+/+	+/+	+/-	-/-	+/-	+/+	+/-	+/+	2010
S. flexneri #53_PNG	+/+	+/-	+/-	-/-	+/-	+/-	+/-	+/-	2010
S. flexneri #52_PNG	+/+	+/-	+/-	-/-	+/-	+/-	+/-	+/-	2010
S. flexneri #2_TS	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #10_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2013
S. flexneri #9_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2013
S. flexneri #55_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #60_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #58_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #57_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #46_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #61_PNG	+/+	+/+	+/-	+/-	+/+	+/+	+/+	+/+	2010
S. flexneri #43_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #42_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #41_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #56_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #45_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #44_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #40_PNG	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	2010
S. flexneri #59_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2010
S. flexneri #14_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2014
S. flexneri #39_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2013
S. flexneri #5_PNG	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	2011
S. sonnei #22_SI	+/+	-/-	-/-	-/-	-/-	-/-	-/-	-/-	2007
S. sonnei #18_VUT	+/+	+/-	+/-	-/-	+/-	+/-	+/-	+/-	2015
S. sonnei #11_VUT	+/+	-/-	-/-	-/-	-/-	-/-	-/-	-/-	2013
S. sonnei #12_WS	+/+	-/-	-/-	-/-	-/-	-/-	-/-	-/-	2013
S. sonnei #6_FJ	+/+	+/-	-/-	-/-	+/-	+/-	+/-	+/-	2012
S. sonnei #7_FJ	+/+	+/-	-/-	-/-	+/-	+/-	+/-	+/-	2012
S. sonnei #8_FJ	+/+	+/-	+/-	-/-	+/-	+/-	+/-	+/-	2012
S. sonnei #30_PNG	+/+	-/-	-/-	-/-	-/-	-/-	-/-	-/-	2008
S. sonnei #28_PNG	+/+	+/-	-/-	-/-	+/-	+/-	+/-	+/-	1999
S. sonnei #16_PNG	+/+	+/-	-/-	-/-	+/-	+/-	+/-	+/-	2014
S. sonnei #3_PNG	+/+	+/-	+/-	-/-	+/-	+/-	+/-	+/-	2011
S. sonnei #29_PNG	+/+	+/+	-/-	-/-	+/+	+/+	+/+	+/+	2010
S. sonnei #15_PNG	+/+	-/-	-/-	-/-	-/-	-/-	-/-	-/-	2014
S. sonnei #17_PNG	+/+	-/-	-/-	-/-	-/-	-/-	-/-	-/-	2015
S. sonnei #31_PNG	+/+	-/-	-/-	-/-	-/-	-/-	-/-	-/-	2011
S. sonnei #13_PNG	+/+	+/-	-/-	-/-	+/-	+/-	+/-	+/-	2014

5.3.3 Detection of *Shigella* AMR genes via WGS in correlation with phenotypic results

Concordance for AMR resistance detected by disc diffusion (CLSI, 2015) and WGS was examined. By disc diffusion (phenotypic testing) 75% of isolates were resistant to the beta-lactam antimicrobial ampicillin, but no resistance to ceftriaxone was detected; resistance to tetracycline (60%), trimethoprim (58%) and chloramphenicol (46%) was common (Figure 5.2).

Genetic traits associated with resistance were commonly detected. A higher proportion of isolates harboured genes associated with resistance than the proportion demonstrating phenotypic resistance for all four classes of antibiotics (Table 5.2). Of the 57 isolates in this study, 48 (84%) had genes predicted to confer resistance to ampicillin, 47 (82%) to tetracycline, 45 (79%) to trimethoprim and 40 (70%) to chloramphenicol. Genotypic results did not correlate well with phenotypic susceptibility, with a total of 48 isolates having discordant results for at least one antimicrobial (Table 5.3). These isolates were genotypically predicted to be resistant to the antimicrobials, but were phenotypically susceptible.

Table 5.4 shows the results for all isolates included in the analyses, with discordant results in blue. Upon visual inspection, there was discordance in the detection of antibiotic resistance in 35 of 57 isolates. A total of 22 isolates had discordant results for a single antibiotic: 10 for trimethoprim, five for beta lactams, four for chloramphenicol and three for tetracycline. Seven isolates were discordant for two antibiotics: five for chloramphenicol and tetracycline, one for chloramphenicol and trimethoprim, and one for

beta lactams and trimethoprim. Four isolates were discordant for three antibiotics and two isolates for four antibiotics.

Table 5.2 The proportion of *S. flexneri* and *S. sonnei* harbouring resistance genes as detected via WGS.

Resistance gene	Antibiotic class	Number (%)
S. flexneri (n=41)		
AMPH_Ecoli	β-lactam	41 (100)
AmpC1_Ecoli	β-lactam	10 (24)
OXA_1	β-lactam	38 (93)
TEM−1D	β-lactam	8 (20)
CatA1	phenicol	36 (88)
TetA	tetracycline	1 (2)
TetB	tetracycline	38 (93)
TetR	tetracycline	1 (2)
DfrA1	trimethoprim	3 (7)
DfrA3	trimethoprim	1 (2)
DfrA5	trimethoprim	27 (66)
DfrA8	trimethoprim	3 (7)
S. sonnei (n=16)		
OXA_1	β-lactam	8 (50)
TEM−1D	β-lactam	12 (75)
CatA1	phenicol	4 (25)
TetA	tetracycline	4 (25)
TetB	tetracycline	7 (44)
DfrA1	trimethoprim	5 (31)
DfrA3	trimethoprim	1 (6)
DfrA5	trimethoprim	10 (63)
DfrA8	trimethoprim	4 (25)

Table 5.3 Phenotypic observations in comparison to genotypic presence/absence of AMR genes in *Shigella* isolates (n=57).

	Phenotype: susceptible		**Phenotype: resistant**	
Antibi-otic	Genotype: resistant (%)	Genotype: susceptible (%)	Genotype: resistant (%)	Genotype: susceptible (%)
AMP	3 (5)	10 (18)	44 (77)	0
C	15 (26)	17 (30)	25 (44)	0
TET	14 (25)	9 (16)	33 (58)	1 (2)
SXT	16 (28)	12 (21)	29 (51)	0

AMP, ampicillin; C, chloramphenicol; TET, tetracycline; SXT, trimethoprim/co-trimoxazole.

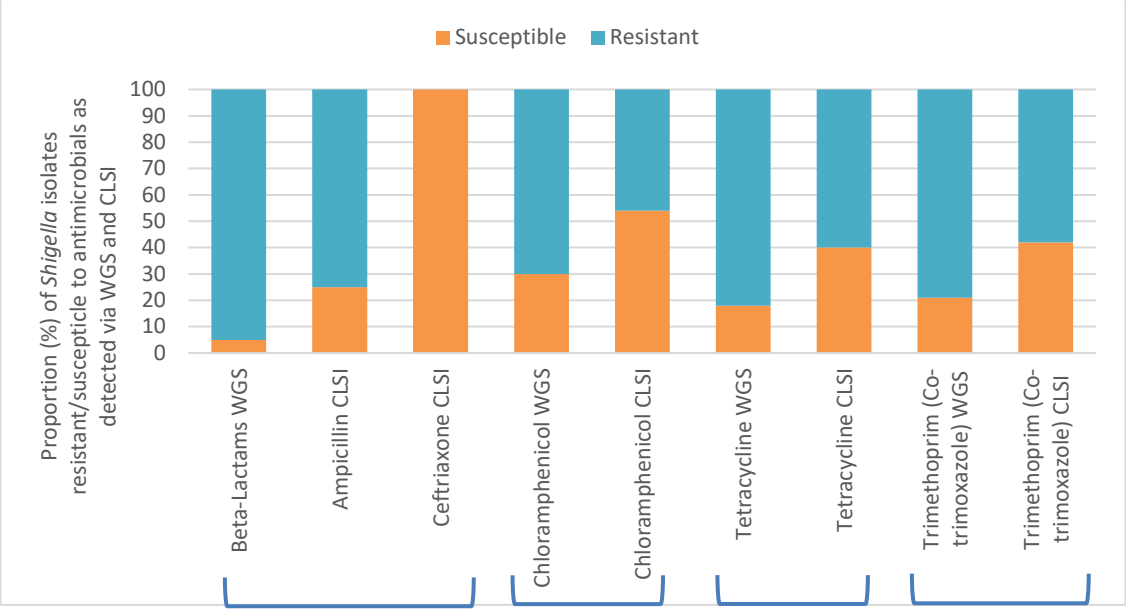


Figure 5.2 The proportion of *Shigella* isolates (n=57) harbouring resistance genes and showing susceptibility to antimicrobials as detected via WGS, in comparison to phynotypic (CLSI) antimicrobial sensitivity testing.

Table 5.4 Concordance of detection of AMR genes and phenotypic resistance (CLSI disc diffusion) in *Shigella* isolates (n=57). WGS data listed first, then CLIS data (i.e. +/- indicates resistance genes detected, but resistance not observed by CLIS). Discordant results are highlighted in blue.

Isolate	Beta-Lactams	Phenicol	Tetracycline	Trimethoprim	Year
S. flexneri #32_PNG	+/-	-/-	-/-	-/-	1990
S. flexneri #20_PNG	+/+	+/+	+/+	-/-	2006
S. flexneri #48_PNG	+/+	+/+	+/+	+/+	2010
S. flexneri #26_PNG	+/+	+/+	+/+	+/-	2007
S. flexneri #47_PNG	+/+	+/-	+/-	-/-	2010
S. flexneri #50_PNG	+/+	+/+	+/+	+/+	2010
S. flexneri #62_PNG	+/+	+/+	+/+	+/-	2010
S. flexneri #49_PNG	+/+	+/+	+/+	+/-	2010
S. flexneri #34_PNG	+/+	+/+	+/+	+/-	1990
S. flexneri #23_PNG	+/-	-/-	-/-	-/-	2004
S. flexneri #33_PNG	+/+	+/+	+/+	+/-	1990
S. flexneri #35_PNG	+/+	+/+	+/+	-/-	1990
S. flexneri #21_FJ	+/+	-/-	+/+	+/-	2007
S. flexneri #24_PNG	+/-	-/-	-/-	-/-	1994
S. flexneri #1_TS	+/-	+/+	+/+	+/+	2009
S. flexneri #19_PNG	+/+	+/-	+/+	+/+	2006
S. flexneri #4_PNG	+/+	+/+	+/+	+/+	2011
S. flexneri #25_PNG	+/+	+/+	+/+	+/-	2009
S. flexneri #51_PNG	+/-	+/-	+/-	+/-	2010
S. flexneri #53_PNG	+/+	+/-	+/-	+/-	2010
S. flexneri #52_PNG	+/+	+/-	+/-	+/+	2010
S. flexneri #2_TS	+/-	+/+	+/+	+/+	2010
S. flexneri #10_PNG	+/+	+/+	+/+	+/+	2013
S. flexneri #9_PNG	+/+	+/-	+/+	+/+	2013
S. flexneri #55_PNG	+/+	+/+	+/+	+/+	2010
S. flexneri #60_PNG	+/+	+/+	+/+	+/-	2010
S. flexneri #58_PNG	+/+	+/-	+/+	+/-	2010
S. flexneri #57_PNG	+/+	+/-	+/+	+/+	2010
S. flexneri #46_PNG	+/+	+/+	+/+	+/+	2010
S. flexneri #61_PNG	+/+	+/-	+/-	-/-	2010
S. flexneri #43_PNG	+/+	+/+	+/-	-/-	2010
S. flexneri #42_PNG	+/+	+/+	+/+	+/+	2010
S. flexneri #41_PNG	+/+	+/+	+/+	-/-	2010
S. flexneri #56_PNG	+/-	+/-	+/-	+/+	2010
S. flexneri #45_PNG	+/+	+/+	+/+	-/-	2010
S. flexneri #44_PNG	+/+	+/+	+/+	+/-	2010
S. flexneri #40_PNG	+/+	+/+	+/+	+/-	2010
S. flexneri #59_PNG	+/+	+/+	+/+	-/-	2010
S. flexneri #14_PNG	+/+	+/-	+/+	+/+	2014
S. flexneri #39_PNG	+/+	+/+	+/+	+/+	2013
S. flexneri #5_PNG	+/+	-/-	+/+	+/+	2011
S. sonnei #22_SI	-/-	-/-	-/-	+/+	2007
S. sonnei #18_VUT	+/-	+/-	+/-	+/+	2015
S. sonnei #11_VUT	-/-	-/-	-/-	-/-	2013
S. sonnei #12_WS	+/+	-/-	+/+	+/+	2013
S. sonnei #6_FJ	+/-	-/-	-/-	+/-	2012
S. sonnei #7_FJ	+/-	-/-	+/-	+/-	2012
S. sonnei #8_FJ	+/-	+/-	+/-	+/-	2012
S. sonnei #30_PNG	-/-	-/-	- /-	+/+	2008
S. sonnei #28_PNG	+/+	+/-	+/-	+/+	1999
S. sonnei #16_PNG	+/+	-/-	-/-	+/+	2014
S. sonnei #3_PNG	+/+	+/-	+/-	+/+	2011
S. sonnei #29_PNG	+/+	-/-	+/-	+/+	2010
S. sonnei #15_PNG	+/+	-/-	+/+	+/+	2014
S. sonnei #17_PNG	+/+	-/-	-/-	+/+	2015
S. sonnei #31_PNG	+/+	-/-	-/-	+/+	2011
S. sonnei #13_PNG	+/+	-/-	+/-	+/+	2014

Table 5.5 Resistance to antimicrobials as detected via the CLSI disc diffusion method in the presence of one or more AMR genes as detected via WGS, suggesting conveyance of antibiotic resistance. All analyses had one degree of freedom and all possible AMR combinations were analysed. Only significant outputs are included in this table.

***Shigella* spp.**	**CLSI resistance**	**AMR gene(s)**	**Chi-Square Test Pearson value**	**Fisher's Exact Test P-value**
S. flexneri	34/41 (AMP_R)	*OXA_1_Bla* (n=38/41)	15.72	<0.05
S. flexneri	34/41 (AMP_R)	*AmpC1_Ecoli_Bla* + *OXA_1_Bla* (n=34/41)	15.72	<0.05
S. flexneri	34/41 (AMP_R)	*TEM-1D_Bla* + *OXA_1_Bla* (n=34/41)	15.72	<0.05
S. sonnei	9/16 (AMP_R)	*TEM−1D_Bla* (n=12/16)	6.86	<0.05
S. flexneri	25/41 (C_R)	*CatA1_Phe* (n=36/41)	8.90	<0.05
S. flexneri	31/41 (TET_R)	*TetB_Tet* (n=38/41)	10.03	<0.05
S. flexneri	17/41 (SXT_R)	*DfrA5_Tmt* (n=27/41)	6.47	<0.05
S. flexneri	17/41(SXT_R)	*DfrA1_Tmt* + *DfrA5_Tmt* (n=2, gene pos. n=27/41)	12.02	<0.05
S. flexneri	17/41 (SXT_R)	*DfrA3_Tmt* + *DfrA5_Tmt* (n=1, <1 gene pos. n=27/41)	8.95	<0.05
S. flexneri	17/41 (SXT_R)	*DfrA8_Tmt* + *DfrA5_Tmt* (n=1, gene pos. n=27/41)	10.32	<0.05

AMP_R, Ampicillin resistant; C_R, Chloramphenicol resistant; TET_R, Tetracycline resistant; SXT_R, Trimethoprim resistant.

Table 5.3 shows correlations between AMR genes and phenotypic resistance. The presence of AMR genes *OXA_1_Bla* alone and in unison with *AmpC1_Ecoli Bla* and *TEM-1D_Bla* genes was associated with resistance to ampicillin in *S. flexneri*. Also in *S. flexneri*, the presence of *CatA1_Phe* was associated with chloramphenicol resistance; the presence of *TetB_Tet* associated with resistance to tetracycline; and the presence of *DfrA5_Tmt* alone, and in unison with *DfrA1_Tmt*, *DfrA3_Tmt* and *DfrA8_Tmt,* with resistance to co-trimoxazole.

In *S. sonnei* the presence of AMR gene *TEM-1D_Bla* was significant for the conveyance of AMR to ampicillin.

5.4 Discussion

The commonly used diagnostic method for detection of *Shigella* in low-resource settings, and indeed the current gold standard, is bacterial culture. However, the low uptake of culture, and the rapidly developing technology and data analysis associated with WGS, suggests alternative methods for *Shigella* diagnosis and characterisation should be investigated. Although this study was conducted in well-resourced laboratories and sequencing facilities in high-income settings, it sought to gain an insight into the applicability of WGS for detection and characterisation of *Shigella* in low-income settings. In doing so, it also sought to detect virulence and AMR genes present in *Shigella* isolates from PNG and other LMICs in Oceania.

This study demonstrates that WGS is able to detect phenotypic AMR, but does so with very poor specificity. Indeed, to determine the suitability of WGS for detecting phenotypic AMR, a diagnostic evaluation was conducted to determine sensitivity and

specificity of selected AMR genes/mutations relative to the gold standard of disc diffusion (phenotypic testing) (Appendix 3). These data clearly demonstrate the poor sensitivity of WGS analysis. While sensitivity was as high as 100% (for *Bla*$_{OXA1}$), the specificity of the same gene was 5%. That is *Bla*$_{OXA1}$ always predicted phenotypic resistance to ampicillin, but the gene was detected in all isolates, irrespective of their phenotypic resistance. Using a combination of genes, where any one gene present was considered predictive of resistance, or all selected genes present was considered predictive for resistance, did not improve outputs. While specificity was often better, sensitivity was poor (data not shown).

The findings of this study, suggesting limitations in the application of WGS to prediction of phenotypic AMR, contradict the finding of two recent studies conducted in Enterobacteriaceae. Sadouki *et al.* (2017) conducted a study on 341 *S. sonnei* isolates, and Neuert *et al.* (2018) investigated 3,491 *Salmonella enterica* isolates. There are many variables involved in these comparative studies. For example, both Sadouki *et al*. and Neuert *et al*. used the EUCAST system (as opposed to the CLSI system used in this study) to determine phenotypic antibiotic sensitivity, and determined minimum inhibitory concentrations using agar diffusion and Epsilometer tests. Moreover, sequencing platforms and pipelines for determination of the presence of AMR genes differed among all three studies. Despite these differences, all studies applied recognised and broadly accepted methods, thus a comparison of results is worthy.

In brief, concordance of phenotypic AMR with detection of AMR genes by WGS was considerably higher in the studies conducted by Sadouki *et al*. (2017) and Neuert *et al*. (2018) in *S. sonnei* and *Salmonella enterica*, respectively. Sadouki and colleagues

reported that for a total of 3350 isolate/antimicrobial combinations (that is 335 isolates of *S. sonnei*, each tested for resistance to 10 antibiotics) there were only 15 discordant results (99.55% concordance). Even more compelling are the results of Neuert and colleagues, where 99.83% of a total of 52,365 isolate/antimicrobial combinations were concordant. These results are in stark contrast to the results of this study. Focusing only on the antibiotics where phenotypic resistance was observed (namely ampicillin, chloramphenicol, tetracycline and co-trimoxazole) only 75%, 44%, 60% and 51% of isolate/antimicrobial combinations respectively were concordant.

The difference in rate of concordance for prediction of AMR by WGS in this study relative to that of Sadouki *et al* (2017), was considerable, but some consistencies were observed between the two studies focusing on *Shigella*. The genes *OXA_1_Bla, CatA1_Phe, DfrA1_Tmt, DfrA5_Tmt, DfrA8_Tmt* and *TetA_Tet* were commonly detected in both studies; though some of the above genes were detected in *S. flexneri* in this study. In looking only at *S. sonnei*, genes detected in both studies were *TEM-1D_Bla, DfrA1_Tmt*, *DfrA3_Tmt* and *DfrA8_Tmt*; and were significantly associated with phenotypic resistance, either singularly or in combination with other genes (Appendix 3). These similarities are welcome, but raise the question of the exact criteria used by Sadouki *et al*. and Neuert *et al*. Despite detailed statistical analyses investigating correlations between genes and phenotypic resistance, neither study clearly states whether the presence of a single AMR gene, or a combination of genes, was used to predict phenotypic AMR. This is an unfortunate omission given the very high concordance observed in those studies. In this study we used the presence of a single gene as a predictor of phenotypic AMR, as combinations of genes did not improve diagnostic outputs (of either sensitivity or specificity).

Most of the discordance observed in this study was in isolates that were predicted to be genotypically resistant to a selection of antimicrobials but were phenotypically susceptible. This suggests that the AMR determinants were present, but were not expressed, as has been observed in other studies (Enne, Delsol, Roe, & Bennett, 2006; Rajakumar, Sasakawa, & Adler, 1997; Sadouki *et al*., 2017; Turner, Luck, Sakellaris, Rajakumar, & Adler, 2001; Walker & Verma, 2002). In this chapter, the focus has been on the application of WGS to replace culture and sensitivity. However, AMR detection is not only important for guiding treatment, but AMR surveillance is of great public health value. For surveillance the presence (regardless of expression) of AMR genes may well be of public health significance.

The WGS approach is well suited to screening and monitoring for AMR, but there are some potential shortcomings associated with its application to guiding treatment. WGS is being used in clinical diagnostic and public health laboratories, and for the most part the newer technology is promising. However, in a study investigating the applicability of WGS for AMR prediction, discrepancies could be traced to relatively easily addressed problems such as sequence quality, database choice or user error (Doyle *et al*., 2020; Koser, Ellington, & Peacock, 2014). In this study it was not possible to ascertain whether discrepancies were the result sequencing and bioinformatics errors, or errors in phenotypic testing. To further investigate these discordant findings MIC testing should be conducted on *Shigella* isolates.

For the detection of virulence genes, the benefits of WGS over PCR are apparent. Where PCR did not detect the presence of some virulence genes, these were picked up using WGS. PCR is an amplification-based method that investigates small, specific fragments

of DNA; and when conducted well is generally considered both sensitive and specific. Thus, PCR is suitable for the detection of particular AMR or virulence genes (Quainoo *et al*., 2017). However, in this study, PCR failed to detect virulence genes that were detected by WGS in some isolates. There are numerous factors that can lead to false negative PCR detection of genes, many of which have been outlined by Schrick & Nitsche (2016). Low copy numbers of a gene can impact on both PCR and WGS, but could conceivably have greater impact on PCR than WGS under some conditions. For example, mutations in PCR primer annealing sites would negatively impact PCR, but not WGS.

Both PCR and WGS have the potential to improve detection of virulence genes and AMR genes. Despite shortcomings in AMR gene detection, it seems likely that WGS may become more widely used in the future for pathogen characterisation. Savings in cost and time are becoming greater as the technology matures. However, this study has demonstrated that there remain some challenges that need to be further investigated. Moreover, specialized bioinformatics tools to analyse WGS data are still required for many pathogens and sequencing platforms. In this analysis, WGS has been has been considered in the context of a replacement for culture and sensitivity and targeted PCR. However, we acknowledge that the logistics and platform used in this study is not ideal for future application in low-income settings. In Chapter 7 this discussion will be furthered, with consideration given to 'field-applicable' sequencing technology.

Chapter 6 - The Phylogenetics of *Shigella sonnei* and *Shigella flexneri* from Oceania with a Focus on Papua New Guinean Isolates

6.1 Introduction

Molecular biology revolutionised microbial phylogenetics, and more recently WGS has enabled microbiologists to comprehensively and accurately determine the within-species relatedness of bacteria, viruses and fungi. While WGS has been applied to a wide range of microbes, the technique has been particularly useful in the characterisation of pathogenic species; vastly improving knowledge of pathogen epidemiology from local levels through to a global scale.

WGS has been applied to *Shigella* spp., with large-scale studies conducted in recent years on the molecular epidemiology of *Shigella* circulating in Africa (Connor *et al*., 2015), Asia (Connor *et al*., 2015; Holt *et al*., 2013), Latin America (Connor *et al*., 2015) and Europe (Connor *et al*., 2015; Holt *et al*., 2012). These and other studies have demonstrated the evolution and regional distribution (at a global scale) of strains of *Shigella*.

In Oceanic LMICs there remains a high burden of infectious disease, but knowledge of circulating pathogens is lacking; this impedes treatment and control of infectious diseases. There is sufficient historical and recent data to suggest that *Shigella* is an important cause of gastrointestinal illness in PNG, the second largest country by population in the Oceania region. Although there only limited data on the burden of shigellosis in other Oceanic LMICs, diarrhoeal diseases remain a high burden throughout the region and *Shigella* may

be an important contributor. However, there are no published data that illustrate the phylogeny of *Shigella* isolates distributed throughout Oceanic LMICs.

The main objectives of this chapter are to determine:

1. The phylogenetic relationship of *Shigella* isolates from PNG and other Oceanic LMICs to each other, and to global isolates.
2. To confirm the presence of *Shigella* virulence genes and compare their findings to detection of virulence genes by targeted PCR (as conducted in Chapter 4).

6.2 Materials and Methods

6.2.1 DNA Sequencing

DNA from 63 *Shigella* isolates were sent to the Wellcome Trust Sanger Institute in the United Kingdom and WGS was conducted using the Illumina Hiseq Genome Analyzer, as described in Section 5.2. For isolates that were successfully sequenced, data were analysed using bioinformatics tools, as described in Section 6.2.2 below.

6.2.2 Mapping and phylogenic methods

All 16 *S. sonnei* genomes from this study, along with 131 global *S. sonnei* genomes from Holt *et al.* (2012), were mapped to reference genome *S. sonnei* 53G (chromosome: HE616528; plasmids: HE616529, HE616530, HE616531 and HE616532) using RedDog v1b10 (https://github.com/katholt/RedDog) with default parameters.

All 47 *S. flexneri* genomes from this study, and 344 global *S. flexneri* genomes from Connor *et al*. (2015), were mapped to reference genome *S. flexneri* 2a str. 301 (Accession

numbers AE005674 and AF386526) (Jin *et al*., 2002), using RedDog v1b10 as described above.

For both *S. sonnei* and *S. flexneri*, the resulting single nucleotide polymorphism (SNP) alignments were used to infer a phylogeny with RAxML v8.2.8 (https://github.com/stamatak/standard-RAxML) (Stamatakis, 2014), using a GTR+G base substitution model.

The sequences of *Shigella* isolates from the Oceania region were then compared to other *Shigella* isolates in the global database.

6.2.3 Detection of AMR and virulence genes

As described in Section 5.2, all 63 genomes from this study were screened for AMR genes in the ARG-ANNOT database (http://www.mediterranee-infection.com/article.php?laref=282&titer=arg-annot) (Gupta *et al.*, 2014). To determine whether prevalence of AMR genes was higher in the Asia-Pacific region than globally, basic prevalence analysis was conducted. Significant differences in prevalence were determined with Fisher's Exact T test using an online interactive calculation tool by (Preacher, 2001).

Shigella virulence genes were detected in the virulence factors database (VFDB) (http://www.mgc.ac.cn/VFs/main.htm) using srst2 v0.2.0 (https://github.com/katholt/srst2) with default parameters; as previously described (Section 5.2).

A mapping approach, using the reads (not assemblies) was used to screen for the presence/absence of genes. The SRST2 program was used to screen for both AMR and virulence genes.

Quality control of SNP calls was handled using the RedDog mapping pipeline (https://github.com/katholt/RedDog).

Phylogenetic trees and heatmaps for AMR genes and virulence genes were visualised using 'ggtree' on R-studio for *S. flexneri* and *S. sonnei*.

6.2.4 WGS serotyping of Oceania *Shigella flexneri* isolates

WGS was used to conduct *in silico* serotyping (H and O antigens), as described in Joensen, Tetzschner, Iguchi, Aarestrup, & Scheutz (2015), using the SerotypeFinder public online gene database (http://cge.cbs.dtu.dk/services/). The detection of the H antigen is confirmed via the presence of the *fliC* gene that encodes for the flagellin protein. O antigen synthesis is inferred via the presence of the sugar transferases genes *wzx*, which codes for flippase, and *wzy*, which codes for O-antigen polymerase.

The presence of serotype-specific genes *gtrI*, *gtrIC*, *gtrII*, *oac*, *gtrIV*, *gtrV* and *gtrX* was used to determine specific *S. flexneri* serotypes. A FASTA file grouping these serotype-specific genes (Accession numbers: AF139596.1, KT988056.1, KT988057.1, NC_000913.3, NC_003444.1 and FJ905303.1) was generated and run against WGS data obtained from the *S. flexneri* (n=47) Oceania isolates, using the NCBI BLAST + blastn tool on the Galaxy web platform (https://usegalaxy.org), as previously described by Sun *et al.* (2011).

6.3 Results

6.3.1 Sequencing

Of the 63 *Shigella* isolates for which sequencing analysis was conducted, 47 were *S. flexneri* and 16 *S. sonnei*.

A total of 1,691 SNPs were detected in the 16 *S. sonnei* Oceania isolates, with a total of 9,710 SNPs detected across the 147 global genomes.

A total of 7,582 SNPs were detected in the 47 *S. flexneri* Oceania isolates, with a total of 83,409 SNPs detected across the 391 global genomes.

6.3.2 Phylogenetics

6.3.2.1 *Shigella sonnei*

The Oceania isolates of *S. sonnei* clustered into two distinct lineages, LII and LIII. Isolates from the Solomon Islands, Vanuatu, Samoa and Fiji were all within the LII lineage and cluster together in their own clade. These isolates are closely related to African isolates from Madagascar and Tanzania (Figure 6.1 and 6.2). Isolates from PNG form their own clade within LIII, and share a common ancestor with isolates predominantly from Asia and the Middle East. A single isolate from New Caledonia sits within the Asia-Middle East isolates in LIII (Figure 6.3).

The estimated time of introduction of the *S. sonnei* LII clade into Oceania was 1986 (95% highest posterior density (HPD) 1975 - 1995), with the LIII clade pre-dating the LII clade by an estimated 2 years (1984, 95% HPD 1977 - 1991).

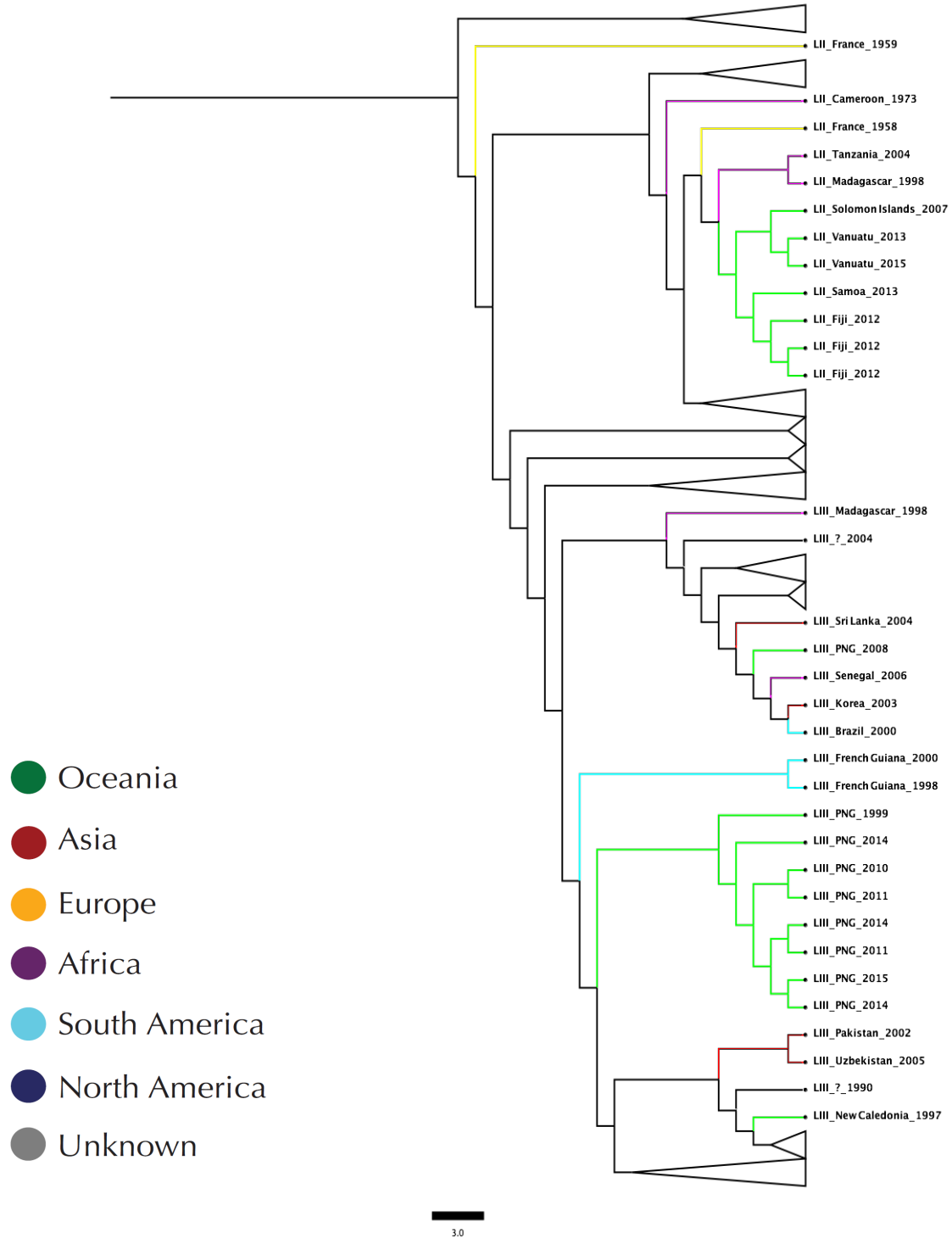


Figure 6.1 *S. sonnei* phylogenetic tree using 'figtree' with collapsed branches, depicts only Oceania isolates with their closest relations.

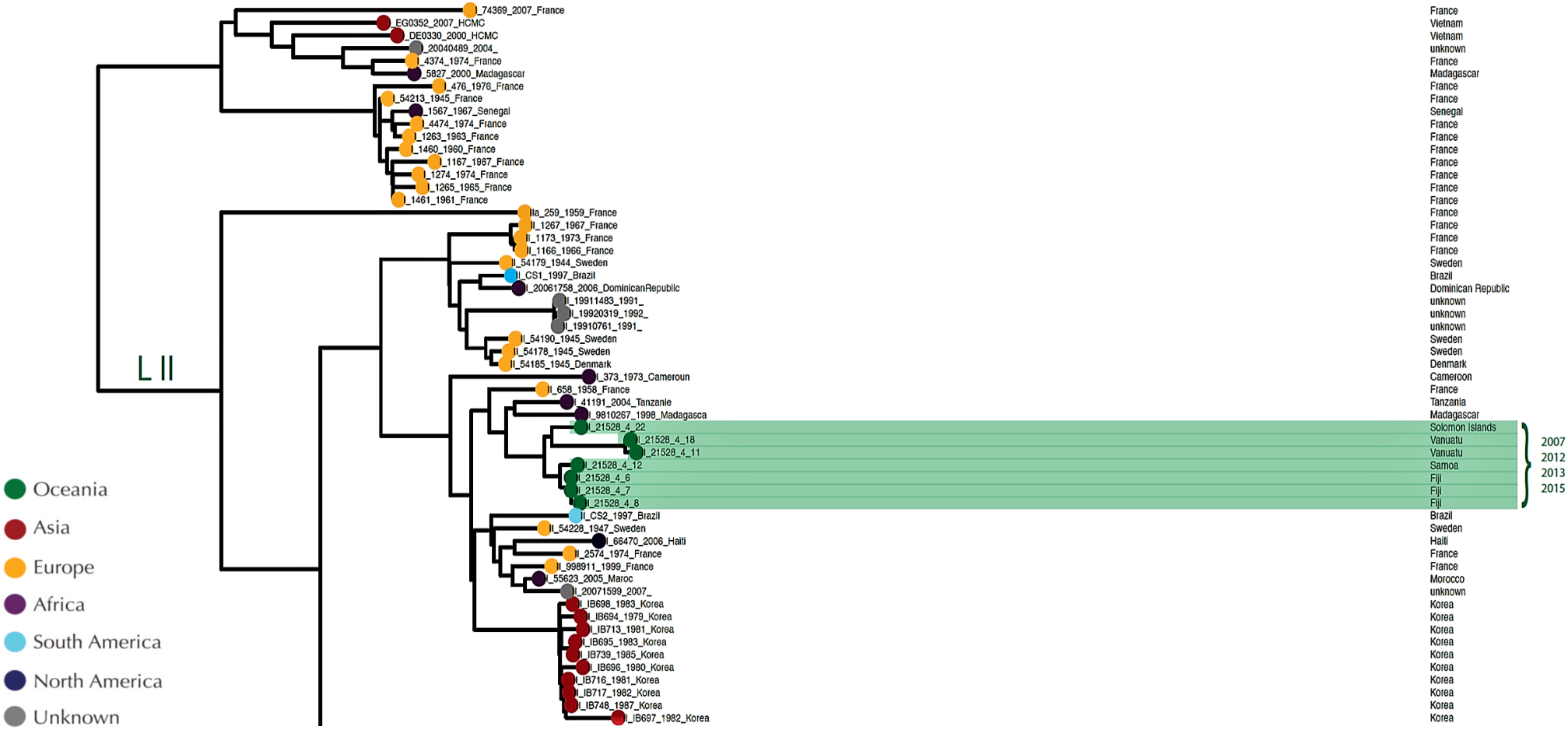


Figure 6.2 A section of the *S. sonnei* phylogenetic tree depicting LII global and Oceania isolates from Solomon Islands, Vanuatu, Samoa and Fiji, with isolation dates to the right (range 2007 – 2015). The full S. sonnei phylogenetic tree is provided in Appendix 4.2.

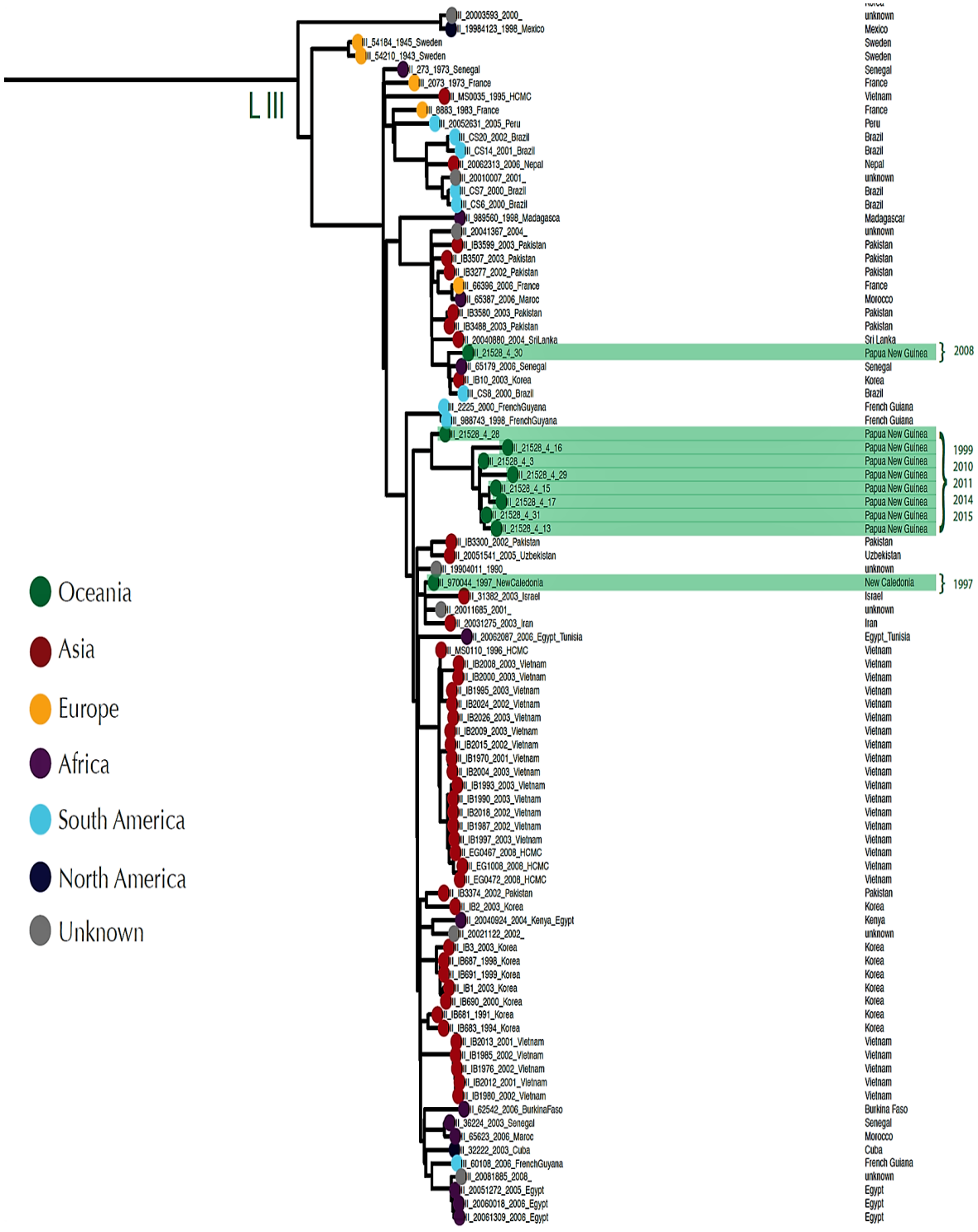


Figure 6.3 A section of *S. sonnei* phylogenetic tree depicting LIII global and Oceania isolates from PNG and New Caledonia, with isolation dates to the right (range 1997 – 2015). The full *S. sonnei* phylogenetic tree is provided in Appendix 4.2.

6.3.2.2 *Shigella flexneri*

All Oceanic *S. flexneri* isolates in this study belong to PG1, PG2 or PG3.

The PG1 isolates from PNG (n=4, near middle of phylogenetic tree in Figure 6.4 and 6.5) cluster tightly together and are closely related to African isolates; another Oceania PG1 isolate from Fiji is located on a separate branch below the PNG isolates and is most closely related to isolates from Central America (Haiti and Dominican Republic) and Vietnam. Due to small sample size, PG1 did not have a sufficiently strong temporal signal to permit estimation of the year of introduction of these isolates into Oceania.

S. flexneri PG2 isolates from PNG (n=7, upper-middle section of tree in Figure 6.6) cluster tightly together and are most closely related to a single isolate from Pakistan. The PG2 isolates from PNG sit amongst a clade of South and Southeast Asian isolates. The introduction of PG2 isolates into Oceania/PNG is estimated to have occurred in 2002 (95% HPD 1995-2004).

The PG3 isolates from PNG (Figure 6.7) are estimated to have been introduced in 1979 (95% HPD 1974-1985). There is a tight cluster of isolates (n=29) from PNG and the Torres Straits Islands (i.e. Australia) that are closely related to a single Asian isolate from Korea. In a separate branch within PG3, one PNG isolate is most closely related to African isolates from Tunisia and Madagascar. Other PG3 isolates (n=3) were closely related to Asian isolates from Bangladesh, and a separate PNG isolate was closely related to Asian isolates from Vietnam.

The distinct clusters of PNG and Oceania *S. flexneri* isolates belonging to PG1, PG2 and PG3 are illustrated in Figures 6.4 – 6.7. The full *S. flexneri* phylogenetic tree is in Appendix 4.3.

Figure 6.4 *S. flexneri* phylogenetic tree using figtree with collapsed branches. Tree depicts only Oceania isolates with closest relations.

Figure 6.5 A section of the *S. flexneri* phylogenetic tree depicting global and Oceania isolates clustered in PG1. Isolates from PNG were from years 1990 and 2004, with the Fiji isolate from year 2007. Year of isolation provided to right of image. The full *S. flexneri* phylogenetic tree is provided in Appendix 4.3.

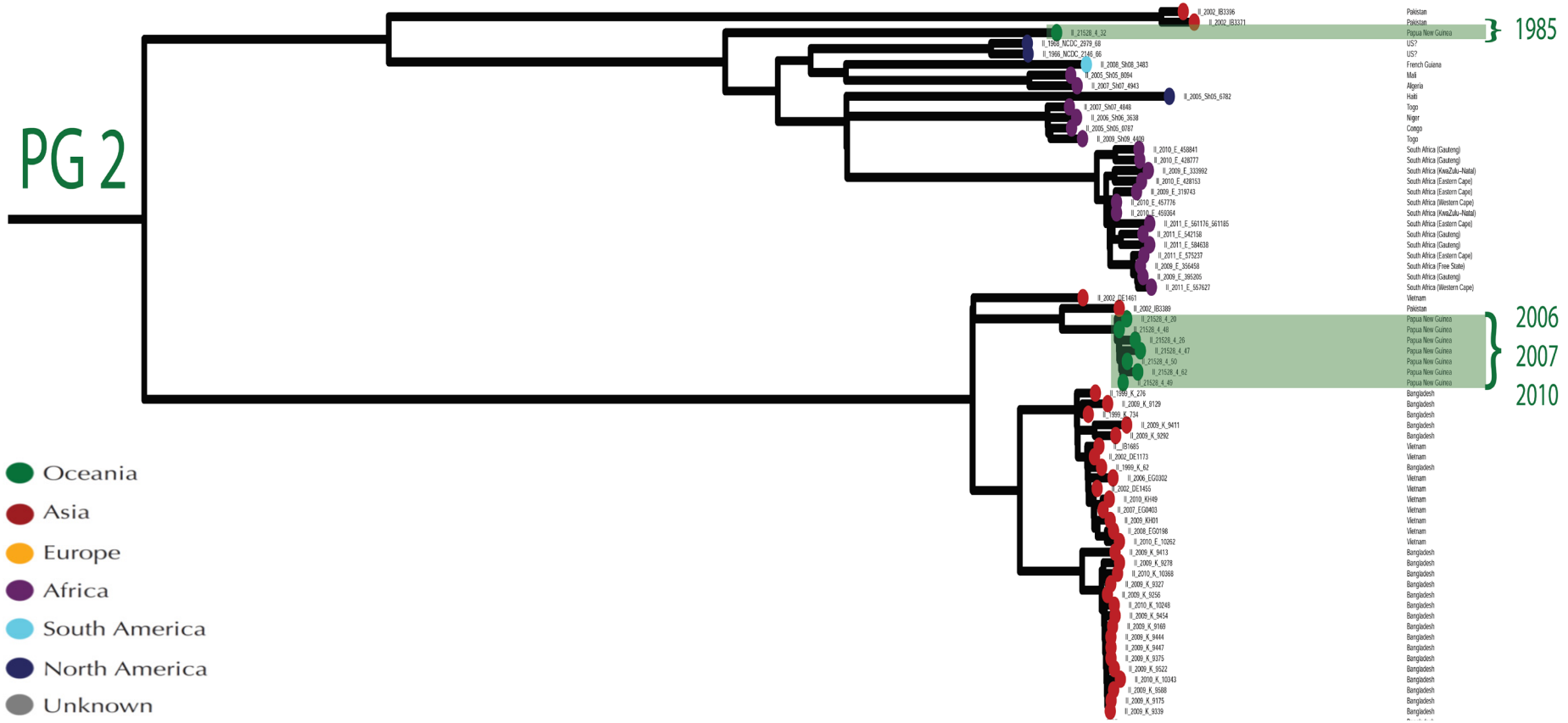


Figure 6.6 A section of *S. flexneri* phylogenetic tree depicting global and Oceania isolates clustered in PG2. Isolation dates range from 1985 to 2010 (year of isolation to right of image). The full *S. flexneri* phylogenetic tree is provided in Appendix 4.3.

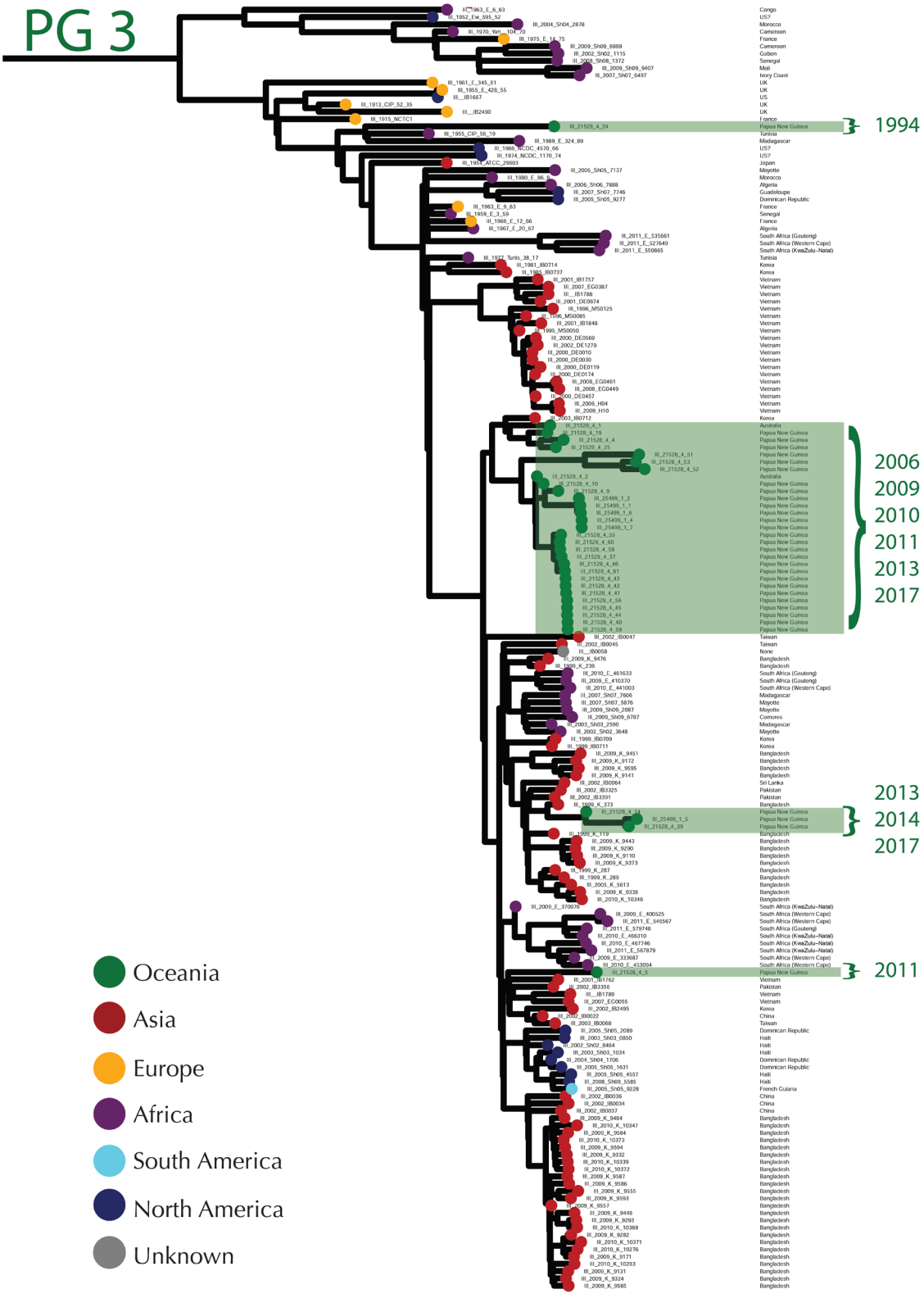


Figure 6.7 A section of *S. flexneri* phylogenetic tree depicting global and Oceania isolates clustered in PG3. Year of isolation ranged from 1994 to 2017 (provided on right of image). The full *S. flexneri* phylogenetic tree is provided in Appendix 4.3.

6.3.3 The presence of AMR and virulence genes in *Shigella sonnei*

Collectively, the 16 Oceania isolates of *S. sonnei* harboured 16 different antimicrobial resistance genes, with prevalence ranging from 6% – 88% for those genes detected (Table 6.1).

The prevalence of *strB* (aminoglycosides), *oxa1* and *tem1d* (beta-lactams) was significantly higher in isolates of *S. sonnei* from Oceania compared to *S. sonnei* global isolates; and the prevalence of *dfrA5*, *dfrA8* (trimethoprim), *tetB* (tetracycline) and *sulII* (sulfonamide) was significantly higher in *S. sonnei* Oceania isolates than both Asian and global isolates. For the other resistance genes detected in *S. sonnei* Oceania isolates, their prevalence was lower than detected in *S. sonnei* Asian and/or global isolates, or no significant difference was observed (Appendix 4.10). A heatmap showing presence/absence of AMR genes detected via WGS can be observed in Appendix 4.8.

The presence/absence of *S. sonnei* virulence genes is presented in Chapter 5. A heatmap showing presence/absence of virulence genes detected via WGS can be observed in Figure 6.8. The *SenB* gene is present in all *S. sonnei* global isolates except for one PNG isolate from LIII isolated in 2008. A total of three *S. sonnei* Oceania isolates, two in LII from Vanuatu and Fiji and one in LIII from PNG, harboured the *set1A* gene. Section 5.3.2 (Table 5.1 and Figure 5.1) depicts in detail the presence/absence of virulence genes in *S. sonnei* Oceania isolates as confirmed via WGS.

ShET-1 (*set1A*) detected in *S. sonnei* (Figure 6.8) by WGS is thought to be a contamination. This gene is not typically observed in *S. sonnei*. There was low coverage of that sequence for the isolates #21528_4_18, #21528_4_8 and #21528_4_3 where the *set1A* gene was detected, most likely the result of carry-over contamination during the sequencing process.

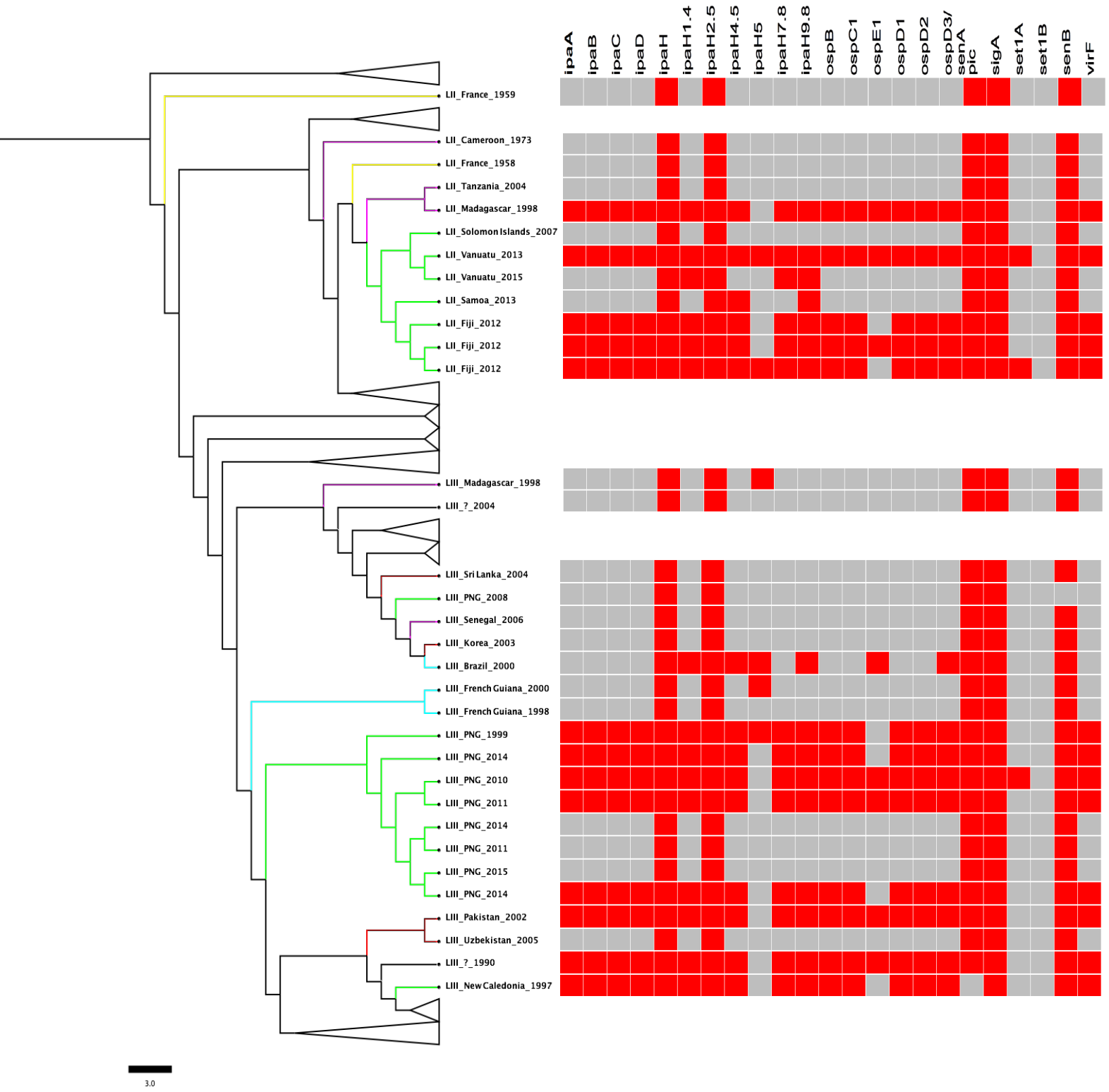


Figure 6.8 *S. sonnei* phylogenetic tree depicts only Oceania isolates with closest relations and heatmap with virulence genes. Red tiles indicate presence of virulence genes, grey tiles indicate absence.

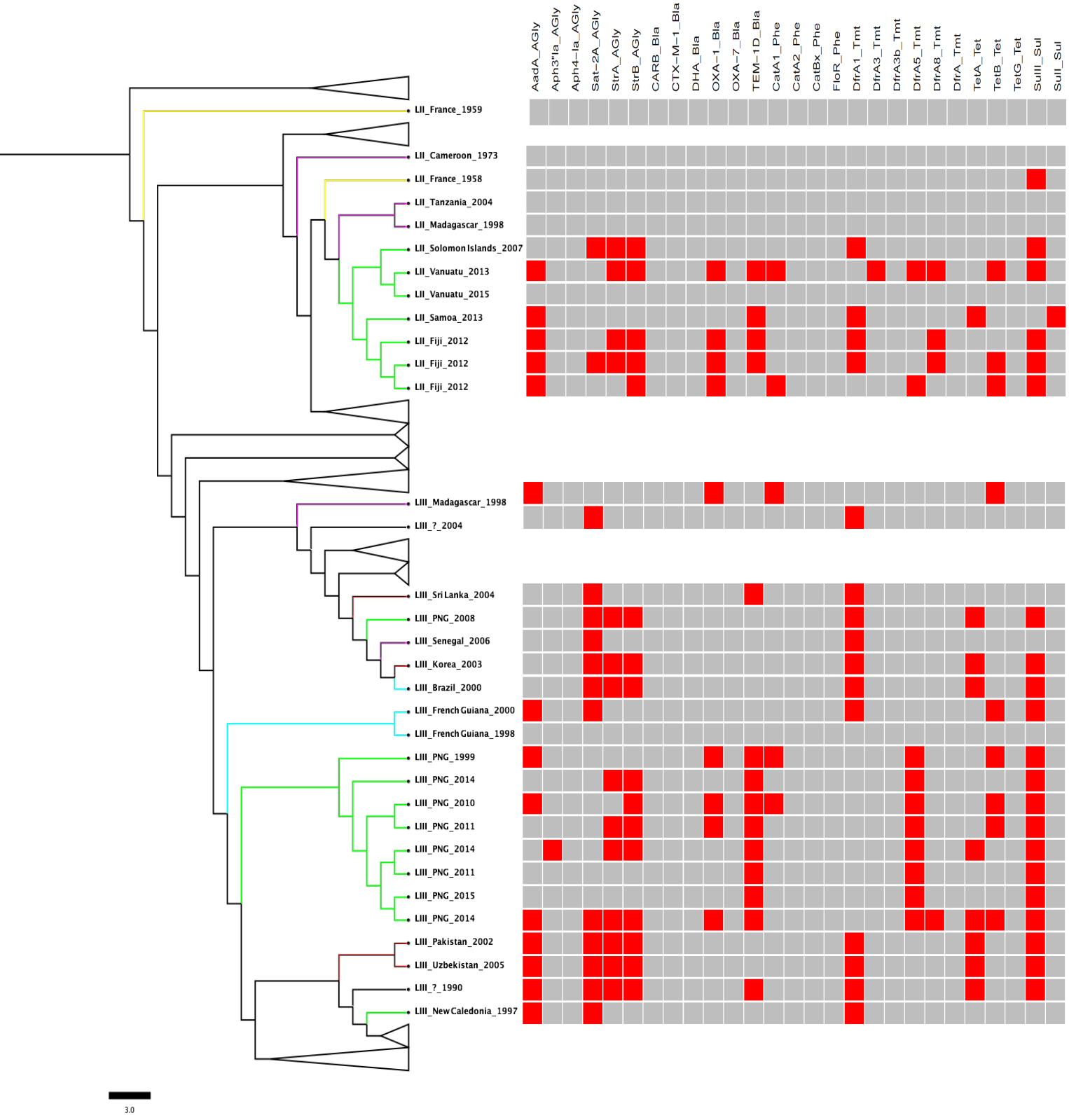


Figure 6.9 Heatmap of *S. sonnei* AMR genes. Red tiles indicate presence of AMR genes and grey tiles indicate gene absence.

6.3.4 The presence of AMR and virulence genes in *Shigella flexneri*

Collectively, the *S. flexneri* Oceania isolates (n=47) harboured 21 different antimicrobial resistance genes with prevalence in the range 2% – 100% for those genes detected (Table 6.1).

The prevalence of *aadA* (aminoglycosides), *AMPC1_Ecoli*, *oxa1* (beta-lactams), catA1 (phenicol), *dfrA5* (trimethoprim) and *tetB* (tetracycline) was significantly higher in *S. flexneri* Oceania/PNG isolates compared to both *S. flexneri* global and Asian isolates; and the prevalence of *sat2A* (aminoglycosides), *dfrA1* (trimethoprim) and *tetR* (tetracycline) was significantly lower in *S. flexneri* Oceania/PNG isolates than both *S. flexneri* Asian and global isolates. For the other resistance genes detected in *S. flexneri* Oceania/PNG isolates, their prevalence was lower than detected in *S. flexneri* Asian and/or global isolates, or no significant difference was observed (Appendix 4.11). A heatmap showing both global and Oceania presence/absence of AMR genes detected via WGS can be observed in Appendix 4.5.

No significant differences were observed in the presence/absence of virulence according to PGs for *S. flexneri* Oceania isolates. However, *S. flexneri* virulence genes coding for SHI-1 were predominantly present in *S. flexneri* Oceania PG3 isolates. A heatmap showing both global and Oceania presence/absence of virulence genes detected via WGS can be observed in Appendix 4.4.

Table 6.1 Presence of AMR genes of *S. sonnei* and *S. flexneri* isolates from Oceania as detected by WGS.

AMR gene	Resistance	Proportion *S. sonnei*	Proportion *S. flexneri*
aac6-Iaa	Aminoglycosides	0/16	1/47
aadA	Aminoglycosides	8/16	44/47
aph3"Ia	Aminoglycosides	1/16	0/16
sat2A	Aminoglycosides	4/16	4/47
strA	Aminoglycosides	9/16	27/47
strB	Aminoglycosides	11/16	32/47
AMPH_Ecoli	Beta-lactams	0/16	47/47
AmpC1_Ecoli	Beta-lactams	0/16	10/47
AmpC2_Ecoli	Beta-lactams	0/16	47/47
oxa1	Beta-lactams	8/16	44/47
tem1D	Beta-lactams	12/16	9/47
Penicillin_Binding	Betal-lactams	0/16	47/47
catA1	Phenicol	4/16	43/47
dfrA1	Trimethoprim	5/16	5/47
dfrA3	Trimethoprim	1/16	1/47
dfrA5	Trimethoprim	10/16	27/47
dfrA8	Trimethoprim	4/16	3/47
sulII	Sulfonamide	14/16	35/47
sulI	Sulfonamide	1/16	3/47
tetA	Tetracycline	4/16	1/47
tetB	Tetracycline	7/16	44/47
tetR	Tetracycline	0/16	1/47

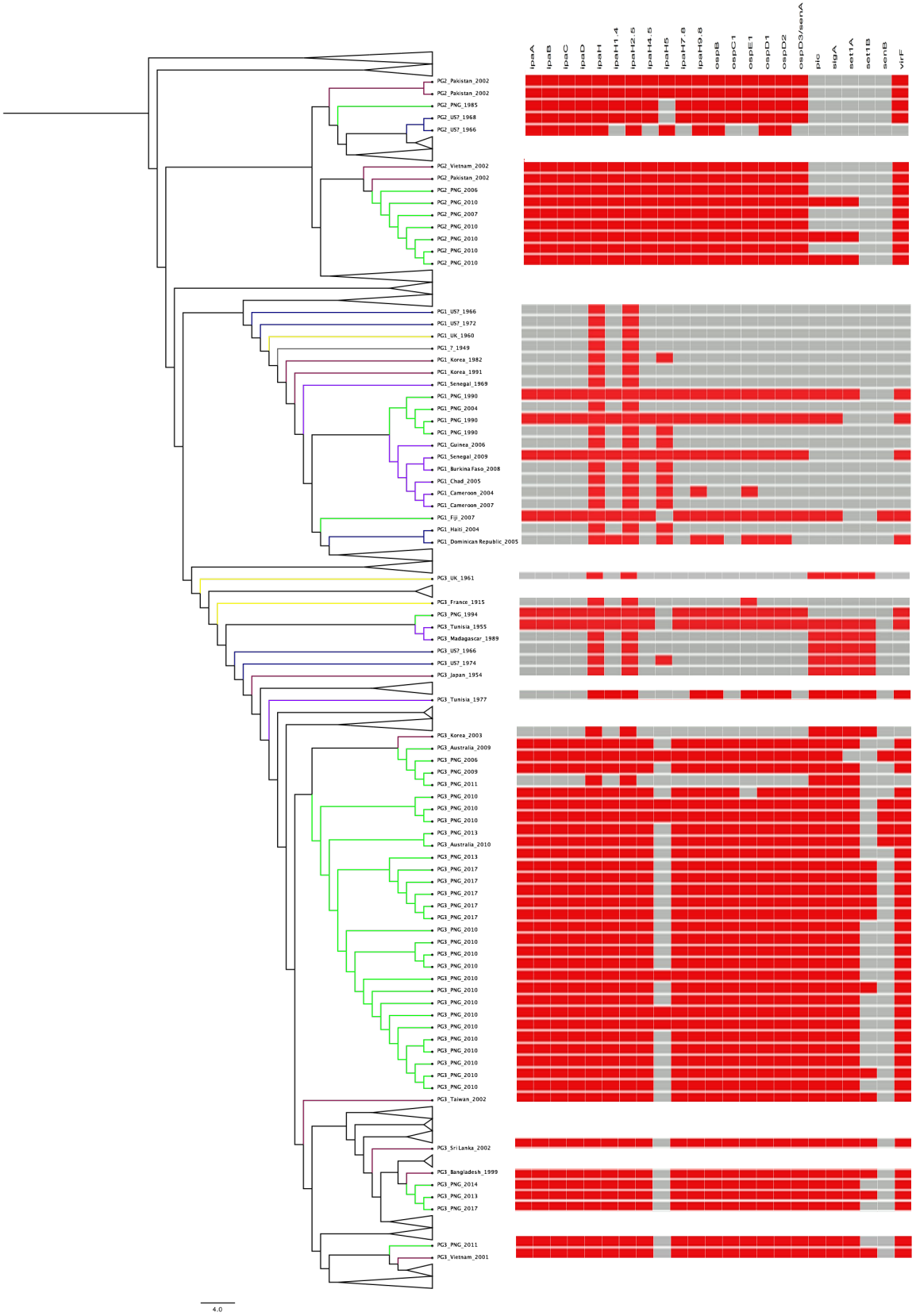


Figure 6.10 *S. flexneri* phylogenetic tree depicts only Oceania isolates with closest relations and heatmap with virulence genes. Red tiles indicate presence of virulence genes, grey tiles indicate absence.

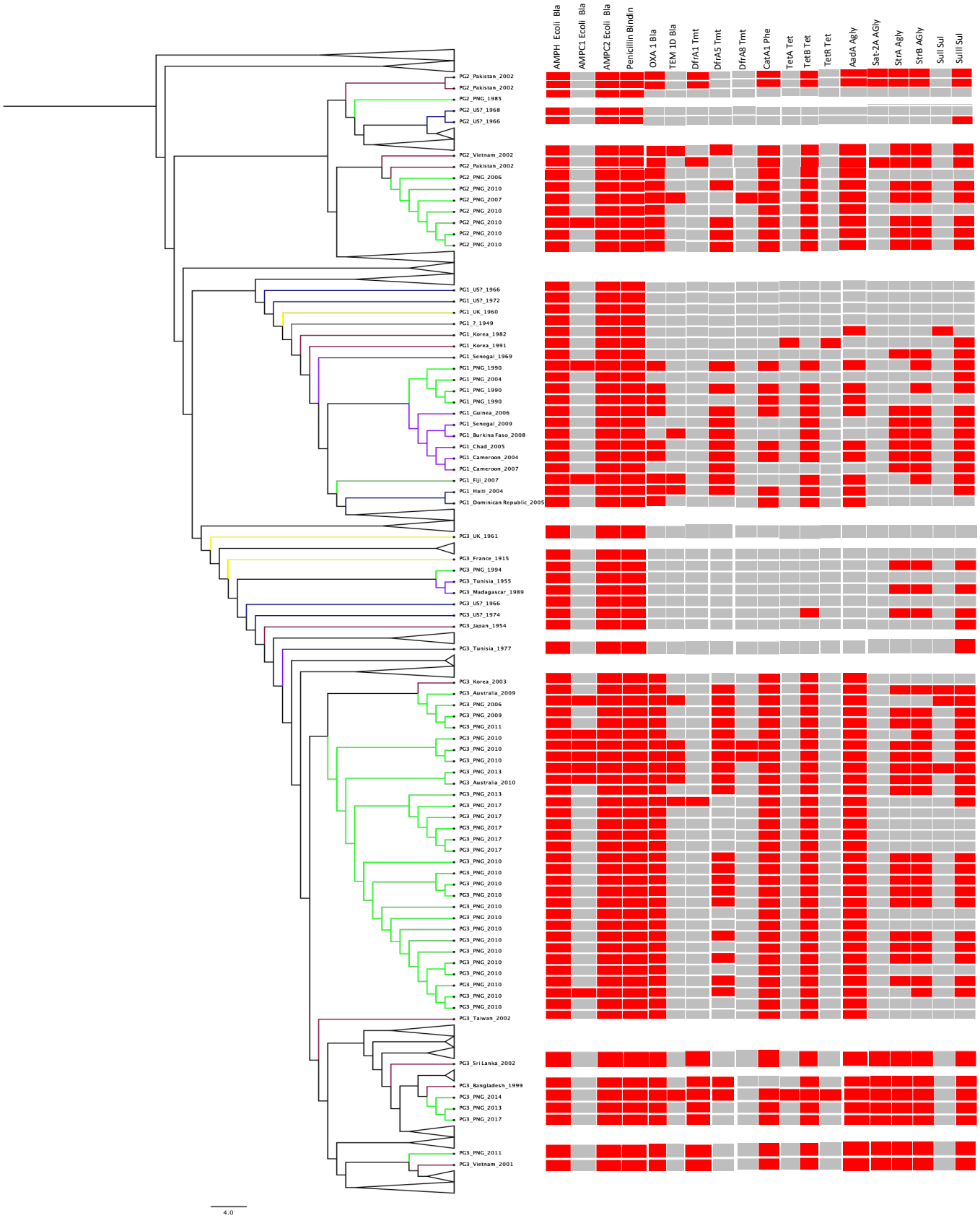


Figure 6.11 Heatmap of *S. flexneri* AMR genes. Red tiles indicate presence of AMR genes and grey tiles indicate gene absence.

6.3.5 *Shigella flexneri* serotypes using WGS data

Some, but not all, isolates were serotyped using traditional methods. WGS provided the means to determine more specifically which O and H antigens and serotype-specific genes were present in *S. flexneri*, thus enabling prediction of each isolate's serotype.

Of the 45 *S. flexneri* serotypes determined by molecular methods for Oceania isolates, the serotype distribution was 1a (1 of 45, 2%), 1b (4 of 45, 9%), 2a (30 of 45, 67%), 2b (3 of 45, 7%) and 3a (7 of 45, 16%). Previous serotyping results acquired via traditional serotyping methods for 12/63 *S. flexneri* isolates are provided in Appendix 1 (Table 2 of first manuscript within Appendix 1). Note that Table 2 contains 19 serotyping results; however, for this section of the study only 12 *S. flexneri* isolates were able to be compared with WGS analysis. Serotypes 1 (3 of 12, 25%), 3 (1 of 12, 8%), X (1 of 12, 8%) and 2a (7 of 12, 58%) were determined by traditional serotyping. Of the 12 isolates serotyped using traditional methods, 11 had the same serotype when determined by WGS analysis, with one isolate differing (formerly serotyped as X, but determined to be 2b using WGS data).

The serotypes of the Oceania isolates correlated to phylogenetic groups (Appendix 4.12). All Oceania isolates determined to be serotypes 1a and 1b belonged to PG1, serotypes 2a and 2b belonged to PG3, and serotype 3a to PG2.

6.4 Discussion

This study demonstrates that multiple strains of *S. flexneri* circulate in PNG. It appears that *S. flexneri* strains have entered PNG periodically, then co-circulate with pre-existing

strains of *S. flexneri*. This scenario is most clearly demonstrated by PNG isolates, where there is a sufficient number of isolates to consider temporal patterns. This study indicates that isolates belonging to PG1, PG2 and PG3 have been co-circulating in PNG since 2007. It is not possible to interpret the likelihood of co-circulation elsewhere in Oceania, due to the low number of non-PNG *S. flexneri* isolates for which genome sequencing data were obtained in this study. However, it can be stated with confidence that the scenario occurs at least in part of the Oceania region, based on PNG findings. Moreover, this supposition is supported by the findings of Connor *et al*. (2015), who analysed over 300 genomes of *S. flexneri* from various global locales (though not Oceania) and determined that *S. flexneri* PGs co-circulate. Thus it appears that, in PNG and the Pacific, the epidemiology of *S. flexneri* follows the same broad patterns as observed elsewhere globally: multiple introductions of the pathogen and co-circulation of long-established and comparatively recently-introduced strains.

Although numerous strains of *S. flexneri* co-exist, there is some geographical clustering of circulating phylogenetic groups. The majority of PNG *S. flexneri* isolates investigated in this study are PG3. In this study a convenience sampling approach was used; only isolates that could be obtained and sequenced to sufficient quality were analysed. On this basis it is not possible to conclude that PG3 strains predominate in PNG, but it does suggest that they are frequently encountered. It could either be a localised PG3 strain in the Eastern Highlands Province or introduced from Asia and has been circulating.

PG3 strains appear to have entered Oceania in 1979, and their common isolation over recent years (2006-17) suggests they remain an important phylogenetic group in PNG. Indeed, further surveillance would inform PNG health authorities whether this strain is an epidemic strain, and its distribution within PNG. The majority of PG3 isolates were

isolated in EHP as a result of the convenience sampling approach required in this study; though it may be that this strain has wider distribution in PNG.

It is likely there has been more than one incursion of PG3 into PNG, based on phylogenetic analysis and time of isolation. There are four different 'groups' (two consisting of a single isolate) of PNG isolates within PG3. Of relevance to the suggestion of multiple incursions of the PG is that one of the groups share an ancestor with African isolates, but the other three groups are closely related to Asian isolates. Moreover, one of the groups within PG3 is clearly related to southern Asian isolates (Bangladesh), but the other two groups are more closely related to eastern and southeastern Asian isolates. The "southern Asian" isolates have circulated alongside "eastern Asian" isolates in recent years. Clearly, though not surprisingly, *S. flexneri* isolates within PGs and from different PGs can co-circulate.

Another noteworthy finding regarding *S. flexneri* PG3 is that isolates from Torres Strait Islands are closely related to one group of PNG isolates. The potential for the cross-border transmission of infectious diseases is well known. Horwood and colleagues termed the Torres Strait Island region the Indo-Pacific conduit, although in their review they sadly overlooked the risk of spread of bacterial infections (with the exception of tuberculosis) (Horwood, McBryde, Peniyamina, & Ritchie, 2018). This study clearly demonstrates the potential for transmission of shigellosis across the PNG and northern Australian borders, despite considerably higher standards of sanitation and hygiene and better access to safe drinking water in Torres Strait Islands relative to PNG (Hall, Gillespie, Rosewell, & Mapira, 2013). Transmission of such infections is typically considered to a greater biosecurity risk for Australia than resource-poor neighbouring

countries, though this study did not seek to determine direction of dispersal. It is of interest that there may be widespread dispersal of the strain within PNG, as most of the PNG isolates from this PG3 subgroup originated from Eastern Highlands Province, which is geographically and topographically distinct from Torres Strait Islands.

As with *S. flexneri* PG3 strains, there appears to have been more than one incursion of PG2 isolates into PNG. Analysis conducted in this study suggested that PG2 was introduced into PNG in the early 2000s; however, this disregards the fact that one *S. flexneri* isolate in PG2 that was isolated in 1985. Notwithstanding the risks associated with drawing conclusions based on a convenience sampling approach, it appears that the earlier PG2 stains no longer circulate in PNG, and the currently circulating strains arrived in PNG in the early 2000s, most likely from southern Asia.

Aside from the Torres Strait Islands isolates, the only non-PNG Oceania isolate of *S. flexneri* to have WGS and phylogenetic analysis conducted belongs to PG1. The Fijian and PNG isolates in PG1 share a recent MRCA, and are estimated to have been introduced into the Oceania region sometime in the 1990s. However, determining their epidemiological links is not easy as these Oceania isolates cluster in a 'global' clade, alongside isolates from Asia and Africa.

S. sonnei forms four distinct phylogenetic lineages, with a common ancestor originating in European countries before spreading to non-European countries (Holt *et al*., 2012). Two of the four lineages, LII and LIII, have reached Oceania.

There appears to be geographical separation of *S. sonnei* lineages in Oceania. Isolates detected in Solomon Islands, Vanuatu, Samoa and Fiji all belong to *S. sonnei* LII, with an estimated date of introduction to Oceania of 1986. These Oceania isolates are closely related to an African isolate; however, that does not necessarily prove that they originated in Africa, as they share a common ancestor with isolates from Central America, Europe, Asia and Africa. All *S. sonnei* isolates from PNG are LIII, and are related to two South American isolates (from French Guiana). Again, this does not necessarily indicate the country of origin, and the PNG isolates share a common ancestor with isolates from the Middle East and Asia. Interestingly, one New Caledonia isolate clusters in this clade. This clade was introduced to Oceania around 1984. Thus, the two clades entered Oceania at approximately the same time, but some countries harbour LII strains and other countries harbour LIII strains.

6.4.1 *Shigella flexneri* serotyping using WGS data

Genetic serotyping data support previous findings in Oceania which were determined using traditional serotyping. In this study 43 of the 47 *S. flexneri* isolates examined by WGS could be genetically serotyped; the serotypes detected (namely serotypes 1a, 1b, 2a, 2b and 3a) are consistent with the serotype distribution reported by by Rosewell *et al.* (2010) in PNG and Storch *et al*. (1980) in the Marshall Islands (i.e. serotypes 1, 1b and 4a). Importantly, serotypes determined using WGS data matched traditional serotyping results for 11 of 12 (92%) of the isolates that had been serotyped using both approaches (to type antigen level, but not group antigen level). Most of the isolates serotyped traditionally were incomplete, with their group antigen (e.g. a or b) not identified (e.g. the traditional method determined the serotype as being 1, but the WGS approach determined it as 1b). One isolate had been assigned to serotype X using traditional serotyping;

however, WGS data predicted that isolate to be serotype 2b. Reasons for why the conventional serotyping method was incorrect could be due to nonspecific cross-reactions, however human error can not be ruled out. Serotype X has not been commonly isolated from Oceania, whereas serotype 2b is expected in this setting. This study suggest that serotyping using WGS data may well be superior to traditional serotyping.

In recent years methods have been developed that enable serotyping to be conducted through genetic analysis for numerous bacterial pathogens. Early genetic serotyping was commonly based on multiplex PCR, but as the cost of WGS has come down, sequence data is now commonly used for serotyping. This scenario is true for *S. flexneri*, for which a suite of PCRs were developed to enable serotyping (Sun *et al*., 2011). Clearly, whole genome data can be used to determine the presence of these target genes, which encode O and H antigens, and serotype-specific genes. This is advantageous because typing sera is costly to produce and transport. It often requires cold-chain transportation, which can be a challenge in low-income settings. It is also prone to contamination and deterioration over time. Only 19% of *S. flexneri* isolates used in this study had been serotyped prior to this study commencing, reflecting in part some of the aforementioned challenges of traditional serotyping.

Historically serotyping has been an important epidemiological tool, though its ongoing importance is difficult to foreshadow with the broader application of WGS. For *S. flexneri* analysed in the current study, serotypes 1a and 1b sit within PG1, serotypes 2a and 2b within PG3 and serotype 3a in PG2. This information potentially provides some insights into historical strains circulating, that have not been subjected to WGS (and whose isolates are unlikely to have been preserved). Serotype 2 has been shown to be the most

commonly detected serotype in a study conducted in the mid-1960s (Morahan, 1968), and also in data analysis conducted on isolates obtained in 2000-9 (Rosewell *et al*., 2010). Based on the findings of Connor *et al*. (2015) and this finding, where serotypes 2a and 2b lie within PG3, it appears that PG3 may be the historical and current predominant phylogenetic group in PNG.

The correlation of serotype and phylogenetic group suggests there could be ongoing utility, if a context exists where WGS analysis is not feasible but serotyping can be readily conducted.

6.4.2 Virulence genes and AMR genes

Detection of AMR genes and its correlation with phenotypic resistance has been addressed in detail in Chapter 5. Despite poor correlation, the detection of AMR genes has public health implications; thus a comparison of AMR genes from Oceania relative to Asian and global isolates was conducted. There was a significantly higher prevalence of some resistance genes in the Oceania/PNG *S. flexneri* isolates compared to Asian and global isolates, whereas for other genes the prevalence was lower in Pacific/PNG isoaltes than in global isoaltes. Almost all *S. flexneri* isolates from the Pacific region (at least 44/47) harboured *AMPH_Ecoli, AmpC2_Ecoli, oxa1, penicillin_binding* (beta-lactams), *CatA1* (phenicol) and *tetB* (tetracycline) resistance genes. The high rate of carriage of these AMR genes in PNG isolates likely reflects the high usage of antibiotics such as beta-lactams, chloramphenicol and tetracycline in PNG.

Of particular interest are the *AMPC1_Ecoli* and *oxa1* genes conferring resistance to beta-lactam antibiotics. The *AMPC1_Ecoli* gene was present in only 11 of 391 (3%) global

isolates of *S. flexneri*; of which one was from Asia and the remaining 10 from Oceania (eight from PNG and one each from from Fiji and the Torres Strait Islands in Australia). There was no specific trend in the presence of the *AMPC1_Ecoli* gene within clades of *S. flexneri* in PNG/Pacific isolates; i.e the gene has not been introduced in a single isolate and remained present in all subsequent decendents within that linage. As this gene has been reported to be present in plasmids (as well as on the chromosome) it appears likely that the gene is picked up from closely related species within the Enterbacteriaceae family within the Pacific.

The *oxa1* gene was observed sporadically throughout *S. flexneri* and *S. sonnei* isolates (Figure 6.9 and Figure 6.11). In this study the gene is present in less than half all global isolates of *S. flexneri* but is present in >90% of Pacific isolates; and for *S. sonnei* the *oxa1* gene was present in 7% of global iosaltes investigated, but 47% of Pacific isolates. This gene is usually present in an integron or plasmid, thus may not be retained if not a competitive advantage to the bacterial host. It is probable that this AMR gene has been introduced into PNG from Asian isolate(s), and there is specific selective pressure to retain the *oxa1* gene in Pacific isolates due to the high rate of beta-lactam antibiotic use in the region.

The Oceania/PNG *S. sonnei* isolates from this study had a significantly higher prevalence of selected genes encoding for resistance to beta-lactams, trimethoprim, tetracycline and sulphonamide compared to the Asian and global isolates. Most Pacific and PNG isolates harbour small MDR plasmid spA[2] containing the *tetAR*, *strAB* and *sul2* genes that encode for resistance to tetracycline, streptomycin and sulphonamides. The study by Holt *et al*.

(2012) reported that these genes are usually found in both LII and LIII isolates of *S. sonnei*, the two lineages present in Oceania.

It is not surprising that there is high resistance to chloramphenicol, trimethoprim and tetracycline observed in the Oceania region, particularly in PNG. Over prescription of antimicrobials in PNG is likely, although conclusive evidence of over-prescription is limited. One recent study revealed that patients with common diseases such as respiratory tract infections, diarrhoea and general body aches are commonly prescribed antibiotics. In short, 73% of the drugs prescribed for such illnesses are antibiotics (Saweri, Hetzel, Mueller, Siba, & Pulford, 2017). The presence of genes linked to resistance to the above antibiotics was commonly detected in Oceania isolates, often at a higher prevalence than those of international isolates analysed in this study. This likely reflects the selction pressure on *Shigella* isoaltes in the Pacific region, and ther eis some evidence ot suggest that Shigella are 'picking up' some of the resistance genes in country (e.g. multiple incursions of *S. flexneri* with no clear correlation between phylogeny and the presence of the aforementioned genes).

The presence of other AMR genes was variable, and thus in agreement with the findings of Connor *et al*. (2015), who observed AMR gene distribution to be variable across and within PGs for *S. flexneri*.

The benefits of applying WGS rather than targeted PCR assays to detect virulence genes has been addressed in Chapter 5. Here, the detection of virulence genes as a proxy for phylogenetic analysis is considered. *S. flexneri* virulence genes coding for the SHI-1 pathogenicity island were predominantly present in PG3 isolates. SHI-1 is considered a

distinguishing feature of serotype 2a (Connor et al., 2015), a serotype that commonly belong to PG3 (as discussed above), SHI-1 is thought to carry three key virulence associated genes: *Pic* and *SigA*, which were present in 74% (35 of 47) of Oceania *S. flexneri* isolates; and ShET-1 (*set1A*) in 68% (32 of 47) of isolates. Moreover, in *S. sonnei*, the SHI-1 pathogenicity island virulence associated genes *Pic* and *SigA* were present in all 16 isolates (in both LII and LIII). Thus, there are some correlations between virulence genes and phylogeny, albeit limited.

Although statistical analyses were not conducted, there appears to be a weak relationship between the presence of virulence genes and phylogeny. On this basis, analyses conducted in Chapter 4 are of some value in determining which virulence genes are present, and the potential for more or less severity of illness based on the presence or absence of virulence genes should not be discounted. Each virulence associated protein is responsible for, or contributes towards, a particular clinical manifestation of *Shigella* infection (Connor *et al*., 2015). However, the presence or absence of virulence genes is a poor indicator of genetic relatedness of *Shigella* isolates.

This study clearly demonstrates the utility of WGS of *Shigella* from resource-poor settings. The application of WGS can both complement traditional epidemiological and antimicrobial susceptibility data; and greatly expand our understanding of the pathogen. The utility of WGS of *Shigella* will be considered further in Chapter 7.

Chapter 7 - General Discussion

The focus of this thesis was to enhance our understanding of *Shigella*, an important cause of diarrhoeal disease in the Oceania region. With limited data on the epidemiology of *Shigella* in the region, understanding social determinants, AMR, virulence trends and phylogenetics of isolates circulating in the region is imperative.

Social determinants have long been important contributing factors towards the burden of diarrhoeal disease. In PNG, where ≥80% of the population reside in rural or urban poor settings, access to WASH is often lacking (PNG National Department of Health, 2010). The work presented in Chapter 3 reiterates the challenges faced in PNG regarding access to WASH, and how that may impact on health outcomes. On the basis of the findings in presented in Chapter 3, safe and reliable water sources and adequate disposal of faecal waste is not available, or is only sporadically available, to many people living in EHP. It is conceivable that people living in other rural and regional locations within Oceania face similar challenges.

The access to WASH observed in PNG likely represent similar scenarios in rural settings around the Pacific, especially in countries such as the Solomon Islands and Vanuatu where an estimated of 77% and 74% of the population lives rurally, respectively (Trading Economics, 2019a, 2019b). Access to WASH is higher in most other Oceania countries than in PNG, but is still sufficiently low to contribute to the transmission of infectious diseases. For example , access to basic 'safe' drinking water in rural Solomon Islands has decreased between 2000 and 2015, and there was lower access to basic sanitation in Vanuatu in 2015 than in 2000 (https://washdata.org/ and refer to Appendix 2.2). In brief,

the conditions are favourable for the transmission of *Shigella* and other gastrointestinal pathogens through much of Oceania.

Given the slow progress in improved WASH in many low- and middle-income countries (including in Oceania), prevention and treatment of infectious diseases needs to occur through means in addition to (and not limited to) WASH. This is particularly important for diarrhoeal diseases such as *Shigella*, which remain an important public health threat globally. A better understanding of important pathogens can improve health outcomes, and this approach is commonly taken for other important causes of infectious diseases in the region such as *Mycobacterium tuberculosis* and *Plasmodium* spp. However, our knowledge beyond basic epidemiology of diarrhoeal pathogens in Oceania is limited; be it at the country level or at the regional level. The molecular epidemiology and phylogenetics of the 2009-11 cholera outbreak in PNG was documented (Greenhill *et al*., 2019; Horwood *et al*., 2011), but that work was reactionary following a sizable cholera outbreak that resulted from the maiden incursion of *V. cholerae* in PNG. The work presented in this thesis represents one of the largest studies conducted on a gastrointestinal pathogen in the Oceania region.

The findings of this study suggest that the population dynamics of the two most important species of *Shigella*, namely *S. flexneri* and *S. sonnei,* reflects global trends. Multiple incursions of *S. flexneri* have occurred over the past 4 decades, with strains from at least some of these incursions continuing to circulate. This is consistent with *S. flexneri* distribution at the global level (Connor *et al*, 2015). By comparison, there is some evidence to support the notion of a single lineage of *S. sonnei* in PNG. The small number of isolates from other countries in Oceania (Fiji, Vanuatu and Solomon Islands) point to

the same occurrence in those countries, albeit a different lineage to that circulating in PNG and New Caledonia. Comparing distribution patterns of *Shigella* in Oceania to global distribution and incursion is more than just an exercise of interest: public health experts can look to global trends and use those trends to predict what might occur in Oceania.

On the basis of the phylogenetic data obtained in this study, it appears that new incursions of *Shigella* from Asia are at greatest risk of occurring. Unlike Asia (Chung The *et al*., 2016), there has yet to be a report of ciprofloxacin resistance in *Shigella* spp. in Oceania countries. Given the close proximity of countries in Oceania to Asia, and increased movement of people between countries in the Asia-Pacific region, it may be only a matter of time before ciprofloxacin resistant *Shigella* makes its way into the Oceania. While preventing such an incursion may be difficult, being able to rapidly respond and potentially curtail the spread of incoming AMR strains could be achievable with adequate laboratory and epidemiological capacity.

Even if Oceanic countries could protect themselves from future incursions of *S. flexneri* and *S. sonnei* (which would take greater biosecurity measures than currently available), there remains a risk of increasing antimicrobial resistance. This study demonstrated a rise of AMR to ampicillin, co-trimoxazole and nalidixic acid in *Shigella* spp. (Chapter 4) and a corresponding high level of AMR genes (e.g. *AMPH_Ecoli*, *AmpC1_Ecoli*, *AmpC2_Ecoli*, *oxa-1*, *tem-1d*, *Penicillin_Binding protein*, *aadA*, *strA*, *strB*, *catA1*, *dfrA1*, *dfrA5*, *tetB*, *sulII* etc.) relative to global strains (Chapter 6). Observation of AMR to the above antimicrobials is not surprising given treatment of suspected bacterial infection in Oceanic countries such as PNG is often based on clinical presentation, in the absence of

laboratory or point-of-care diagnostics (PNG National Department of Health, 2012; PNG Paediatric Society, 2011). This leads to the misuse or overuse of antimicrobials, which is a driving force in the development of AMR in the region (Foxlee, Townell, McIver, & Lau, 2019; Holmes *et al*., 2016; Saweri, Hetzel, Mueller, Siba, & Pulford, 2017; Williams & Berkley, 2018; World Health Organization, 2005). Various other factors likely contribute to increasing AMR in the Oceania region. Self-medication and the purchasing of less expensive treatments (of unknown quality and often insufficient quantity) from the general market is thought to be commonplace. Additionally, the remoteness of rural communities, limited resources and fragile health infrastructure make it difficult for government agencies and healthcare providers to properly regulate supply chains (Foxlee *et al*., 2019; Hetzel *et al*., 2014). A widespread and proactive approach to data collection and analysis of AMR in *Shigella* and other important pathogens is required in the Oceania region.

Utilising CLSI and WGS methods to determine AMR, this study demonstrated that resistance to antibiotics in *Shigella* isolates from Oceania is common, and potentially are becoming more prevalent. Collectively, *Shigella* isolates from Oceania isolated in recent years are more commonly resistant to first-line antibiotics than were bacteria isolated prior to 2010. Moreover, both *S. flexneri* and *S. sonnei* had a significantly higher prevalence of resistance genes to beta-lactams, trimethoprim, tetracyclines and sulphonamides in comparison to their global and Asian counterparts. The presence of genetic determinants of resistance correlate with phenotypic resistance. *Shigella* isolates, particularly from PNG, are commonly resistant to ampicillin, tetracycline, co-trimoxazole and chloramphenicol, with similar trends observed in at least one other Oceanic country, namely Fiji (Foxlee *et al*., 2019).

The study detected the presence of virulence genes among *S. flexneri* isolates and *S. sonnei* isolates, with *S. flexneri* isolates harbouring the majority of the *Shigella* enterotoxin genes included in this study (i.e. *set1A*/*set1B* and *sen/ospD3*). The true significance of the presence/absence of virulence genes is difficult to ascertain, as some isolates may have lost mobile genetic elements during culture (for initial isolation and characterisation), long-term storage and/or re-culture during the current study. Nonetheless, this work does raise avenues for future work. Assuming the outcomes of Chapter 4 accurately represent temporal prevalence of virulence genes, it raises the question of whether there are public health implications of *Shigella* isolates with more virulence genes circulating in Oceania. A previous study by Lluque *et al*. (2015) in South America found a similar occurrence, with heterogeneity in the genes present among *S. flexneri* isolates. The relevance of the finding could be determined by a case-control study, or even the collection of excellent clinical histories as part of enhanced surveillance.

The challenge of how best to diagnose infectious diseases is very real, be it in low-, middle- or high-income settings. With the advent of molecular biology, in particular PCR, the use of culture as a diagnostic method has long been thought by many to be headed for extinction. However, at this point of the 21st century there remains a high dependency on culture and sensitivity for diagnosis of bacterial and fungal infections. Nonetheless, current molecular diagnostic techniques such as conventional and/or real-time PCR can greatly improve the diagnosis of shigellosis and other diarrhoeal disease in developing, resource-poor countries, (Lindsay *et al*., 2013; Liu *et al*., 2016). Moreover, such methods can contribute to the characterisation of *Shigella* spp. and other pathogens. As with all diagnostic methods, there are challenges associated with PCR and related technologies,

some of which are addressed below. Even accounting for challenges and shortcomings, it is likely that a rapid and expansive uptake of molecular biology in clinical diagnostic laboratories in low-income settings could greatly improve infectious disease diagnosis.

The approach gaining perhaps the most attention for its perceived potential is the use of metagenomic sequencing as a diagnostic tool. In theory, metagenomic approaches can detect any pathogen present in a diagnostic sample, rather than targeting specific pathogens, as is the case with (real-time) PCR and culture. Metagenomic sequencing has been applied as a proof of principle to various diagnostic specimens, including diarrhoeal samples (Zhou *et al*, 2016; Joensen *et al*, 2017), but is yet to have widespread application for the diagnosis of any infectious diseases. Numerous reviews highlight the potential benefits of metagenomic sequencing as a diagnostic tool. However, Greninger (2018) provides perhaps the most grounded appraisal by acknowledging that there are numerous challenges associated with cost, interpretation (impacted by laboratory workflows) and turnaround time that need to be addressed before metagenomic sequencing replaces currently used approaches, even in well-resourced settings.

A large focus of this study was the application of WGS to better understand the epidemiology and AMR of *Shigella* from low and middle-income countries in Oceania. WGS has the capacity to provide detailed characterisation of microbial pathogens and insights into their evolutionary lineages (Franco-Duarte *et al*., 2019), and has been applied to almost all recognised bacterial pathogens. For *Shigella*, as with other bacterial pathogens, WGS can reveal which virulence genes are present, and which genetic determinants of AMR are present (Papaventsis *et al*., 2017; Sadouki *et al*., 2017), as was applied in this study. WGS allows for observation of vertical and horizontal gene transfer

such as transfer of mobile genetic elements in virulence/AMR plasmids and chromosomal islands which will be important in monitoring disease outbreaks caused from AMR strains of *Shigella* (Franco-Duarte *et al*., 2019; Papaventsis *et al*., 2017; Sadouki *et al*., 2017). There is little doubt that the emergence of antimicrobial resistance needs further scrutiny.

Despite the positive implications of WGS, there are some aspects of traditional methods that outweigh molecular methods. For instance, phenotypic AMR testing, despite limitations of its own, more accurately reflect resistance to antibiotics, whereas WGS and PCR which detect the presence of resistance genes are detecting the potential for resistance. In this study poor concordance between phenotypic resistance and genetic markers for resistance was observed. Similarly, Doyle *et al*. (2019) observed discordant bioinformatic predictions across various clinical microbiology laboratories. This held true for various bacterial isolates to an array of antimicrobials, when compared to the use of phenotypic methods. These discrepancies could well be overcome in due course, and overall there is no doubt that WGS is a valuable tool in clinical microbiology.

To date, WGS is limited to well-resourced laboratories, and still requires considerable technical capability. For instance, sufficient DNA concentrations are required for optimal sequence outputs, and library preparation and maintenance of sequencers remains complex (Deurenberg *et al*., 2017; Koser *et al*., 2012). Thereafter, data generated from the next generation sequencing needs to be analysed using bioinformatic platforms. To do so requires access to reliable and fast internet services; however, most developing countries such as PNG and other countries in Oceania have at times unreliable and slow internet (Deurenberg *et al*., 2017; Koser *et al*., 2012). An advantage of PCR over WGS is that it is considerably cheaper and it is easier to set up and analyse in resource-poor settings

compared to WGS (though the need for sufficient quality template is true for both PCR and WGS).

There are many challenges that need to be considered for the implementation and maintenance of next generation WGS methods for diagnosis and/or pathogen characterisation in low resource settings like PNG and similar settings around Oceania. Essential facilities such as access to fast internet and telecommunication services, reliable and consistent electricity supply are often lacking in these settings. In addition, the maintenance and upkeep of WGS equipment and accessories is often expensive and requires frequent support from service technicians. Finally, experienced and skilled laboratory technicians are required to conduct sequencing runs. There are two potential approaches that might be suitable for the Oceania region, as addressed below.

One approach would be to collaborate with other well equipped diagnostic and/or research laboratories in the region. There are numerous examples of one or more laboratories in the region acting as a regional reference centre for various bacterial and viral pathogens in Oceania, perhaps most notably influenza virus. However, there are well-credentialed laboratories that commonly support surveillance of other pathogens such as rotavirus, *Mycobacterium tuberculosis* and *Streptococcus pneumoniae*, to name but three. A reference laboratory, or a network of laboratories, to support surveillance of gastrointestinal pathogens such as (though not necessarily limited to) *Shigella* warrants consideration.

A second approach, perhaps in concert with the above approach, would be to explore technologies such as the MinION Nanopore sequencing platform. The MinION is a

portable sequencing unit that is powered through a computer USB port, and works by detecting the electric current of single-stranded DNA as it passes through the small protein channels (nanopores) of the disposable flow cells (Loman & Watson, 2015; Wei, Weiss, & Williams, 2018). Open source software has been developed and is undergoing constant refinement which allows the MinION system to provide identification of microorganisms and detection of virulence genes and AMR genes almost in real time. This system is considered to have great potential for field based studies and resource limited settings, and can be applied to both WGS and metagenomic detection of pathogens. However, many of the challenges that exist for commonly used (Illumina) sequencing still apply to MinION technology, namely the requirement for high-quality DNA (arguably more important in nanopore sequencing due to the long reads), technical expertise, and good internet access.

There were a number of limitations to this study. First, when acquiring isolates for analysis, there was no established network for systematic isolation and storage of *Shigella* spp. In the Oceania region. Consequently, a rather piecemeal approach was taken, with *Shigella* isolates acquired from a number of our collaborators. Hence, there was an under-representation of isolates from the Oceania/Pacific region, particularly from outside of PNG. Ideally more *Shigella* isolates acquired from different areas of the Oceania region (apart from PNG) would give us a better representation of the burden of *Shigella* in the region.

Secondly, there is a need to expand AMR testing. Only seven antibiotics were used in this study, and important antibiotics were omitted, such amoxicillin and gentamicin (both recommended for treatment of severe diarrhoea in children in PNG), doxycycline and

tigecycline. Additionally, minimum inhibitory concentrations could be determined to both confirm disc diffusion results and provide data of extent of resistance to a given antibiotic.

Thirdly, there were some minor challenges associated with some of the laboratory. For examples, the conventional PCR results for the *Shigella* virulence genes *set1A* and *set1B* provided results that differed from WGS data. These discrepancies illustrate the challenges associated with laboratory work, and are not dissimilar in magnitude to what might be expected in any laboratory where such work is conducted. Notwithstanding the above shortcomings, the study has shed new light on a long-standing health problem.

In summary, this study sets a foundation for providing the means to further investigate methods to better manage and treat diarrhoeal diseases in the Oceania region, with a focus on one of the most important aetiological agents, *Shigella*. For example, setting up a system for countries in the Oceania region to isolate, characterise and archive *Shigella* isolates detected in their respective countries would be ideal. Clearly, the establishment of such a network would present with a number of logistical and political issues. For instance, where would the samples be processed and stored? Where would the WGS data generated from these samples be stored and who would be granted access to them? What funding sources could be accessed to support such an undertaking? These issues would need to be discussed and agreed upon by the governing bodies. However, data generated from such a study could provide a means to help prevent diarrhoeal diseases. This is especially important for pathogens such as *Shigella*, for which no vaccine is currently available, but work is being conducted to develop one. If future *Shigella* vaccines are

serotype specific, a comprehensive understanding of circulating strains will be imperative to ensure vaccines can be effective.

References

Anandan, S., Muthuirulandi Sethuvel, D. P., Gajendiren, R., Verghese, V. P., Walia, K., & Veeraraghavan, B. (2017). Molecular characterization of antimicrobial resistance in clinical *Shigella* isolates during 2014 and 2015: trends in South India. *Germs, 7*(3), 115-122. doi:10.18683/germs.2017.1116

Ashkenazi, S., May-Zahav, M., Dinari, G., Gabbay, U., Zilberberg, R., & Samra, Z. (1993). Recent trends in the epidemiology of *Shigella* species in Israel. *Clin Infect Dis, 17*(5), 897-899.

Bainomugisa, A., Pandey, S., Donnan, E., Simpson, G., Foster, J., Lavu, E., . . . Coulter, C. (2019). Cross-Border Movement of Highly Drug-Resistant *Mycobacterium tuberculosis* from Papua New Guinea to Australia through Torres Strait Protected Zone, 2010-2015. *Emerg Infect Dis, 25*(3), 406-415. doi:10.3201/eid2503.181003

Barrington, D. J. (2016). Inadequate water, sanitation and hygiene in the South Pacific: how might it be impacting children? *Rev Environ Health, 31*(1), 159-162. doi:10.1515/reveh-2015-0034

Barrington, D. J., Sridharan, S., Saunders, S. G., Souter, R. T., Bartram, J., Shields, K. F., . . . Hughes, R. K. (2016). Improving community health through marketing exchanges: A participatory action research study on water, sanitation, and hygiene in three Melanesian countries. *Soc Sci Med, 171*, 84-93. doi:10.1016/j.socscimed.2016.11.003

Benny, E., Mesere, K., Pavlin, B. I., Yakam, L., Ford, R., Yoannes, M., . . . Horwood, P. F. (2014). A large outbreak of shigellosis commencing in an internally displaced population, Papua New Guinea, 2013. *Western Pac Surveill Response J, 5*(3), 18-21. doi:10.5365/WPSAR.2014.5.2.003

Bergsbaken, T., Fink, S. L., & Cookson, B. T. (2009). Pyroptosis: host cell death and inflammation. *Nat Rev Microbiol, 7*(2), 99-109. doi:10.1038/nrmicro2070

Bhattacharya, D., Bhattacharya, H., Sayi, D. S., Bharadwaj, A. P., Singhania, M., Sugunan, A. P., & Roy, S. (2015). Changing patterns and widening of antibiotic resistance in *Shigella* spp. over a decade (2000-2011), Andaman Islands, India. *Epidemiol Infect, 143*(3), 470-477. doi:10.1017/s0950268814000958

Bhattacharya, D., Bhattacharya, H., Thamizhmani, R., Sayi, D. S., Reesu, R., Anwesh, M., . . . Roy, S. (2014). Shigellosis in Bay of Bengal Islands, India: clinical and seasonal patterns, surveillance of antibiotic susceptibility patterns, and molecular characterization of multidrug-resistant Shigella strains isolated during a 6-year period from 2006 to 2011. *Eur J Clin Microbiol Infect Dis, 33*(2), 157-170. doi:10.1007/s10096-013-1937-2

Bliss, J., Bouhenia, M., Hale, P., Couturier, B. A., Iyer, A. S., Rumunu, J., . . . Leung, D. T. (2018). High prevalence of *Shigella* or Enteroinvasive *Escherichia coli* carriage among residents of an internally displaced persons camp in South Sudan. *Am J Trop Med Hyg, 98*(2), 595-597. doi:10.4269/ajtmh.17-0339

Broach, W. H., Egan, N., Wing, H. J., Payne, S. M., & Murphy, E. R. (2012). VirF-independent regulation of *Shigella* virB transcription is mediated by the small RNA RyhB. *PLoS One, 7*(6), e38592. doi:10.1371/journal.pone.0038592

Cairncross, S., Hunt, C., Boisson, S., Bostoen, K., Curtis, V., Fung, I. C., & Schmidt, W. P. (2010). Water, sanitation and hygiene for the prevention of diarrhoea. *Int J Epidemiol, 39 Suppl 1*, i193-205. doi:10.1093/ije/dyq035

Casabonne, C., Gonzalez, A., Aquili, V., & Balague, C. (2016). Prevalence and virulence genes of *Shigella* spp. isolated from patients with diarrhea in Rosario, Argentina. *Jpn J Infect Dis, 69*(6), 477-481. doi:10.7883/yoken.JJID.2015.459

Chen, L., Yang, J., Yu, J., Yao, Z., Sun, L., Shen, Y., & Jin, Q. (2005). VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res, 33*(Database issue), D325-328. doi:10.1093/nar/gki008

Chung The, H., Rabaa, M. A., Pham Thanh, D., De Lappe, N., Cormican, M., Valcanis, M., . . . Baker, S. (2016). South Asia as a Reservoir for the Global Spread of Ciprofloxacin-Resistant *Shigella sonnei*: A Cross-Sectional Study. *PLoS Med, 13*(8), e1002055. doi:10.1371/journal.pmed.1002055

CLSI. (2015). *Clinical and Laboratory Standards Institute. Performance Standards for Antimicrobial Susceptibility Testing: 25th Informational Supplement. CLSI Document M100-S25.* Wayne, Pennsylvania, USA: Clinical and Laboratory Standards Institute.

Connor, T. R., Barker, C. R., Baker, K. S., Weill, F. X., Talukder, K. A., Smith, A. M., . . . Thomson, N. R. (2015). Species-wide whole genome sequencing reveals historical global spread and recent local persistence in *Shigella flexneri*. *Elife, 4*, e07335. doi:10.7554/eLife.07335

Curtis, P. G. (1964). The isolation, incidence and sensitivity of *Shigella* organisms. *P N G Med J, 7*(1), 23-26.

da Cruz, C. B., de Souza, M. C., Serra, P. T., Santos, I., Balieiro, A., Pieri, F. A., . . . Orlandi, P. P. (2014). Virulence factors associated with pediatric shigellosis in Brazilian Amazon. *Biomed Res Int, 2014*, 539697. doi:10.1155/2014/539697

Department of Health and Human Services. (2011). The blue book. Retrieved from https://www2.health.vic.gov.au/about/publications/researchandreports/The%20blue%20book

Deurenberg, R. H., Bathoorn, E., Chlebowicz, M. A., Couto, N., Ferdous, M., Garcia-Cobos, S., . . . Rossen, J. W. (2017). Application of next generation sequencing in

clinical microbiology and infection prevention. *J Biotechnol, 243*, 16-24. doi:10.1016/j.jbiotec.2016.12.022

Dhital, S., Sherchand, J. B., Pokharel, B. M., Parajuli, K., Mishra, S. K., Sharma, S., . . . Rijal, B. (2017). Antimicrobial susceptibility pattern of *Shigella* spp. isolated from children under 5 years of age attending tertiary care hospitals, Nepal along with first finding of ESBL-production. *BMC Res Notes, 10*(1), 192. doi:10.1186/s13104-017-2512-1

Doyle, R. M., O'Sullivan, D. M., Aller, S. D., Bruchmann, S., Clark, T., Pelegrin, A. C., . . . Harris, K. A. (2019). Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: An inter-laboratory study. *bioRxiv*, 793885. doi:10.1101/793885

Doyle, R. M., O'Sullivan, D. M., Aller, S. D., Bruchmann, S., Clark, T., Coello Pelegrin, A., . . . Harris, K. A. (2020). Discordant bioinformatic predictions of antimicrobial resistance from whole-genome sequencing data of bacterial isolates: an inter-laboratory study. *Microb Genom, 6*(2). doi:10.1099/mgen.0.000335

Duke, T. (2011). Zinc sulphate for treatment and prevention of diarrhoea and other conditions in children in Papua New Guinea. *P N G Med J, 54*(1-2), 17-22.

DuPont, H. L., Levine, M. M., Hornick, R. B., & Formal, S. B. (1989). Inoculum size in shigellosis and implications for expected mode of transmission. *J Infect Dis, 159*(6), 1126-1128.

Enne, V. I., Delsol, A. A., Roe, J. M., & Bennett, P. M. (2006). Evidence of antibiotic resistance gene silencing in *Escherichia coli*. *Antimicrob Agents Chemother, 50*(9), 3003-3010. doi:10.1128/aac.00137-06

Faherty, C. S., Wu, T., Morris, C. R., Grassel, C. L., Rasko, D. A., Harper, J. M., . . . Barry, E. M. (2016). The synthesis of OspD3 (ShET2) in *Shigella flexneri* is independent of OspC1. *Gut Microbes, 7*(6), 486-502. doi:10.1080/19490976.2016.1239682

Fan, W., Qian, H., Shang, W., Ying, C., Zhang, X., Cheng, S., . . . Ma, P. (2017). Low distribution of genes encoding virulence factors in *Shigella flexneri* serotypes 1b clinical isolates from eastern Chinese populations. *Gut Pathog, 9*, 76. doi:10.1186/s13099-017-0222-9

Faruque, S. M., Khan, R., Kamruzzaman, M., Yamasaki, S., Ahmad, Q. S., Azim, T., . . . Sack, D. A. (2002). Isolation of *Shigella dysenteriae* type 1 and *S. flexneri* strains from surface waters in Bangladesh: comparative molecular analysis of environmental Shigella isolates versus clinical strains. *Appl Environ Microbiol, 68*(8), 3908-3913.

Fischer Walker, C. L., Perin, J., Aryee, M. J., Boschi-Pinto, C., & Black, R. E. (2012). Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. *BMC Public Health, 12*, 220. doi:10.1186/1471-2458-12-220

Fleming, L., Anthonj, C., Thakkar, M. B., Tikoisuva, W. M., Manga, M., Howard, G., . . . Bartram, J. (2019). Urban and rural sanitation in the Solomon Islands: How resilient are these to extreme weather events? *Sci Total Environ, 683*, 331-340. doi:10.1016/j.scitotenv.2019.05.253

Foxlee, N. D., Townell, N., McIver, L., & Lau, C. L. (2019). Antibiotic resistance in Pacific Island countries and territories: A Systematic Scoping Review. *Antibiotics (Basel), 8*(1). doi:10.3390/antibiotics8010029

Franco-Duarte, R., Cernakova, L., Kadam, S., Kaushik, K. S., Salehi, B., Bevilacqua, A., . . . Rodrigues, C. F. (2019). Advances in Chemical and Biological Methods to

Identify Microorganisms-From Past to Present. *Microorganisms, 7*(5). doi:10.3390/microorganisms7050130

Germani, Y., Morillon, M., Begaud, E., Dubourdieu, H., Costa, R., & Thevenon, J. (1994). Two-year study of endemic enteric pathogens associated with acute diarrhea in New Caledonia. *J Clin Microbiol, 32*(6), 1532-1536.

Ghosh, S., Pazhani, G. P., Niyogi, S. K., Nataro, J. P., & Ramamurthy, T. (2014). Genetic characterization of *Shigella* spp. isolated from diarrhoeal and asymptomatic children. *J Med Microbiol, 63*(Pt 7), 903-910. doi:10.1099/jmm.0.070912-0

Gibney, K. B., Cheng, A. C., Hall, R., & Leder, K. (2017). Sociodemographic and geographical inequalities in notifiable infectious diseases in Australia: a retrospective analysis of 21 years of national disease surveillance data. *Lancet Infect Dis, 17*(1), 86-97. doi:10.1016/s1473-3099(16)30309-7

Gomez-Duarte, O. G., Bai, J., & Newell, E. (2009). Detection of *Escherichia coli*, *Salmonella* spp., *Shigella* spp., *Yersinia enterocolitica*, *Vibrio cholerae*, and *Campylobacter* spp. enteropathogens by 3-reaction multiplex polymerase chain reaction. *Diagn Microbiol Infect Dis, 63*(1), 1-9. doi:10.1016/j.diagmicrobio.2008.09.006

Greenhill, A., Guwada, C., Siba, V., Michael, A., Yoannes, M., Wawarie, Y., . . . Horwood, P. F. (2014). Antibiotic resistant *Shigella* is a major cause of diarrhoea in the Highlands of Papua New Guinea. *J Infect Dev Ctries, 8*(11), 1391-1397. doi:10.3855/jidc.4396

Greenhill, A., Mutreja, A., Bulach, D., Belousoff, M. J., Jonduo, M. H., Collins, D. A., . . . Horwood, P. F. (2019). Wave 2 strains of atypical *Vibrio cholerae* El Tor caused the 2009-2011 cholera outbreak in Papua New Guinea. *Microb Genom, 5*(3). doi:10.1099/mgen.0.000256

Greenhill, A., Rosewell, A., Kas, M., Manning, L., Latorre, L., Siba, P., & Horwood, P. (2012). Improved laboratory capacity is required to respond better to future cholera outbreaks in Papua New Guinea. *Western Pac Surveill Response J, 3*(2), 30-32. doi:10.5365/WPSAR.2011.2.4.016

Greninger, A.L. (2018). The challenge of diagnostic metagenomics. *Expert Rev Mol Diagn*, *18*(7):605-615. doi: 10.1080/14737159.2018.1487292

Gu, B., Cao, Y., Pan, S., Zhuang, L., Yu, R., Peng, Z., . . . Tong, M. (2012). Comparison of the prevalence and changing resistance to nalidixic acid and ciprofloxacin of *Shigella* between Europe-America and Asia-Africa from 1998 to 2009. *Int J Antimicrob Agents, 40*(1), 9-17. doi:10.1016/j.ijantimicag.2012.02.005

Gunzburg, S., Gracey, M., Burke, V., & Chang, B. (1992). Epidemiology and microbiology of diarrhoea in young Aboriginal children in the Kimberley region of Western Australia. *Epidemiol Infect, 108*(1), 67-76.

Gupta, S. K., Padmanabhan, B. R., Diene, S. M., Lopez-Rojas, R., Kempf, M., Landraud, L., & Rolain, J. M. (2014). ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob Agents Chemother, 58*(1), 212-220. doi:10.1128/aac.01310-13

Hall, J. J., Gillespie, J. A., Rosewell, A., & Mapira, P. (2013). The Papua New Guinea cholera outbreak: implications for PNG, Australia and the Torres Strait. *Med J Aust, 199*(9), 576-577. doi:10.5694/mja13.10516

Hetzel, M. W., Page-Sharp, M., Bala, N., Pulford, J., Betuela, I., Davis, T. M., & Lavu, E. K. (2014). Quality of antimalarial drugs and antibiotics in Papua New Guinea: a survey of the health facility supply chain. *PLoS One, 9*(5), e96810. doi:10.1371/journal.pone.0096810

Heymann, D. (2014). *Control of Communicable Diseases Manual, 20th edition*

Holmes, A. H., Moore, L. S., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., . . . Piddock, L. J. (2016). Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet, 387*(10014), 176-187. doi:10.1016/s0140-6736(15)00473-0

Holt, K. E., Baker, S., Weill, F. X., Holmes, E. C., Kitchen, A., Yu, J., . . . Thomson, N. R. (2012). *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. *Nat Genet, 44*(9), 1056-1059. doi:10.1038/ng.2369

Holt, K. E., Thieu Nga, T. V., Thanh, D. P., Vinh, H., Kim, D. W., Vu Tra, M. P., . . . Baker, S. (2013). Tracking the establishment of local endemic populations of an emergent enteric pathogen. *Proc Natl Acad Sci U S A, 110*(43), 17522-17527. doi:10.1073/pnas.1308632110

Horwood, P., Barrington, D. J., & Greenhill, A. R. (2013). More resources need to be committed to the provision of safe water sources and sanitation in Papua New Guinea. *P N G Med J, 56*(3-4), 103-109.

Horwood, P., & Greenhill, A. (2012). Cholera in Papua New Guinea and the importance of safe water sources and sanitation. *Western Pac Surveill Response J, 3*(1), 3-5. doi:10.5365/WPSAR.2011.2.4.014

Horwood, P., Luang-Suarkia, D., Bebes, S., Boniface, K., Datta, S. S., Siba, P. M., & Kirkwood, C. D. (2012). Surveillance and molecular characterization of group A rotaviruses in Goroka, Papua New Guinea. *Am J Trop Med Hyg, 87*(6), 1145-1148. doi:10.4269/ajtmh.2012.12-0234

Horwood, P. F., Collins, D., Jonduo, M. H., Rosewell, A., Dutta, S. R., Dagina, R., . . . Greenhill, A. R. (2011). Clonal origins of *Vibrio cholerae* O1 El Tor strains, Papua New Guinea, 2009-2011. *Emerg Infect Dis, 17*(11), 2063-2065. doi:10.3201/eid1711.110782

Horwood, P. F., McBryde, E. S., Peniyamina, D., & Ritchie, S. A. (2018). The Indo-Papuan conduit: a biosecurity challenge for Northern Australia. *Aust N Z J Public Health, 42*(5), 434-436. doi:10.1111/1753-6405.12808

Horwood, P. F., Soli, K. W., Maure, T., Naito, Y. I., Morita, A., Natsuhara, K., . . . Greenhill, A. R. (2017). A High Burden of Asymptomatic Gastrointestinal Infections in Traditional Communities in Papua New Guinea. *Am J Trop Med Hyg, 97*(6), 1872-1875. doi:10.4269/ajtmh.17-0282

Howard, P., Alexander, N. D., Atkinson, A., Clegg, A. O., Gerega, G., Javati, A., . . . Alpers, M. P. (2000). Bacterial, viral and parasitic aetiology of paediatric diarrhoea in the highlands of Papua New Guinea. *J Trop Pediatr, 46*(1), 10-14.

Hoy, D., Roth, A., Viney, K., Souares, Y., & Lopez, A. D. (2014). Findings and implications of the Global Burden of Disease 2010 Study for the Pacific Islands. *Prev Chronic Dis, 11*, E75. doi:10.5888/pcd11.130344

IHME. (2013a). Fiji. Retrieved from http://www.healthdata.org/fiji

IHME. (2013b). Samoa. Retrieved from http://www.healthdata.org/samoa

IHME. (2013c). Solomon Islands. Retrieved from http://www.healthdata.org/solomon-islands

IHME. (2013d). Vanuatu. Retrieved from http://www.healthdata.org/vanuatu

Inouye, M., Dashnow, H., Raven, L. A., Schultz, M. B., Pope, B. J., Tomita, T., . . . Holt, K. E. (2014). SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med, 6*(11), 90. doi:10.1186/s13073-014-0090-6

International Monetary Fund. (2018). Regional Economic Outlook: Asia and Pacific. Retrieved from https://www.imf.org/en/Publications/REO/APAC/Issues/2018/04/16/areo0509

Islam, M. S., Hossain, M. A., Khan, S. I., Khan, M. N., Sack, R. B., Albert, M. J., . . . Colwell, R. R. (2001). Survival of *Shigella dysenteriae* type 1 on fomites. *J Health Popul Nutr, 19*(3), 177-182.

Jenkins, C. (1995). Changing hygiene behaviour in Papua New Guinea. *P N G Med J, 38*(4), 320-324.

Jin, Q., Yuan, Z., Xu, J., Wang, Y., Shen, Y., Lu, W., . . . Yu, J. (2002). Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Res, 30*(20), 4432-4441. doi:10.1093/nar/gkf566

Joensen, K. G., Tetzschner, A. M., Iguchi, A., Aarestrup, F. M., & Scheutz, F. (2015). Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. *J Clin Microbiol, 53*(8), 2410-2426. doi:10.1128/jcm.00008-15

Joensen, K. G., Engsbro, A.L., Lukjancenko, O.. Kaas, R.S., Lund, O., Westh, H., & Aarestrup, F. M. (2017). Evaluating next-generation sequencing for direct clinical diagnostics in diarrhoeal disease. *Eur J Clin Microbiol Infect Dis*, 36(7), 1325-1338. doi: 10.1007/s10096-017-2947-2

Kim, J. S., Kim, J. J., Kim, S. J., Jeon, S. E., Seo, K. Y., Choi, J. K., . . . Kim, J. (2015). Outbreak of ciprofloxacin-resistant *Shigella sonnei* associated with travel to Vietnam, Republic of Korea. *Emerg Infect Dis, 21*(7), 1247-1250. doi:10.3201/eid2107.150363

Kodish, S. R., Grey, K., Matean, M., Palaniappan, U., Gwavuya, S., Gomez, C., . . . Erasmus, W. (2019). Socio-ecological factors that influence infant and young child nutrition in Kiribati: a biocultural perspective. *Nutrients, 11*(6). doi:10.3390/nu11061330

Korff, J. (2015). Aboriginal population in Australia. Retrieved from http://www.creativespirits.info/aboriginalculture/people/aboriginal-population-in-australia

Koser, C. U., Ellington, M. J., Cartwright, E. J., Gillespie, S. H., Brown, N. M., Farrington, M., . . . Peacock, S. J. (2012). Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog, 8*(8), e1002824. doi:10.1371/journal.ppat.1002824

Koser, C. U., Ellington, M. J., & Peacock, S. J. (2014). Whole-genome sequencing to control antimicrobial resistance. *Trends Genet, 30*(9), 401-407. doi:10.1016/j.tig.2014.07.003

Kotloff, K. L., Nataro, J. P., Blackwelder, W. C., Nasrin, D., Farag, T. H., Panchalingam, S., . . . Levine, M. M. (2013). Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet, 382*(9888), 209-222. doi:10.1016/S0140-6736(13)60844-2

Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., Sansonetti, P. J., . . . Levine, M. M. (1999). Global burden of *Shigella* infections: implications for vaccine development and implementation of control strategies. *Bull World Health Organ, 77*(8), 651-666.

Kramer, A., Schwebke, I., & Kampf, G. (2006). How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. *BMC Infect Dis, 6*, 130. doi:10.1186/1471-2334-6-130

Lan, R., & Reeves, P. R. (2002). *Escherichia coli* in disguise: molecular origins of *Shigella*. *Microbes Infect, 4*(11), 1125-1132.

Li, S., Sun, Q., Wei, X., Klena, J. D., Wang, J., Liu, Y., . . . Tang, G. (2015). Genetic characterization of *Shigella flexneri* isolates in Guizhou Province, China. *PLoS One, 10*(1), e0116708. doi:10.1371/journal.pone.0116708

Lin, W. S., Cheng, C. M., & Van, K. T. (2010). A quantitative PCR assay for rapid detection of *Shigella* species in fresh produce. *J Food Prot, 73*(2), 221-233.

Lindsay, B., Ochieng, J. B., Ikumapayi, U. N., Toure, A., Ahmed, D., Li, S., . . . Stine, O. C. (2013). Quantitative PCR for detection of *Shigella* improves ascertainment of *Shigella* burden in children with moderate-to-severe diarrhea in low-income countries. *J Clin Microbiol, 51*(6), 1740-1746. doi:10.1128/jcm.02713-12

Liu, H., Zhu, B., Qiu, S., Xia, Y., Liang, B., Yang, C., . . . Song, H. (2018). Dominant serotype distribution and antimicrobial resistance profile of *Shigella* spp. in Xinjiang, China. *PLoS One, 13*(4), e0195259. doi:10.1371/journal.pone.0195259

Liu, J., Platts-Mills, J. A., Juma, J., Kabir, F., Nkeze, J., Okoi, C., . . . Houpt, E. R. (2016). Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. *Lancet, 388*(10051), 1291-1301. doi:10.1016/s0140-6736(16)31529-x

Liu, L., Johnson, H. L., Cousens, S., Perin, J., Scott, S., Lawn, J. E., . . . Unicef. (2012). Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet, 379*(9832), 2151-2161. doi:10.1016/S0140-6736(12)60560-1

Lluque, A., Mosquito, S., Gomes, C., Riveros, M., Durand, D., Tilley, D. H., . . . Ruiz, J. (2015). Virulence factors and mechanisms of antimicrobial resistance in *Shigella* strains from periurban areas of Lima (Peru). *Int J Med Microbiol, 305*(4-5), 480-490. doi:10.1016/j.ijmm.2015.04.005

Loman, N. J., & Watson, M. (2015). Successful test launch for nanopore sequencing. *Nat Methods, 12*(4), 303-304. doi:10.1038/nmeth.3327

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., . . . Memish, Z. A. (2012). Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet, 380*(9859), 2095-2128. doi:10.1016/S0140-6736(12)61728-0

Mamatha, B., Pusapati, B. R., & Rituparna, C. (2007). Changing patterns of antimicrobial susceptibility of *Shigella* serotypes isolated from children with acute diarrhea in Manipal, South India, a 5 year study. *Southeast Asian J Trop Med Public Health, 38*(5), 863-866.

Mandomando, I., Jaintilal, D., Pons, M. J., Valles, X., Espasa, M., Mensa, L., . . . Ruiz, J. (2009). Antimicrobial susceptibility and mechanisms of resistance in *Shigella* and *Salmonella* isolates from children under five years of age with diarrhea in rural Mozambique. *Antimicrob Agents Chemother, 53*(6), 2450-2454. doi:10.1128/aac.01282-08

Marteyn, B., Gazi, A., & Sansonetti, P. (2012). *Shigella*: a model of virulence regulation in vivo. *Gut Microbes, 3*(2), 104-120. doi:10.4161/gmic.19325

Mattock, E., & Blocker, A. J. (2017). How Do the Virulence Factors of *Shigella* Work Together to Cause Disease? *Front Cell Infect Microbiol, 7*, 64. doi:10.3389/fcimb.2017.00064

McCrickard, L. S., Crim, S. M., Kim, S., & Bowen, A. (2018). Disparities in severe shigellosis among adults - Foodborne diseases active surveillance network, 2002-2014. *BMC Public Health, 18*(1), 221. doi:10.1186/s12889-018-5115-4

Medeiros, P., Lima, A. A. M., Guedes, M. M., Havt, A., Bona, M. D., Rey, L. C., . . . Lima, I. F. N. (2018). Molecular characterization of virulence and antimicrobial

resistance profile of *Shigella* species isolated from children with moderate to severe diarrhea in northeastern Brazil. *Diagn Microbiol Infect Dis, 90*(3), 198-205. doi:10.1016/j.diagmicrobio.2017.11.002

Mitobe, J., Morita-Ishihara, T., Ishihama, A., & Watanabe, H. (2009). Involvement of RNA-binding protein Hfq in the osmotic-response regulation of invE gene expression in Shigella sonnei. *BMC Microbiol, 9*, 110. doi:10.1186/1471-2180-9-110

Morahan, R. J. (1968). *Salmonella, Shigella* and enteropathogenic *Escherichia coli* isolations in the East and West Sepik Districts, Territory of Papua and New Guinea. *Med J Aust, 2*(10), 438-440.

Muller, D., Greune, L., Heusipp, G., Karch, H., Fruth, A., Tschape, H., & Schmidt, M. A. (2007). Identification of unconventional intestinal pathogenic *Escherichia coli* isolates expressing intermediate virulence factor profiles by using a novel single-step multiplex PCR. *Appl Environ Microbiol, 73*(10), 3380-3390. doi:10.1128/AEM.02855-06

Murray, C. J., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., . . . Memish, Z. A. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet, 380*(9859), 2197-2223. doi:10.1016/S0140-6736(12)61689-4

Neuert, S., Nair, S., Day, M. R., Doumith, M., Ashton, P. M., Mellor, K. C., . . . Dallman, T. J. (2018). Prediction of Phenotypic Antimicrobial Resistance Profiles From Whole Genome Sequences of Non-typhoidal *Salmonella enterica*. *Front Microbiol, 9*, 592. doi:10.3389/fmicb.2018.00592

Niyogi, S. K. (2005). Shigellosis. *J Microbiol, 43*(2), 133-143.

Niyogi, S. K. (2007). Increasing antimicrobial resistance--an emerging problem in the treatment of shigellosis. *Clin Microbiol Infect, 13*(12), 1141-1143. doi:10.1111/j.1469-0691.2007.01829.x

NNDSS Annual Report Writing Group. (2013). *Australia's Notifiable Disease Status, 2013: Annual Report of the National Notifiable Diseases Surveillance System*. Retrieved from http://www.health.gov.au/internet/main/publishing.nsf/content/cda-cdi3903-pdf-cnt.htm/$FILE/cdi3903i.pdf

Painter, J. E., Walker, A. T., Pytell, J., Nua, M. T., Soliai-Lemusu, S., Mintz, E., . . . Prevention. (2015). Notes from the field: outbreak of diarrheal illness caused by *Shigella flexneri* - American Samoa, May-June 2014. *MMWR Morb Mortal Wkly Rep, 64*(1), 30.

Papaventsis, D., Casali, N., Kontsevaya, I., Drobniewski, F., Cirillo, D. M., & Nikolayevskyy, V. (2017). Whole genome sequencing of *Mycobacterium tuberculosis* for detection of drug resistance: a systematic review. *Clin Microbiol Infect, 23*(2), 61-68. doi:10.1016/j.cmi.2016.09.008

Passey, M. (1995). The new problem of typhoid fever in Papua New Guinea: how do we deal with it? *P N G Med J, 38*(4), 300-304.

Phantouamath, B., Sithivong, N., Insisiengmay, S., Ichinose, Y., Higa, N., Song, T., & Iwanaga, M. (2005). Pathogenicity of *Shigella* in healthy carriers: a study in Vientiane, Lao People's Democratic Republic. *Jpn J Infect Dis, 58*(4), 232-234.

Phuanukoonnon, S., Namosha, E., Kua, L., Siba, P. M., & Greenhill, A. R. (2013). Evaluation of a WASH intervention demonstrates the potential for improved hygiene practices in Hiri District, Central Province. *P N G Med J, 56*(3-4), 126-135.

Pires, S. M., Fischer-Walker, C. L., Lanata, C. F., Devleesschauwer, B., Hall, A. J., Kirk, M. D., . . . Angulo, F. J. (2015). Aetiology-Specific Estimates of the Global and Regional Incidence and Mortality of Diarrhoeal Diseases Commonly Transmitted through Food. *PLoS One, 10*(12), e0142927. doi:10.1371/journal.pone.0142927

Platts-Mills, J. A., Liu, J., Rogawski, E. T., Kabir, F., Lertsethtakarn, P., Siguas, M., . . . Houpt, E. R. (2018). Use of quantitative molecular diagnostic methods to assess the aetiology, burden, and clinical characteristics of diarrhoea in children in low-resource settings: a reanalysis of the MAL-ED cohort study. *Lancet Glob Health, 6*(12), e1309-e1318. doi:10.1016/s2214-109x(18)30349-8

PNG National Department of Health. (2010). *Papua New Guinea National Health Plan 2011-2020, Policies and Strategies Vol. 1*. Retrieved from Papua New Guinea: http://www.wpro.who.int/countries/png/PNGNHP_Part1.pdf

PNG National Department of Health. (2012). Standard Treatment Guidelines for Adults. 6. Retrieved from http://www.adi.org.au/wp-content/uploads/2013/07/2012_PNG-Dept-Health_Adult-Standard-Treatment-Manual.pdf

PNG Paediatric Society. (2016). Standard Treatment for Common Illnesses of Children in Papua New Guinea, Tenth Edition. Retrieved from https://pngpaediatricsociety.org/wp-content/uploads/2016/11/PNG-Standard-Treatment-Book-10th-edition-2016.pdf

Poka, H., & Duke, T. (2013). Clinical management of diarrhoea in children. *P N G Med J, 56*(3-4), 156-161.

Poramathikul, K., Bodhidatta, L., Chiek, S., Oransathid, W., Ruekit, S., Nobthai, P., . . . Swierczewski, B. (2016). Multidrug-resistant *Shigella* infections in patients with diarrhea, Cambodia, 2014-2015. *Emerg Infect Dis, 22*(9), 1640-1643. doi:10.3201/eid2209.152058

Preacher, K. J. (2001). Calculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence [Computer software]. Retrieved from http://quantpsy.org/calc.htm

Psutka, R., Priest, P., Davies, T., Rakunuea, T., Iddings, S., & Reiffer, A. (2013). Assessing the demographic, behavioural and environmental characteristics and the potential effectiveness of a household water filter in the Republic of Kiribati. *Journal of Water, Sanitation and Hygiene for Development, 3*(4), 530-540. doi:10.2166/washdev.2013.139

Qu, M., Zhang, X., Liu, G., Huang, Y., Jia, L., Liang, W., . . . Wang, Q. (2014). An eight-year study of *Shigella* species in Beijing, China: serodiversity, virulence genes, and antimicrobial resistance. *J Infect Dev Ctries, 8*(7), 904-908. doi:10.3855/jidc.3692

Quainoo, S., Coolen, J. P. M., van Hijum, S., Huynen, M. A., Melchers, W. J. G., van Schaik, W., & Wertheim, H. F. L. (2017). Whole-Genome Sequencing of Bacterial Pathogens: the Future of Nosocomial Outbreak Analysis. *Clinical Microbiology Reviews, 30*(4), 1015-1063. doi:10.1128/cmr.00016-17

Rajakumar, K., Sasakawa, C., & Adler, B. (1997). Use of a novel approach, termed island probing, identifies the *Shigella flexneri* she pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. *Infect Immun, 65*(11), 4606-4614.

Roser, M., & Ritchie, H. (2019). Burden of Disease. Retrieved from https://ourworldindata.org/burden-of-disease

Rosewell, A., Ropa, B., Posanai, E., Dutta, S. R., Mola, G., Zwi, A., & MacIntyre, C. R. (2010). *Shigella* spp. antimicrobial drug resistance, Papua New Guinea, 2000-2009. *Emerg Infect Dis, 16*(11), 1797-1799. doi:10.3201/eid1611.101025

Sack, D. A., Lyke, C., McLaughlin, C., & Suwanvanichkij, V. (2001). *Antimicrobial resistance in shigellosis, cholera and campylobacteriosis* (pp. 1-51). Retrieved from https://www.who.int/drugresistance/Antimicrobial_resistance_in_shigellosis_cholera_and_cam.pdf

Sadouki, Z., Day, M. R., Doumith, M., Chattaway, M. A., Dallman, T. J., Hopkins, K. L., . . . Jenkins, C. (2017). Comparison of phenotypic and WGS-derived antimicrobial resistance profiles of *Shigella sonnei* isolated from cases of diarrhoeal disease in England and Wales, 2015. *J Antimicrob Chemother, 72*(9), 2496-2502. doi:10.1093/jac/dkx170

Sansonetti, P. J. (2001). Microbes and microbial toxins: paradigms for microbial-mucosal interactions III. Shigellosis: from symptoms to molecular pathogenesis. *Am J Physiol Gastrointest Liver Physiol, 280*(3), G319-323.

Saweri, O. P., Hetzel, M. W., Mueller, I., Siba, P. M., & Pulford, J. (2017). The treatment of non-malarial febrile illness in Papua New Guinea: findings from cross sectional and longitudinal studies of health worker practice. *BMC Health Serv Res, 17*(1), 10. doi:10.1186/s12913-016-1965-6

Schrick, L., & Nitsche, A. (2016). Pitfalls in PCR troubleshooting: Expect the unexpected? *Biomol Detect Quantif, 6*, 1-3. doi:10.1016/j.bdq.2015.11.001

Schroeder, G. N., & Hilbi, H. (2008). Molecular pathogenesis of *Shigella* spp.: Controlling host cell signaling, invasion, and death by type III secretion. *Clinical Microbiology Reviews, 21*(1), 134-+. doi:10.1128/Cmr.00032-07

Schuurkamp, G. J., Bulungol, P. K., & Kereu, R. K. (1990). The monitoring of foodhandlers in Papua New Guinea. *P N G Med J, 33*(2), 99-106.

Shiferaw, B., Solghan, S., Palmer, A., Joyce, K., Barzilay, E. J., Krueger, A., & Cieslak, P. (2012). Antimicrobial susceptibility patterns of *Shigella* isolates in Foodborne Diseases Active Surveillance Network (FoodNet) sites, 2000-2010. *Clin Infect Dis, 54 Suppl 5*, S458-463. doi:10.1093/cid/cis230

Shiff, C. (2002). Integrated approach to malaria control. *Clinical Microbiology Reviews, 15*(2), 278-293. doi:10.1128/cmr.15.2.278-293.2002

Singh, R. B., Hales, S., de Wet, N., Raj, R., Hearnden, M., & Weinstein, P. (2001). The influence of climate variation and change on diarrheal disease in the Pacific Islands. *Environ Health Perspect, 109*(2), 155-159.

Soli, K. W., Maure, T., Kas, M. P., Bande, G., Bebes, S., Luang-Suarkia, D., . . . Horwood, P. F. (2014). Detection of enteric viral and bacterial pathogens associated with paediatric diarrhoea in Goroka, Papua New Guinea. *Int J Infect Dis, 27*, 54-58. doi:10.1016/j.ijid.2014.02.023

SPEHIS. (1999a). *Annual report for the year 1995*. Retrieved from New Caledonia: http://www.spc.int/phs/index.php?option=com_docman&task=cat_view&gid=54&dir=ASC&order=date&Itemid=102&limit=5&limitstart=0

SPEHIS. (1999b). *Annual report for the year 1996*. Retrieved from New Caledonia: http://www.spc.int/phs/index.php?option=com_docman&task=cat_view&gid=54&dir=ASC&order=date&Itemid=102&limit=5&limitstart=0

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313. doi:10.1093/bioinformatics/btu033

Storch, G. A., Gunn, R. A., Martin, W. T., Pollard, R. A., & Sinclair, S. P. (1980). Shigellosis in the Marshall Islands: epidemiologic aspects of an outbreak. *Am J Trop Med Hyg, 29*(3), 456-463.

Sun, Q., Lan, R., Wang, Y., Zhao, A., Zhang, S., Wang, J., . . . Xu, J. (2011). Development of a multiplex PCR assay targeting O-antigen modification genes for molecular serotyping of *Shigella flexneri*. *J Clin Microbiol, 49*(11), 3766-3770. doi:10.1128/jcm.01259-11

Talukder, K. A., Mondol, A. S., Islam, M. A., Islam, Z., Dutta, D. K., Khajanchi, B. K., . . . Sack, D. A. (2007). A novel serovar of *Shigella dysenteriae* from patients with diarrhoea in Bangladesh. *J Med Microbiol, 56*(Pt 5), 654-658. doi:10.1099/jmm.0.46999-0

The World Bank. (2019). Classifying countries by income - GNI per Capita & Income Thresholds. Retrieved from http://datatopics.worldbank.org/world-development-indicators/stories/the-classification-of-countries-by-income.html

Tickell, K. D., Brander, R. L., Atlas, H. E., Pernica, J. M., Walson, J. L., & Pavlinac, P. B. (2017). Identification and management of *Shigella* infection in children with diarrhoea: a systematic review and meta-analysis. *Lancet Glob Health, 5*(12), e1235-e1248. doi:10.1016/s2214-109x(17)30392-3

Tornieporth, N. G., John, J., Salgado, K., de Jesus, P., Latham, E., Melo, M. C., . . . Riley, L. W. (1995). Differentiation of pathogenic *Escherichia coli* strains in Brazilian children by PCR. *J Clin Microbiol, 33*(5), 1371-1374.

Trading Economics. (2019a). Solomon Islands - Rural population. Retrieved from https://tradingeconomics.com/solomon-islands/rural-population-percent-of-total-population-wb-data.html

Trading Economics. (2019b). Vanuatu - Rural population. Retrieved from https://tradingeconomics.com/vanuatu/rural-population-percent-of-total-population-wb-data.html

Turner, S. A., Luck, S. N., Sakellaris, H., Rajakumar, K., & Adler, B. (2001). Nested deletions of the SRL pathogenicity island of *Shigella flexneri* 2a. *J Bacteriol, 183*(19), 5535-5543. doi:10.1128/jb.183.19.5535-5543.2001

Vargas, M., Gascon, J., Jimenez De Anta, M. T., & Vila, J. (1999). Prevalence of *Shigella* enterotoxins 1 and 2 among *Shigella* strains isolated from patients with traveler's diarrhea. *J Clin Microbiol, 37*(11), 3608-3611.

Veenendaal, A. K., Hodgkinson, J. L., Schwarzer, L., Stabat, D., Zenk, S. F., & Blocker, A. J. (2007). The type III secretion system needle tip complex mediates host cell sensing and translocon insertion. *Mol Microbiol, 63*(6), 1719-1730. doi:10.1111/j.1365-2958.2007.05620.x

Vince, J. D. (1995). Diarrhoea in children in Papua New Guinea. *P N G Med J, 38*(4), 262-271.

Vinh, H., Nhu, N. T., Nga, T. V., Duy, P. T., Campbell, J. I., Hoang, N. V., . . . Baker, S. (2009). A changing picture of shigellosis in southern Vietnam: shifting species dominance, antimicrobial susceptibility and clinical presentation. *BMC Infect Dis, 9*, 204. doi:10.1186/1471-2334-9-204

von Seidlein, L., Kim, D. R., Ali, M., Lee, H., Wang, X., Thiem, V. D., . . . Clemens, J. (2006). A multicentre study of Shigella diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. *PLoS Med, 3*(9), e353. doi:10.1371/journal.pmed.0030353

Walker, C. L., Rudan, I., Liu, L., Nair, H., Theodoratou, E., Bhutta, Z. A., . . . Black, R. E. (2013). Global burden of childhood pneumonia and diarrhoea. *Lancet, 381*(9875), 1405-1416. doi:10.1016/S0140-6736(13)60222-6

Walker, J. C., & Verma, N. K. (2002). Identification of a putative pathogenicity island in *Shigella flexneri* using subtractive hybridisation of the *S. flexneri* and *Escherichia*

coli genomes. *FEMS Microbiol Lett, 213*(2), 257-264. doi:10.1111/j.1574-6968.2002.tb11315.x

Watson, C. (2006). Death from multi-resistant shigelloses: a case study from Fiji. *Pac Health Dialog, 13*(2), 111-114.

Wei, S., Weiss, Z. R., & Williams, Z. (2018). Rapid Multiplex Small DNA Sequencing on the MinION Nanopore Sequencing Platform. *G3 (Bethesda), 8*(5), 1649-1657. doi:10.1534/g3.118.200087

WHO/UNICEF. (2019). *Progress on household drinking water, sanitation and hygiene 2000-2017. Special focus on inequalities*. Retrieved from New York:

Williams, P., & Berkley, J. A. (2016). Dysentery (Shigellosis) - Current WHO guidelines and the WHO essential medicine list for children. *World Health Organization*.

Williams, P., & Berkley, J. A. (2018). Guidelines for the treatment of dysentery (shigellosis): a systematic review of the evidence. *Paediatr Int Child Health, 38*(sup1), S50-s65. doi:10.1080/20469047.2017.1409454

World Health Organization. (2005). Guidelines for the control of shigellosis, including epidemics due to *Shigella dysenteriae* type 1. Retrieved October 24, 2019 https://www.who.int/maternal_child_adolescent/documents/9241592330/en/

World Health Organization. (2011). Water, sanitation and hygiene interventions and the prevention of diarrhoea. Retrieved from https://www.who.int/elena/titles/bbc/wsh_diarrhoea/en/

World Health Organization. (2013). Global Health Observatory (GHO) data - Country Statistics. *WHO*. Retrieved 2015-02-01 22:47:00 from http://www.who.int/gho/countries/en/#P

Yaghoubi, S., Ranjbar, R., Dallal, M. M. S., Fard, S. Y., Shirazi, M. H., & Mahmoudi, M. (2017). Profiling of Virulence-associated Factors in *Shigella* Species Isolated from

Acute Pediatric Diarrheal Samples in Tehran, Iran. *Osong Public Health Res Perspect, 8*(3), 220-226. doi:10.24171/j.phrp.2017.8.3.09

Yang, F., Yang, J., Zhang, X., Chen, L., Jiang, Y., Yan, Y., . . . Jin, Q. (2005). Genome dynamics and diversity of *Shigella* species, the etiologic agents of bacillary dysentery. *Nucleic Acids Res, 33*(19), 6445-6458. doi:10.1093/nar/gki954

Yavzori, M., Cohen, D., & Orr, N. (2002). Prevalence of the genes for *Shigella* enterotoxins 1 and 2 among clinical isolates of shigella in Israel. *Epidemiol Infect, 128*(3), 533-535.

Zaidi, M. B., & Estrada-Garcia, T. (2014). : A Highly Virulent and Elusive Pathogen. *Curr Trop Med Rep, 1*(2), 81-87. doi:10.1007/s40475-014-0019-6

Zhang, J., Qian, L., Wu, Y., Cai, X., Li, X., Cheng, X., & Qu, D. (2013). Deletion of pic results in decreased virulence for a clinical isolate of *Shigella flexneri* 2a from China. *BMC Microbiol, 13*, 31. doi:10.1186/1471-2180-13-31

Zhou, Y., Wylie, K.M., El Feghaly, R.E., Mihindukulasuriya, K.A., Elward, A., Haslam, D.B., Storch, G.A., & Weinstock, G.M. (2016) Metagenomic Approach for Identification of the Pathogens Associated with Diarrhea in Stool Specimens. *J Clin Microbiol*. 2016 Feb;54(2):368-75. doi: 10.1128/JCM.01965-15

Appendices

Appendix 1 Manuscripts accepted for publication

Appendix 1.1 Manuscript published in the journal of Infection, Genetics and Evolution. *64*, 52-56

Title: Antimicrobial sensitivity trends and virulence genes in *Shigella* spp. from the Oceania region

Authors: Elisheba Malau[1], Rebecca Ford[2], Mary Valcanis[3], Amy V Jennison[4], Jenny Mosse[1], David Bean[1] Mition Yoannes[2], William Pomat[2], Paul F Horwood[1, 5], Andrew R Greenhill[1]

Affiliations:

[1] School of Health and Life Sciences, Federation University Australia, Churchill, Australia, e.malau@federation.edu.au; jennifer.mosse@federation.edu.au; d.bean@federation.edu.au; andrew.greenhill@federation.edu.au

[2] Papua New Guinea Institute of Medical Research, Goroka, Papua New Guinea, rebecca.ford@pngimr.org.pg; mition.yoannes@pngimr.org.pg; william.pomat@pngimr.org.pg

[3] Microbiological Diagnostic Unit Public Health Laboratory, Doherty Institute, Melbourne, Australia, valcanis@unimelb.edu.au

[4]Public Health Microbiology, Forensic and Scientific Services, Queensland Department of Health, Brisbane, Australia, Amy.Jennison@health.qld.gov.au

[5] Australian Institute of Tropical Health and Medicine, James Cook University, Cairns, Australia, paul.horwood@jcu.edu.au

Corresponding author:

Dr Andrew Greenhill

Senior Lecturer in Microbiology, School of Applied and Biomedical Sciences, Federation University, Churchill, 3842, Victoria, Australia, andrew.greehill@federation.edu.au; (03) 51226439

Abstract

Shigella is a common cause of diarrhoea in Papua New Guinea (PNG) and other Oceania countries. However, little is known about the strains causing infection. Archived *Shigella* isolates (n=72) were obtained from research laboratories in PNG and reference laboratories in Australia. *Shigella* virulence genes were detected by PCR, and antimicrobial susceptibility was determined by disk diffusion. The *ipaH* virulence gene was present in all 72 isolates. The prevalence of other virulence genes was variable, with *ial*, *invE*, *ipaBCD*, *sen/ospD3* and *virF* present in 60% of isolates and *set1A* and *set1B* genes present in 42% of isolates. Most *S. flexneri* isolates contained genes encoding enterotoxin 1 and/ or enterotoxin 2. Resistance to antibiotics was common, with 51/72 isolates resistant to 2-4 antimicrobials. A greater proportion of bacteria isolated since 2010 (relative to pre-2010 isolates) were resistant to commonly used antibiotics such as ampicillin, chloramphenicol, tetracycline, and trimethoprim-sulfamethoxazole; suggesting that antimicrobial resistance (AMR) in *Shigella* is increasing over time in the Oceania region. There is a need for improved knowledge regarding *Shigella* circulation in the Oceania region and further monitoring of AMR patterns.

Abbreviations:

AMR: antimicrobial resistance

DNA: Deoxyribonucleic acid

MDR: multi-drug resistant

PCR: polymerase chain reaction

PNG: Papua New Guinea

Keywords:

Shigellosis, *Shigella*, antimicrobial resistance, virulence genes, surveillance, Oceania, Papua New Guinea

Main text

Shigellosis is a leading global cause of moderate-severe diarrhoea in children (Bardhan et al. 2010) and adults (Lozano et al. 2012), affecting ~125 million people per year. There are four species of *Shigella*, each with their own epidemiological characteristics. *S. flexneri* is responsible for most of the shigellosis burden in developing countries globally; with *S. boydii* also important, but largely confined to the Indian subcontinent. *S. dysenteriae* causes sporadic, epidemic outbreaks of diarrhoea, especially in areas experiencing conflict and natural disasters (Kotloff et al. 1999). *S. sonnei* occurs predominantly in developed countries and in countries transitioning from low- to middle-income (Niyogi 2005; Taneja and Mewara 2016).

Papua New Guinea (PNG) is the only low/middle income country in the Oceania region for which recent reports on shigellosis exist, with evidence that *Shigella* is an important cause of enteric disease in adults (Greenhill et al. 2014), children (Howard et al. 2000; Soli et al. 2014) and vulnerable populations (Benny et al. 2014). Little is known about the

genetic traits or virulence of *Shigella* strains circulating in the Oceania region, beyond the sporadic documentation of antibiotic resistance to many commonly used antibiotics such as ampicillin, chloramphenicol, tetracycline and trimethoprim-sulfamethoxazole (Greenhill et al. 2014; Rosewell et al. 2010; Storch et al. 1980; Watson 2006). In this study, we sought to gain a greater understanding of *Shigella* isolated from PNG and nearby Oceania countries, looking specifically at the presence of virulence genes and antibiotic resistance.

Seventy-two *Shigella* isolates from PNG (n=60) and neighbouring Pacific Island nations (n=12) were analysed. Of the PNG isolates: 30 were from a study previously conducted by our research team (Greenhill et al. 2014); one was from a recent outbreak of shigellosis in PNG (Benny et al. 2014); seven isolates were from a previous case-control study conducted in PNG (Howard et al. 2000); and the remaining isolates (n = 22) were from travellers returning to Australia from PNG. The 12 non-PNG isolates were from travellers returning to Australia from Pacific Island nations (Fiji, Vanuatu, Samoa, Solomon Islands), including three isolates from people who lived in or had spent time in the Torres Strait Islands (situated north of mainland Australia). The study collection comprised 53 *S. flexneri* isolates (from 1985-2014), 16 *S. sonnei* isolates (from 1999-2015) and three *S. dysenteriae* isolates (from 1985 and 2010).

Isolates were cultured on nutrient agar and incubated overnight at 37°C prior to DNA extraction (FavorPrep Tissue Genomic DNA Extraction Mini Kit, Favorgen, Taiwan), and antibiotic susceptibility testing. Isolates had previously been confirmed as *Shigella* and speciated using poly-O antisera; all 72 isolates were confirmed as *Shigella* by real-time

PCR detection of the *ipaH* gene (Lin et al. 2010) prior to further analyses. Approximately half (56%) the isolates were serotyped/biotyped.

Conventional PCR (Table 1) was used to confirm the presence/absence of known *Shigella* virulence genes: *ipaBCD* and *ipaH* (invasion plasmid antigen BCD and H genes); *ial* (invasion-associated locus); *virF* and *invE* (regulators for transcription activation on the virulence plasmid); *set1A* and *set1B* (*Shigella* enterotoxin 1 genes); and *sen/ospD3* (*Shigella* enterotoxin 2 gene). Amplicons were visualised by gel electrophoresis.

Isolates were tested for antibiotic susceptibility by disk diffusion following the Clinical Laboratory Standards Institute guidelines (CLSI 2015), using *Escherichia coli* ATCC 25922 as a control. Antibiotics assayed were ampicillin (10 μg), ceftriaxone (30 μg), chloramphenicol (30 μg), ciprofloxacin (5 μg), nalidixic acid (30 μg), tetracycline (30 μg) and trimethoprim-sulfamethoxazole (25 μg).

Isolates of each species were grouped according to their virulence gene profiles (presence/absence of genes), producing four groups of *S. flexneri*, two groups of *S. sonnei* and a single group of *S. dysenteriae* (Table 2). Detailed statistical analyses were not conducted due to opportunistic sample collection, but a trend for more recent *S. flexneri* isolates to harbour more virulence genes was noted. In particular, genes coding for *Shigella* enterotoxin 1 (*set1A* and *set1B*) were detected only in isolates obtained since 2009. All isolates positive for *set1A* and *set1B* were from PNG, or from the Torres Strait Islands - a region of Australia with close geographical and cultural links to PNG.

The enterotoxin genes are primarily responsible for the clinical manifestation of acute-watery diarrhoea and dysentery in shigellosis. Genes encoding *Shigella* enterotoxin 1 (*set1A* and *set1B*) were present in *S. flexneri* serotype 2a isolates collected from 2009 onwards (~57% of *S. flexneri* isolates) but not detected in either *S. sonnei* or *S. dysenteriae*. These genes have been detected at similar rates in paediatric cases in the Brazilian Amazon (da Cruz et al. 2014) and in northeastern Brazil (Mendeiros et al, 2018), but at lower rates in comparable studies in Argentina (Casabonne et al. 2016). No published studies have investigated virulence genes in *Shigella* from the Oceania region, and relatively few in Asia. In China a similar species distribution of the s*et* gene was observed in Beijing, being commonly detected in *S. flexneri* (88% of isolates, thus more frequently than in PNG isolates), but infrequently in other species (Qu et al, 2014). In a study conducted in Eastern China *set-1A* and *set-1B* genes were detected in 74.5% and 78% of *S. flexneri*, respectively (Fan et al, 2017) In our study *Shigella* enterotoxin 2 genes (*sen/ospD3*) were present in 79% of *S. flexneri* isolates, a detection rate higher than two comparable studies conducted in South America (da Cruz et al. 2014; Casabonne et al. 2016) but lower than detected in *S. flexneri* isolates from northeastern Brazil (Medeiros et al, 2018). One isolate of *S. sonnei* harboured *sen/ospD3*, which has been reported previously (da Cruz et al. 2014; Casabonne et al. 2016); and indeed appears common in parts of Brazil (Medeiros et al, 2018) and China (Qu et al, 2014; Fan et al, 2017). The high prevalence of both enterotoxin encoding genes in circulating strains of *S. flexneri* could have public health ramifications in PNG, where access to healthcare is often limited.

The *Shigella* invasion plasmid is a large unstable plasmid that encodes several virulence genes, including *ipaH*, *ipaBCD*, *virF*, *invE* and *ial*; the *ipaH* gene is also chromosomally

encoded, so is considered a stable target gene. The *ipaH* gene was present in all isolates, while 60% of isolates contained all four *ipaBCD*, *virF*, *invE* and *ial* genes (Table 2). It is important to note that the latter virulence genes *(ipaBCD*, *virF*, *invE* and *ial)* all contribute to facilitating and maintaining the spread and survival of *Shigella* during infection (Broach et al, 2012; Lluque et al. 2015; Schroeder et al, 2008). Our results suggest that there is a strong association between all of these genes, as they were invariably present/absent as a set (Table 2). In regards to the stability of these invasion plasmid virulence genes, some studies have shown that the *ial* gene in particular is prone to deletion (Zhang et al, 2013).

While recently circulating isolates tended to have an increased number of virulence genes, *S. flexneri* isolates with different virulence gene profiles were isolated as recently as 2012-14, suggesting co-circulation of strains. A study of global isolates of *S. flexneri* using whole genome sequencing revealed old and new strains of *S. flexneri* to persist alongside one another and continue causing infection (Connor et al. 2015).

S. flexneri isolates were frequently resistant to ampicillin (77%), tetracycline (74%), chloramphenicol (60%) and trimethoprim-sulfamethoxazole (49%). *S. sonnei* resistance to trimethoprim-sulfamethoxazole (75%) and ampicillin (56%) was common, followed by tetracycline (19%) and nalidixic acid (6%). *S. dysenteriae* also displayed resistance to ampicillin, tetracycline and trimethoprim-sulfamethoxazole (33%) (Table 3).

An increase in the prevalence of antimicrobial drug resistance to older antimicrobials, namely ampicillin, tetracycline, chloramphenicol and trimethoprim-sulfamethoxazole (Figure 1), was observed in post-2010 isolates compared to pre-2010 isolates. Few

comparable data exist in the Oceania region, but this increase in antimicrobial drug resistance coincides with findings in Fiji, where there is evidence of ampicillin and chloramphenicol AMR rates increasing over time, especially in *S. flexneri* (Watson 2006). Our observed trend of increasing AMR in *Shigella* is also reflected in other low-income settings. In Mozambique, *Shigella* isolates are commonly resistant to the same antibiotics to which resistance was observed in this study (Mandomando et al, 2009). Multidrug resistance (MDR), defined as resistance to at least two antimicrobials, was detected in 71% (n=51) of isolates. Using a more stringent definition (resistance to at least three antimicrobials), MDR was detected in 54% (n=39) of isolates (Table 4); a rate also similar to that observed in Mozambique (Mandomando et al, 2009).

Resistance to ciprofloxacin or ceftriaxone was not observed in any *Shigella* isolates in this study; however, one *S. sonnei* isolate was resistant to nalidixic acid. Ciprofloxacin resistance in *S. flexneri* strains and nalidixic acid resistance in *S. sonnei* is increasingly reported in Asia and Africa (Gu et al. 2012). Given the close proximity of Asia to Oceania, the possibility of AMR strains being introduced into the Oceania region is a cause for concern and warrants monitoring.

This study provides initial data on the distribution of *Shigella* virulence genes in PNG and neighbouring Pacific Islands and provides insight into the AMR trends. It is notable that many of the recently isolated *S. flexneri* harbour more virulence genes than most pre-2009 isolates. Similarly, there appears to be increasing resistance to commonly used antibiotics. Further genetic typing is required to determine whether these represent new strains; or they are the same strains which have acquired genetic elements associated with virulence and antibiotic resistance. There is a need to monitor *Shigella* strains in the region to

enable better management of clinical cases and to reduce morbidity and mortality in vulnerable populations of the Oceania region.

Acknowledgements

We thank staff at the Microbiology Diagnostic Unit Public Health Microbiology, (MDU PHL) located in the Doherty Institute at the University of Melbourne, Public Health Microbiology, Forensic and Scientific Services, Queensland Health in Brisbane and PNG Institute of Medical Research (PNGIMR) in Goroka, PNG, for provision of isolates and analysis in this study. This study was funded by internal research funding from Federation University Australia.

Accompanying Tables and Figures published with the manuscript

Table 1. List of standard PCR primers used for the detection of *Shigella* virulence genes.

Gene target	Primers	Annealing temp. °C	Amplicon size (bp)	Reference
ipaH	ipaH_F1: GCTGGAAAAACTCAGTGCCT ipaH_R1: CCAGTCCGTAAATTCATTCT	56	424	Tornieporth et al., 1995
ial	ial_F: CTGGATGGTATGGTGAGG ial_R: GGAGGCCAACAATTATTTCC	58	320	Talukder et al., 2007
invE	invE_F: CGATAGATGGCGAGAAATTATATCCCG invE_R: CGATCAAGAATCCCTAACAGAAGAATCAC	57	766	Muller et al., 2007
ipaBCD	ipaBCD_F: GCTATAGCAGTGACATG ipaBCD_R: ACGAGTTCGAAGCACTC	55	500	Faruque et al., 2002
sen/ospD3	sen/ospD3_F: ATGTGCCTGCTATTATTTAT sen/ospD3_R: CATAATAATAAGCGGTCAGC	54	799	Talukder et al., 2007
set1A	set1A_F: TCACGCTACCATCAAAGA set1A_R: TATCCCCCTTTGGTGGTA	57	309	Talukder et al., 2007
set1B	set1B_F: GTGAACCTGCTGCCGATATC set1B_R: ATTAGTGGATAAAAATGACG	57	147	Talukder et al., 2007
virF	virF_F: TCAGGCAATGAAACTTTGAC virF_R: TGGGCTTGATATTCCGATAAGTC	58	618	Gomez-Duarte et al., 2009

Table 2. Virulence profiles based on presence/absence of virulence genes in *Shigella* by species

Species	*ipaH*	*sen/ospD 3*	*set1A*	*set1B*	*invE*	*ial*	*ipaBCD*	*virF*	Year of Isolation	No. Isolates	Serotype	Biotype	Origin
S. flexneri	+	+	-	-	+	+	+	+	1985	1	3		PNG
									1994	1	2a		PNG
									2002	1	6		PNG
									2006	1	3a		PNG
									2006	1	2a		PNG
									2007	1	3a		PNG
									2007	1	1b		Fiji
									2010	4	nt		PNG
									2011	1	nt		PNG
									2012	1	X		Samoa
S. flexneri	+	-	-	-	-	-	-	-	1985	3	1		PNG
									1985	1	3		PNG
									1992	1	nt		PNG
									2000	1	2a		PNG
									2004	1	1b		PNG
									2010	3	nt		PNG
S. flexneri	+	-	+	+	-	-	-	-	2009	1	2a		PNG
S. flexneri	+	+	+	+	+	+	+	+	2009	1	2a		Torres Strait
									2010	1	2a		Torres Strait
									2010	20	nt		PNG
									2011	2	2a		PNG

													Torres
									2013	1	2a		Strait
									2013	3	2a		PNG
									2014	1	X		PNG
									1999	1		a	PNG
													Solomon
									2007	1		a	Islands
									2008	1		g	PNG
									2011	2		g	PNG
S. sonnei	+	-	-	-	-	-	-	-	2012	3		a	Fiji
									2013	1		a	Samoa
									2013	1		a	Vanuatu
									2014	3		g	PNG
									2015	1		g	PNG
									2015	1		a	Vanuatu
S. sonnei	+	+	-	-	+	+	+	+	2010	1		f	PNG
S. dysenteriae	+	-	-	-	-	-	-	-	1985	1	nt		PNG
									2010	2	nt		PNG

Legend: nt not tested; PNG Papua New Guinea

Table 3. Proportion of *Shigella* isolates resistant to selected antimicrobials.

Number of *Shigella* species (%) with – antimicrobial resistance

Antimicrobial	*S. flexneri* (n=53)	*S. sonnei* (n=16)	*S. dysenteriae* (n=3)	Total (%) (n=72)
AMP	41 (77)	9 (56)	1 (33)	51 (71)
CHL	32 (60)	0	0	32 (44)
CIP	0	0	0	0
CRO	0	0	0	0
NAL	0	1 (6)	0	1 (1)
TET	39 (74)	3 (19)	1 (33)	43 (60)
SXT	26 (49)	12 (75)	1 (33)	39 (54)

Legend: Ampicillin (AMP), Chloramphenicol (CHL), Ciprofloxacin (CIP), Nalidixic acid (NAL), Tetracycline (TET), trimethoprim-sulfamethoxazole (SXT).

Table 4. Antibiotic resistance profiles of *Shigella* isolates.

	S. flexneri	*S. sonnei*	*S. dysenteriae*	Total (%)
AMP/CHL/TET/ SXT	17 (32)	0	0	17 (24)
AMP/CHL/TET	12 (23)	0	0	12 (17)
AMP/SXT	1 (2)	7 (44)	1 (33)	9 (13)
AMP/TET/SXT	4 (8)	2 (13)	0	6 (8)
AMP/TET	3 (6)	0	0	3 (4)
CHL/TET/SXT	2 (4)	0	0	2 (3)
AMP/CHL/SXT	1 (2)	0	0	1 (1)
NAL/TET/SXT	0	1 (6)	0	1 (1)
AMP	3 (6)	0	0	3 (4)
TET	1 (2)	0	1 (33)	2 (3)
SXT	1 (2)	2 (13)	0	3 (4)
Susceptible	8 (15)	4 (25)	1 (33)	13 (18)
Total	53	16	3	72

Figure 1. Proportion of isolates (%) isolated before and after 2010 that are susceptible or resistant to selected antibiotics. The post-2010 group includes isolates from 2010.

References

Bardhan, P., Faruque, A.S.G., Naheed, A., Sack, D.A., 2010. Decrease in shigellosis-related deaths without Shigella spp.-specific interventions, Asia. *Emerg Infect Dis.* 16, 1718-23.

Benny, E., Mesere, K., Pavlin, B. I., Yakam, L., Ford, R., Yoannes, M., Kisa, D., Abdad, M. Y., Menda, L., Greenhill, A. R., Horwood, P. F., 2014. A large outbreak of shigellosis commencing in an internally displaced population, Papua New Guinea, 2013. *Western Pac Surveill Response J.* 5, 18-21.

Broach, W. H., Egan, N., Wing, H. J., Payne, S. M., Murphy, E. R., 2012. VirF-independent regulation of Shigella virB transcription is mediated by the small RNA RyhB. *PLoS One*. 7(6), e38592.

Casabonne, C., Gonzalez, A., Aquili, V., Balague, C., 2016. Prevalence and Virulence Genes of Shigella spp. Isolated from Patients with Diarrhea in Rosario, Argentina. *Jpn J Infect Dis*. 69, 477-81.

CLSI., 2015. *Performance Standards for Antimicrobial Susceptibility Testing: 25th Informational Supplement. CLSI Document M100-S25.* Clinical and Laboratory Standards Institute: Wayne, USA.

Connor, T. R., Barker, C. R., Baker, K. S., Weill, F. X., Talukder, K. A., Smith, A. M., Baker, S., Gouali, M., Pham Thanh, D., Jahan Azmi, I., Dias da Silveira, W., Semmler, T., Wieler, L. H., Jenkins, C., Cravioto, A., Faruque, S. M., Parkhill, J., Wook Kim, D., Keddy, K. H.,Thomson, N. R., 2015. Species-wide whole genome sequencing reveals historical global spread and recent local persistence in Shigella flexneri. *Elife*. 4, e07335.

da Cruz, C. B., de Souza, M. C., Serra, P. T., Santos, I., Balieiro, A., Pieri, F. A., Nogueira, P. A., Orlandi, P. P., 2014. Virulence factors associated with pediatric shigellosis in Brazilian Amazon. *Biomed Res Int*., 539697.

Fan, W., Qian, H., Shang, W., Ying, C., Zhang, X., Cheng, S., Gu, B., Ma, P., 2017. Low distribution of genes encoding virulence factors in *Shigella flexneri* serotypes 1b clinical isolates from eastern Chinese populations. *Gut Pathog.* 9, 76.

Greenhill, A. R., Guwada, C., Siba, V., Michael, A., Yoannes, M., Wawarie, Y., Ford, R., Siba, P. M., Horwood, P. F., 2014. Antibiotic resistant Shigella is a major cause of diarrhoea in the Highlands of Papua New Guinea. *J Infect Dev Ctries*. 8, 1391-7.

Gu, B., Cao, Y., Pan, S., Zhuang, L., Yu, R., Peng, Z., Qian, H., Wei, Y., Zhao, L., Liu, G., Tong, M., 2012. Comparison of the prevalence and changing resistance to nalidixic acid and ciprofloxacin of Shigella between Europe-America and Asia-Africa from 1998 to 2009. *Int J Antimicrob Agents*. 40, 9-17.

Howard, P., Alexander, N. D., Atkinson, A., Clegg, A. O., Gerega, G., Javati, A., Kajoi, M., Lupiwa, S., Lupiwa, T., Mens, M., Saleu, G., Sanders, R. C., West, B., Alpers, M. P., 2000. Bacterial, viral and parasitic aetiology of paediatric diarrhoea in the highlands of Papua New Guinea. *J Trop Pediatr*. 46, 10-4.

Kotloff, K. L., Winickoff, J. P., Ivanoff, B., Clemens, J. D., Swerdlow, D. L., Sansonetti, P. J., Adak, G. K., Levine, M. M., 1999. Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. *Bull World Health Organ*. 77, 651-66.

Lin, W. S., Cheng, C. M., Van, K. T., 2010. A quantitative PCR assay for rapid detection of Shigella species in fresh produce. *J Food Prot*. 73, 221-33.

Lluque, A., Mosquito, S., Gomes, C., Riveros, M., Durand, D.,Tilley, D. H., Bernal, M., Prada, A., Ochoa, T. J., Ruiz, J., 2015. Virulence factors and mechanisms of antimicrobial resistance in Shigella strains from periurban areas of Lima (Peru). *Int J Med Microbiol*. 305, 480-90.

Lozano, R., Naghavi, M., Foreman, K., Lim, S., Shibuya, K., Aboyans, V., Abraham, J., Adair, T., Aggarwal, R., Ahn, S. Y., Alvarado, M., Anderson, H. R., Anderson, L. M., Andrews, K. G., Atkinson, C., Baddour, L. M., Barker-Collo, S., Bartels, D. H., Bell, M. L., Benjamin, E. J., Bennett, D., Bhalla, K., Bikbov, B., Bin Abdulhak, A., Birbeck, G., Blyth, F., Bolliger, I., Boufous, S., Bucello, C., Burch, M., Burney, P., Carapetis, J., Chen, H., Chou, D., Chugh, S. S., Coffeng, L. E., Colan, S. D., Colquhoun, S., Colson, K. E., Condon, J., Connor, M. D., Cooper, L. T., Corriere, M., Cortinovis, M., de Vaccaro, K. C., Couser, W., Cowie, B. C., Criqui, M. H., Cross, M., Dabhadkar, K. C., Dahodwala, N., De Leo, D., Degenhardt, L., Delossantos, A., Denenberg, J., Des Jarlais, D. C., Dharmaratne, S. D., Dorsey, E. R., Driscoll, T., Duber, H., Ebel, B., Erwin, P. J., Espindola, P., Ezzati, M., Feigin, V., Flaxman, A. D., Forouzanfar, M. H., Fowkes, F. G., Franklin, R., Fransen, M., Freeman, M. K., Gabriel, S. E., Gakidou, E., Gaspari, F., Gillum, R. F., Gonzalez-Medina, D., Halasa, Y. A., Haring, D., Harrison, J. E., Havmoeller, R., Hay, R. J., Hoen, B., Hotez, P. J., Hoy, D., Jacobsen, K. H., James, S. L., Jasrasaria, R., Jayaraman, S., Johns, N., Karthikeyan, G., Kassebaum, N., Keren, A., Khoo, J. P., Knowlton, L. M., Kobusingye, O., Koranteng, A., Krishnamurthi, R., Lipnick, M., Lipshultz, S. E., Ohno, S. L., Mabweijano, J., MacIntyre, M. F., Mallinger, L., March, L., Marks, G. B., Marks, R., Matsumori, A., Matzopoulos, R., Mayosi, B. M., McAnulty, J. H., McDermott, M. M., McGrath, J., Mensah, G. A., Merriman, T. R., Michaud,

C., Miller, M., Miller, T. R., Mock, C., Mocumbi, A. O., Mokdad, A. A., Moran, A., Mulholland, K., Nair, M. N., Naldi, L., Narayan, K. M., Nasseri, K., Norman, P., O'Donnell, M., Omer, S. B., Ortblad, K., Osborne, R., Ozgediz, D., Pahari, B., Pandian, J. D., Rivero, A. P., Padilla, R. P., Perez-Ruiz, F., Perico, N., Phillips, D., Pierce, K., Pope III, C. A., Porrini, E., Pourmalek, F., Raju, M., Ranganathan, D., Rehm, J. T., Rein, D. B., Remuzzi, G., Rivara, F. P., Roberts, T., De Leon, F. R., Rosenfeld, L. C., Rushton, L., Sacco, R. L., Salomon, J. A., Sampson, U., Sanman, E., Schwebel, D. C., Segui-Gomez, M., Shepard, D. S., Singh, D., Singleton, J., Sliwa, K., Smith, E., Steer, A., Taylor, J. A., Thomas, B., Tleyjeh, I. M., Towbin, J. A., Truelsen, T., Undurraga, E. A., Venketasubramanian, N., Vijayakumar, L., Vos, T., Wagner, G. R., Wang, M., Wang, W., Watt, K., Weinstock, M. A., Weintraub, R., Wilkinson, J. D., Woolf, A. D., Wulf, S., Yeh, P. H., Yip P., Zabetian, A., Zheng, Z. J., Lopez, A. D., Murray, C. J., AlMazroa, M. A., Memish, Z. A., 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 380, 2095-128.

Mandomando, I, Jaintilal, D, Pons, M.J., Vallès, X., Espasa, M., Mensa, L. Sigaúque, B., Sanz, S., Sacarlal, J., Macete, E., Abacassamo, F., Alonso, P.L., Ruiz, J. 2009. Antimicrobial susceptibility and mechanisms of resistance in *Shigella* and *Salmonella* isolates from children under five years of age with diarrhea in rural Mozambique. Antimicrob Agents Chemother. 53, 2450-4.

Mendeiros, P.H.Q.S., Lima, A.A.M, Guedes, M.M., Havt, A., Bona, M.D., Rey, L.C., Soares, A.M., Guerrant, R.L., Weigl, B.H., Lima, I.F.N. 2018. Molecular characterization of virulence and antimicrobial resistance profile of *Shigella*

species isolated from children with moderate to severe diarrhea in northeastern Brazil. *Diagn Microbiol Infect Dis*. 90, 198-205.

Niyogi, S. K. 2005. Shigellosis. *J Microbiol*. 43, 133-43.

Qu, M., Zhang, X., Guirong, L., Huang, Y., Jia, L., Liang, W., Li, X., Wu, X., Li, J., Yan, H., Kan, B., Wang, Q. 2014. An eight-year study of *Shigella* species in Beijing, China: serodiversity, virulence genes, and antimicrobial resistance. *J Infect Dev Ctries*. 8, 904-8.

Rosewell, A., Ropa, B., Posanai, E., Dutta, S. R., Mola, G., Zwi, A., MacIntyre, C. R., 2010. Shigella spp. antimicrobial drug resistance, Papua New Guinea, 2000-2009. *Emerg Infect Dis*. 16, 1797-9.

Schroeder, G. N., Hilbi, H., 2008. Molecular pathogenesis of Shigella spp.: Controlling host cell signaling, invasion, and death by type III secretion. *Clinical Microbiology Reviews*. 21, 134-156.

Soli, K. W., Maure, T., Kas, M. P., Bande, G., Bebes, S., Luang-Suarkia, D., Siba, P. M., Morita, A., Umezaki, M., Greenhill, A. R., Horwood, P. F., 2014. Detection of enteric viral and bacterial pathogens associated with paediatric diarrhoea in Goroka, Papua New Guinea. *Int J Infect Dis*. 27, 54-8.

Storch, G. A., Gunn, R. A., Martin, W. T., Pollard, R. A., Sinclair, S. P., 1980. Shigellosis in the Marshall Islands: epidemiologic aspects of an outbreak. *Am J Trop Med Hyg*. 29, 456-63.

Taneja, N., Mewara, A., 2016. Shigellosis: Epidemiology in India. *Indian J Med Res*. 143, 565-76.

Thong, K. L., Hoe, S. L., Puthucheary, S. D., Yasin, R., 2005. Detection of virulence genes in Malaysian Shigella species by multiplex PCR assay. *BMC Infect Dis*. 5, 8.

Watson, C., 2006. Death from multi-resistant shigelloses: a case study from Fiji. *Pac Health Dialog*. 13, 111-4.

Zhang, J., Qian, L., Wu, Y., Cai, X., Li, X., Cheng, X., Qu, D., 2013. Deletion of pic results in decreased virulence for a clinical isolate of Shigella flexneri 2a from China. *BMC Microbiol*. 13: 31.

Appendix 1.2 Manuscript published in the PNG medical journal *59*(3-4), 147-154

TITLE: Shigellosis: A truly neglected disease in Papua New Guinea

AUTHORS: Elisheba Malau[1], Jenny Mosse[1], Paul F. Horwood[2] and Andrew R. Greenhill[1]

Affiliations:

[1] School of Applied and Biomedical Sciences, Federation University Australia, Churchill, Victoria, Australia

[2] Institut Pasteur in Cambodia, Institut Pasteur International Network, Phnom Penh, Cambodia

SUMMARY

Diarrhoeal diseases still affect many people, especially children living in impoverished and under-developed settings. In Papua New Guinea (PNG) diarrhoea remains one of the leading causes of hospitalisation and a major cause of death. Here, we focus on the role of *Shigella* in diarrhoeal illness in PNG, and provide an overview of the causative organism and the illness. A review of the available data on the aetiology of diarrhoea in PNG suggests that shigellosis is a major cause of diarrhoeal illness. Shigellosis can cause protracted and life-threatening illness, thus an appreciation of the burden of shigellosis in important to aid in the development of optimal prevention and control strategies. Treatment strategies for all cases of moderate-severe diarrhoeal illness should centre on rehydration, but if antimicrobial treatment is required consideration should be given to the increasing antimicrobial resistance observed in *Shigella* isolates in PNG.

<u>The global burden of diarrhoeal diseases and shigellosis</u>

Diarrhoeal diseases remain one of the greatest contributors to morbidity and mortality globally, particularly among children less than 5 years of age. Of all infectious diseases, diarrhoea remains the second leading contributor to disability adjusted life years (DALYs) – only ranking behind pneumonia. This high burden remains the case despite an approximate 50% decline in the burden of diarrhoeal illness between 1990 and 2010 (1, 2). It is pertinent to note in the context of 'tropical' diseases that diarrhoea causes more all-age illness than malaria or tuberculosis (2), yet has attracted considerably less attention in recent years. Diarrhoeal diseases, along with pneumonia, appear to fall into a group of what might be described as truly neglected diseases. These two diseases have not received the attention and funding that malaria, tuberculosis and HIV have over the past decade, and neither do they feature in the WHO list of neglected diseases (3).

One of the challenges in addressing the high burden of diarrhoeal diseases is the myriad of aetiological agents that cause gastrointestinal infections. It is difficult to ascertain which pathogens are the greatest contributors to the burden of diarrhoeal disease, and thus which pathogens warrant targeted intervention. Despite the challenges of determining the cause of diarrhoeal diseases, recent studies and analyses of the data suggest that *Shigella* is one of the most important causes of diarrhoea (2, 4). Indeed, *Shigella* has been recognised as a major contributor to the burden of diarrhoeal disease for almost two decades (2, 5); although until recently there was a lack of strong aetiological data to support the epidemiological modelling. With the recent global enteric multi-centre study (GEMS) (4) we now have irrefutable evidence that *Shigella* is a major contributor to diarrhoeal disease in children in low-income countries.

A recent study conducted by Kotloff and colleagues (2013) investigated the burden and aetiology of diarrhoeal disease in infants and young children in developing countries. *Shigella* is one of the four leading pathogens responsible for causing moderate-severe diarrhoea in children. Moreover, unlike other leading causes of diarrhoea such as rotavirus, *Cryptosporidium* and *E. coli*, a child's chances of being infected with *Shigella* increased with age within the first 5 years of life (4). Although older children and adults are generally less susceptible to moderate-severe diarrhoeal illness, shigellosis does occur in these populations as well. Indeed, shigellosis is the second leading cause of diarrhoeal mortality amongst all ages globally (6). Thus, *Shigella* is an important cause of moderate-

severe gastrointestinal illness from infancy all the way through to, and including, adulthood.

While the greatest burden is in low-income settings where sanitation and hygiene is lacking, diarrhoeal diseases can also affect people in high-income settings. In countries such as the United States and Europe, *Shigella* is often found amongst children attending day-care centres, migrant workers, people living in correctional facilities, men who have sex with men, travellers visiting developing countries, and in communities struck by natural disasters or conflict (7).

Based on these global findings and the limited country-specific data currently available for Papua New Guinea (PNG), it is likely that *Shigella* is a major contributor to the burden of diarrhoeal illness in PNG. In this review we provide an overview of *Shigella* and shigellosis, provide an insight into the possible burden of shigellosis in PNG, and review other issues pertinent to shigellosis such as treatment, sanitation and hygiene.

An overview of *Shigella* and shigellosis

Shigella is a Gram negative, non-motile, non-spore forming, facultative anaerobic bacilli belonging to the Family Enterobacteriaceae. *Shigella* is closely related to *Escherichia coli*, an organism commonly present in the gut of many birds and mammals (7, 8). Most strains of *E. coli* are non-pathogenic, though some strains cause disease of the gastrointestinal tract and other organ systems. *Shigella* spp. have a genome akin to that of *E. coli*, though with the deletion of some segments of the genome and the acquisition of numerous insertion sequences. There are similarities in the pathogenesis of *Shigella* and some of the diarrhoeaogenic strains of *E. coli*, in particular the enteroinvasive *E. coli* (8).

The *Shigella* genus consists of four species*: S. dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei* (7). Based on genetic analysis these four species *Shigella* could be more accurately considered different varieties of a single species; however, the four species remain due to historical, epidemiological and clinical considerations. While the greatest burden is in children living in low-income settings (5), the pathogen can readily infect people of all ages living in both developing and industrialized settings. Humans are the only natural host, and therefore main reservoir, for *Shigella*; however, the organism will survive in

water and on foods. Reflecting this, transmission is primarily through the faecal-oral route from person-to-person contact (predominantly from inadequate basic hygiene practices), and also through the consumption of contaminated food and water supplies.

The distribution of each species varies according to socio-economic circumstances: *S. sonnei* is the most common of the four species to be found in high-income settings (7, 9). In low-income settings *S. flexneri* is considered the most common (5, 7), with *S. dysenteriae* and *S. boydii* also present. Moreover, it appears that as countries transition to middle-income status the epidemiology of circulating strains also transitions. This supposition is based, in part, on the findings of von Seidlein *et al*: *Shigella* surveillance across six Asian countries over 4 years found *S. flexneri* to predominate in Bangladesh, Pakistan, Indonesia, China and Vietnam, while *S. sonnei* accounted for most cases in Thailand. It was postulated that this difference in distribution may correspond to Thailand's rapid progression towards being a middle-income country (10). The findings of Vinh and colleagues seem to concur with this theory, with a shift in species distribution from *S. flexneri* to *S. sonnei* in southern Vietnam (11).

A similar phenomenon was observed as early as the 1960s in countries such as Israel, the United States of America and parts of Europe. Circulating *Shigella* species transitioned from the once predominant *S. dysenteriae* to *S. flexneri* and *S. sonnei*; with the latter (*S. sonnei*) gradually becoming the predominant species in developed countries (12). It was postulated that this gradual rise of *S. sonnei* infections may be related to the increase of antimicrobial resistance (12), though other factors also likely contribute.

Shigella is highly infectious; ingestion of as few as 10-100 bacteria can cause infection. Incubation time depends on the species serotype causing infection; typically 1-3 days, but can range from 12 hours to 7 days (13, 14). Infection with *Shigella* is often characterised by blood and/or mucus in stools (dysentery). However, these symptoms are not always present in infected people, with the disease often resulting in diarrhoea without dysentery. A mild case of shigellosis may result in fever, fatigue, malaise, anorexia and watery, non-bloody diarrhoea. These symptoms are similar to those caused by other enteric pathogens, and therefore can be difficult to clinically distinguish one from the other (7). In severe cases, severe abdominal cramps, dysentery and sometimes seizures occur (7, 9). Shigellosis can result in death. There are multiple virulence factors and mechanisms of

pathogenesis that can lead to mortality, including dehydration, intestinal complications and decreased immune function and nutritional status leading to co-infection (7).

The burden of shigellosis in Papua New Guinea

According to PNG's current national health plan for 2011-2020, diarrheal diseases are one of the five most frequent illnesses reported, with approximately 40 people per 1,000 presenting to hospital outpatients and/or health centres annually with diarrhoeal related illness (30). Over 200 people per 100,000 of the population are admitted for diarrheal related illnesses. Age specific data on the burden of diarrhoea are not provided in current health plan, but the burden is likely highest in children. Data published by the World Health Organization (WHO) (16) state that 9% of deaths occurring in children less than 5 years of age in 2012 were diarrhoea related. However, this could be an underestimation of the true burden of diarrhoeal disease in PNG because of poorly recorded and documented outpatient records. Indeed, other WHO published data state that diarrhoeal illness accounted for 14% of deaths in children aged 1 month to 5 years of age in 2013, with no appreciable decrease in the burden over the past 14 years (17). Moreover, the impact of diarrhoeal diseases on the general health of children, and therefore their susceptibility to other causes of death, may not be fully accounted for in these estimates.

Despite recognition that diarrhoea is an important cause of morbidity and mortality in PNG, particularly among children, data on the aetiology of diarrheal disease in PNG are lacking. The most robust data comes from a case-control study by Howard and colleagues (18); although the data are now dated with recruitment completed a quarter of a century ago. The study found *Shigella* to be one of the most important aetiologies for children hospitalised with diarrhoea in Goroka, Eastern Highlands Province; being detected in 13% of cases. Recently, surveillance by our research team, conducted in the same setting and a similar cohort (children hospitalised with diarrhoea in Goroka, EHP), found *Shigella* spp. and rotavirus to be the most commonly detected pathogens; with *Shigella* detected in 26.6% of children (19). In another recent study, focusing on both children and adults presenting with diarrhoea to hospital outpatients or an urban health clinic, *Shigella* was isolated from 22% of all study participants (20). On the basis of the aforementioned studies, it appears that *Shigella* remains an important pathogen in the Eastern Highlands of PNG.

Other studies provide evidence of the nationwide distribution of *Shigella*. The first case report of *Shigella* infection in PNG (which was then the territory of Papua & New Guinea under Australian administration) was by Morahan (1968). This study occurred over a 33 month period from March 1965 to November 1967. Faecal specimens were collected from the Wewak Hospital and other hospitals in the Sepik districts. A total of 907 stool samples were obtained from 848 persons, of which 90 were *Shigella* positive. *S. flexneri* was the most common species isolated (and was responsible for two fatal cases of shigellosis), followed by *S. sonnei* and then *S. boydii* (23).

In early 1987, Schuurkamp and colleagues (1990) monitored food handlers in PNG working at Ok Tedi Mine. The mining company had to re-enforce their pre-medical health checks on all potential food handlers following two cases of typhoid fever imported from the Highlands. An initial screen of 155 food handlers and 85 non-food handlers resulted in the isolation of *Shigella* sp. from 2.6% of food handlers and 3.5% of non-food handlers; all were asymptomatic carriers (24). A second survey of 160 food handlers including those from private fast food establishments detected one *S. boydii* infection in a local fast food shop staff member (24). *S. boydii* and *S. sonnei* were responsible for the majority of *Shigella* infections experienced on site (24). This study demonstrates that asymptomatic carriage of *Shigella* occurs in PNG: a finding supported by our recently obtained unpublished data.

In 2009, an epidemic caused by multi-drug resistant *S. flexneri* was reported in four provinces (22). This led to a review of antimicrobial susceptibility of *Shigella* spp. isolated at the Port Moresby General Hospital between 2000 and 2009. Of 3,419 fecal samples cultured, approximately 4% were positive for *Shigella*, with a majority of *S. flexneri* isolates showing antimicrobial resistance to amoxicillin, chloramphenicol and co-trimoxazole (22). However, there are little data from other parts of the country. In 2013 there was an outbreak of shigellosis in a settlement camp near Bulolo, Morobe Province. There were an estimated 1200 cases, with five fatalities (21), making it one of the largest outbreaks of shigellosis ever reported in the scientific literature.

Shigella is likely to be under-recognised as a cause of illness in PNG

As with many infectious diseases, the exact burden of shigellosis in PNG is unknown. Detailed aetiological data pertaining to two of the most important infectious diseases in PNG, namely pneumonia and diarrhoea, are lacking: as highlighted in previous editions of the Medical Journal in recent years (25-28). This is largely attributable to the country's lack of resources and diagnostic capabilities (29, 30).

Diagnosis is a challenge in PNG given the limited resources available in most health facilities; issues with storage and transportation of samples are also problematic, which contributes to the challenges faced with performing proper laboratory diagnosis. These issues are very important for *Shigella* diagnosis, particularly to the cultivation and isolation of the organism, which appears to be more susceptible to loss of viability in transport than many other enteric pathogens. Once faecal samples are collected they need to be processed within a few hours to maximise chances of recovery of the organism. Additionally, *Shigella's* close relationship with other members of the Enterobacteriaceae family makes it difficult to differentiate *Shigella* spp. from other closely related species including *E. coli*. Training for local health and laboratory staff in the proper collection, handling and storage of faecal samples, in conjunction with improved diagnostic procedures and techniques, could help improve *Shigella* spp. detection (30).

Approximately 87% of PNG's population live in rural settings (15); diarrhoeal diseases and acute respiratory infections occur at higher incidence in these rural communities (16). There is poor access to clean water and proper sanitation facilities, which increases the risk of communities acquiring gastro-intestinal illnesses (20, 28, 29). Access to general health services is difficult and nearly a third of all aid posts in the country have closed since independence in 1975 (20, 29, 31). Communities living in the urban poor regions of major towns and cities also experience similar problems.

On the basis of recent data, it would seem that: (a) *Shigella* remains an important gastrointestinal pathogen in PNG; (b) it is difficult to ascertain the exact burden of *Shigella* throughout the country due to the range of detection rates in recent studies (4% of stool samples in Port Moresby, compared to ~25% detection rate in Goroka) (19,20,22,32); however, when specific efforts are made to detect the pathogen (by culture or PCR, as conducted in Goroka) it is commonly isolated (19,20); and (c) there has been no appreciable decrease in the burden of shigellosis over the past two decades. Our recent

studies suggest shigellosis is the second leading contributor to diarrhoea in children hospitalised with diarrhoeal illness (19), and a leading cause of diarrhoea in children and adults (20).

Treatment

The optimal treatment and management of diarrhoeal diseases requires an understanding of the cause of the disease, which unfortunately is not currently obtainable in most settings in PNG. However, irrespective of the cause hydration is central in the treatment of diarrhoea, and is especially important in children. With this in mind, treating diarrhoea based on severity of illness is routinely conducted in PNG (33, 34) and can ensure adequate health outcomes in the absence of aetiological data. In children with mild diarrhoea with no signs of dehydration parents are advised to give extra fluids; oral rehydration solution (ORS) is recommended for children with moderate cases of diarrhoea and signs of dehydration; while intravenous fluids is recommended for patients with severe diarrhoea and severe dehydration. Continued breastfeeding of babies is strongly advised to reduce the severity of diarrhoea and complications of malnutrition, which can be experienced in infants and young children; and zinc supplementation is also recommended (33, 35, 36).

Antibiotics are overused in the treatment of diarrhoea (35, 36); however, evidence of shigellosis warrants their use. Such evidence usually comes in the form of dysentery. Ciprofloxacin is the recommended antibiotic for both children and adults with dysentery (33-36). Previously it was cotrimoxazole (or nalidixic acid or ampicillin) (37), but with the rise of antimicrobial resistance reported globally and also observed in recent studies in PNG the recommendations for the antibiotic treatment of shigellosis was updated (33,34). Two recent studies have reiterated the need for the change in antimicrobial treatment regimes. Rosewell and colleagues reported that *Shigella* was commonly resistant to amoxicillin, chloramphenicol and co-trimoxazole; these findings also concurred with the more recent study conducted by Greenhill and colleagues (20, 22).

The current practise of recommending treatment of dysentery with ciprofloxacin is appropriate given *Shigella* is the most common bacterial cause of dysentery; and circulating strains are commonly resistant to other antibiotics (20, 22). However,

shigellosis does not always result in dysenteric stools. Indeed, in our recent surveillance of children hospitalised with acute watery diarrhoea, children with dysentery were excluded from the surveillance; yet *Shigella* was detected in 26.6% of children. The standard treatment manual for children recommends chloramphenicol, if antibiotics are required. Our recent data (19, 20) suggests that the most common bacterial cause of diarrhoea (in the absence of dysentery) is *Shigella*. Given the high rates of resistance to chloramphenicol in *Shigella* (~80% resistance or intermediate resistance; Rosewell et al; Greenhill et al) (20, 22), treatment with chloramphenicol may be of little benefit to the patient.

Future considerations

Increased and sustained effort is required to lower the burden of all causes of diarrhoea in PNG and of shigellosis in particular. The importance of access to safe water, sanitation and hygiene (i.e. WASH) cannot be overstated. Challenges in the uptake of such measures exist (28, 38); but the impact on diarrhoeal illness, and indeed overall health, would be significant (28). Indeed, two recent large outbreaks of gastrointestinal illness in PNG, namely cholera and shigellosis, could have been avoided or drastically reduced with access and uptake of WASH in the country's rural majority and urban poor communities (21, 28). Vaccines, currently under development for *Shigella*, may form an important component of control strategies in the future.

Routine diagnosis of diarrhoeal disease is rarely conducted in PNG. Port Moresby General Hospital laboratory, is the only laboratory to regularly conduct diagnosis; even there an average of ~350 samples were tested annually between 2000 and 2009 (22). From a public health perspective there would be benefits for PNG if there was improved diagnostic capacity, as witnessed during the cholera outbreak (29, 30). Culture and sensitivity should be conducted in the major provincial capitals; with a view to introducing technologies suitable for use in resource poor settings in the future to supplement traditional diagnostic methods. Given the resistance of *Shigella* to multiple antibiotics, ongoing antimicrobial susceptibility testing is imperative in PNG.

Conclusions

A major impediment to control of diarrhoeal disease is the multiple aetiologies responsible. However, current evidence suggests that *Shigella* is a leading cause of diarrhoea (and dysentery) in PNG. Increasing our understanding of the epidemiology of shigellosis throughout PNG will help further our understanding of the burden of disease, and could assist patient management. Control of diarrhoeal diseases needs to be both broad and targeted in approach. Improved hygiene and sanitation would result in a decrease in the burden of all-cause diarrhoea and shigellosis alike; however such WASH interventions will take considerable time to implement. Treatment of diarrhoea according to the relevant guidelines is recommended. The primary focus of treatment should remain rehydration; however, the judicious use of antibiotics, preferably ciprofloxacin, is appropriate in probable cases of shigellosis (which may or may not present with blood or mucus in the stools).

References:

1 **Walker CL, Rudan I, Liu L, Nair H, Theodoratou E, Bhutta ZA, O'Brien KL, Campbell H, Black RE.** Global burden of childhood pneumonia and diarrhoea. *Lancet* 2013;381(9875):1405-1416.

2 **Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V, Abraham J, Ackerman I, Aggarwal R, Ahn SY, Ali MK, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Bahalim AN, Barker-Collo S, Barrero LH, Bartels DH, Basanez MG, Baxter A, Bell ML, Benjamin EJ, Bennett D, Bernabe E, Bhalla K, Bhandari B, Bikbov B, Bin Abdulhak A, Birbeck G, Black JA, Blencowe H, Blore JD, Blyth F, Bolliger I, Bonaventure A, Boufous S, Bourne R, Boussinesq M, Braithwaite T, Brayne C, Bridgett L, Brooker S, Brooks P, Brugha TS, Bryan-Hancock C, Bucello C, Buchbinder R, Buckle G, Budke CM, Burch M, Burney P, Burstein R, Calabria B, Campbell B, Canter CE, Carabin H, Carapetis J, Carmona L, Cella C, Charlson F, Chen H, Cheng AT, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahiya M, Dahodwala N, Damsere-Derry J, Danaei G, Davis A, De Leo D, Degenhardt L, Dellavalle R, Delossantos A, Denenberg J, Derrett S, Des Jarlais DC, Dharmaratne SD, Dherani M, Diaz-Torne C, Dolk**

H, Dorsey ER, Driscoll T, Duber H, Ebel B, Edmond K, Elbaz A, Ali SE, Erskine H, Erwin PJ, Espindola P, Ewoigbokhan SE, Farzadfar F, Feigin V, Felson DT, Ferrari A, Ferri CP, Fevre EM, Finucane MM, Flaxman S, Flood L, Foreman K, Forouzanfar MH, Fowkes FG, Fransen M, Freeman MK, Gabbe BJ, Gabriel SE, Gakidou E, Ganatra HA, Garcia B, Gaspari F, Gillum RF, Gmel G, Gonzalez-Medina D, Gosselin R, Grainger R, Grant B, Groeger J, Guillemin F, Gunnell D, Gupta R, Haagsma J, Hagan H, Halasa YA, Hall W, Haring D, Haro JM, Harrison JE, Havmoeller R, Hay RJ, Higashi H, Hill C, Hoen B, Hoffman H, Hotez PJ, Hoy D, Huang JJ, Ibeanusi SE, Jacobsen KH, James SL, Jarvis D, Jasrasaria R, Jayaraman S, Johns N, Jonas JB, Karthikeyan G, Kassebaum N, Kawakami N, Keren A, Khoo JP, King CH, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Laden F, Lalloo R, Laslett LL, Lathlean T, Leasher JL, Lee YY, Leigh J, Levinson D, Lim SS, Limb E, Lin JK, Lipnick M, Lipshultz SE, Liu W, Loane M, Ohno SL, Lyons R, Mabweijano J, MacIntyre MF, Malekzadeh R, Mallinger L, Manivannan S, Marcenes W, March L, Margolis DJ, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM, McAnulty JH, McDermott MM, McGill N, McGrath J, Medina-Mora ME, Meltzer M, Mensah GA, Merriman TR, Meyer AC, Miglioli V, Miller M, Miller TR, Mitchell PB, Mock C, Mocumbi AO, Moffitt TE, Mokdad AA, Monasta L, Montico M, Moradi-Lakeh M, Moran A, Morawska L, Mori R, Murdoch ME, Mwaniki MK, Naidoo K, Nair MN, Naldi L, Narayan KM, Nelson PK, Nelson RG, Nevitt MC, Newton CR, Nolte S, Norman P, Norman R, O'Donnell M, O'Hanlon S, Olives C, Omer SB, Ortblad K, Osborne R, Ozgediz D, Page A, Pahari B, Pandian JD, Rivero AP, Patten SB, Pearce N, Padilla RP, Perez-Ruiz F, Perico N, Pesudovs K, Phillips D, Phillips MR, Pierce K, Pion S, Polanczyk GV, Polinder S, Pope CA 3rd, Popova S, Porrini E, Pourmalek F, Prince M, Pullan RL, Ramaiah KD, Ranganathan D, Razavi H, Regan M, Rehm JT, Rein DB, Remuzzi G, Richardson K, Rivara FP, Roberts T, Robinson C, De Leon FR, Ronfani L, Room R, Rosenfeld LC, Rushton L, Sacco RL, Saha S, Sampson U, Sanchez-Riera L, Sanman E, Schwebel DC, Scott JG, Segui-Gomez M, Shahraz S, Shepard DS, Shin H, Shivakoti R, Singh D, Singh GM, Singh JA, Singleton J, Sleet DA, Sliwa K, Smith E, Smith JL, Stapelberg NJ, Steer A, Steiner T, Stolk WA, Stovner LJ,

Sudfeld C, Syed S, Tamburlini G, Tavakkoli M, Taylor HR, Taylor JA, Taylor WJ, Thomas B, Thomson WM, Thurston GD, Tleyjeh IM, Tonelli M, Towbin JA, Truelsen T, Tsilimbaris MK, Ubeda C, Undurraga EA, van der Werf MJ, van Os J, Vavilala MS, Venketasubramanian N, Wang M, Wang W, Watt K, Weatherall DJ, Weinstock MA, Weintraub R, Weisskopf MG, Weissman MM, White RA, Whiteford H, Wiebe N, Wiersma ST, Wilkinson JD, Williams HC, Williams SR, Witt E, Wolfe F, Woolf AD, Wulf S, Yeh PH, Zaidi AK, Zheng ZJ, Zonies D, Lopez AD, AlMazroa MA, Memish ZA. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380(9859):2197-2223.

3 **World Health Organization.** Working to overcome the global impact of neglected tropical diseases. First WHO report on neglected tropical diseases. Geneva: World Health Organization, 2010.

4 **Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omore R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acacio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM.** Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 2013;382(9888):209-222.

5 **Kotloff KL, Winickoff JP, Ivanoff B, Clemens JD, Swerdlow DL, Sansonetti PJ, Adak GK, Levine MM.** Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. *Bull World Health Organ* 1999;77(8):651-666.

6 **Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, Bennett D, Bhalla K, Bikbov B, Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S, Bucello C, Burch M, Burney P, Carapetis J, Chen H, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson**

KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahodwala N, De Leo D, Degenhardt L, Delossantos A, Denenberg J, Des Jarlais DC, Dharmaratne SD, Dorsey ER, Driscoll T, Duber H, Ebel B, Erwin PJ, Espindola P, Ezzati M, Feigin V, Flaxman AD, Forouzanfar MH, Fowkes FG, Franklin R, Fransen M, Freeman MK, Gabriel SE, Gakidou E, Gaspari F, Gillum RF, Gonzalez-Medina D, Halasa YA, Haring D, Harrison JE, Havmoeller R, Hay RJ, Hoen B, Hotez PJ, Hoy D, Jacobsen KH, James SL, Jasrasaria R, Jayaraman S, Johns N, Karthikeyan G, Kassebaum N, Keren A, Khoo JP, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Lipnick M, Lipshultz SE, Ohno SL, Mabweijano J, MacIntyre MF, Mallinger L, March L, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM, McAnulty JH, McDermott MM, McGrath J, Mensah GA, Merriman TR, Michaud C, Miller M, Miller TR, Mock C, Mocumbi AO, Mokdad AA, Moran A, Mulholland K, Nair MN, Naldi L, Narayan KM, Nasseri K, Norman P, O'Donnell M, Omer SB, Ortblad K, Osborne R, Ozgediz D, Pahari B, Pandian JD, Rivero AP, Padilla RP, Perez-Ruiz F, Perico N, Phillips D, Pierce K, Pope CA 3rd, Porrini E, Pourmalek F, Raju M, Ranganathan D, Rehm JT, Rein DB, Remuzzi G, Rivara FP, Roberts T, De Leon FR, Rosenfeld LC, Rushton L, Sacco RL, Salomon JA, Sampson U, Sanman E, Schwebel DC, Segui-Gomez M, Shepard DS, Singh D, Singleton J, Sliwa K, Smith E, Steer A, Taylor JA, Thomas B, Tleyjeh IM, Towbin JA, Truelsen T, Undurraga EA, Venketasubramanian N, Vijayakumar L, Vos T, Wagner GR, Wang M, Wang W, Watt K, Weinstock MA, Weintraub R, Wilkinson JD, Woolf AD, Wulf S, Yeh PH, Yip P, Zabetian A, Zheng ZJ, Lopez AD, Murray CJ, AlMazroa MA, Memish ZA. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380(9859):2095-2128.

7 **Niyogi SK.** Shigellosis. *J Microbiol* 2005;43(2):133-143.

8 **Lan R, Reeves PR.** Escherichia coli in disguise: molecular origins of Shigella. *Microbes Infect* 2002;4(11):1125-1132.

9 **Zaidi MB, Estrada-Garcia T.** *Shigella*: A Highly Virulent and Elusive Pathogen. *Curr Trop Med Rep* 2014;1(2):81-87.

10 **von Seidlein L, Kim DR, Ali M, Lee H, Wang X, Thiem VD, Canh do G, Chaicumpa W, Agtini MD, Hossain A, Bhutta ZA, Mason C, Sethabutr O, Talukder K, Nair GB, Deen JL, Kotloff K, Clemens J.** A multicentre study of *Shigella* diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. *PLoS Med* 2006;3(9):e353.

11 **Vinh H, Nhu NT, Nga TV, Duy PT, Campbell JI, Hoang NV, Boni MF, My PV, Parry C, Nga TT, Van Minh P, Thuy CT, Diep TS, Phuong le T, Chinh MT, Loan HT, Tham NT, Lanh MN, Mong BL, Anh VT, Bay PV, Chau NV, Farrar J, Baker S.** A changing picture of shigellosis in southern Vietnam: shifting species dominance, antimicrobial susceptibility and clinical presentation. *BMC Infect Dis* 2009;9:204.

12 **Ashkenazi S, May-Zahav M, Dinari G, Gabbay U, Zilberberg R, Samra Z.** Recent trends in the epidemiology of Shigella species in Israel. *Clin Infect Dis* 1993; 17(5):897-899.

13 **Department of Health and Human services, Victoria, Australia.** Bluebook - Infectious Diseases Epidemiology & Surveillance. Melbourne: Victorian Government Department of Human Services, 2005.

14 **Heymann DL.** Control of Communicable Diseases Manual, 20th edition. America: American Public Health Association, 2014.

15 **World Health Organization.** Papua New Guinea: WHO statistical profile. WHO/UN, 2012 (http://www.who.int/gho/countries/png.pdf?ua=1, accessed on 3 April 2015).

16 **World Health Organization,** Global Health Observatory Data Repository**.** Estimates for the cause of death among children aged <5 years, Papua New Guinea. WHO, 2013 (http://apps.who.int/gho/data/view.main.ghe300-PNG?lang=en, accessed on 17 July 2015).

17 **Howard P, Alexander ND, Atkinson A, Clegg AO, Gerega G, Javati A, Kajoi M, Lupiwa S, Lupiwa T, Mens M, Saleu G, Sanders RC, West B, Alpers MP.** Bacterial, viral and parasitic aetiology of paediatric diarrhoea in the highlands of Papua New Guinea. *J Trop Pediatr* 2000;46(1):10-14.

18 **Soli KW, Maure T, Kas MP, Bande G, Bebes S, Luang-Suarkia D, Siba PM, Morita A, Umezaki M, Greenhill AR, Horwood PF.** Detection of enteric viral and bacterial pathogens associated with paediatric diarrhoea in Goroka, Papua New Guinea. *Int J Infect Dis* 2014;27:54-58.

19 **Greenhill AR, Guwada C, Siba V, Michael A, Yoannes M, Wawarie Y, Ford R, Siba PM, Horwood PF.** Antibiotic resistant *Shigella* is a major cause of diarrhoea in the Highlands of Papua New Guinea. *J Infect Dev Ctries* 2014;8(11):1391-1397.

20 **Benny E, Mesere K, Pavlin BI, Yakam L, Ford R, Yoannes M, Kisa D, Abdad MY, Menda L, Greenhill AR, Horwood PF.** A large outbreak of shigellosis commencing in an internally displaced population, Papua New Guinea, 2013. *Western Pac Surveill Response J* 2014;5(3):18-21.

21 **Rosewell A, Ropa B, Posanai E, Dutta SR, Mola G, Zwi A, MacIntyre CR.** *Shigella* spp. antimicrobial drug resistance, Papua New Guinea, 2000-2009. *Emerg Infect Dis* 2010;16(11):1797-1799.

22 **Morahan RJ.** *Salmonella*, *Shigella* and enteropathogenic *Escherichia coli* isolations in the East and West Sepik Districts, Territory of Papua and New Guinea. *Med J Aust* 1968;2(10):438-440.

23 **Schuurkamp GJ, Bulungol PK, Kereu RK.** The monitoring of foodhandlers in Papua New Guinea. *P N G Med J* 1990;33(2):99-106.

24 **Toliman PJ, Guwada C, Soli KW.** A review of diarrhoea aetiology in Papua New Guinea, 1995-2012. *P N G Med J* 2013;56(3-4):145-155.

25 **Pomat WS, Greenhill AR, Lehmann D.** Pneumonia in Papua New Guinea: lessons learnt for the way forward. *P N G Med J* 2010;53(3-4):89-93.

26 **Kirkham LA, Smith-Vaughan HC, Greenhill AR.** Improving the aetiological diagnosis of bacterial pneumonia and meningitis in Papua New Guinea. *P N G Med J* 2010;53(3-4):139-146.

27 **Horwood PF, Barrington DJ, Greenhill AR.** More resources need to be committed to the provision of safe water sources and sanitation in Papua New Guinea. *P N G Med J* 2013;56(3-4):103-109.

28 **Horwood P, Greenhill A.** Cholera in Papua New Guinea and the importance of safe water sources and sanitation. *Western Pac Surveill Response J* 2012;3(1):3-5.

29 **Greenhill A, Rosewell A, Kas M, Manning L, Latorre L, Siba P, Horwood, P.** Improved laboratory capacity is required to respond better to future cholera outbreaks in Papua New Guinea. *Western Pac Surveill Response J* 2012;3(2):30-32.

30 **Papua New Guinea National Department of Health.** Papua New Guinea National Health Plan 2011-2020. Policies and Strategies Vol. 1. Port Moresby (Papua New Guinea): Department of Health, 2010.

31 **Kakazo M, Lehmann D, Coakley K, Gratten H, Saleu G, Taime J, Riley ID, Alpers MP.** Mortality rates and the utilization of health services during terminal illness in the Asaro Valley, Eastern Highlands Province, Papua New Guinea. *P N G Med J* 1999;42(1-2):13-26.

32 **Rosewell A, Dagina R, Murhekar M, Ropa B, Posanai E, Dutta S, Barr I, Mola G, Zwi A, MacIntyre CR.** Concurrent influenza and shigellosis outbreaks, Papua New Guinea, 2009. *Emerg Infect Dis* 2011;17(4):756-758.

33 **Papua New Guinea National Department of Health.** Standard treatment for common illnesses of children in Papua New Guinea. A manual for nurses, community health workers, health extension officers and doctors. Port Moresby: Department of Health, 2011.

34 **Papua New Guinea National Department of Health.** Standard treatment guidelines for adults. Port Moresby: Department of Health, 2012.

35 **Vince JD.** Diarrhoea in children in Papua New Guinea. *P N G Med J* 1995;38(4):262-271.

36 **Poka H, Duke T.** Clinical management of diarrhoea in children. 2013;56(3-4):156-161.

37 **Duke T.** Zinc sulphate for treatment and prevention of diarrhoea and other conditions in children in Papua New Guinea. *P N G Med J* 2011;54(1-2):17-22.

38 **Phuanukoonnon S, Namosha E, Kua L, Siba PM, Greenhill AR.** Evaluation of a WASH intervention demonstrates the potential for improved hygiene practices in Hiri District, Central Province. *P N G Med J* 2013;56(3-4):126-135.

Appendix 2 Chapter 3 Supplementary Material

Appendix 2.1 Questionnaire

Diarrhoea Questionnaire

Patient name: Bleeding code:

Date:

Section 1: General Sociological Data

No.	Questions	Coding categories	Codes
1.01	Patient sex	• Male • Female	1 2
1.02	Patient date of birth (if patient does not know day or month, please fill in year if possible)	• Date of birth (dd/mm/yy) • Don't know • No response • Estimated age (if don't know or no response)	___/___/___ 2 3 _____ years
1.03	Marital status	• Single (never married) • Married • Divorced/separated • Widow(er) • Other • No response	1 2 3 4 5 99
1.04	Which village/hamlet or town do you live in?	…… No response	99

1.05	Please provide details of where you live. For example: town section, street name and/or number, or description of where you live relative to well known landmark (in the village a store, tank or church might be a good landmark. In town a bank, a specific shop, the post office or Mt Kiss would be appropriate).	…… …… …… …… …… No response	99
1.06	Which languages are you able to speak, read or write (please circle as many as are appropriate)?	• Speak Pidgin • Read Pidgin • Write Pidgin • Speak English • Read English • Write English • No response	1 2 3 4 5 6 99
1.07	How much schooling have you have completed?	• No formal education • Grade 1 – 6 • Grade 7 – 10 • Grade 11 – 12 • Vocational/technical • University • Other • No response	1 2 3 4 5 6 7 99

Section 1: General Sociological Data (continued)

No.	Questions	Coding categories	Codes
1.08	Do you work for pay?	Yes	1
		No	2
		No response	99

Section 2: History of Illness (symptomatic participants only)

No.	Questions	Coding categories	Codes
2.01	How long have you had diarrhoea for?	Constantly for less than 2 days	1
		Constantly for 2 – 6 days	2
		Constantly for 1 week or longer	3
		On and off for less than 1 week	4
		On and off for 1 week or longer	5
		Don't know	88
		No response	99
2.02	Please describe the appearance/characteristics of your pekpek since the diarrhoea began (select as many as appropriate)	Very watery	1
		Blood in stools	2
		Mucus and/or pus in stools	3
		Green colour	4
		Very dark (black) colour	5
		Foul smell	6
		Don't know	88
		No response	99

2.03	Have you had any of the following symptoms during the time of this illness? (N.B. the patient might not be suffering from the symptom at the time of presentation, but may have been yesterday). **1**: Patient has the symptom now **2**: Patient had the symptom during this illness but not now **3**: Patient has not had the symptom **4**: Patient does not know if he/she has had that symptom, or does not respond	a) Vomiting b) Stomach pains c) Headache d) Nausea e) General aches and pains f) Decreased appetite g) Dehydration h) Jaundice i) Pallor j) Other (please describe) …… …… ……	1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4

Section 3: Treatment History

No.	Questions	Coding categories	Codes
3.01	Have you taken antibiotics since this diarrhoea started?	Yes No **(skip to Q 3.05)** Don't know **(skip to Q 3.05)** No response **(skip to Q 3.05)**	1 2 88 99

3.02	What was the name of the antibiotic?	Penicillin Amoxyl Chloramphenicol Septrim Other ……. Don't know No response	1 2 3 4 5 88 99
3.03	How long have you been taking antibiotics for?	<2 days 2 full days or longer Don't know No response	1 2 88 99
3.04	Where did you get the antibiotics from?	• Already had it at home • From a friend, relative, work colleague, etc. • Chemist • Store • Market/street vendor • Other…… • Don't know • No response	1 2 3 4 5 6 88 99
3.05	Have you taken oral rehydration solution since this diarrhoea started?	Yes No **(skip to Q 3.7)** Don't know **(skip to Q 3.7)** No response **(skip to Q 3.7)**	1 2 3 4
3.06	How long have you been taking the oral	<2 days 2 full days or longer Don't know	1 2 88

	rehydration solution for?	No response	99
3.07	Where did you get the oral rehydration solution from?	• Already had it at home • From a friend, relative, work colleague, etc. • Chemist • Store • Clinic/outpatients • Already had it at home • Homemade • Other…….. • Don't know • No response	1 2 3 4 5 6 7 8 88 99
3.08	Have you taken any other medicines (including traditional medicines)?	Yes No **(skip to Q 4.01)** Don't know **(skip to Q 4.01)** No response **(skip to Q 4.01)**	1 2 3 4
3.09	Provide the name or a brief description of the medication?	……. ……. ……. ……. Don't know No response	 88 99

Section 4: Sanitation and Hygiene

No.	Questions	Coding categories	Codes
4.01	Do you think any of these factors **increase** the risk of you getting diarrhoea (select as many as you think are important)?	• Eating certain food (see 4.02) • Poor quality drinking water • Traditional spells • Unprotected sex (sex without a condom) • Being unfaithful to your partner • Contact with pekpek • Not washing hands • There are no specific causes, it just happens from time to time • Don't know • No response	1 2 3 4 5 6 7 8 88 99
4.02	If you identified certain foods as the cause of diarrhoea, which of these foods are important? If study participant did not identify eating certain foods as important (in previous question), circle "no response" and go to question 4.03.	• Foods high in sugar (lollies, softdrinks, sweet biscuits, etc) Foods high in fat (fried foods, meat with a lot of fat on them e.g. lambflaps) • Food that was cooked a long time ago and is stored at room temperature • Salad and other uncooked food • Food purchased from street sellers • Food purchased from Kia bars • Food that smells bad (off food) • Other foods (please list)……. …… ……	1 2 3 4 5 6 7

		• Don't know • No response	88 99

Section 4 Sanitation and Hygiene (continued)

No.	Questions	Coding categories	Codes
4.03	Do you think any of these factors **decrease** the risk of you getting diarrhoea (select as many as you think are important)?	Personal hygiene (bathing) Hand washing Access to clean and safe drinking water Using latrines Access to health services Access to health education Mosquito nets Having a clean and tidy yard Don't know No response	1 2 3 4 5 6 7 8 88 99
4.04	How do you rate your own level of personal hygiene?	Excellent Good Not bad/Could be better Bad Don't know No response	1 2 3 4 88 99

4.05	How many people (in addition to yourself) do you share a house with? **INTERVIEWER PLEASE NOTE:** If the number of people living in the house changes, please obtain an approximation of what is the "usual number". Please ensure children are included in the total number.	0 (live alone) 1-2 3-5 6-9 10 or more Don't know No response	1 2 3 4 5 88 99
4.06	Have any of the people you share a house with suffered the same sickness in the two weeks?	Yes No **(skip to Q 4.08)** Don't know **(skip to Q 4.08)** No response **(skip to Q 4.08)**	1 2 88 99
4.07	How many people you share a house with have had this similar illness in the two weeks?	1 2 3-4 5 or more Don't know No response	1 2 3 4 88 99
4.08	Do you look after pigs in your house?	Yes No Don't know No response	1 2 88 99

4.09	Where do you get your drinking water from?	Piped water Tank water River/stream water Well water Spring water Other …….. Don't know No response	1 2 3 4 5 6 88 99

Section 4 Sanitation and Hygiene (continued)

No.	Questions	Coding categories	Codes
4.10	How far do you have to travel to collect water?	• It is accessible within the house • Just outside the house • Short walk • Long walk • Don't know • No response	1 2 3 4 88 99
4.11	Where do you go to the toilet for pekpek?	• Pit toilet • Bush disposal • Into river/stream • Flushing toilet • Other …….. • No response	1 2 3 4 5 99

4.12	How often do you wash your hands after going to the toilet (for pekpek)?	Usually Sometimes Never No response	1 2 3 99
4.13	How often do you wash your hands before preparing or eating a meal?	Usually Sometimes Never No response	1 2 3 99
4.14	When you wash your hands, how often do you use soap?	Usually Sometimes Never No response	1 2 3 99
4.15	How often is soap available in the house?	Usually Sometimes Never No response	1 2 3 99
4.16	When soap is available, do you always use it?	Usually Sometimes Never No response	1 2 3 99
4.17	When you wash your hands, how do you normally dry them?	On a towel On a cloth/rag On your own clothes Other …….. Don't know No response	1 2 3 4 88 99

4.18	How often do you have a wash (was was)?	Every day Every second day More than once a week Less than once a week Don't know No response	1 2 3 4 88 99

Section 4 Sanitation and Hygiene (continued)

No.	Questions	Coding categories	Codes
4.19	How often do you wash your clothes?	• Every 1-2 days • More than once a week • Weekly • Less than once a week • Don't know • No response	1 2 3 4 88 99
4.20	Which of the following would help you improve your own personal hygiene (mark as many as appropriate)?	• Improved access to clean safe drinking water • Improved access to toilets • Free soap • Education and public awareness campaigns • Don't know • No response	1 2 3 4 88 99
4.21	How/where do you think public awareness campaigns should be given (mark as many as appropriate)?	Schools Radio, newspaper, television, etc	1 2

		At stores where you buy soap	3
		By community leaders	4
		Church	5
		Don't know	88
		No response	99
4.22	How often do you attend church?	Every week	1
		Most weeks	2
		About half the time	3
		Not often	4
		Never, or very rarely	5
		Don't know	88
		No response	99
4.23	If your pastor told you it was God's will for you to have good personal hygiene and to wash regularly, would you be more likely to do so?	Definitely	1
		Maybe	2
		No	4
		Don't know	88
		No response	99

PNG IMR or GBH Staff use

This questionnaire was conducted by:

_____

Printed Name of Person Conducting Questionnaire

_____ _____

Signature of Person Conducting Questionnaire Date

Data entry checked by:

First entry by:

Second entry by:

Double checked by:

Appendix 3 Chapter 5 Supplementary Material

Appendix 3.1a Table with CLSI/Kirby Bauer and WGS results for AMR *S. flexneri* Oceania isolates

	Resistant	Beta-lactamases resistance genes presence/absence (WGS)				Resistant	Phenicol resistance genes presence/absence (WGS)	Resistant	Tetracycline resistance genes presence/absence (WGS)			Resistant	Trimethoprim resistance genes presence/absence (WGS)				
Sample ID	CLSI/Kirby Bauer	*AMPH_Ecoli_Bla*	*AmpC1_Ecoli_Bla*	*OXA_1_Bla*	*TEM−1D_Bla*	CLSI/Kirby Bauer	*CatA1_Phe*	CLSI/Kirby Bauer	*TetA_Tet*	*TetB_Tet*	*TetR_Tet*	CLSI/Kirby Bauer	*DfrA1_Tmt*	*DfrA3_Tmt*	*DfrA5_Tmt*	*DfrA8_Tmt*	Country
21528_4_32	-	+	-	-	-												PNG
21528_4_20	+	+	-	+	-	+	+	+	-	+	-						PNG
21528_4_48	+	+	-	+	-	+	+	+	-	+	-	+	-	-	+	-	PNG
21528_4_26	+	+	-	+	+	+	+	+	-	+	-	-	-	-	-	+	PNG
21528_4_47	+	+	-	+	-	-	+	-	-	+	-						PNG
21528_4_50	+	+	+	+	-	+	+	+	-	+	-	+	-	-	+	-	PNG
21528_4_62	+	+	-	+	-	+	+	+	-	+	-	-	-	-	+	-	PNG
21528_4_49	+	+	-	+	-	+	+	+	-	+	-	-	-	-	+	-	PNG
21528_4_34	+	+	+	+	-	+	+	+	-	+	-	-	-	-	+	-	PNG
21528_4_23	-	+	-	-	-												PNG
21528_4_33	+	+	-	+	-	+	+	+	-	+	-	-	-	-	+	-	PNG
21528_4_35	+	+	-	+	-	+	+	+	-	+	-						PNG
21528_4_21	+	+	+	+	+			+	-	+	-	-	-	-	+	-	Fiji
21528_4_24	-	+	-	-	-												PNG
21528_4_1	-	+	-	+	-	+	+	+	-	+	-	+	-	-	+	-	Torres Strait
21528_4_19	+	+	+	+	+	-	+	+	-	+	-	+	-	-	+	-	PNG
21528_4_4	+	+	-	+	-	+	+	+	-	+	-	+	-	-	+	-	PNG
21528_4_25	+	+	-	+	-	+	+	+	-	+	-	-	-	-	+	-	PNG
21528_4_51	-	+	+	+	-	-	+	-	-	+	-	-	-	-	+	-	PNG
21528_4_53	+	+	+	+	+	-	+	-	-	+	-	-	-	-	+	+	PNG
21528_4_52	+	+	+	+	+	-	+	-	-	+	-	+	-	-	+	+	PNG
21528_4_2	-	+	+	+	+	+	+	+	-	+	-	+	-	-	+	-	Torres Strait
21528_4_10	+	+	+	+	+	+	+	+	-	+	-	+	-	+	+	-	PNG
21528_4_9	+	+	-	+	-	-	+	+	-	+	-	+	-	-	+	-	PNG
21528_4_55	+	+	-	+	-	+	+	+	-	+	-	+	-	-	+	-	PNG
21528_4_60	+	+	-	+	-	+	+	+	-	+	-	-	-	-	+	-	PNG
21528_4_58	+	+	-	+	-	-	+	+	-	+	-	-	-	-	+	-	PNG
21528_4_57	+	+	-	+	-	-	+	+	-	+	-	+	-	-	+	-	PNG
21528_4_46	+	+	-	+	-	+	+	+	-	+	-	+	-	-	+	-	PNG
21528_4_61	+	+	-	+	-	-	+	-	-	+	-						PNG
21528_4_43	+	+	-	+	-	+	+	-	-	+	-						PNG
21528_4_42	+	+	-	+	-	+	+	+	-	+	-	+	-	-	+	-	PNG
21528_4_41	+	+	-	+	-	+	+	+	-	+	-						PNG
21528_4_56	-	+	-	+	-	-	+	-	-	+	-	+	-	-	+	-	PNG
21528_4_45	+	+	-	+	-	+	+	+	-	+	-						PNG
21528_4_44	+	+	-	+	-	+	+	+	-	+	-	-	-	-	+	-	PNG
21528_4_40	+	+	+	+	-	+	+	+	-	+	-	-	-	-	+	-	PNG
21528_4_59	+	+	-	+	-	+	+	+	-	+	-						PNG
21528_4_14	+	+	-	+	+	-	+	+	+	+	+	+	+	-	+	-	PNG
21528_4_39	+	+	-	+	-	+	+	+	-	+	-	+	+	-	-	-	PNG
21528_4_5	+	+	-	+	-			+	-	+	-	+	+	-	-	-	PNG

*Note: Isolates in 'highlighted' cells had no results.

Appendix 3.1b Table with CLSI/Kirby Bauer and WGS results for AMR *S. sonnei* Oceania isolates

		Resistant	Beta-lactamases resistance genes presence/absence (WGS)		Resistant	Trimethoprim resistance genes presence/absence (WGS)				Resistant	Phenicol resistance genes +/- (WGS)	Resistant	Tetracycline resistance genes presence/absence (WGS)		
Old sample ID	Sample ID	CLSI/Kirby Bauer	*OXA_1_Bla*	*TEM−1D_Bla*	CLSI/Kirby Bauer	*DfrA1_Tmt*	*DfrA3_Tmt*	*DfrA5_Tmt*	*DfrA8_Tmt*	CLSI/Kirby Bauer	*CatA1_Phe*	CLSI/Kirby Bauer	*TetA_Tet*	*TetB_Tet*	Country
QLD_26	21528_4_22	-	-	-	+	+	-	-	-	-	-	-	-	-	Solomon Islands
QLD_22	21528_4_18	-	+	+	+	-	+	+	+	-	+	-	-	+	Vanuatu
QLD_13	21528_4_11	-	-	-	-	-	-	-	-	-	-	-	-	-	Vanuatu
QLD_14	21528_4_12	+	-	+	+	+	-	-	-	-	-	+	+	-	Samoa
QLD_6	21528_4_6	-	+	+	-	+	-	-	+	-	-	-	-	-	Fiji
QLD_8	21528_4_7	-	+	+	-	+	-	-	+	-	-	-	-	+	Fiji
QLD_9	21528_4_8	-	+	-	-	-	-	+	-	-	+	-	-	+	Fiji
MDU_21	21528_4_30	-	-	-	+	+	-	-	-	-	-	+	-	-	PNG
MDU_19	21528_4_28	+	+	+	+	-	-	+	-	-	+	-	+	+	PNG
QLD_20	21528_4_16	+	-	+	+	-	-	+	-	-	-	-	-	-	PNG
QLD_3	21528_4_3	+	+	+	+	-	-	+	-	-	+	-	-	+	PNG
MDU_20	21528_4_29	+	+	+	+	-	-	+	-	-	-	-	-	+	PNG
QLD_19	21528_4_15	+	-	+	+	-	-	+	-	-	-	+	+	-	PNG
QLD_21	21528_4_17	+	-	+	+	-	-	+	-	-	-	-	-	-	PNG
MDU_22	21528_4_31	+	-	+	+	-	-	+	-	-	-	-	-	-	PNG
QLD_17	21528_4_13	+	+	+	+	-	-	+	+	-	-	-	+	+	PNG

Appendix 3.2 "Diagnostic" evaluation of WGS data for determination of AMR using CLSI testing as the gold standard

B lactam ampicillin

B-lactam	CLSI (GS)	WGS	
res	43	54	97
sus	14	3	17
	57	57	0
Sensitivity	0.754386	0.754386	0.754386
Specificity	0.052632	0.052632	0.052632
PPV	0.443299	0.443299	0.443299
NPV	0.176471	0.176471	0.176471
PDLR	0.796296	0.796296	
NDLR		4.666667	

B lactam ceft

B-lactam	CLSI (GS)	WGS	
res	0	54	54
sus	57	3	60
	57	57	0
Sensitivity	0	0	0
Specificity	0.052632	0.052632	0.052632
PPV	0	0	0
NPV	0.05	0.05	0.05
PDLR	0	0	
NDLR		19	

S. flex

Amp pb OXA & Bla

B-lactam	CLSI (GS)	WGS	
res	35	34	69
sus	6	7	13
	41	41	0
Sensitivity	0.853659	0.853659	0.853659
Specificity	0.170732	0.170732	0.170732
PPV	0.507246	0.507246	0.507246
NPV	0.538462	0.538462	0.538462
PDLR	1.029412	1.029412	
NDLR		0.857143	

Phenicol

Chloramphenicol	CLSI (GS)	WGS	
res	25	40	65
sus	32	17	49
	57	57	0
Sensitivity	0.438596	0.438596	0.438596
Specificity	0.298246	0.298246	0.298246
PPV	0.384615	0.384615	0.384615
NPV	0.346939	0.346939	0.346939
PDLR	0.625	0.625	
NDLR		1.882353	

S. flex

Amp predicted by Bla (OXA)

B-lactam	CLSI res +	CLSI Res - (sus)	
WGS res +	34	4	38
WGS Res - (sus)	0	3	3
	34	7	0
Sensitivity	1	1	1
Specificity	0.428571	0.428571	0.428571
PPV	0.894737	0.894737	0.894737
NPV	1	1	1
PDLR	1.75	1.75	
NDLR		0	

S. flex

Amp predicted by Bla (OXA) & Bla (AMPH E. coli)

B-lactam	CLSI res +	CLSI Res - (sus)	
WGS res +	34	4	38
WGS Res - (sus)	0	3	3
	34	7	0
Sensitivity	1	1	1
Specificity	0.428571	0.428571	0.428571
PPV	0.894737	0.894737	0.894737
NPV	1	1	1
PDLR	1.75	1.75	
NDLR		0	

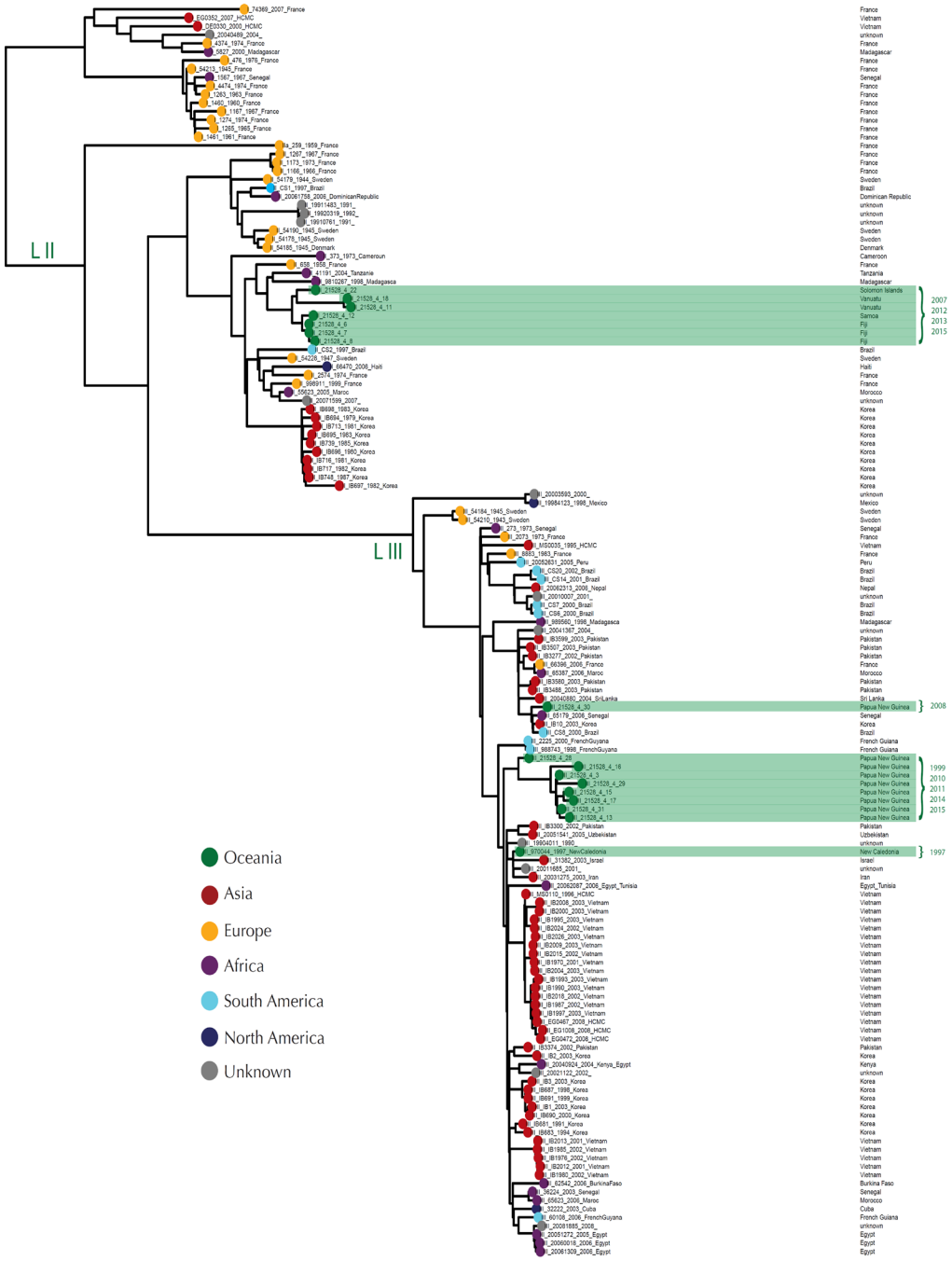
Tetracycline

Tetracylcine	CLSI (GS)	WGS	
res	33	47	80
sus	24	10	34
	57	57	0
Sensitivity	0.578947	0.578947	0.578947
Specificity	0.175439	0.175439	0.175439
PPV	0.4125	0.4125	0.4125
NPV	0.294118	0.294118	0.294118
PDLR	0.702128	0.702128	
NDLR		2.4	

Appendix 4 Chapter 6 Supplementary Material

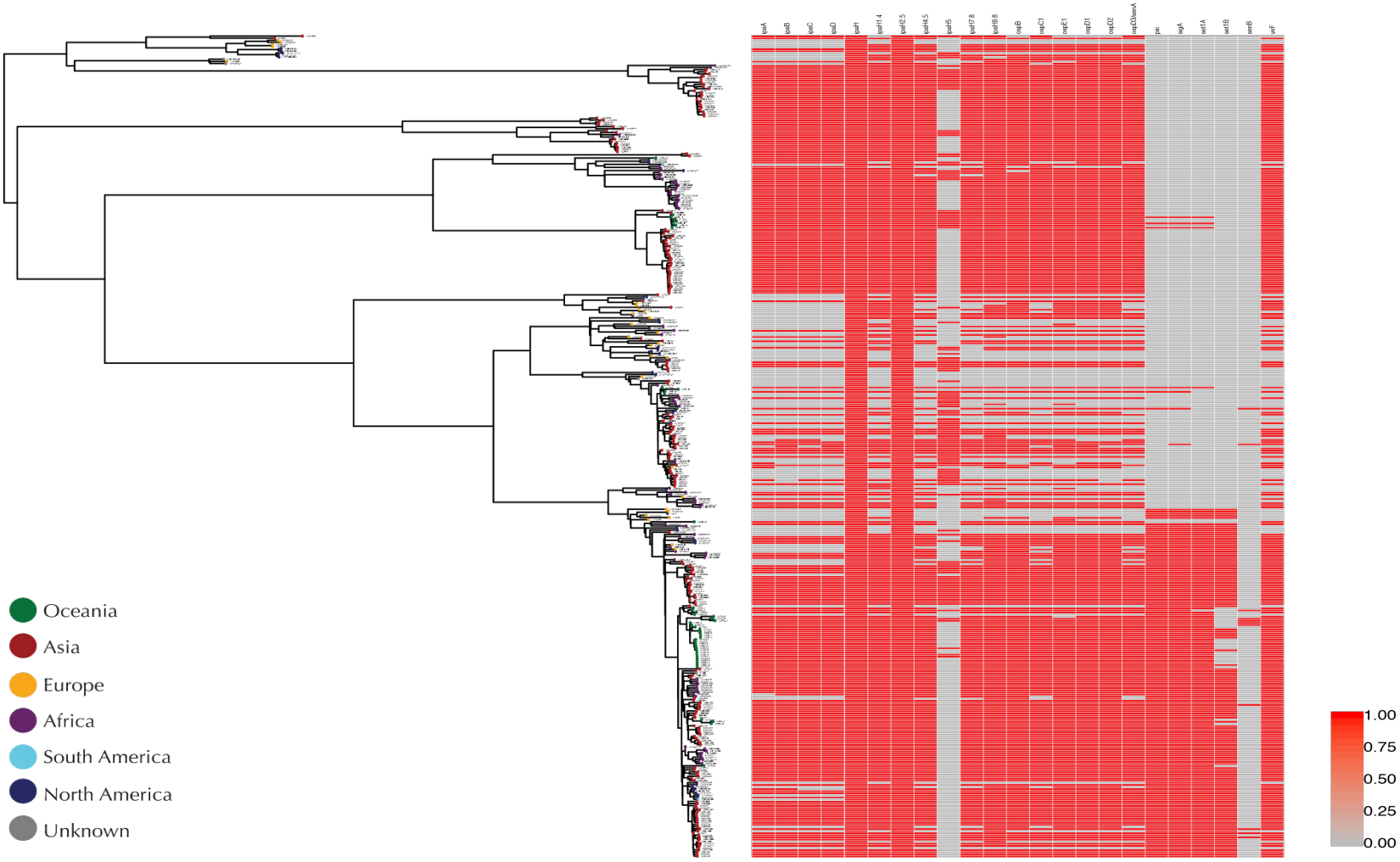
Appendix 4.1 Colour legend of regions globally for which *Shigella* isolates were identified to interpret the phylogenetic trees constructed in Chapter 6 and Appendices 4.2 – 4.5 and 4.7 – 4.8.

Appendix 4.2 *Shigella sonnei* phylogenetic tree featuring global and Oceania isolates

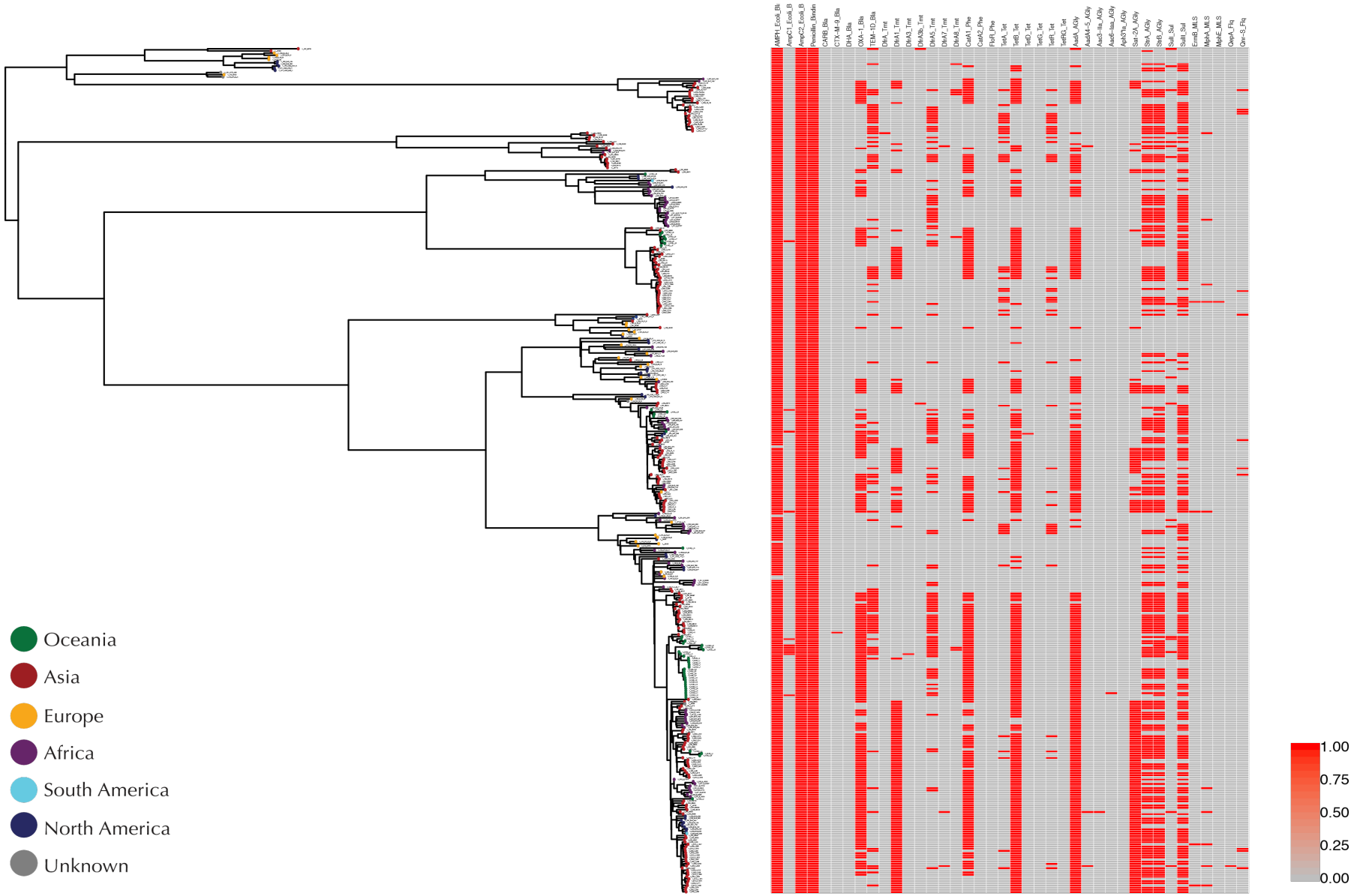


Appendix 4.3 *Shigella flexneri* phylogenetic tree featuring global and Oceania isolates

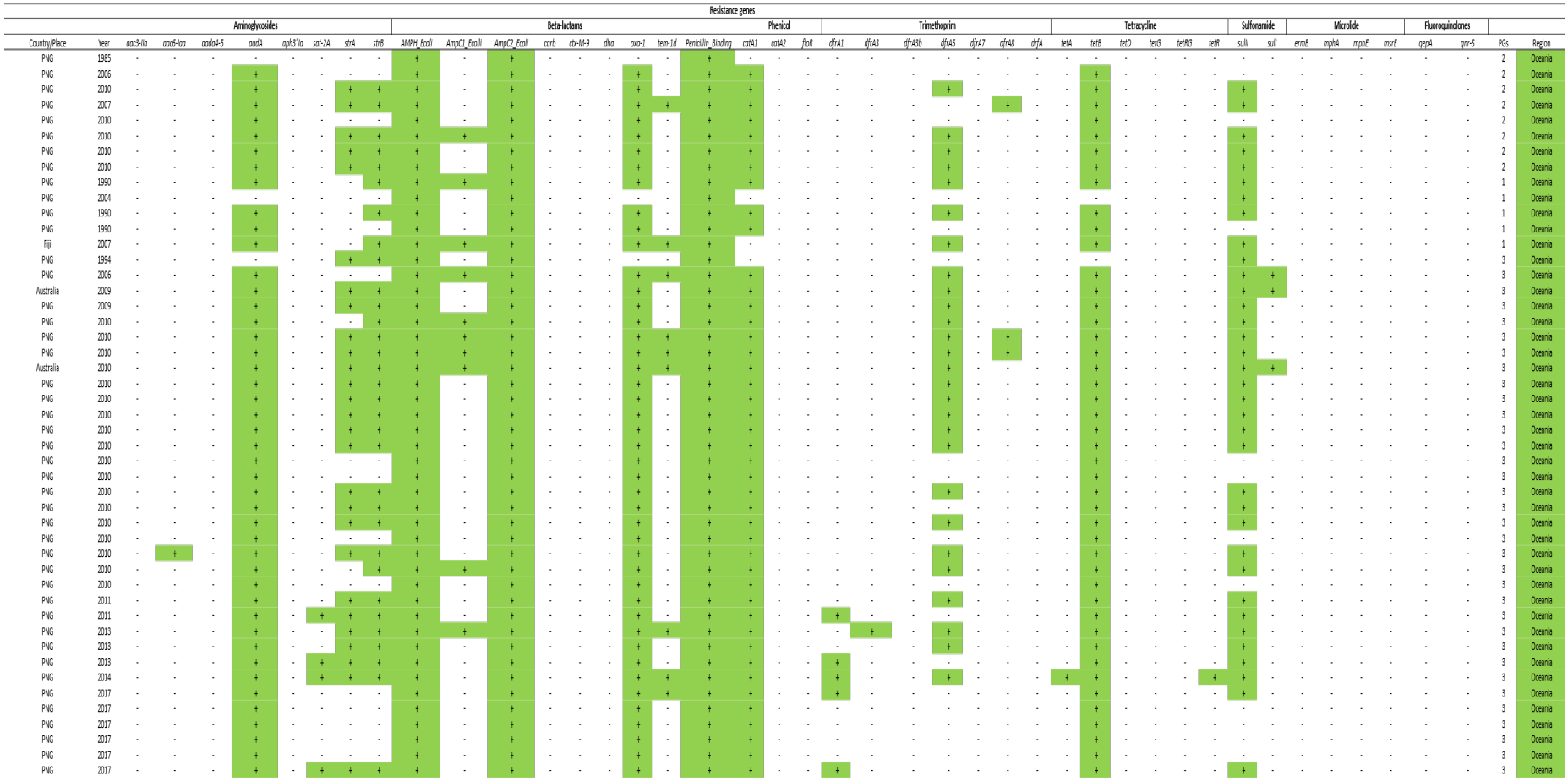
Appendix 4.4 *Shigella flexneri* phylogenetic tree with heatmap of presence/absence of virulence genes



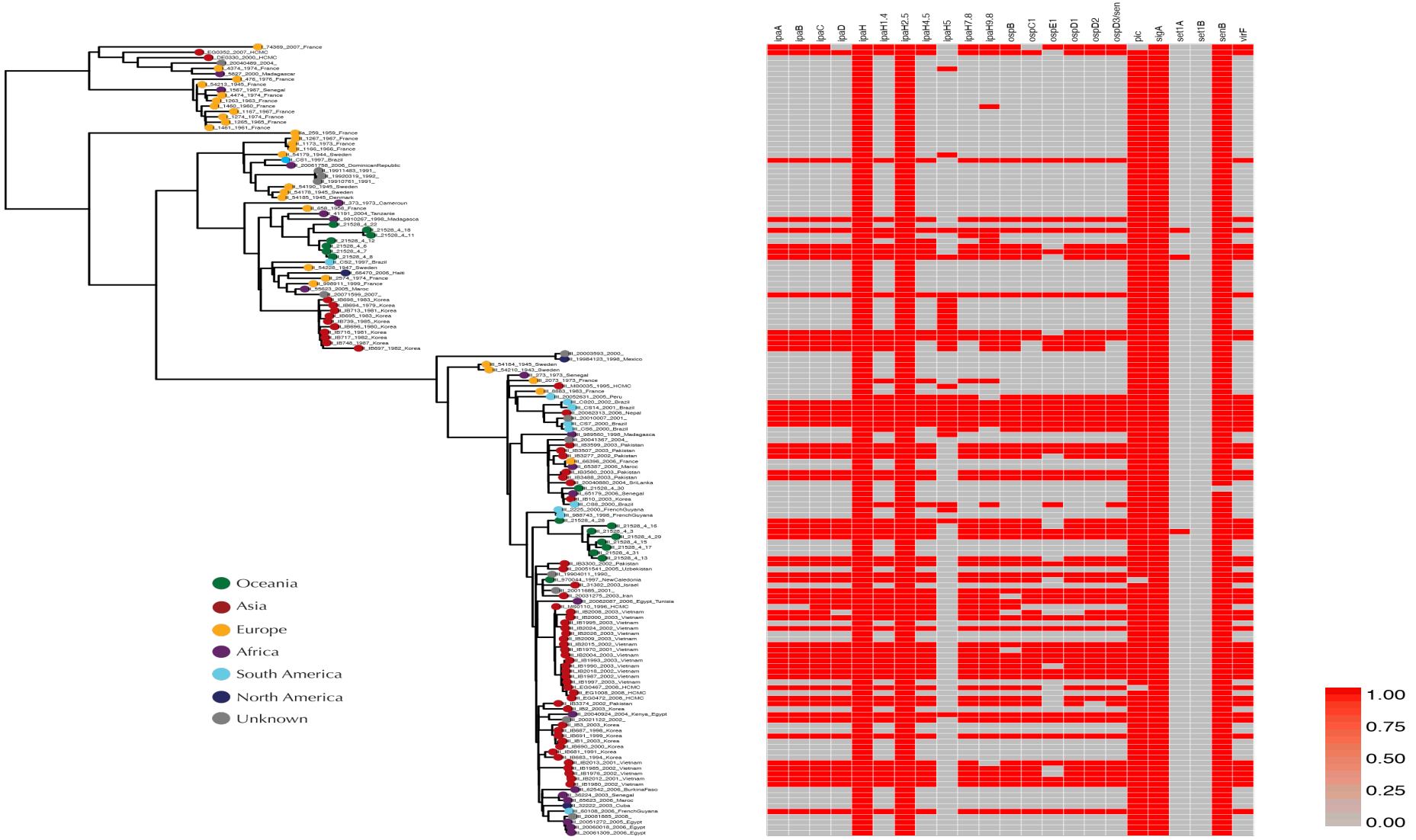
Appendix 4.5 *Shigella flexneri* phylogenetic tree with heatmap of presence/absence of AMR genes



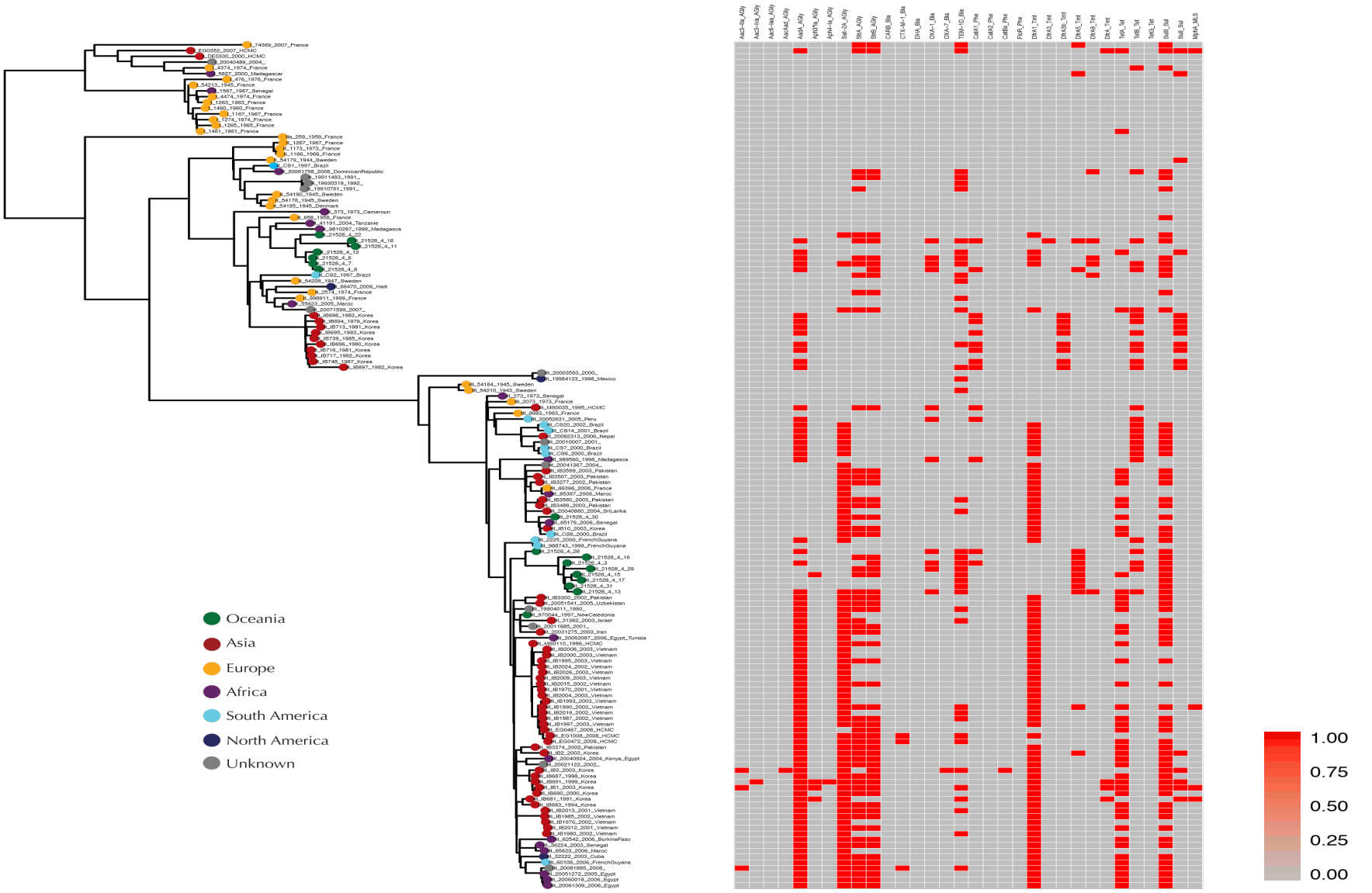
Appendix 4.6 Summary of AMR heatmap results in Appendix 3.4 for *S. flexneri* Oceania isolates



Appendix 4.7 *Shigella sonnei* phylogenetic tree with heatmap of presence/absence of virulence genes



Appendix 4.8 *Shigella sonnei* phylogenetic tree with heatmap of presence/absence of AMR genes



Appendix 4.9 Summary of AMR heatmap results in appendix 3.8 for *S. sonnei* Oceania isolates

		Resistance genes																																	
		Aminoglycosides										Beta-lactams						Phenicol				Trimethoprim						Tetracycline			Sulfonamide		Microlide		
Country/Place	Year	*aac3-IIa*	*aac3-Iva*	*aac6-Iaa*	*aacaad*	*aadA*	*aph3"Ia*	*aph4-Ia*	*sat-2A*	*strA*	*strB*	*carb*	*ctx-M-1*	*dha*	*oxa-1*	*oxa-7*	*tem-1d*	*catA1*	*catA2*	*catBx*	*floR*	*dfrA1*	*dfrA3*	*dfrA3b*	*dfrA5*	*dfrA8*	*dfrA*	*tetA*	*tetB*	*tetG*	*sulII*	*sul*	*mphA*	Lineage	Region
Solomon Islands	2007	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	2	Oceania
Vanuatu	2015	-	-	-	-	+	-	-	-	+	+	-	-	-	+	-	+	+	-	-	-	-	+	-	+	+	-	-	+	-	+	-	-	2	Oceania
Vanuatu	2013	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	Oceania
Samoa	2013	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	+	-	-	-	+	-	2	Oceania
Fiji	2012	-	-	-	-	+	-	-	-	+	+	-	-	-	+	-	+	-	-	-	-	+	-	-	-	+	-	-	-	-	+	-	-	2	Oceania
Fiji	2012	-	-	-	-	+	-	-	+	+	+	-	-	-	+	-	+	-	-	-	-	+	-	-	-	+	-	-	+	-	+	-	-	2	Oceania
Fiji	2012	-	-	-	-	+	-	-	-	-	+	-	-	-	+	-	-	+	-	-	-	-	-	-	+	-	-	-	+	-	+	-	-	2	Oceania
PNG	2008	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	+	-	-	3	Oceania
PNG	1999	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-	+	+	-	-	-	-	-	-	+	-	-	-	+	-	+	-	-	3	Oceania
PNG	2014	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	3	Oceania
PNG	2011	-	-	-	-	+	-	-	-	-	+	-	-	-	+	-	+	+	-	-	-	-	-	-	+	-	-	-	+	-	+	-	-	3	Oceania
PNG	2010	-	-	-	-	-	-	-	-	+	+	-	-	-	+	-	+	-	-	-	-	-	-	-	+	-	-	-	+	-	+	-	-	3	Oceania
PNG	2014	-	-	-	-	-	+	-	-	+	+	-	-	-	-	-	+	-	-	-	-	-	-	-	+	-	+	-	-	-	+	-	-	3	Oceania
PNG	2015	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	3	Oceania
PNG	2011	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	-	3	Oceania
PNG	2014	-	-	-	-	+	-	-	+	+	+	-	-	-	+	-	+	-	-	-	-	-	-	-	+	+	+	+	-	-	+	-	-	3	Oceania
New Caledonia	1997	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	3	Oceania

Appendix 4.10 Fisher's Exact Test (1 tail) for *Shigella sonnei* Global vs Pacific, Global vs PNG and Asian vs Pacific isolates statistic results

									Fisher's Exact Test (1 tail)			
AMR genes	Global isolates (n=147)		Asian isolates (n=57)		Pacific isolates (n=17)		PNG isolates (n=9)		Glob vs Pacific	Glob vs PNG	Asian vs Pacific	Asian vs PNG
Aminoglycosides	Total Positive	Total Negative	Total Positive	Total Negative	Total Positive	Total Negative	Total Positive	Total Negative		P - values		
aac3-IIa	3	144	2	55	0	17	0	9	0.72	0.84	0.59	0.74
aac3-IVa	1	146	1	56	0	17	0	9	0.9	0.94	0.77	0.86
aac6-Iaa	0	147	0	57	0	17	0	9	1	1	1	1
aacaad	1	146	1	56	0	17	0	9	0.9	0.94	0.77	0.86
aadA	78	69	47	10	9	8	3	6	0.61	0.21	0.02	0.005
aph3''Ia	4	143	3	54	1	16	1	8	0.43	0.26	0.78	0.45
aph4-Ia	1	146	1	56	0	17	0	9	0.9	0.94	0.77	0.86
sat-2A	74	73	43	14	5	12	2	7	0.083	0.10	0.0008	0.0034
strA	60	87	31	26	9	8	5	4	0.24	0.3	0.65	0.62
strB	61	86	31	26	11	6	6	3	0.06	0.13	0.32	0.38
Beta-lactams												
carb	0	147	0	57	0	17	0	9	1	1	1	1
ctx-M-1	3	144	2	55	0	17	0	9	0.72	0.84	0.59	0.74
dha	0	147	0	57	0	17	0	9	1	1	1	1
oxa-1	11	136	1	56	8	9	4	5	0.0001	0.0050	0.0000	0.0008
oxa-7	1	146	1	56	0	17	0	9	0.9	0.94	0.77	0.86
tem-1d	41	106	16	41	12	5	8	1	0.0007	0.0004	0.002	0.0009
Phenicol												
catA1	14	133	8	49	4	13	2	7	0.1	0.23	0.28	0.41
catA2	0	147	0	57	0	17	0	9	1	1	1	1
catBx	1	146	1	56	0	17	0	9	0.9	0.94	0.77	0.86
floR	0	147	0	57	0	17	0	9	1	1	1	1
Trimethoprim												
dfrA1	74	73	42	15	6	11	1	8	0.18	0.02	0.005	0.0006
dfrA3	1	146	0	57	1	16	0	9	0.2	0.94	0.23	1
dfrA3b	8	139	8	49	0	17	0	9	0.41	0.61	0.11	0.29
dfrA5	14	133	2	55	10	7	8	1	0.0000	0.0000	0.0000	0.0000
dfrA8	6	141	0	57	4	13	1	8	0.01	0.35	0.002	0.14
dfrA	4	143	4	53	0	17	0	9	0.64	0.79	0.34	0.55
Tetracycline												
tetA	47	100	28	29	4	13	3	6	0.34	0.69	0.05	0.3
tetB	27	120	9	48	7	10	4	5	0.04	0.08	0.03	0.07
tetG	0	147	0	57	0	17	0	9	1	1	1	1
Sulfonamide												
sulII	74	73	31	26	14	3	9	0	0.01	0.003	0.03	0.007
sulI	17	130	14	43	1	16	0	9	0.42	0.34	0.08	0.1
Microlide												
mphA	4	143	4	53	0	17	0	9	0.64	0.79	0.34	0.55

*Note: Cells highlighted in 'yellow' have a p-value <0.05 and cells highlighted in 'orange' have a p-value slightly over 0.05.

Appendix 4.11 Fisher's Exact Test (1 tail) for *Shigella flexneri* Global vs Pacific, Global vs PNG and Asian vs Pacific isolates statistic results

AMR genes	Global isolates (n=391)		Asia isolates (n=198)		Pacific isolates (n=47)		PNG isolates (n=44)		Fisher's Exact Test (1 tail)			
Aminoglycosides	Total Positive	Total Negative	Total Positive	Total Negative	Total Positive	Total Negative	Total Positive	Total Negative	Glob vs Pacific	Glob vs PNG	Asian vs Pacific	Asian vs PNG
									P - values			
aac3-IIa	1	390	1	197	0	47	0	44	0.89	0.90	0.81	0.82
aac6-Iaa	1	390	0	198	1	46	1	43	0.20	0.19	0.19	0.18
aada4-5	3	388	3	195	0	47	0	44	0.71	0.73	0.53	0.55
aadA	251	140	155	43	44	3	41	3	0.0000	0.0000	0.0089	0.0138
aph3"Ia	0	391	0	198	0	47	0	44	1	1	1	1
sat-2A	126	265	90	108	4	43	4	40	0.0003	0.0006	0.0000	0.0000
strA	218	173	131	67	27	20	25	19	0.48	0.51	0.17	0.16
strB	222	169	130	68	32	15	29	15	0.0911	0.16	0.45	0.56
Beta-lactams												
AMPH_Ecoli	386	5	198	0	47	0	44	0	0.57	0.59	1	1
AmpC1_Ecolli	11	380	1	197	10	37	8	36	0.0000	0.0002	0.0000	0.0000
AmpC2_Ecoli	391	0	198	0	47	0	44	0	1	1	1	1
carb	391	0	0	198	0	47	0	44	1	1	1	1
ctx-M-9	1	390	1	197	0	47	0	44	0.89	0.90	0.81	0.82
dha	0	391	0	198	0	47	0	44	1	1	1	1
oxa-1	194	197	107	91	44	3	41	3	0.0000	0.0000	0.0000	0.0000
tem-1d	91	300	74	124	9	38	7	37	0.33	0.18	0.0118	0.0040
Penicillin_Binding	391	0	198	0	47	0	44	0	1	1	1	1
Phenicol												
catA1	213	178	125	73	43	4	41	3	0.0000	0.0000	0.0001	0.0000
catA2	0	391	0	198	0	47	0	44	1	1	1	1
floR	0	391	0	198	0	47	0	44	1	1	1	1
Trimethoprim												
dfrA1	144	247	106	92	5	42	5	39	0.0001	0.0003	0.0000	0.0000
dfrA3	1	390	0	198	1	46	1	43	0.20	0.19	0.19	0.18
dfrA3b	2	389	2	196	0	47	0	44	0.80	0.81	0.65	0.67
dfrA5	110	281	50	148	27	20	24	20	0.0001	0.0005	0.0000	0.0002
dfrA7	3	388	3	195	0	47	0	44	0.71	0.73	0.53	0.55
dfrA8	7	384	3	195	3	44	3	41	0.0811	0.0700	0.0866	0.0753
drfA	1	390	1	197	0	47	0	44	0.89	0.90	0.81	0.82
Tetracycline												
tetA	44	347	36	162	1	46	1	43	0.0320	0.0419	0.0021	0.0033
tetB	244	147	143	55	44	3	41	3	0.0000	0.0000	0.0008	0.0014
tetD	1	390	0	198	0	47	0	44	0.89	0.90	1	1
tetG	0	391	0	198	0	47	0	44	1	1	1	1
tetRG	0	391	0	198	0	47	0	44	1	1	1	1
tetR	44	347	36	162	1	46	1	43	0.0320	0.0419	0.0021	0.0033
Sulfonamide												
sulII	249	142	143	55	35	12	32	12	0.0947	0.15	0.46	0.55
sulI	18	373	11	187	3	44	1	43	0.40	0.41	0.52	0.32
Microlide												
ermB	4	387	4	194	0	47	0	44	0.63	0.65	0.42	0.45
mphA	10	381	8	190	0	47	0	44	0.32	0.34	0.18	0.20
mphE	1	390	1	197	0	47	0	44	0.89	0.90	0.81	0.82
msrE	1	390	1	197	0	47	0	44	0.89	0.90	0.81	0.82
Fluoroquinolones												
qepA	1	390	1	197	0	47	0	44	0.89	0.90	0.81	0.82
qnr-S	12	379	12	186	0	47	0	44	0.25	0.27	0.0726	0.0845

*Note: Cells highlighted in 'yellow' have a p-value <0.05 and cells highlighted in 'orange' have a p-value slightly over 0.05.

Appendix 4.12 Table depicting Serotyping of *S. flexneri* Oceania isolates detecting presence of O antigens and observing the presence of serotype-specific genes.

Sample ID	*wzx*	*gtrI*	*gtrIC*	*gtrII*	*oac*	*gtrIV*	*gtrV*	*gtrX*	*Trad	*Mol	PGs	Country	Year
21528_4#21	+	+	-	-	+	-	-	-	-	1b	PG1	FIJI	2007
21528_4#23	+	+	-	-	-	-	-	-	-	1a	PG1	PNG	2004
21528_4#33	+	+	-	-	+	-	-	-	1	1b	PG1	PNG	1990
21528_4#34	+	+	-	-	+	-	-	-	1	1b	PG1	PNG	1990
21528_4#35	+	+	-	-	+	-	-	-	1	1b	PG1	PNG	1990
21528_4#20	+	-	-	-	+	-	-	+	-	3a	PG2	PNG	2006
21528_4#26	+	-	-	-	+	-	-	+	-	3a	PG2	PNG	2007
21528_4#32	+	-	-	-	+	-	-	+	3	3a	PG2	PNG	1990
21528_4#47	+	-	-	-	+	-	-	+	-	3a	PG2	PNG	2010
21528_4#48*	+	-	-	+	+	-	-	+	-	Unknown	PG2	PNG	2010
21528_4#49	+	-	-	-	+	-	-	+	-	3a	PG2	PNG	2010
21528_4#50	+	-	-	-	+	-	-	+	-	3a	PG2	PNG	2010
21528_4#62	+	-	-	-	+	-	-	+	-	3a	PG2	PNG	2010
21528_4#1	+	-	-	+	-	-	-	-	2a	2a	PG3	AUST	2009
21528_4#10	+	-	-	+	-	-	-	-	2a	2a	PG3	PNG	2013
21528_4#14	+	-	-	+	-	-	-	+	X	2b	PG3	PNG	2014

21528_4#19	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2006
21528_4#2	+	-	-	+	-	-	-	-	2a	2a	PG3	AUST	2010
21528_4#24	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	1994
21528_4#25	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2009
21528_4#39	+	-	-	+	-	-	-	+	2a	2b	PG3	PNG	2013
21528_4#4	+	-	-	+	-	-	-	-	2a	2a	PG3	PNG	2011
21528_4#40	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#41	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#42	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#43	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#44*	+	-	-	+	+	-	-	-	-	Unknown	PG3	PNG	2010
21528_4#45	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#46	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#5	+	-	-	+	-	-	-	-	2a	2a	PG3	PNG	2011
21528_4#51	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#52	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#53	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#55	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#56	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010

21528_4#57	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#58	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#59	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#60	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#61	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2010
21528_4#9	+	-	-	+	-	-	-	-	2a	2a	PG3	PNG	2013
25499_1#1	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2017
25499_1#2	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2017
25499_1#4	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2017
25499_1#5	+	-	-	+	-	-	-	+	-	2b	PG3	PNG	2017
25499_1#6	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2017
25499_1#7	+	-	-	+	-	-	-	-	-	2a	PG3	PNG	2017

Confirmed via *Trad = Traditional serotyping methods and *Mol = Molecular serotyping methods

Appendix 4.13 *Shigella flexneri* Oceania and Global WGS metadata

TreeID	Isolate	country	year	region	Cover%_AE005674	Depth_AE005674	Mapped%_AE005674	Mapped%_Total	Total_Reads	SNPs	Indels
III_21528_4_44	21528_4#44	Papua New Guinea	2010	Oceania	99.62	166.40	81.37	81.37	7656188	1199	125
II_21528_4_48	21528_4#48	Papua New Guinea	2010	Oceania	99.91	196.14	87.53	87.53	8521414	904	119
III_21528_4_41	21528_4#41	Papua New Guinea	2010	Oceania	99.68	196.87	90.51	90.51	8136580	1197	122
I_21395_4_23	21395_4#23	Papua New Guinea	2004	Oceania	92.34	197.10	88.06	88.06	8163286	2908	156
II_21395_4_47	21395_4#47	Papua New Guinea	2010	Oceania	93.97	186.45	84.85	84.85	8257444	8041	326
III_21395_4_24	21395_4#24	Papua New Guinea	1994	Oceania	97.17	153.66	64.52	64.52	9413898	1025	104
III_21395_4_46	21395_4#46	Papua New Guinea	2010	Oceania	99.84	194.93	84.27	84.27	8995772	1360	115
III_21528_4_46	21528_4#46	Papua New Guinea	2010	Oceania	99.85	194.27	86.35	86.35	8443480	1450	121
II_21395_4_32	21395_4#32	Papua New Guinea	1990	Oceania	94.25	210.97	83.77	83.77	9385238	7702	313
III_21395_4_51	21395_4#51	Papua New Guinea	2010	Oceania	99.54	160.55	81.89	81.89	7826120	1299	143
III_21395_4_19	21395_4#19	Papua New Guinea	2006	Oceania	99.52	175.14	87.55	87.55	7813042	1147	113
III_21528_4_1	21528_4#1	Australia	2009	Oceania	98.96	186.30	80.40	80.40	8616896	1140	121
III_21395_4_4	21395_4#4	Papua New Guinea	2011	Oceania	99.30	183.11	83.28	83.28	8506146	1118	112
I_21528_4_21	21528_4#21	Fiji	2007	Oceania	99.12	150.27	82.71	82.71	6827558	3750	195
II_21528_4_32	21528_4#32	Papua New Guinea	1990	Oceania	94.25	209.46	85.94	85.94	8788638	7860	364
III_21528_4_59	21528_4#59	Papua New Guinea	2010	Oceania	99.10	165.23	92.71	92.71	6610002	1202	130
III_21528_4_5	21528_4#5	Papua New Guinea	2011	Oceania	99.05	172.03	83.99	83.99	7624280	1200	131
III_21395_4_43	21395_4#43	Papua New Guinea	2010	Oceania	99.31	176.58	85.34	85.34	8187320	1146	119
II_21395_4_50	21395_4#50	Papua New Guinea	2010	Oceania	99.90	162.45	83.20	83.20	7869744	4868	235
III_21528_4_42	21528_4#42	Papua New Guinea	2010	Oceania	99.47	204.90	89.86	89.86	8511064	1187	115
I_21528_4_35	21528_4#35	Papua New Guinea	1990	Oceania	97.05	199.88	91.47	91.47	7933492	3548	184
III_21528_4_10	21528_4#10	Papua New Guinea	2013	Oceania	99.63	163.05	85.94	85.94	7211498	1104	104
III_21395_4_1	21395_4#1	Australia	2009	Oceania	98.96	187.27	78.41	78.41	9208954	1104	114
II_21395_4_49	21395_4#49	Papua New Guinea	2010	Oceania	99.91	163.13	84.18	84.18	7741640	6335	246
II_21528_4_49	21528_4#49	Papua New Guinea	2010	Oceania	99.90	165.07	86.97	86.97	7264812	6200	258

I_21395_4_21	21395_4#21	Fiji	2007	Oceania	99.11	147.59	80.00	80.00	7253034	3712	175
III_21395_4_41	21395_4#41	Papua New Guinea	2010	Oceania	99.69	190.31	87.06	87.06	8644744	1151	118
I_21395_4_34	21395_4#34	Papua New Guinea	1990	Oceania	99.94	182.93	87.66	87.66	8243208	1286	128
II_21528_4_50	21528_4#50	Papua New Guinea	2010	Oceania	99.90	166.60	86.59	86.59	7366532	4748	244
III_21395_4_52	21395_4#52	Papua New Guinea	2010	Oceania	99.79	147.80	72.35	72.35	8145644	1421	145
III_21528_4_14	21528_4#14	Papua New Guinea	2014	Oceania	99.68	193.07	85.92	85.92	8440894	1039	119
III_21395_4_61	21395_4#61	Papua New Guinea	2010	Oceania	99.31	175.62	90.06	90.06	7564346	1148	119
II_21528_4_47	21528_4#47	Papua New Guinea	2010	Oceania	94.00	187.92	87.62	87.62	7734282	8287	377
III_21528_4_58	21528_4#58	Papua New Guinea	2010	Oceania	99.30	185.76	91.96	91.96	7501346	1191	124
III_21528_4_57	21528_4#57	Papua New Guinea	2010	Oceania	99.30	201.03	92.08	92.08	8106868	1182	123
II_21528_4_62	21528_4#62	Papua New Guinea	2010	Oceania	94.36	183.88	86.68	86.68	7659006	8342	377
III_21395_4_2	21395_4#2	Australia	2010	Oceania	99.59	179.24	85.77	85.77	8266072	1088	111
III_21395_4_25	21395_4#25	Papua New Guinea	2009	Oceania	96.41	194.72	89.44	89.44	8436040	1118	104
III_21528_4_39	21528_4#39	Papua New Guinea	2013	Oceania	98.67	200.32	90.03	90.03	8226544	1574	129
II_21395_4_48	21395_4#48	Papua New Guinea	2010	Oceania	99.91	198.27	85.58	85.58	9080118	842	107
III_21528_4_55	21528_4#55	Papua New Guinea	2010	Oceania	99.31	212.49	90.85	90.85	8691946	1192	112
III_21528_4_51	21528_4#51	Papua New Guinea	2010	Oceania	99.53	165.39	85.11	85.11	7366670	1307	155
III_21528_4_52	21528_4#52	Papua New Guinea	2010	Oceania	99.78	150.22	74.95	74.95	7616398	1384	165
III_21528_4_24	21528_4#24	Papua New Guinea	1994	Oceania	97.20	155.24	67.06	67.06	8784762	1081	108
III_21395_4_60	21395_4#60	Papua New Guinea	2010	Oceania	99.30	168.07	89.82	89.82	7217804	1137	121
III_21528_4_9	21528_4#9	Papua New Guinea	2013	Oceania	99.18	182.11	91.20	91.20	7427914	1186	130
I_21395_4_35	21395_4#35	Papua New Guinea	1990	Oceania	97.08	199.35	88.73	88.73	8528332	3377	159
III_21528_4_25	21528_4#25	Papua New Guinea	2009	Oceania	96.41	208.72	94.41	94.41	7944678	1176	107
III_21395_4_9	21395_4#9	Papua New Guinea	2013	Oceania	99.18	173.30	86.93	86.93	7927774	1133	125
I_21395_4_33	21395_4#33	Papua New Guinea	1990	Oceania	99.90	206.44	89.11	89.11	9162032	3299	146
III_21528_4_40	21528_4#40	Papua New Guinea	2010	Oceania	99.56	203.30	91.41	91.41	8292174	1207	118
III_21528_4_19	21528_4#19	Papua New Guinea	2006	Oceania	99.50	177.10	90.38	90.38	7315266	1167	117
I_21528_4_33	21528_4#33	Papua New Guinea	1990	Oceania	99.90	212.18	92.69	92.69	8559264	3392	171
III_21528_4_4	21528_4#4	Papua New Guinea	2011	Oceania	99.30	181.32	85.22	85.22	7962828	1150	124

III_21528_4_43	21528_4#43	Papua New Guinea	2010	Oceania	99.32	186.01	88.88	88.88	7818810	1203	126
II_21395_4_62	21395_4#62	Papua New Guinea	2010	Oceania	94.34	161.18	80.12	80.12	8035386	8024	333
II_21395_4_26	21395_4#26	Papua New Guinea	2007	Oceania	94.50	161.97	83.07	83.07	7397242	7991	339
III_21395_4_5	21395_4#5	Papua New Guinea	2011	Oceania	99.03	175.36	81.69	81.69	8323920	1162	116
III_21395_4_59	21395_4#59	Papua New Guinea	2010	Oceania	99.10	160.94	89.52	89.52	7015350	1152	122
III_21395_4_58	21395_4#58	Papua New Guinea	2010	Oceania	99.30	181.12	88.75	88.75	7984982	1140	110
III_21395_4_14	21395_4#14	Papua New Guinea	2014	Oceania	99.71	189.05	83.26	83.26	8911794	1008	110
III_21395_4_40	21395_4#40	Papua New Guinea	2010	Oceania	99.55	202.01	88.89	88.89	8824560	1153	108
III_21528_4_45	21528_4#45	Papua New Guinea	2010	Oceania	99.31	179.71	91.11	91.11	7344732	1173	125
I_21528_4_34	21528_4#34	Papua New Guinea	1990	Oceania	99.94	187.60	90.93	90.93	7730662	1269	126
III_21528_4_2	21528_4#2	Australia	2010	Oceania	99.59	183.26	88.91	88.91	7763318	1142	110
III_21395_4_57	21395_4#57	Papua New Guinea	2010	Oceania	99.33	198.59	89.32	89.32	8635016	1134	118
III_21395_4_45	21395_4#45	Papua New Guinea	2010	Oceania	99.31	175.21	87.92	87.92	7809650	1143	119
III_21395_4_56	21395_4#56	Papua New Guinea	2010	Oceania	99.30	194.33	88.19	88.19	8653838	1144	113
III_21395_4_42	21395_4#42	Papua New Guinea	2010	Oceania	99.47	199.58	86.56	86.56	9073996	1136	116
III_21395_4_44	21395_4#44	Papua New Guinea	2010	Oceania	99.62	161.79	78.29	78.29	8152722	1146	118
I_21528_4_23	21528_4#23	Papua New Guinea	2004	Oceania	92.47	202.81	91.51	91.51	7653268	2994	168
III_21395_4_55	21395_4#55	Papua New Guinea	2010	Oceania	99.31	210.14	88.17	88.17	9257978	1133	109
II_21528_4_20	21528_4#20	Papua New Guinea	2006	Oceania	94.14	176.83	86.50	86.50	7405952	8325	393
III_21395_4_53	21395_4#53	Papua New Guinea	2010	Oceania	99.86	137.93	72.70	72.70	7604898	1112	123
II_21528_4_26	21528_4#26	Papua New Guinea	2007	Oceania	94.50	166.58	85.95	85.95	7040974	8253	384
III_21528_4_53	21528_4#53	Papua New Guinea	2010	Oceania	99.86	141.81	75.64	75.64	7127536	1125	133
III_21528_4_56	21528_4#56	Papua New Guinea	2010	Oceania	99.32	200.49	91.62	91.62	8138254	1202	115
III_21528_4_60	21528_4#60	Papua New Guinea	2010	Oceania	99.30	168.63	92.08	92.08	6799618	1166	125
III_21395_4_10	21395_4#10	Papua New Guinea	2013	Oceania	99.63	160.25	82.95	82.95	7691292	1066	100
II_21395_4_20	21395_4#20	Papua New Guinea	2006	Oceania	94.22	171.37	82.87	82.87	7919122	8039	349
III_21528_4_61	21528_4#61	Papua New Guinea	2010	Oceania	99.30	179.41	92.75	92.75	7186902	1185	127
III_21395_4_39	21395_4#39	Papua New Guinea	2013	Oceania	98.68	195.11	86.92	86.92	8728478	1494	119
III_2002_Sh02_1115	ERR042796	Gabon	2002	Africa	98.07	97.31	95.68	95.68	6464586	1814	142

III_2002_Sh02_3648	ERR042799	Mayotte	2002	Africa	97.60	136.15	97.37	97.37	8823968	1096	113
III_2002_Sh02_8484	ERR042801	Haiti	2002	North America	97.85	108.25	96.21	96.21	7148406	1090	105
III_2003_Sh03_0850	ERR042802	Haiti	2003	North America	97.42	107.14	88.25	88.25	7685384	1072	106
III_2003_Sh03_1034	ERR042803	Haiti	2003	North America	98.40	134.62	96.14	96.14	8991064	1085	105
III_2003_Sh03_2590	ERR042804	Madagascar	2003	Africa	98.63	126.20	96.69	96.69	8334222	1048	103
III_2004_Sh04_1706	ERR042805	Dominican Republic	2004	North America	98.66	123.09	93.63	93.63	8459040	1081	111
I_2004_Sh04_4915	ERR042807	Algeria	2004	Africa	94.48	79.07	92.20	92.20	5275650	3260	196
I_2004_Sh04_9300	ERR042809	Haiti	2004	North America	94.86	107.23	92.51	92.51	7209340	3278	192
I_2005_Sh05_1382	ERR042812	Mali	2005	Africa	97.92	93.74	95.57	95.57	6213874	3429	213
III_2005_Sh05_2089	ERR042815	Dominican Republic	2005	North America	98.15	108.66	98.04	98.04	7068438	1122	119
III_2005_Sh05_4557	ERR042816	Haiti	2005	North America	98.36	105.71	96.48	96.48	7029326	1101	112
II_2005_Sh05_6782	ERR042819	Haiti	2005	North America	94.11	103.23	90.61	90.61	7082402	7941	371
III_2005_Sh05_7137	ERR042821	Mayotte	2005	Africa	98.55	105.05	97.43	97.43	6911348	1092	115
II_2005_Sh05_8094	ERR042822	Mali	2005	Africa	90.43	109.71	90.57	90.57	7178706	7500	332
III_2005_Sh05_9228	ERR042824	French Guiana	2005	South America	98.90	98.30	95.63	95.63	6643274	1120	119
VII_2006_Sh06_1976	ERR042827	Ivory Coast	2006	Africa	92.97	113.22	86.98	86.98	8126566	8387	370
II_2006_Sh06_3638	ERR042829	Niger	2006	Africa	94.22	99.97	90.98	90.98	6805944	7910	376
VI_2007_Sh07_0130	ERR042832	Madagascar	2007	Africa	92.75	153.60	86.35	86.35	10957646	8635	376
II_2007_Sh07_4848	ERR042836	Togo	2007	Africa	94.31	125.16	91.83	91.83	8409006	8026	379
II_2007_Sh07_4943	ERR042837	Algeria	2007	Africa	94.28	122.28	92.98	92.98	8119224	7885	355
III_2007_Sh07_5876	ERR042839	Mayotte	2007	Africa	98.94	122.07	94.10	94.10	8351398	1191	112
III_2007_Sh07_6497	ERR042840	Ivory Coast	2007	Africa	98.20	123.52	93.80	93.80	8410798	1819	141
III_2007_Sh07_7606	ERR042841	Madagascar	2007	Africa	99.04	139.71	95.70	95.70	9390226	1118	107
III_2007_Sh07_7746	ERR042842	Guadeloupe	2007	North America	99.32	187.19	95.59	95.59	12595422	1102	114
IV_2008_Sh08_0350	ERR042843	Dominican Republic	2008	North America	96.05	118.43	87.75	87.75	8531028	7136	328
III_2008_Sh08_1372	ERR042845	Senegal	2008	Africa	97.99	132.09	92.57	92.57	9097260	1944	139
I_2008_Sh08_2354	ERR042846	Burkina Faso	2008	Africa	96.82	126.90	95.58	95.58	8312222	3309	200
II_2008_Sh08_3483	ERR042848	French Guiana	2008	South America	95.04	56.85	92.38	92.38	3834382	7998	385
I_2008_Sh08_4122	ERR042849	Madagascar	2008	Africa	96.39	123.54	94.82	94.82	8111164	3238	193

III_2008_Sh08_5585	ERR042850	Haiti	2008	North America	99.06	138.98	92.64	92.64	9645340	1123	120
I_2009_Sh09_0167	ERR042851	Senegal	2009	Africa	96.77	117.37	92.16	92.16	8006696	3384	199
I_2009_Sh09_1522	ERR042852	Algeria	2009	Africa	97.68	113.04	94.84	94.84	7517148	3123	209
I_2009_Sh09_1985	ERR042853	Ivory Coast	2009	Africa	96.63	147.01	95.74	95.74	9590530	3335	207
II_2009_Sh09_4409	ERR042855	Togo	2009	Africa	95.24	111.49	87.75	87.75	7911116	7951	374
VII_2009_Sh09_5516	ERR042858	Congo	2009	Africa	93.99	113.05	83.67	83.67	8342738	8735	392
III_2009_Sh09_6787	ERR042860	Comores	2009	Africa	99.03	112.65	92.63	92.63	7771934	1116	113
III_2009_Sh09_6999	ERR042861	Cameroon	2009	Africa	98.89	102.90	91.85	91.85	7176190	1834	147
III_2009_Sh09_9407	ERR042863	Mali	2009	Africa	98.03	98.73	90.12	90.12	6937978	1785	151
I_2010_H66	ERR047236	Vietnam	2010	Asia	96.65	111.29	61.90	61.90	11180092	3213	191
II_2010_KH49	ERR047280	Vietnam	2010	Asia	94.92	121.25	82.10	82.10	9355814	8046	369
I_1995_MS0019	ERR047294	Vietnam	1995	Asia	96.88	149.87	89.88	89.88	10687242	3345	200
VII_1995_MS0022	ERR047297	Vietnam	1995	Asia	93.40	147.03	86.76	86.76	10530066	8328	378
III_1995_MS0050	ERR047306	Vietnam	1995	Asia	99.55	74.06	91.78	91.78	5252392	1067	115
VII_1995_MS0052	ERR047307	Vietnam	1995	Asia	96.32	181.41	85.39	85.39	13673068	8257	342
I_1995_MS0059	ERR047309	Vietnam	1995	Asia	97.55	228.36	86.47	86.47	16848352	3355	131
III_2000_DE0010	ERR047346	Vietnam	2000	Asia	99.63	141.58	78.45	78.45	12083226	1074	116
III_2000_DE0030	ERR047348	Vietnam	2000	Asia	99.62	153.53	84.85	84.85	12043900	1064	107
III_2000_DE0569	ERR047365	Vietnam	2000	Asia	99.57	171.80	84.52	84.52	13451100	1072	107
VII_2001_DE0804	ERR047372	Vietnam	2001	Asia	94.37	133.96	70.16	70.16	12303248	8362	380
III_2001_DE0874	ERR047374	Vietnam	2001	Asia	99.32	324.91	84.94	84.94	25984088	1097	18
II_2002_DE1173	ERR047376	Vietnam	2002	Asia	95.32	200.25	75.62	75.62	17088684	7912	266
II_2002_DE1455	ERR047384	Vietnam	2002	Asia	95.88	19.50	53.40	53.40	2355288	7034	304
II_2008_EG0198	ERR047394	Vietnam	2008	Asia	94.38	141.19	87.97	87.97	10291928	7999	355
II_2006_EG0302	ERR047396	Vietnam	2006	Asia	95.08	271.69	89.10	89.10	19850630	8106	126
III_2007_EG0387	ERR047401	Vietnam	2007	Asia	99.49	95.63	90.30	90.30	6938682	1099	128
II_2007_EG0403	ERR047402	Vietnam	2007	Asia	95.22	123.94	90.53	90.53	8742064	7969	364
III_2008_EG0449	ERR047406	Vietnam	2008	Asia	99.48	62.61	94.32	94.32	4345514	1092	118
III_2008_EG0461	ERR047407	Vietnam	2008	Asia	99.53	36.86	93.25	93.25	2604282	1091	120

III_2009_H10	ERR047436	Vietnam	2009	Asia	99.51	138.80	86.21	86.21	10578270	1131	122
I_1996_MS0084	ERR048234	Vietnam	1996	Asia	96.74	141.14	85.89	85.89	10161018	3320	200
III_1996_MS0085	ERR048235	Vietnam	1996	Asia	99.58	165.82	90.99	90.99	11626846	1069	118
I_1996_MS0118	ERR048245	Vietnam	1996	Asia	94.35	211.38	88.43	88.43	14558252	3181	169
III_1996_MS0125	ERR048246	Vietnam	1996	Asia	97.54	186.02	95.75	95.75	12129578	1120	118
V_1951_IB1657	ERR048265	UK	1951	Europe	97.02	151.84	95.80	95.80	9793462	4739	260
V_1951_IB1658	ERR048266	UK	1951	Europe	96.75	123.84	89.76	89.76	8470328	4766	256
V__IB1664	ERR048272	US		North America	96.63	142.84	88.66	88.66	9840858	4802	252
V__IB1665	ERR048273	US		North America	95.41	154.03	94.34	94.34	9904918	4517	251
III__IB1667	ERR048275	US		North America	98.27	151.35	94.27	94.27	9961222	1119	118
II__IB1685	ERR048281	Vietnam		Asia	94.43	166.37	89.18	89.18	12248678	7945	371
III__IB1788	ERR048283	Vietnam		Asia	98.27	128.39	77.10	77.10	10445262	1129	118
IV__IB1835	ERR048285	Vietnam		Asia	96.05	153.74	87.73	87.73	11199724	7005	333
I__IB1885	ERR048286	Vietnam		Asia	95.01	135.27	83.99	83.99	9693358	3211	199
V_1998_IB4230	ERR048287	India	1998	Asia	96.22	119.97	78.95	78.95	10587228	5190	254
VI_2000_IB4235	ERR048288	India	2000	Asia	92.34	131.51	82.80	82.80	9423024	8911	396
I_2002_IB3328	ERR048290	Pakistan	2002	Asia	96.76	119.04	62.38	62.38	13008270	3240	194
I_2002_IB3350	ERR048291	Pakistan	2002	Asia	97.12	93.63	64.88	64.88	10070960	3298	195
III_2002_IB3356	ERR048292	Pakistan	2002	Asia	99.10	154.37	92.74	92.74	11334508	1075	111
II_2002_IB3389	ERR048294	Pakistan	2002	Asia	94.96	231.73	90.54	90.54	15692656	7998	286
III_2002_IB3391	ERR048295	Pakistan	2002	Asia	98.73	162.16	91.98	91.98	10996654	1414	119
II_2002_IB3396	ERR048296	Pakistan	2002	Asia	94.89	157.55	80.73	80.73	12158128	7967	352
I_1983_IB0700	ERR048299	Korea	1983	Asia	95.80	131.17	86.25	86.25	9209802	3224	199
IV_1987_IB0708	ERR048300	Korea	1987	Asia	96.22	163.36	88.82	88.82	11264026	7174	362
III_1999_IB0711	ERR048302	Korea	1999	Asia	98.27	52.80	84.76	84.76	4029778	1073	112
III_2003_IB0712	ERR048303	Korea	2003	Asia	99.27	133.27	97.67	97.67	8666252	1080	116
III_1981_IB0714	ERR048304	Korea	1981	Asia	99.14	129.99	98.34	98.34	10445700	1026	103
I_1982_IB0719	ERR048305	Korea	1982	Asia	94.38	95.20	72.71	72.71	8165060	2857	182
III_1985_IB0737	ERR048306	Korea	1985	Asia	98.59	161.51	96.64	96.64	10362144	1027	109

VII_2002_IB1808	ERR048309	Vietnam	2002	Asia	93.38	121.76	85.41	85.41	8673096	8303	382
III_2001_IB1848	ERR048310	Vietnam	2001	Asia	99.13	156.67	91.51	91.51	10734140	1076	117
VII_2003_IB1896	ERR048311	Vietnam	2003	Asia	92.16	165.48	85.37	85.37	11446190	8251	401
VII_2002_IB2494	ERR048312	Korea	2002	Asia	93.32	151.07	80.58	80.58	11216894	8645	394
III_2002_IB2495	ERR048313	Korea	2002	Asia	99.08	161.41	94.54	94.54	10765270	1143	117
VII_1991_IB0017	ERR048317	Korea	1991	Asia	89.67	99.53	79.83	79.83	7367302	7851	372
III_2002_IB0022	ERR048318	China	2002	Asia	98.12	131.20	86.00	86.00	12502786	1051	107
III_2002_IB0034	ERR048319	China	2002	Asia	98.79	111.49	83.21	83.21	8966122	1128	115
III_2002_IB0036	ERR048320	China	2002	Asia	98.95	172.16	89.26	89.26	12053192	1343	128
III_2002_IB0037	ERR048321	China	2002	Asia	98.65	155.57	92.76	92.76	10992100	1093	117
III_2002_IB0047	ERR048326	Taiwan	2002	Asia	99.25	97.00	69.40	69.40	10044230	1107	112
I_1977_NCDC_9768	ERR832494	US?	1977	North America	95.31	83.13	97.16	97.16	4442460	2862	193
III_2003_IB0068	ERR048327	Taiwan	2003	Asia	98.95	187.01	94.89	94.89	12295408	1076	114
I_1949_CIP_52_37	ERR832491	None	1949	None	96.87	85.74	97.70	97.70	4498176	2770	188
VII_2002_IB1716	ERR048328	Vietnam	2002	Asia	92.47	145.90	83.60	83.60	10333138	8263	401
IV_1977_NCDC_2825_H	ERR832488	US?	1977	North America	96.27	102.20	89.14	89.14	6042504	7082	318
III_1913_CIP_52_35	ERR832489	UK	1913	Europe	98.65	92.63	99.00	99.00	4872598	994	108
III_2001_IB1757	ERR048330	Vietnam	2001	Asia	99.02	197.42	92.39	92.39	13355104	1089	121
I_1989_E_179_89	ERR832486	Egypt	1989	Africa	97.76	88.27	95.61	95.61	4926412	2985	197
III_1989_E_324_89	ERR832487	Madagascar	1989	Africa	97.95	92.53	96.94	96.94	5199732	925	99
III_2001_IB1762	ERR048331	Vietnam	2001	Asia	99.02	131.49	92.46	92.46	8921714	1093	116
VII_2002_IB1773	ERR048332	Vietnam	2002	Asia	93.72	158.64	81.91	81.91	11750108	8431	365
I_1975_NCDC_361_75	ERR832482	US?	1975	North America	98.38	100.17	96.92	96.92	5356680	2925	201
IV_1949_CIP_52_24	ERR832490	None	1949	None	96.82	82.03	89.42	89.42	4750932	6593	358
VII__IB1743	ERR048334	Vietnam		Asia	92.71	162.90	79.14	79.14	12237040	8232	389
III_2002_IB3325	ERR048337	Pakistan	2002	Asia	99.13	193.08	91.87	91.87	13197426	1361	118
III__IB0058	ERR048338	None		None	99.16	141.26	90.23	90.23	9822418	1093	118
III_2002_IB0064	ERR048339	Sri Lanka	2002	Asia	98.79	207.15	92.04	92.04	14068544	1367	116
III_1975_E_14_75	ERR832481	France	1975	Europe	96.43	113.85	99.05	99.05	5972362	1514	132

II_2010_E_10262	ERR049152	Vietnam	2010	Asia	96.25	430.49	83.89	83.89	31744204	8066	30
III_2009_E_333687	ERR126957	South Africa (Western Cape)	2009	Africa	99.04	108.29	93.79	93.79	5970806	1115	113
III_2009_E_370076	ERR126958	South Africa (KwaZulu-Natal)	2009	Africa	99.35	93.14	91.90	91.90	5160610	905	109
III_1974_NCDC_1170_74	ERR832480	US?	1974	North America	99.68	82.39	95.21	95.21	4660014	1036	109
III_2009_E_400525	ERR126960	South Africa (Western Cape)	2009	Africa	99.05	97.18	91.02	91.02	5583228	1153	115
III_2009_E_410370	ERR126961	South Africa (Gauteng)	2009	Africa	99.20	104.01	93.85	93.85	5713870	1127	113
I_1978_E_262_78	ERR832483	France	1978	Europe	97.67	109.09	97.40	97.40	5829392	2921	193
II_2009_E_319743	ERR126963	South Africa (Eastern Cape)	2009	Africa	94.27	90.37	88.99	88.99	5117630	7777	356
IV_1970_NCDC_5797_70	ERR832473	US?	1970	North America	96.10	83.10	87.43	87.43	5086792	6975	317
III_1970_Yah__104_70	ERR832474	Cameroon	1970	Africa	98.44	107.15	94.47	94.47	6056858	1634	131
I_1971_NCDC_1862_71	ERR832475	US?	1971	North America	97.89	103.28	86.46	86.46	6360828	2937	204
I_1971_NCDC_1921_71	ERR832476	US?	1971	North America	98.14	87.31	91.55	91.55	4987930	2891	201
I_1972_NCDC_5612_72	ERR832479	US?	1972	North America	96.99	108.55	97.60	97.60	5777424	2895	195
II_2009_E_395205	ERR126967	South Africa (Gauteng)	2009	Africa	93.90	102.17	83.78	83.78	6155994	7608	360
III_2010_E_441003	ERR126986	South Africa (Western Cape)	2010	Africa	99.20	96.33	94.21	94.21	5380962	1128	115
III_2010_E_466310	ERR126987	South Africa (KwaZulu-Natal)	2010	Africa	99.13	89.76	86.55	86.55	5463918	1135	118
III_2010_E_461633	ERR126988	South Africa (Gauteng)	2010	Africa	99.05	82.47	92.06	92.06	4789720	1127	118
I_1970_NCDC_1242_70	ERR832472	US?	1970	North America	97.70	93.60	86.26	86.26	5747708	3056	194
III_2010_E_453094	ERR126989	South Africa (Western Cape)	2010	Africa	99.04	79.90	77.74	77.74	5476932	1132	119
I_1969_NCDC_880_69	ERR832471	US?	1969	North America	97.88	87.44	86.89	86.89	5280212	3116	205
III_2010_E_467746	ERR126990	South Africa (KwaZulu-Natal)	2010	Africa	99.02	86.56	90.59	90.59	4942320	1132	117
III_1966_NCDC_4570_66	ERR832467	US?	1966	North America	97.96	86.72	93.70	93.70	4926014	977	107
II_1968_NCDC_2979_68	ERR832469	US?	1968	North America	93.96	103.10	89.37	89.37	6002622	7407	353
I_1969_Dakar_948_69	ERR832470	Senegal	1969	Africa	96.41	103.52	95.95	95.95	5661124	2853	188
II_2010_E_457776	ERR126994	South Africa (Western Cape)	2010	Africa	94.10	98.32	89.03	89.03	5672882	7702	362
I_1966_NCDC_4173_66	ERR832466	US?	1966	North America	95.28	107.04	98.49	98.49	5612758	2809	179
II_2010_E_459364	ERR126995	South Africa (KwaZulu-Natal)	2010	Africa	94.18	112.20	88.17	88.17	6338938	7767	352
II_1966_NCDC_2146_66	ERR832465	US?	1966	North America	94.54	85.32	88.48	88.48	5094642	7366	354
III_2011_E_579748	ERR127012	South Africa (Gauteng)	2011	Africa	99.29	105.14	92.52	92.52	6015836	1146	116

V_1966_CIP_67_61	ERR832464	Iraq	1966	Asia	96.60	68.88	91.48	91.48	4309886	4692	258
III_2011_E_550865	ERR127013	South Africa (KwaZulu-Natal)	2011	Africa	99.29	116.31	93.43	93.43	6533840	1149	113
III_2011_E_527649	ERR127014	South Africa (Western Cape)	2011	Africa	99.26	103.87	92.76	92.76	5989768	1182	127
V_1961_NCDC_6154_61	ERR832459	US?	1961	North America	95.66	93.99	90.79	90.79	5359706	4668	247
I_1963_E_5_63	ERR832460	Vietnam	1963	Asia	96.03	100.06	96.85	96.85	5330596	2774	183
III_2011_E_546567	ERR127016	South Africa (Western Cape)	2011	Africa	99.13	107.48	91.69	91.69	6190566	1161	113
III_2011_E_567879	ERR127017	South Africa (KwaZulu-Natal)	2011	Africa	98.87	97.09	92.34	92.34	5492680	1148	114
II_2011_E_561176_561185	ERR127018	South Africa (Eastern Cape)	2011	Africa	94.27	101.75	88.48	88.48	5845754	7637	352
I_1918_CIP_106164	ERR832451	Europe?	1918	Europe	97.45	121.65	96.53	96.53	7112182	2802	197
I_1918_CIP_106171	ERR832452	Europe?	1918	Europe	97.73	69.02	96.62	96.62	4096228	2731	192
III_1952_Ew_595_52	ERR832453	US?	1952	North America	97.45	81.40	97.63	97.63	4431164	1433	127
III_1954_ATCC_29903	ERR832456	Japan	1954	Asia	99.57	87.41	97.87	97.87	4775576	944	112
II_2011_E_557627	ERR127021	South Africa (Western Cape)	2011	Africa	93.87	110.55	90.72	90.72	6186534	7601	344
I_1999_K_147	ERR217084	Bangladesh	1999	Asia	98.25	60.23	87.64	87.64	3391024	3448	220
III_1915_NCTC1	ERR559526	France	1915	Europe	98.14	73.24	99.29	99.29	2309646	1117	124
I_1952_NCDC_5380_52	ERR832454	US?	1952	North America	97.67	91.85	92.96	92.96	5170976	2933	200
I_1960_E_509_60	ERR127033	UK	1960	Europe	95.92	105.60	93.34	93.34	5693740	2752	179
I_2003_KP_32	ERR217081	Bangladesh	2003	Asia	98.27	55.06	87.61	87.61	3091376	3484	224
I_2009_K_9708	ERR217082	Bangladesh	2009	Asia	98.19	40.99	86.57	86.57	2273728	3448	216
I_2000_K_820	ERR217086	Bangladesh	2000	Asia	96.83	131.01	89.98	89.98	7203388	3294	198
III_1961_E_345_61	ERR127035	UK	1961	Europe	98.70	108.65	97.56	97.56	5669976	1149	121
I_1972_E2610_72	ERR127036	UK	1972	Europe	97.69	94.55	91.81	91.81	5280378	2945	209
I_2003_KP_44	ERR217078	Bangladesh	2003	Asia	97.89	67.06	91.52	91.52	3598754	3380	208
I_2003_KP_45	ERR217079	Bangladesh	2003	Asia	96.88	74.94	91.90	91.90	3985912	3314	207
IV_1958_E_403_58	ERR127038	UK	1958	Europe	94.59	95.31	87.25	87.25	5468362	6851	338
I_2009_K_9590	ERR217076	Bangladesh	2009	Asia	96.50	63.70	91.77	91.77	3406766	3326	214
IV_1972_E_535	ERR127039	UK	1972	Europe	94.04	112.21	87.24	87.24	6381010	6779	345
I_2009_K_9442	ERR217075	Bangladesh	2009	Asia	96.50	63.64	92.01	92.01	3400166	3326	210
I_1953_E_34_53	ERR127041	UK	1953	Europe	95.30	100.57	91.93	91.93	5529264	2828	196

I_1999_K_211	ERR217047	Bangladesh	1999	Asia	97.75	43.41	88.95	88.95	2326694	3025	206
III_1999_K_238	ERR217048	Bangladesh	1999	Asia	99.14	69.16	91.95	91.95	3733206	1093	120
I_2003_KP_15	ERR217073	Bangladesh	2003	Asia	97.05	67.91	93.13	93.13	3530392	3326	212
V_1951_E_192_51	ERR127043	UK	1951	Europe	96.84	104.32	94.76	94.76	5732476	4684	247
I_2000_K_915	ERR217046	Bangladesh	2000	Asia	96.53	57.83	93.43	93.43	2961456	3375	209
V_1951_E_245_51	ERR127044	UK	1951	Europe	96.46	101.75	94.40	94.40	5525692	4497	243
I_1999_K_314	ERR217045	Bangladesh	1999	Asia	94.58	70.06	92.64	92.64	3540548	3274	202
IV_1961_E_531_61	ERR127046	UK	1961	Europe	94.14	311.53	90.91	90.91	17266810	6787	50
III_1955_E_428_55	ERR127047	UK	1955	Europe	98.93	346.81	96.81	96.81	18508538	1154	13
VI_2003_IVI_77_AR3444	ERR217034	Bangladesh	2003	Asia	92.26	68.22	87.68	87.68	3647246	8865	402
I__E126923	ERR127048	UK		Europe	97.90	407.49	87.07	87.07	23971578	3377	16
I_2010_K_10267	ERR200344	Bangladesh	2010	Asia	97.42	63.54	94.23	94.23	3357884	3263	205
I_2003_MS_341	ERR217033	Bangladesh	2003	Asia	96.92	73.97	89.36	89.36	3977828	2747	182
I_2010_K_10302	ERR200345	Bangladesh	2010	Asia	96.72	49.86	85.29	85.29	2939902	3218	202
VI_2007_K_8263	ERR217029	Bangladesh	2007	Asia	91.59	59.08	86.41	86.41	3139938	8584	404
VI_2003_MS_565	ERR217030	Bangladesh	2003	Asia	91.53	41.77	81.15	81.15	2308072	8500	419
I_1999_K_345	ERR200347	Bangladesh	1999	Asia	96.74	69.06	85.23	85.23	4091204	3307	196
VI_2009_K_9144_b	ERR217032	Bangladesh	2009	Asia	91.39	36.84	81.41	81.41	2031572	8509	411
I_2009_K_9138	ERR200348	Bangladesh	2009	Asia	96.90	69.30	93.84	93.84	3696572	3254	194
VI_2003_MS_182	ERR217028	Bangladesh	2003	Asia	91.98	43.34	80.74	80.74	2461436	8910	410
I_2009_K_9326	ERR200350	Bangladesh	2009	Asia	97.01	69.09	86.06	86.06	3999480	3288	197
VI_2009_K_9485	ERR217023	Bangladesh	2009	Asia	92.20	55.83	81.36	81.36	3206612	8493	407
VI_2009_K_9144_a	ERR217025	Bangladesh	2009	Asia	91.56	64.40	88.02	88.02	3386694	8552	405
VI_2003_LS_47317	ERR217026	Bangladesh	2003	Asia	96.67	51.40	86.12	86.12	2900002	6759	333
I_2009_K_9477	ERR200352	Bangladesh	2009	Asia	97.04	63.43	92.67	92.67	3420562	3252	200
VI_2003_MS_228_b	ERR217021	Bangladesh	2003	Asia	92.20	68.23	87.79	87.79	3643722	8900	403
III_2010_K_10276	ERR200353	Bangladesh	2010	Asia	99.03	54.79	80.71	80.71	3487974	1131	121
III_2010_K_10283	ERR200354	Bangladesh	2010	Asia	98.78	62.03	87.66	87.66	3610002	1120	120
III_2010_K_10339	ERR200355	Bangladesh	2010	Asia	99.03	59.09	81.45	81.45	3704962	1112	121

III_2010_K_10347	ERR200356	Bangladesh	2010	Asia	99.06	40.67	84.01	84.01	2467552	1133	117
III_2010_K_10369	ERR200357	Bangladesh	2010	Asia	99.05	69.06	86.57	86.57	4083010	1129	125
VI_2003_MS_607	ERR217012	Bangladesh	2003	Asia	91.68	61.60	87.33	87.33	3259866	8488	402
VI_2004_K_4142	ERR217013	Bangladesh	2004	Asia	92.73	43.60	85.90	85.90	2547074	8508	372
VI_2003_KD_945	ERR217016	Bangladesh	2003	Asia	91.59	79.34	87.16	87.16	4231784	8524	402
II_1999_K_62	ERR217018	Bangladesh	1999	Asia	94.12	58.44	86.58	86.58	3227354	7939	393
III_1999_K_119	ERR217019	Bangladesh	1999	Asia	98.93	54.40	89.95	89.95	2994170	1381	121
III_2010_K_10373	ERR200360	Bangladesh	2010	Asia	99.30	75.44	90.29	90.29	4182860	1119	123
III_2009_K_9131	ERR200361	Bangladesh	2009	Asia	98.87	78.49	91.48	91.48	4265698	1124	122
VI_2005_K_5898	ERR217011	Bangladesh	2005	Asia	91.57	49.31	86.27	86.27	2719586	8476	412
III_2009_K_9141	ERR200362	Bangladesh	2009	Asia	98.81	56.28	91.36	91.36	3112202	1417	124
VI_1998_K_1372	ERR217010	Bangladesh	1998	Asia	95.05	63.87	87.18	87.18	3512014	8901	393
III_2009_K_9171	ERR200363	Bangladesh	2009	Asia	99.02	65.49	96.49	96.49	3419072	1145	123
III_2009_K_9172	ERR200364	Bangladesh	2009	Asia	99.05	55.53	90.42	90.42	3113004	1395	121
III_2009_K_9282	ERR200365	Bangladesh	2009	Asia	97.78	81.76	96.06	96.06	4212028	1136	122
VI_2000_K_1347	ERR200414	Bangladesh	2000	Asia	91.77	43.56	76.72	76.72	2687846	8480	418
VI_2004_SH_4	ERR217014	Bangladesh	2004	Asia	93.75	74.46	86.64	86.64	4160636	8438	392
III_2009_K_9293	ERR200366	Bangladesh	2009	Asia	98.77	89.96	87.49	87.49	5105680	1125	119
III_2009_K_9324	ERR200367	Bangladesh	2009	Asia	98.91	57.79	90.18	90.18	3208080	1120	128
II_2009_K_9327	ERR200368	Bangladesh	2009	Asia	93.93	54.91	89.99	89.99	2953590	7708	383
V_2009_K_9597	ERR200413	Bangladesh	2009	Asia	96.62	58.60	90.82	90.82	3157282	4978	282
III_2009_K_9332	ERR200369	Bangladesh	2009	Asia	99.00	54.32	90.91	90.91	3026640	1119	118
II_2009_K_9375	ERR200405	Bangladesh	2009	Asia	93.74	53.59	86.16	86.16	3007346	7699	392
II_2009_K_9413	ERR200407	Bangladesh	2009	Asia	93.77	67.91	85.33	85.33	3844618	7714	380
II_2009_K_9444	ERR200408	Bangladesh	2009	Asia	93.82	50.17	85.60	85.60	2853544	7679	386
III_2009_K_9451	ERR200371	Bangladesh	2009	Asia	98.70	48.03	90.92	90.92	2688238	1403	123
II_2009_K_9339	ERR200404	Bangladesh	2009	Asia	93.96	45.83	86.47	86.47	2559652	7686	395
II_2009_K_9411	ERR200406	Bangladesh	2009	Asia	91.59	88.58	85.83	85.83	4790936	7822	356
III_2009_K_9484	ERR200373	Bangladesh	2009	Asia	99.04	45.53	85.36	85.36	2691748	1112	120

II_2009_K_9292	ERR200403	Bangladesh	2009	Asia	94.16	46.96	86.28	86.28	2614026	7908	381
III_2009_K_9555	ERR200374	Bangladesh	2009	Asia	99.29	85.70	91.21	91.21	4732956	1139	124
III_2009_K_9557	ERR200375	Bangladesh	2009	Asia	99.42	65.85	91.36	91.36	3660522	1056	113
II_2009_K_9278	ERR200402	Bangladesh	2009	Asia	93.26	63.03	86.73	86.73	3460294	7630	381
III_2009_K_9584	ERR200376	Bangladesh	2009	Asia	99.05	63.13	89.74	89.74	3559858	1117	125
II_2009_K_9175	ERR200400	Bangladesh	2009	Asia	93.86	52.51	84.32	84.32	2968038	7735	384
II_2009_K_9256	ERR200401	Bangladesh	2009	Asia	93.60	45.89	86.96	86.96	2539706	7658	383
III_2009_K_9586	ERR200378	Bangladesh	2009	Asia	99.04	75.93	91.26	91.26	4241228	1115	122
III_2009_K_9587	ERR200379	Bangladesh	2009	Asia	98.97	43.18	88.59	88.59	2498450	1104	126
II_2009_K_9129	ERR200398	Bangladesh	2009	Asia	92.52	43.53	81.98	81.98	2527870	7803	375
III_2009_K_9593	ERR200380	Bangladesh	2009	Asia	99.11	49.39	94.23	94.23	2637156	1143	123
II_1999_K_734	ERR200397	Bangladesh	1999	Asia	92.69	42.37	82.60	82.60	2470104	7804	374
III_2009_K_9594	ERR200381	Bangladesh	2009	Asia	99.39	52.21	88.26	88.26	3037250	1104	125
III_2009_K_9595	ERR200382	Bangladesh	2009	Asia	98.81	38.23	83.74	83.74	2265782	1412	124
III_2010_K_10346	ERR200383	Bangladesh	2010	Asia	98.52	46.90	86.73	86.73	2749738	1407	126
II_2010_K_10368	ERR200395	Bangladesh	2010	Asia	93.63	45.11	85.65	85.65	2564418	7702	386
III_1999_K_287	ERR200384	Bangladesh	1999	Asia	98.76	54.33	89.73	89.73	3042002	1382	124
II_2010_K_10343	ERR200394	Bangladesh	2010	Asia	94.46	87.36	87.31	87.31	4849906	7716	382
II_2009_K_9169	ERR200399	Bangladesh	2009	Asia	93.93	75.09	87.99	87.99	4091188	7685	375
III_1999_K_373	ERR200386	Bangladesh	1999	Asia	99.04	69.14	88.03	88.03	3933518	1386	120
III_2005_K_5613	ERR200387	Bangladesh	2005	Asia	98.58	58.34	82.20	82.20	3580614	1420	119
III_2009_K_9443	ERR200392	Bangladesh	2009	Asia	98.75	86.70	90.23	90.23	4769582	1427	120
III_2009_K_9110	ERR200388	Bangladesh	2009	Asia	98.80	47.56	87.70	87.70	2686338	1431	125
III_2009_K_9338	ERR200390	Bangladesh	2009	Asia	98.83	68.86	89.89	89.89	3810372	1457	127
IV_2002_Sh02_1564	ERR042797	Dominican Republic	2002	North America	94.55	93.81	87.62	87.62	6607916	7079	339
III_2004_Sh04_2878	ERR042806	Morocco	2004	Africa	96.44	103.18	92.98	92.98	6921902	1707	139
I_2004_Sh04_9462	ERR042810	Cameroon	2004	Africa	96.08	402.18	95.78	95.78	26222642	3281	23
II_2005_Sh05_0787	ERR042811	Congo	2005	Africa	95.66	106.69	91.77	91.77	7296838	7960	379
I_2005_Sh05_1396	ERR042813	Chad	2005	Africa	95.14	110.52	96.03	96.03	7074892	3240	203

III_2005_Sh05_1631	ERR042814	Dominican Republic	2005	North America	98.19	111.05	93.95	93.95	7543350	1077	115
I_2005_Sh05_5540	ERR042818	Dominican Republic	2005	North America	95.52	82.50	94.07	94.07	5471308	3343	197
III_2005_Sh05_9277	ERR042825	Dominican Republic	2005	North America	98.85	105.19	95.32	95.32	7078620	1107	112
I_2006_Sh06_2028	ERR042828	Guinea	2006	Africa	95.96	113.92	96.38	96.38	7359274	3280	190
III_2006_Sh06_7888	ERR042830	Algeria	2006	Africa	97.28	104.00	93.60	93.60	7048002	1096	110
I_2006_Sh06_8628	ERR042831	Algeria	2006	Africa	97.31	117.67	93.03	93.03	8093836	3024	193
VI_2007_Sh07_1096	ERR042833	Mali	2007	Africa	92.34	159.77	91.32	91.32	10695742	8910	375
I_2007_Sh07_3008	ERR042835	Cameroon	2007	Africa	94.63	122.28	97.12	97.12	7760944	3185	179
VII_2007_Sh07_5519	ERR042838	Cameroon	2007	Africa	92.48	123.13	87.21	87.21	8708308	8164	385
III_2009_Sh09_2087	ERR042854	Mayotte	2009	Africa	98.68	153.26	92.07	92.07	10579950	1135	112
II_2009_KH01	ERR047239	Vietnam	2009	Asia	94.48	121.02	77.06	77.06	9789582	8023	376
I_1995_MS0021	ERR047296	Vietnam	1995	Asia	96.60	152.99	89.82	89.82	10741058	3361	196
III_2000_DE0119	ERR047352	Vietnam	2000	Asia	99.59	77.29	88.04	88.04	5875728	1073	116
III_2000_DE0174	ERR047353	Vietnam	2000	Asia	99.61	152.68	81.87	81.87	12485080	1082	111
III_2000_DE0457	ERR047364	Vietnam	2000	Asia	99.61	114.91	88.49	88.49	8605504	1077	117
III_2002_DE1279	ERR047382	Vietnam	2002	Asia	98.62	201.21	84.30	84.30	15769128	1063	86
II_2002_DE1461	ERR047385	Vietnam	2002	Asia	94.81	90.84	84.17	84.17	6924996	8050	368
III_2007_EG0055	ERR047390	Vietnam	2007	Asia	99.26	213.59	91.03	91.03	15505182	1317	87
III_2008_H04	ERR047430	Vietnam	2008	Asia	99.46	198.62	78.22	78.22	16606294	1124	104
IV_1949_IB1651	ERR048259	UK	1949	Europe	96.25	157.34	76.46	76.46	12642822	6613	348
III__IB2490	ERR048279	UK		Europe	98.71	160.73	98.62	98.62	10189602	1110	141
III__IB1789	ERR048284	Vietnam		Asia	98.92	139.29	94.40	94.40	9443210	1120	112
II_2002_IB3371	ERR048293	Pakistan	2002	Asia	96.71	143.32	85.03	85.03	11240918	8030	367
III_1999_IB0709	ERR048301	Korea	1999	Asia	96.59	90.28	91.10	91.10	6254308	1329	102
VII_2001_IB1709	ERR048308	Vietnam	2001	Asia	91.38	106.33	76.58	76.58	8885678	8045	384
III_1977_Tunis_38_17	ERR832495	Tunisia	1977	Africa	98.98	92.82	98.25	98.25	5092308	976	110
IV_1951_NCTC_4839	ERR832492	None	1951	None	95.98	92.16	91.52	91.52	5225354	6541	349
I_1991_IB0012	ERR048316	Korea	1991	Asia	96.49	99.70	60.03	60.03	11579512	2893	188
III_1980_E_86_8	ERR832485	Morocco	1980	Africa	98.38	99.32	96.03	96.03	5408992	988	107

III_2002_IB0045	ERR048324	Taiwan	2002	Asia	98.51	153.23	94.23	94.23	10641184	1075	105
VII_2003_IB1753	ERR048329	Vietnam	2003	Asia	93.49	119.46	82.51	82.51	8774404	8252	373
IV_1971_NCDC_2783_71	ERR832477	US?	1971	North America	96.04	80.87	80.90	80.90	5259430	6997	331
VII_1998_IB4229	ERR048333	India	1998	Asia	93.10	209.91	86.19	86.19	14535944	8380	362
III_1967_E_20_67	ERR832468	Algeria	1967	Africa	96.83	102.65	96.47	96.47	5734002	960	107
III_1966_E_12_66	ERR832463	France	1966	Europe	97.02	104.43	98.28	98.28	5532600	951	99
III_1959_E_3_59	ERR832458	Senegal	1959	Africa	97.55	91.26	98.41	98.41	4823992	917	100
III_1963_E_6_63	ERR832461	Congo	1963	Africa	97.64	90.50	98.77	98.77	5018924	1392	124
III_1963_E_9_63	ERR832462	France	1963	Europe	96.27	97.93	97.42	97.42	5313470	912	92
III_1955_CIP_56_19	ERR832457	Tunisia	1955	Africa	98.04	96.59	94.77	94.77	5499884	857	104
II_2009_E_333992	ERR126965	South Africa (KwaZulu-Natal)	2009	Africa	96.03	92.27	85.90	85.90	5563888	7707	359
II_2009_E_356458	ERR126966	South Africa (Free State)	2009	Africa	93.77	112.11	90.37	90.37	6696742	7556	355
I_2000_K_716	ERR217085	Bangladesh	2000	Asia	98.25	61.17	89.88	89.88	3365818	3452	217
II_2010_E_458841	ERR126991	South Africa (Gauteng)	2010	Africa	94.04	99.85	92.16	92.16	5472276	7699	376
II_2010_E_428153	ERR126992	South Africa (Eastern Cape)	2010	Africa	94.21	122.39	89.77	89.77	6798136	7749	361
II_2010_E_428777	ERR126993	South Africa (Gauteng)	2010	Africa	93.84	80.55	89.12	89.12	4581252	7836	378
I_2009_K_9108	ERR217083	Bangladesh	2009	Asia	98.27	59.58	86.35	86.35	3383432	3516	226
I_2003_KP_55	ERR217080	Bangladesh	2003	Asia	97.88	157.22	91.85	91.85	8531964	3267	192
III_2011_E_535661	ERR127015	South Africa (Gauteng)	2011	Africa	99.20	115.74	93.81	93.81	6537084	1187	118
I_2000_K_923	ERR217077	Bangladesh	2000	Asia	96.77	60.32	87.59	87.59	3398728	3346	212
II_2011_E_584638	ERR127019	South Africa (Gauteng)	2011	Africa	94.08	102.59	89.70	89.70	5715478	7753	344
II_2011_E_542158	ERR127020	South Africa (Gauteng)	2011	Africa	94.26	95.84	89.52	89.52	5421842	7709	361
II_2011_E_575237	ERR127022	South Africa (Eastern Cape)	2011	Africa	93.77	120.70	88.53	88.53	6844260	7607	352
I_1949_E_109_49	ERR127032	UK	1949	Europe	97.70	107.13	94.01	94.01	5920382	2882	203
I__E_1c	ERR127034	UK		Europe	96.36	107.46	91.56	91.56	6086022	2954	200
IV_1963_E_5_63_2	ERR127037	UK	1963	Europe	95.60	89.38	87.86	87.86	5184206	6914	343
VI_1998_K_1376	ERR217031	Bangladesh	1998	Asia	92.01	60.40	87.50	87.50	3218146	8903	408
I_1959_E_15_59	ERR127040	UK	1959	Europe	96.15	103.00	85.05	85.05	6241342	2776	182
V_1955_E_13_55	ERR127042	UK	1955	Europe	95.41	103.90	93.40	93.40	5712348	4444	251

I_2001_K_2535	ERR217027	Bangladesh	2001	Asia	97.35	67.17	90.98	90.98	3582656	3379	209
VI_2009_K_9165	ERR217024	Bangladesh	2009	Asia	92.41	48.11	86.76	86.76	2585460	8519	398
I_1999_K_294	ERR200346	Bangladesh	1999	Asia	96.85	64.40	89.09	89.09	3616358	3364	205
VI_1999_K_139	ERR217020	Bangladesh	1999	Asia	92.19	56.95	83.15	83.15	3212498	8915	412
VI_2003_MS_568	ERR217022	Bangladesh	2003	Asia	92.36	56.04	88.45	88.45	2940192	8512	410
I_2009_K_9286	ERR200349	Bangladesh	2009	Asia	97.78	77.41	93.66	93.66	4143010	3000	192
I_2009_K_9342	ERR200351	Bangladesh	2009	Asia	97.84	46.33	90.43	90.43	2561974	3182	189
VI_2003_MS_228_a	ERR217017	Bangladesh	2003	Asia	91.96	62.47	87.59	87.59	3320238	8859	405
VI_2004_KD_1776	ERR217015	Bangladesh	2004	Asia	91.55	71.55	84.96	84.96	3922852	8499	400
III_2010_K_10371	ERR200358	Bangladesh	2010	Asia	99.03	61.52	87.44	87.44	3576910	1138	118
III_2010_K_10372	ERR200359	Bangladesh	2010	Asia	99.02	76.37	86.90	86.90	4568550	1133	115
VII_1999_K_319	ERR200415	Bangladesh	1999	Asia	93.35	82.02	71.54	71.54	5589640	8357	398
II_2009_K_9588	ERR200412	Bangladesh	2009	Asia	93.74	52.96	92.80	92.80	2715820	7725	386
II_2009_K_9522	ERR200411	Bangladesh	2009	Asia	94.88	48.15	83.82	83.82	2843262	7611	377
II_2009_K_9454	ERR200410	Bangladesh	2009	Asia	93.86	55.90	88.71	88.71	3049588	7722	391
II_2009_K_9447	ERR200409	Bangladesh	2009	Asia	94.16	87.69	87.72	87.72	4809536	7742	386
III_2009_K_9448	ERR200370	Bangladesh	2009	Asia	99.17	77.30	92.66	92.66	4205288	1129	124
III_2009_K_9476	ERR200372	Bangladesh	2009	Asia	99.19	54.15	91.54	91.54	2985744	1059	114
III_2009_K_9585	ERR200377	Bangladesh	2009	Asia	99.03	45.32	85.80	85.80	2702134	1113	116
II_1999_K_276	ERR200396	Bangladesh	1999	Asia	94.09	39.25	80.96	80.96	2334094	7997	390
II_2010_K_10248	ERR200393	Bangladesh	2010	Asia	94.04	45.38	86.44	86.44	2549092	7730	391
III_1999_K_289	ERR200385	Bangladesh	1999	Asia	98.61	59.35	88.70	88.70	3370254	1116	128
III_2009_K_9373	ERR200391	Bangladesh	2009	Asia	97.98	38.25	87.18	87.18	2177288	1390	120
III_2009_K_9290	ERR200389	Bangladesh	2009	Asia	98.70	39.46	92.17	92.17	2129954	1431	124

Appendix 4.14 *Shigella sonnei* Oceania and Global WGS metadata

Isolate	treeID	Country	Year	Region	data_source	Cover%_NC_016822	Depth_NC_016822	Mapped%_NC_016822	Mapped%_Total	Total_Reads	SNPs	Indels
pool16_tag9	I_1167_1967_France	France	1967	Western Europe	global	95.22	17.84	93.31	93.31	1731618	1087	68
pool16_tag6	I_1263_1963_France	France	1963	Western Europe	global	94.51	26.38	93.92	93.92	2522514	1082	64
pool16_tag7	I_1265_1965_France	France	1965	Western Europe	global	91.87	30.57	92.66	92.66	2878524	1109	76
pool17_tag5	I_1274_1974_France	France	1974	Western Europe	global	94.20	15.74	93.06	93.06	1513078	1505	70
pool16_tag3	I_1460_1960_France	France	1960	Western Europe	global	94.97	24.59	93.13	93.13	2394740	1097	75
pool16_tag4	I_1461_1961_France	France	1961	Western Europe	global	95.26	11.78	91.68	91.68	1164918	974	53
pool16_tag11	I_1567_1967_Senegal	Senegal	1967	Western Africa	global	94.70	12.89	94.69	94.69	1228018	1058	55
pool21_tag6	I_20040489_2004_	unknown	2004	unknown	global	96.70	69.49	91.91	91.91	4923630	1687	87
pool17_tag7	I_4374_1974_France	France	1974	Western Europe	global	96.21	40.82	92.90	92.90	4011472	1550	83
pool17_tag8	I_4474_1974_France	France	1974	Western Europe	global	95.15	52.24	93.86	93.86	5051630	1202	90
pool17_tag9	I_476_1976_France	France	1976	Western Europe	global	93.90	38.76	94.38	94.38	3653372	1191	83
pool15_tag10	I_54213_1945_France	France	1945	Western Europe	global	95.25	12.18	94.28	94.28	1164636	967	53
pool18_tag6	I_5827_2000_Madagascar	Madagascar	2000	Eastern Africa	global	96.96	40.54	94.30	94.30	3936778	1467	77
pool10_tag11	I_74369_2007_France	France	2007	Western Europe	global	97.04	71.14	93.45	93.45	6999538	1999	93
pool5_tag2	I_DE0330_2000_HCMC	Vietnam	2000	South-Eastern Asia	global	95.94	15.79	95.14	95.14	1564504	1098	59
pool7_tag6	I_EG0352_2007_HCMC	Vietnam	2007	South-Eastern Asia	global	95.84	10.91	84.24	84.24	1198062	1043	63
pool16_tag8	II_1166_1966_France	France	1966	Western Europe	global	96.35	32.35	98.33	98.33	2991438	456	33
pool17_tag1	II_1173_1973_France	France	1973	Western Europe	global	96.24	62.52	98.08	98.08	5772040	447	40
pool16_tag10	II_1267_1967_France	France	1967	Western Europe	global	95.40	17.44	95.53	95.53	1641776	458	32
pool20_tag2	II_19910761_1991_	unknown	1991	unknown	global	97.57	31.53	88.81	88.81	2329676	522	41
pool20_tag11	II_19911483_1991_	unknown	1991	unknown	global	97.84	224.16	87.14	87.14	17114192	529	44
pool20_tag12	II_19920319_1992_	unknown	1992	unknown	global	97.69	70.10	86.31	86.31	5434350	541	43
pool20_tag10	II_20061758_2006_DominicanRepublic	Dominican Republic	2006	Caribbean	global	97.43	30.51	86.92	86.92	2338828	466	43
pool22_tag3	II_20071599_2007_	unknown	2007	unknown	global	98.53	85.05	90.62	90.62	6287026	265	29

pool17_tag6	II_2574_1974_France	France	1974	Western Europe	global	97.05	62.61	98.81	98.81	5790006	232	21
pool17_tag4	II_373_1973_Cameroun	Cameroon	1973	Middle Africa	global	97.06	23.14	96.12	96.12	2202544	709	35
pool10_tag1	II_41191_2004_Tanzanie	Tanzania	2004	Eastern Africa	global	99.78	23.42	99.25	99.25	2205328	441	29
pool15_tag3	II_54178_1945_Sweden	Sweden	1945	Northern Europe	global	98.39	27.03	96.13	96.13	2604042	848	40
pool15_tag4	II_54179_1944_Sweden	Sweden	1944	Northern Europe	global	98.00	22.66	96.02	96.02	2173102	444	35
pool15_tag7	II_54185_1945_Denmark	Denmark	1945	Northern Europe	global	98.46	16.19	94.15	94.15	1607920	803	31
pool15_tag8	II_54190_1945_Sweden	Sweden	1945	Northern Europe	global	98.39	24.29	96.12	96.12	2350392	870	35
pool15_tag12	II_54228_1947_Sweden	Sweden	1947	Northern Europe	global	98.96	47.37	99.50	99.50	4431186	362	21
pool10_tag2	II_55623_2005_Maroc	Morocco	2005	Northern Africa	global	99.18	52.82	96.23	96.23	5118132	206	32
pool16_tag1	II_658_1958_France	France	1958	Western Europe	global	97.13	35.42	95.61	95.61	3394104	210	29
pool10_tag10	II_66470_2006_Haiti	Haiti	2006	Caribbean	global	98.61	67.88	97.68	97.68	6443768	295	27
pool18_tag3	II_9810267_1998_Madagasca	Madagascar	1998	Eastern Africa	global	99.21	42.49	96.35	96.35	4109086	277	29
pool18_tag4	II_998911_1999_France	France	1999	Western Europe	global	99.99	14.60	94.32	94.32	1461658	535	26
pool22_tag9	II_CS1_1997_Brazil	Brazil	1997	South America	global	97.51	64.14	87.49	87.49	4820014	457	45
pool20_tag7	II_CS2_1997_Brazil	Brazil	1997	South America	global	99.01	27.64	88.52	88.52	2125076	362	22
pool11_tag12	II_IB694_1979_Korea	Korea	1979	Eastern Asia	global	99.16	33.50	95.95	95.95	3251230	247	22
pool12_tag1	II_IB695_1983_Korea	Korea	1983	Eastern Asia	global	99.16	28.09	97.11	97.11	2699002	236	20
pool12_tag2	II_IB696_1980_Korea	Korea	1980	Eastern Asia	global	99.16	39.87	97.71	97.71	3810762	254	20
pool12_tag3	II_IB697_1982_Korea	Korea	1982	Eastern Asia	global	99.16	23.03	90.92	90.92	2505916	305	29
pool12_tag4	II_IB698_1983_Korea	Korea	1983	Eastern Asia	global	98.93	12.29	90.94	90.94	1266784	206	14
pool12_tag5	II_IB713_1981_Korea	Korea	1981	Eastern Asia	global	99.16	15.33	98.16	98.16	1455422	258	23
pool12_tag6	II_IB716_1981_Korea	Korea	1981	Eastern Asia	global	98.87	22.20	93.46	93.46	2216588	190	10
pool12_tag7	II_IB717_1982_Korea	Korea	1982	Eastern Asia	global	99.16	33.56	93.68	93.68	3347544	238	22
pool12_tag8	II_IB739_1985_Korea	Korea	1985	Eastern Asia	global	99.16	31.04	99.04	99.04	2927386	242	25
pool12_tag9	II_IB748_1987_Korea	Korea	1987	Eastern Asia	global	99.16	34.65	94.57	94.57	3427438	233	24

pool21_tag8	III_19904011_1990_	unknown	1990	unknown	global	96.05	52.14	89.74	89.74	3944668	1271	79
pool21_tag5	III_19984123_1998_Mexico	Mexico	1998	Central America	global	96.12	29.91	92.58	92.58	2097654	1247	76
pool21_tag2	III_20003593_2000_	unknown	2000	unknown	global	96.12	47.18	87.24	87.24	3523386	1262	79
pool22_tag5	III_20010007_2001_	unknown	2001	unknown	global	96.06	175.29	88.48	88.48	13315376	1368	81
pool21_tag9	III_20011685_2001_	unknown	2001	unknown	global	95.45	69.51	92.68	92.68	5020962	1288	78
pool21_tag3	III_20021122_2002_	unknown	2002	unknown	global	95.35	88.16	92.31	92.31	6209808	1307	79
pool21_tag4	III_20031275_2003_Iran	Iran	2003	Southern Asia	global	95.80	41.53	90.71	90.71	2995818	1353	84
pool20_tag6	III_20040880_2004_SriLanka	Sri Lanka	2004	Southern Asia	global	95.93	42.03	91.06	91.06	2984190	1305	77
pool22_tag1	III_20040924_2004_Kenya_Egypt	Kenya	2004	Eastern Africa	global	95.77	128.08	90.54	90.54	9296004	1332	80
pool20_tag4	III_20041367_2004_	unknown	2004	unknown	global	95.29	31.66	89.55	89.55	2280574	1264	75
pool20_tag1	III_20051272_2005_Egypt	Egypt	2005	Northern Africa	global	95.70	24.06	89.30	89.30	1745766	1229	65
pool20_tag8	III_20051541_2005_Uzbekistan	Uzbekistan	2005	Central Asia	global	95.38	21.68	88.21	88.21	1655056	1239	70
pool21_tag7	III_20052631_2005_Peru	Peru	2005	South America	global	96.31	47.54	87.60	87.60	3637944	1855	82
pool20_tag3	III_20060018_2006_Egypt	Egypt	2006	Northern Africa	global	96.06	51.30	89.73	89.73	3715432	1316	83
pool21_tag1	III_20061309_2006_Egypt	Egypt	2006	Northern Africa	global	95.67	63.22	92.33	92.33	4479526	1323	81
pool22_tag2	III_20062087_2006_Egypt_Tunisia	Egypt_Tunisia	2006	Northern Africa	global	96.13	169.40	92.34	92.34	12097902	1675	75
pool22_tag6	III_20062313_2006_Nepal	Nepal	2006	Southern Asia	global	95.44	200.00	89.32	89.32	14486272	1331	74
pool20_tag5	III_20081885_2008_	unknown	2008	unknown	global	96.13	23.26	86.79	86.79	1741564	1235	63
pool17_tag2	III_2073_1973_France	France	1973	Western Europe	global	92.60	60.78	81.37	81.37	6529984	1183	67
pool18_tag5	III_2225_2000_FrenchGuyana	French Guiana	2000	South America	global	95.52	12.16	96.52	96.52	1138950	1452	59
pool17_tag3	III_273_1973_Senegal	Senegal	1973	Western Africa	global	94.75	44.58	95.58	95.58	4176906	1140	73
pool18_tag8	III_31382_2003_Israel	Israel	2003	Western Asia	global	95.66	54.05	93.64	93.64	5208876	1447	76
pool18_tag9	III_32222_2003_Cuba	Cuba	2003	Caribbean	global	94.73	29.37	97.66	97.66	2687858	1282	71
pool18_tag10	III_36224_2003_Senegal	Senegal	2003	Western Africa	global	95.10	25.36	96.50	96.50	2357924	1266	74
pool15_tag6	III_54184_1945_Sweden	Sweden	1945	Northern Europe	global	95.67	26.22	94.86	94.86	2501080	1006	53

pool15_tag9	III_54210_1943_Sweden	Sweden	1943	Northern Europe	global	97.55	99.14	88.61	88.61	10311090	1097	69
pool10_tag3	III_60108_2006_FrenchGuyana	French Guiana	2006	South America	global	94.39	77.33	90.76	90.76	7592110	1250	75
pool10_tag4	III_62542_2006_BurkinaFaso	Burkina Faso	2006	Western Africa	global	95.04	74.86	96.72	96.72	6964654	1350	76
pool10_tag5	III_65179_2006_Senegal	Senegal	2006	Western Africa	global	95.44	49.95	97.00	97.00	4641274	1291	81
pool10_tag6	III_65387_2006_Maroc	Morocco	2006	Northern Africa	global	95.53	43.03	98.36	98.36	3931962	1320	72
pool10_tag7	III_65623_2006_Maroc	Morocco	2006	Northern Africa	global	94.74	46.14	98.21	98.21	4189090	1300	69
pool10_tag9	III_66396_2006_France	France	2006	Western Europe	global	95.50	62.04	97.05	97.05	5767262	1344	82
pool17_tag10	III_8883_1983_France	France	1983	Western Europe	global	93.11	31.85	97.26	97.26	2883984	1245	70
pool17_tag11	III_970044_1997_NewCaledonia	New Caledonia	1997	Melanesia	global	92.42	14.66	96.07	96.07	1334706	1133	51
pool18_tag1	III_988743_1998_FrenchGuyana	French Guiana	1998	South America	global	95.88	52.03	98.38	98.38	4774402	1640	81
pool18_tag2	III_989560_1998_Madagasca	Madagascar	1998	Eastern Africa	global	96.43	59.39	97.36	97.36	5547928	1304	76
pool21_tag12	III_CS14_2001_Brazil	Brazil	2001	South America	global	95.79	62.03	90.50	90.50	4607960	1374	81
pool21_tag10	III_CS20_2002_Brazil	Brazil	2002	South America	global	95.88	92.86	91.76	91.76	6834616	1369	81
pool21_tag11	III_CS6_2000_Brazil	Brazil	2000	South America	global	95.92	37.33	88.63	88.63	2805176	1337	80
pool22_tag7	III_CS7_2000_Brazil	Brazil	2000	South America	global	96.05	113.73	88.86	88.86	8398540	1354	84
pool22_tag8	III_CS8_2000_Brazil	Brazil	2000	South America	global	95.91	116.27	90.15	90.15	8439526	1333	82
pool5_tag8	III_EG0467_2008_HCMC	Vietnam	2008	South-Eastern Asia	global	95.44	13.93	88.26	88.26	1494254	1279	64
pool5_tag10	III_EG0472_2008_HCMC	Vietnam	2008	South-Eastern Asia	global	95.47	14.27	91.31	91.31	1477530	1287	59
pool9_tag1	III_EG1008_2008_HCMC	Vietnam	2008	South-Eastern Asia	global	95.12	40.77	97.08	97.08	3798644	1269	77
pool11_tag1	III_IB1_2003_Korea	Korea	2003	Eastern Asia	global	95.19	47.79	83.48	83.48	5137476	1312	74
pool11_tag4	III_IB10_2003_Korea	Korea	2003	Eastern Asia	global	95.46	55.83	95.30	95.30	5297878	1353	80
pool12_tag10	III_IB1970_2001_Vietnam	Vietnam	2001	South-Eastern Asia	global	95.65	20.46	93.90	93.90	1979026	1274	73
pool12_tag11	III_IB1976_2002_Vietnam	Vietnam	2002	South-Eastern Asia	global	94.91	11.27	95.81	95.81	1067510	1130	49
pool12_tag12	III_IB1980_2002_Vietnam	Vietnam	2002	South-Eastern Asia	global	95.06	48.12	95.09	95.09	4566476	1301	67
pool13_tag1	III_IB1985_2002_Vietnam	Vietnam	2002	South-Eastern Asia	global	94.35	20.01	96.69	96.69	1874244	1225	52

pool13_tag2	III_IB1987_2002_Vietnam	Vietnam	2002	South-Eastern Asia	global	95.48	42.85	96.59	96.59	4040468	1304	73
pool13_tag3	III_IB1990_2003_Vietnam	Vietnam	2003	South-Eastern Asia	global	95.42	16.86	92.42	92.42	1670500	1242	58
pool13_tag4	III_IB1993_2003_Vietnam	Vietnam	2003	South-Eastern Asia	global	94.62	33.95	95.47	95.47	3199644	1292	76
pool13_tag5	III_IB1995_2003_Vietnam	Vietnam	2003	South-Eastern Asia	global	95.44	13.29	97.54	97.54	1238292	1230	60
pool13_tag6	III_IB1997_2003_Vietnam	Vietnam	2003	South-Eastern Asia	global	95.47	36.57	98.35	98.35	3386146	1274	67
pool11_tag2	III_IB2_2003_Korea	Korea	2003	Eastern Asia	global	95.46	45.65	94.41	94.41	4359412	1340	80
pool13_tag7	III_IB2000_2003_Vietnam	Vietnam	2003	South-Eastern Asia	global	95.48	28.50	94.51	94.51	2733660	1282	66
pool13_tag8	III_IB2004_2003_Vietnam	Vietnam	2003	South-Eastern Asia	global	94.64	32.52	94.59	94.59	3111768	1253	73
pool13_tag9	III_IB2008_2003_Vietnam	Vietnam	2003	South-Eastern Asia	global	95.48	25.04	95.29	95.29	2391826	1309	73
pool13_tag10	III_IB2009_2003_Vietnam	Vietnam	2003	South-Eastern Asia	global	95.31	11.26	96.53	96.53	1067686	1484	42
pool13_tag11	III_IB2012_2001_Vietnam	Vietnam	2001	South-Eastern Asia	global	93.41	15.64	88.86	88.86	1716652	1199	62
pool13_tag12	III_IB2013_2001_Vietnam	Vietnam	2001	South-Eastern Asia	global	94.94	50.51	89.95	89.95	5136550	1296	72
pool14_tag1	III_IB2015_2002_Vietnam	Vietnam	2002	South-Eastern Asia	global	94.52	17.55	93.42	93.42	1705740	1629	62
pool14_tag2	III_IB2018_2002_Vietnam	Vietnam	2002	South-Eastern Asia	global	95.66	56.53	93.45	93.45	5540266	1331	80
pool14_tag3	III_IB2024_2002_Vietnam	Vietnam	2002	South-Eastern Asia	global	95.50	51.34	94.01	94.01	4989778	1335	76
pool14_tag4	III_IB2026_2003_Vietnam	Vietnam	2003	South-Eastern Asia	global	94.93	32.25	96.51	96.51	3030150	1321	78
pool11_tag3	III_IB3_2003_Korea	Korea	2003	Eastern Asia	global	95.13	50.07	93.36	93.36	4810966	1306	73
pool14_tag7	III_IB3277_2002_Pakistan	Pakistan	2002	Southern Asia	global	95.46	44.32	92.72	92.72	4413538	1319	79
pool14_tag8	III_IB3300_2002_Pakistan	Pakistan	2002	Southern Asia	global	95.14	49.52	92.80	92.80	4964914	1318	75
pool14_tag9	III_IB3374_2002_Pakistan	Pakistan	2002	Southern Asia	global	95.49	47.19	97.44	97.44	4445432	1305	78
pool14_tag10	III_IB3488_2003_Pakistan	Pakistan	2003	Southern Asia	global	95.44	25.48	93.27	93.27	2492354	1316	76
pool14_tag11	III_IB3507_2003_Pakistan	Pakistan	2003	Southern Asia	global	95.39	33.97	95.18	95.18	3272808	1312	81
pool14_tag12	III_IB3580_2003_Pakistan	Pakistan	2003	Southern Asia	global	95.46	67.29	94.09	94.09	6548204	1341	79
pool15_tag1	III_IB3599_2003_Pakistan	Pakistan	2003	Southern Asia	global	96.05	21.52	95.31	95.31	2053486	1241	56
pool11_tag7	III_IB681_1991_Korea	Korea	1991	Eastern Asia	global	95.52	37.69	97.35	97.35	3491008	1295	74

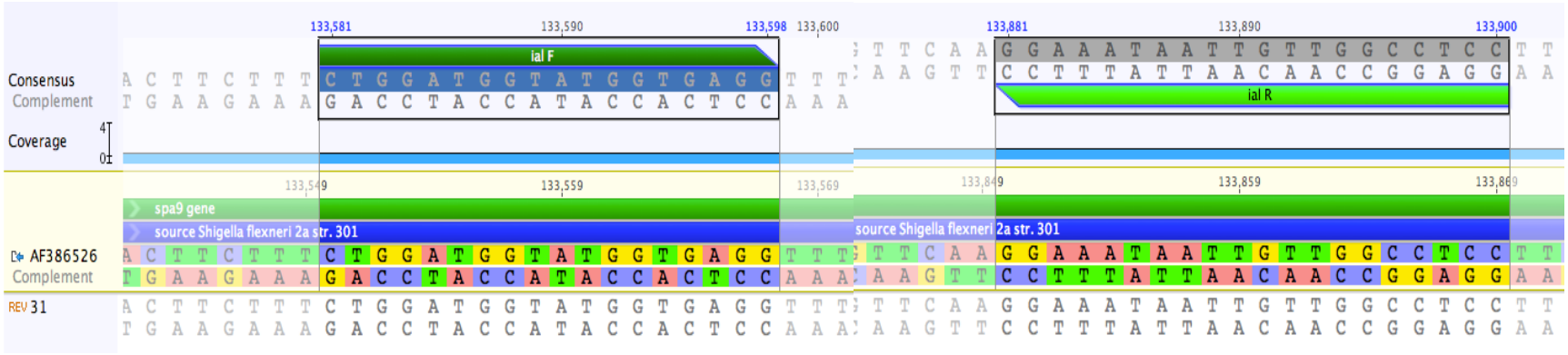
pool11_tag8	III_IB683_1994_Korea	Korea	1994	Eastern Asia	global	95.56	44.93	94.54	94.54	4290962	1260	81
pool11_tag9	III_IB687_1998_Korea	Korea	1998	Eastern Asia	global	94.41	45.73	96.47	96.47	4224332	1284	73
pool11_tag10	III_IB690_2000_Korea	Korea	2000	Eastern Asia	global	95.47	53.46	95.85	95.85	5031350	1323	74
pool11_tag11	III_IB691_1999_Korea	Korea	1999	Eastern Asia	global	95.52	44.17	92.35	92.35	4333958	1318	78
pool1_tag5	III_MS0035_1995_HCMC	Vietnam	1995	South-Eastern Asia	global	95.40	20.38	95.01	95.01	2007372	1428	66
pool2_tag5	III_MS0110_1996_HCMC	Vietnam	1996	South-Eastern Asia	global	95.47	22.92	95.61	95.61	2267982	1310	75
pool16_tag2	IIa_259_1959_France	France	1959	Western Europe	global	97.70	31.90	95.40	95.40	3090178	1527	75
21528_4#18	II_21528_4_18	Vanuatu	2015	Melanesia	png	99.09	173.54	92.43	92.43	7562074	242	31
21528_4#16	III_21528_4_16	Papua New Guinea	2014	Melanesia	png	96.21	211.80	90.76	90.76	9140446	1764	77
21395_4#11	II_21395_4_11	Vanuatu	2013	Melanesia	png	98.95	159.08	91.70	91.70	7365390	287	28
21528_4#11	II_21528_4_11	Vanuatu	2013	Melanesia	png	99.00	167.55	95.88	95.88	6956750	286	31
21395_4#15	III_21395_4_15	Papua New Guinea	2014	Melanesia	png	95.48	209.92	91.76	91.76	9269746	1361	57
21395_4#18	II_21395_4_18	Vanuatu	2015	Melanesia	png	99.09	175.19	90.00	90.00	8147116	245	26
21395_4#31	III_21395_4_31	Papua New Guinea	2011	Melanesia	png	94.65	186.91	89.08	89.08	8443648	1359	64
21395_4#6	II_21395_4_6	Fiji	2012	Melanesia	png	99.11	174.62	90.41	90.41	8184246	190	28
21395_4#22	II_21395_4_22	Solomon Islands	2007	Melanesia	png	98.89	197.62	93.94	93.94	8783768	274	27
21528_4#17	III_21528_4_17	Papua New Guinea	2015	Melanesia	png	95.45	202.46	95.77	95.77	8181164	1403	75
21528_4#12	II_21528_4_12	Samoa	2013	Polynesia	png	98.96	163.31	92.37	92.37	7048522	275	33
21528_4#8	II_21528_4_8	Fiji	2012	Melanesia	png	99.08	179.34	96.71	96.71	7396330	306	31
21395_4#3	III_21395_4_3	Papua New Guinea	2011	Melanesia	png	96.43	180.01	91.32	91.32	8058556	1584	68
21528_4#13	III_21528_4_13	Papua New Guinea	2014	Melanesia	png	98.99	188.43	92.57	92.57	8202782	1431	75
21528_4#15	III_21528_4_15	Papua New Guinea	2014	Melanesia	png	95.48	211.98	94.62	94.62	8678470	1367	69
21395_4#12	II_21395_4_12	Samoa	2013	Polynesia	png	98.96	160.27	89.24	89.24	7544470	275	25
21395_4#16	III_21395_4_16	Papua New Guinea	2014	Melanesia	png	96.19	210.15	88.20	88.20	9731878	1676	63
21395_4#13	III_21395_4_13	Papua New Guinea	2014	Melanesia	png	98.99	186.58	89.22	89.22	8883208	1417	58

21395_4#8	II_21395_4_8	Fiji	2012	Melanesia	png	99.08	178.87	94.17	94.17	7887002	329	31
21395_4#29	III_21395_4_29	Papua New Guinea	2010	Melanesia	png	97.19	153.63	84.99	84.99	7490594	1758	64
21395_4#28	III_21395_4_28	Papua New Guinea	1999	Melanesia	png	98.00	153.75	87.72	87.72	7408106	1563	63
21395_4#17	III_21395_4_17	Papua New Guinea	2015	Melanesia	png	95.51	203.24	93.47	93.47	8736730	1385	62
21528_4#31	III_21528_4_31	Papua New Guinea	2011	Melanesia	png	94.68	189.83	91.94	91.94	7928402	1374	80
21528_4#28	III_21528_4_28	Papua New Guinea	1999	Melanesia	png	98.09	158.30	91.20	91.20	6946692	1533	72
21528_4#3	III_21528_4_3	Papua New Guinea	2011	Melanesia	png	96.46	184.35	93.81	93.81	7721396	1612	75
21395_4#7	II_21395_4_7	Fiji	2012	Melanesia	png	99.12	148.45	90.76	90.76	7050792	170	26
21395_4#30	III_21395_4_30	Papua New Guinea	2008	Melanesia	png	94.73	176.39	90.58	90.58	7975886	1336	67
21528_4#7	II_21528_4_7	Fiji	2012	Melanesia	png	99.12	159.77	95.77	95.77	6676906	164	30
21528_4#29	III_21528_4_29	Papua New Guinea	2010	Melanesia	png	97.14	155.19	87.41	87.41	7055192	1881	69
21528_4#30	III_21528_4_30	Papua New Guinea	2008	Melanesia	png	94.73	179.50	94.70	94.70	7295474	1349	76
21528_4#6	II_21528_4_6	Fiji	2012	Melanesia	png	99.11	178.69	94.11	94.11	7591108	196	31
21528_4#22	II_21528_4_22	Solomon Islands	2007	Melanesia	png	98.90	202.11	97.09	97.09	8267800	278	28

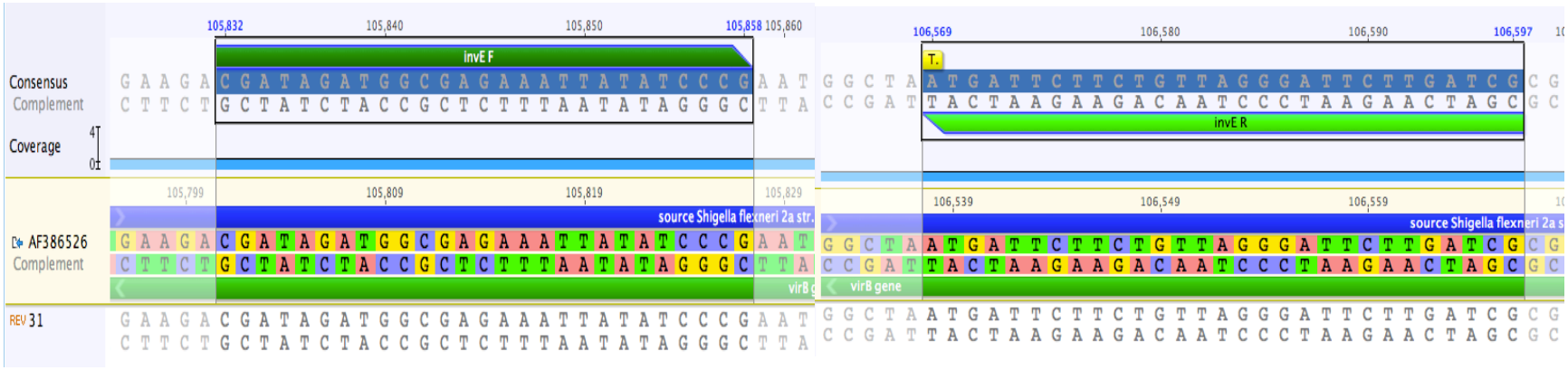
Appendix 4.15 *Shigella flexneri* PCR primer checks using Geneious

S. flexneri #34_PNG isolate

ial Forward and Reverse primers

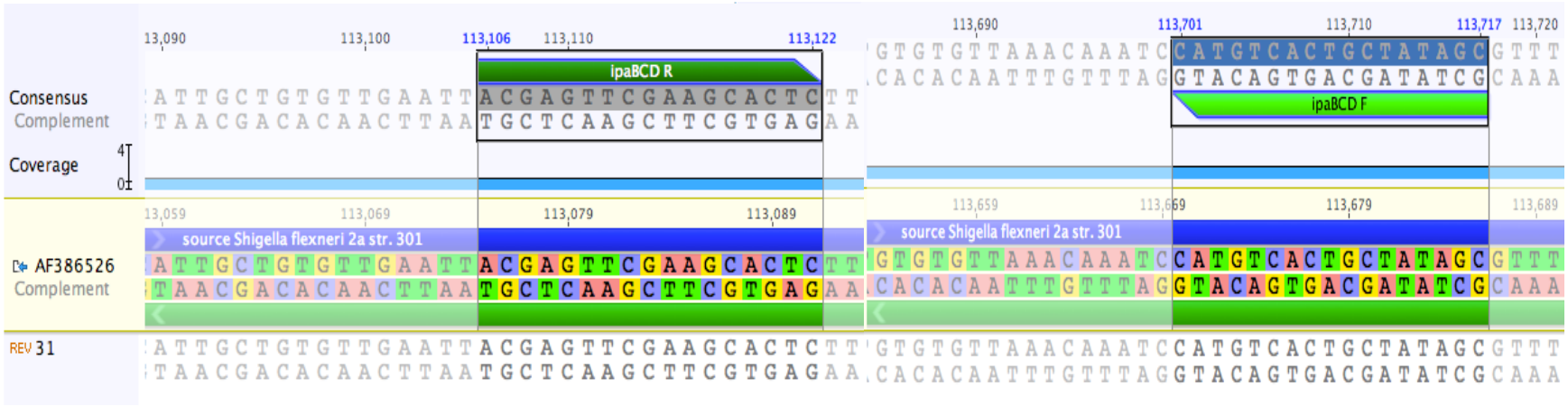


invE Forward and Reverse primers

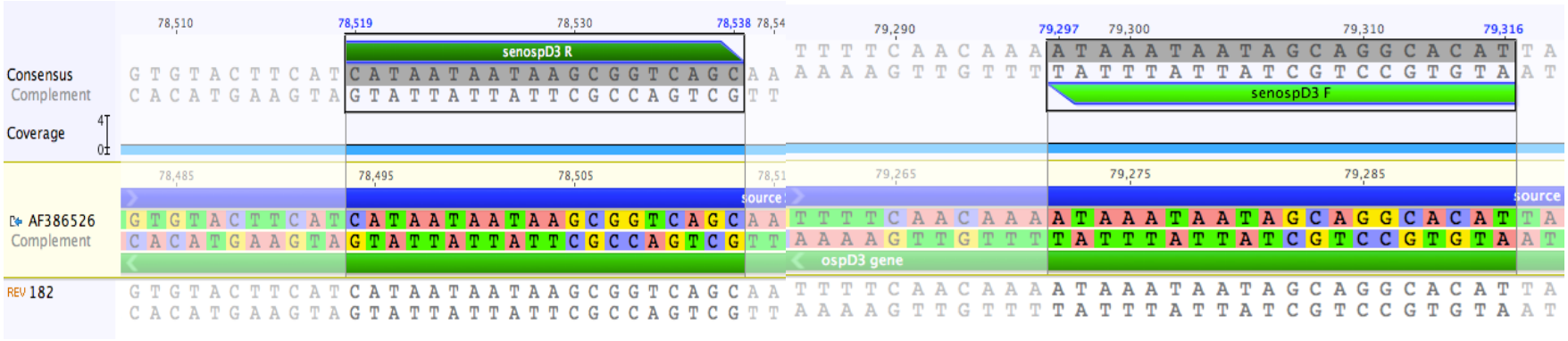


S. flexneri #34_PNG isolate

ipaBCD Forward and Reverse primers

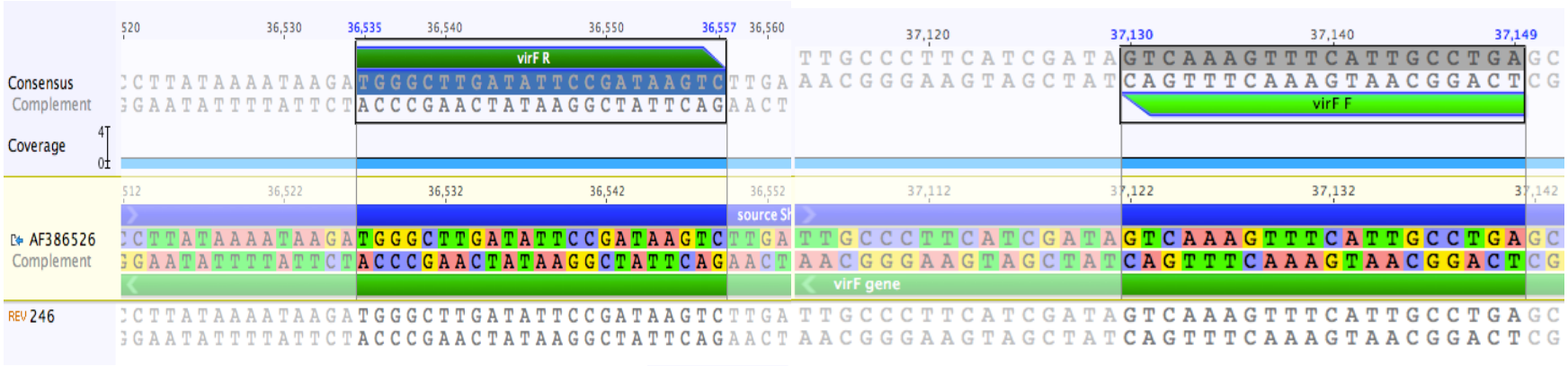


senospD3 Forward and Reverse primers

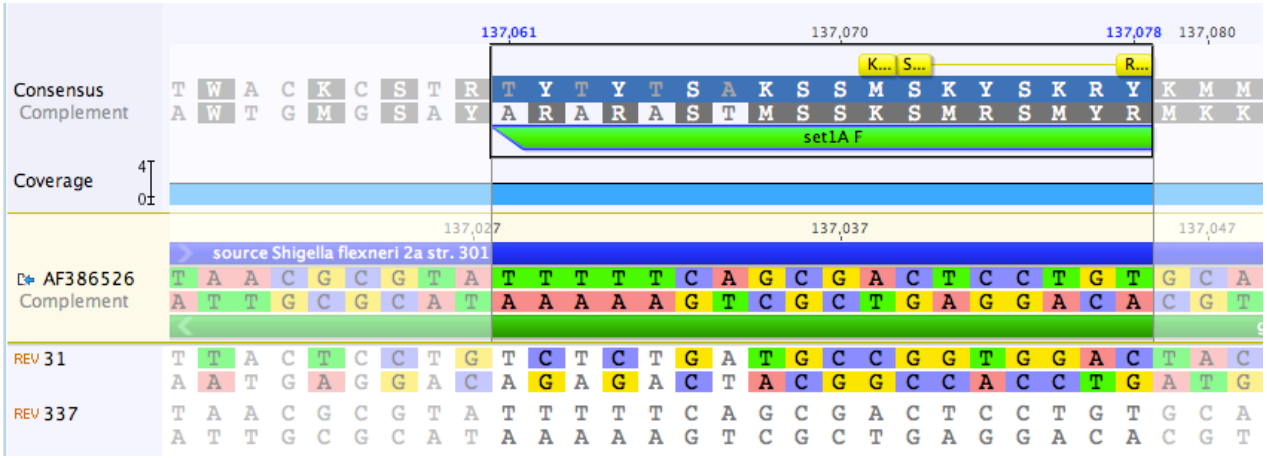


S. flexneri #34_PNG isolate

virF Forward and Reverse primers

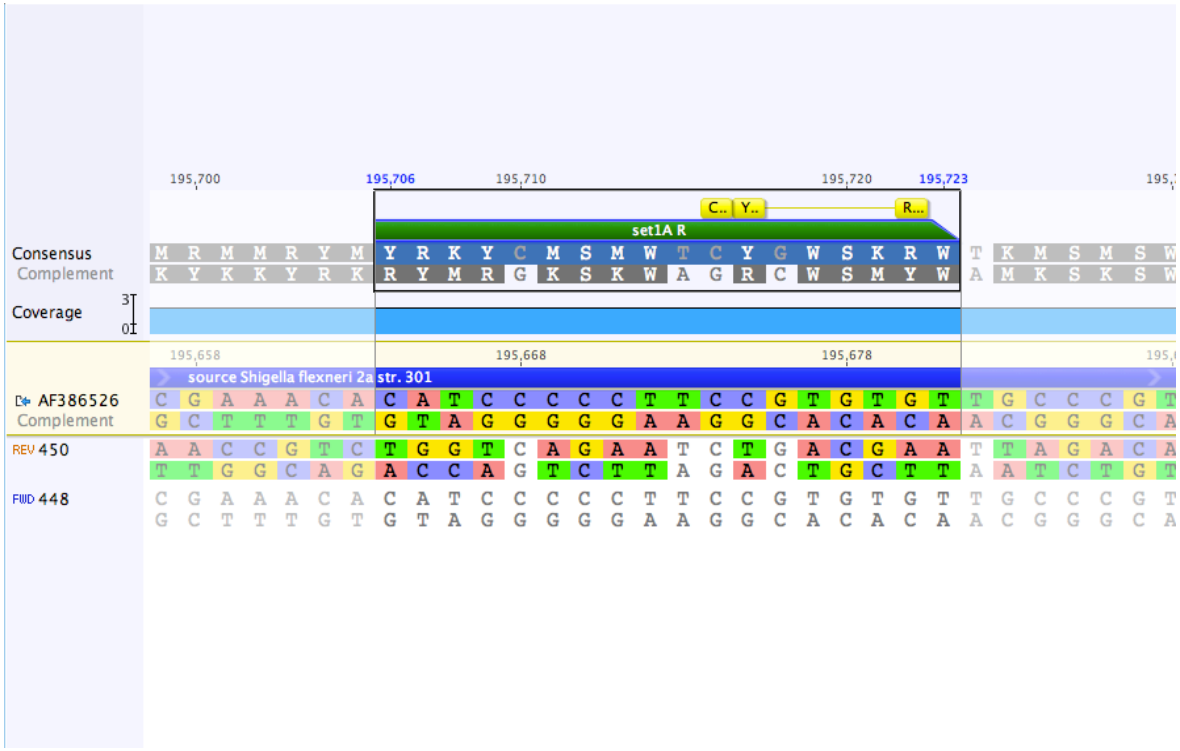


set1A Forward primer with 3 miss-matches (Note: Only the Forward primer was detected in this isolate)



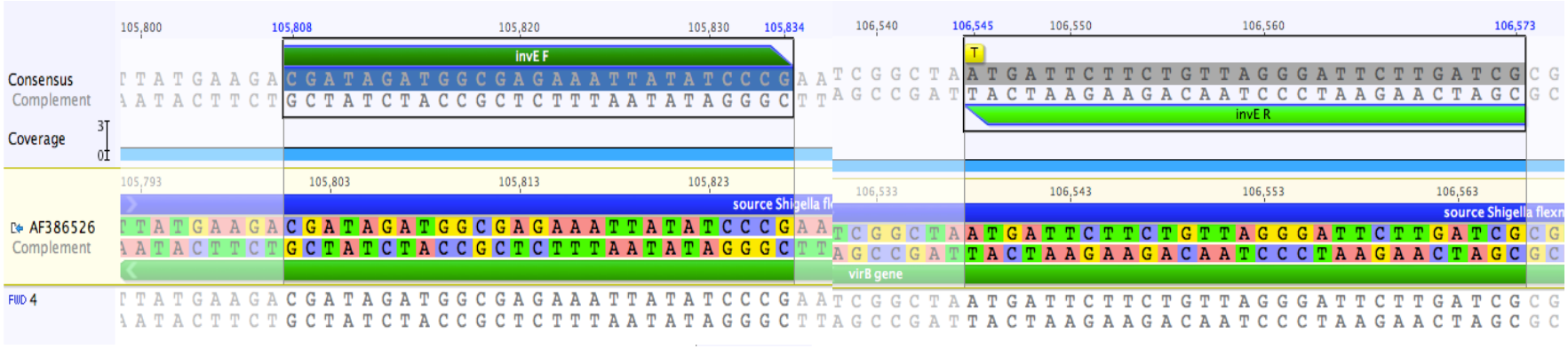
S. flexneri #48_PNG isolate

set1A Reverse primer with 3 miss-matches (Note: Only the Reverse primer was detected in this isolate)

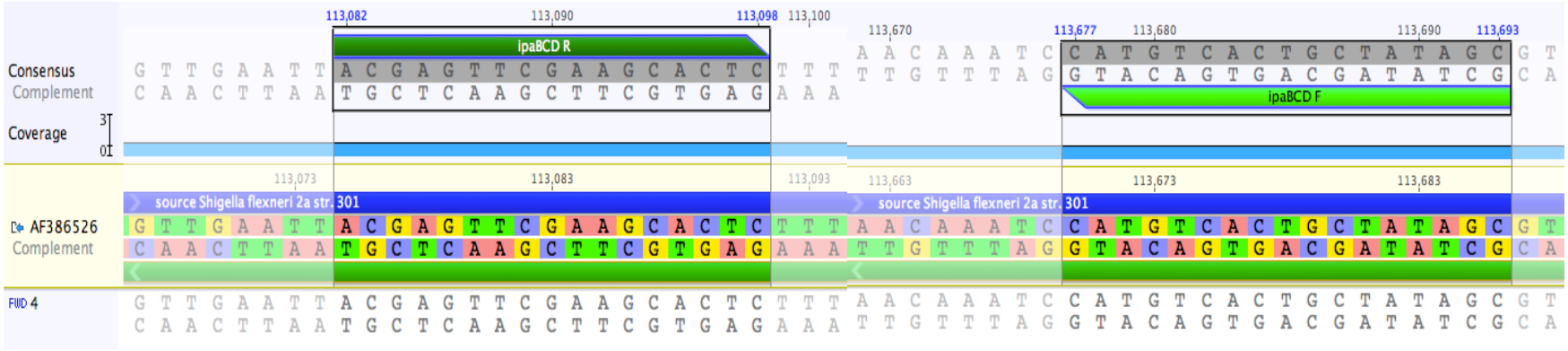


S. flexneri #51_PNG isolate

invE Forward and Reverse primers

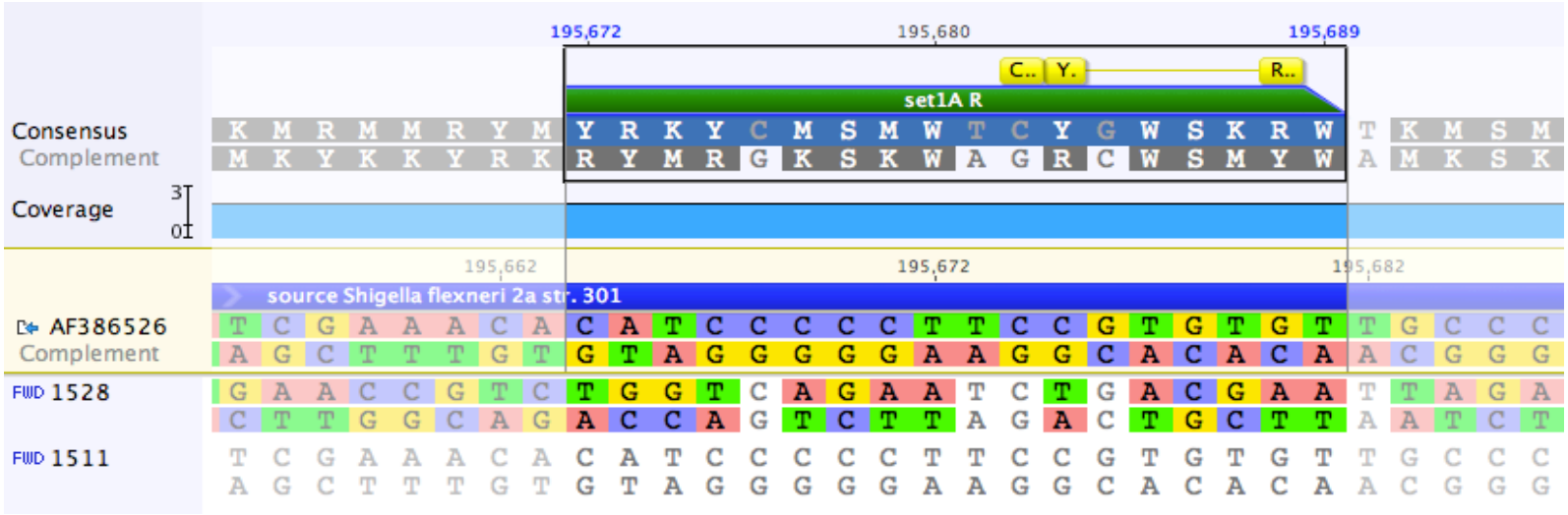


ipaBCD Forward and Reverse primers



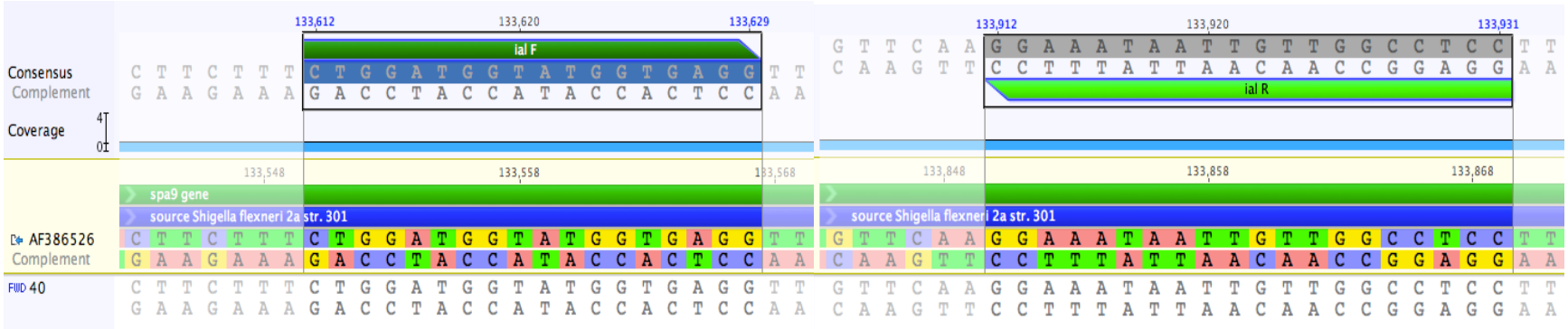
S. flexneri #51_PNG isolate

Set1A Reverse primer with 3 miss-matches (Note: Only the Reverse primer was detected in this isolate)

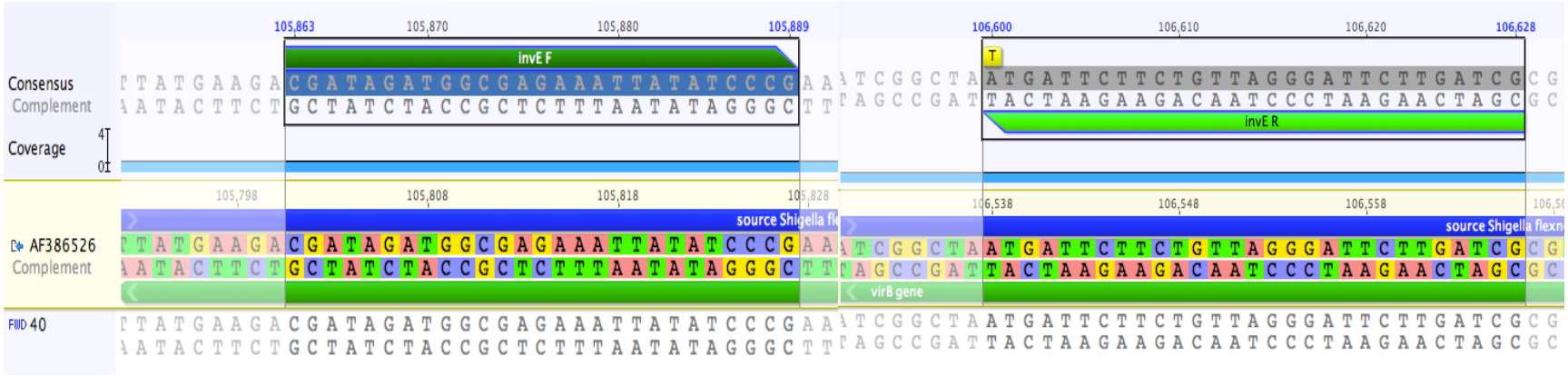


S. flexneri #52_PNG isolate

ial Forward and Reverse primers

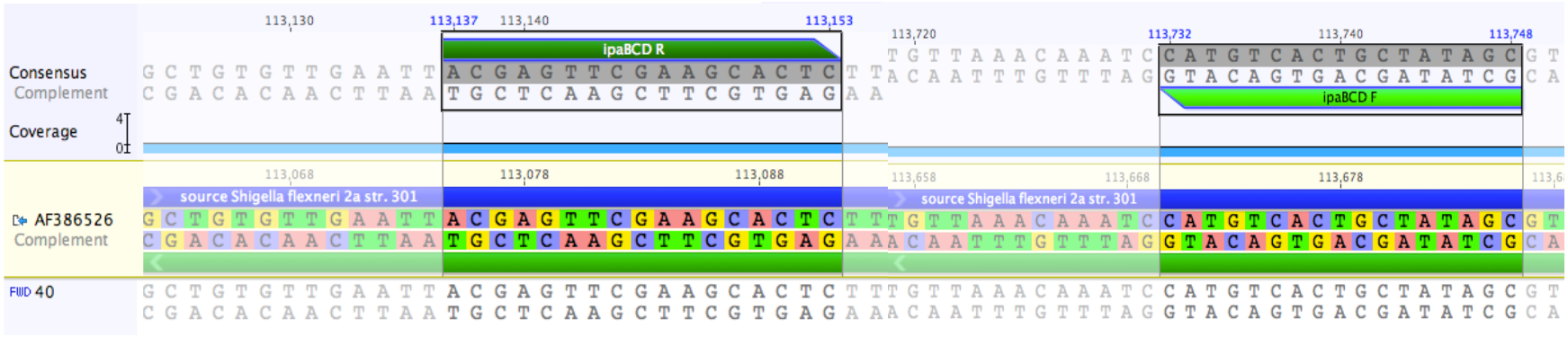


invE Forward and Reverse primers

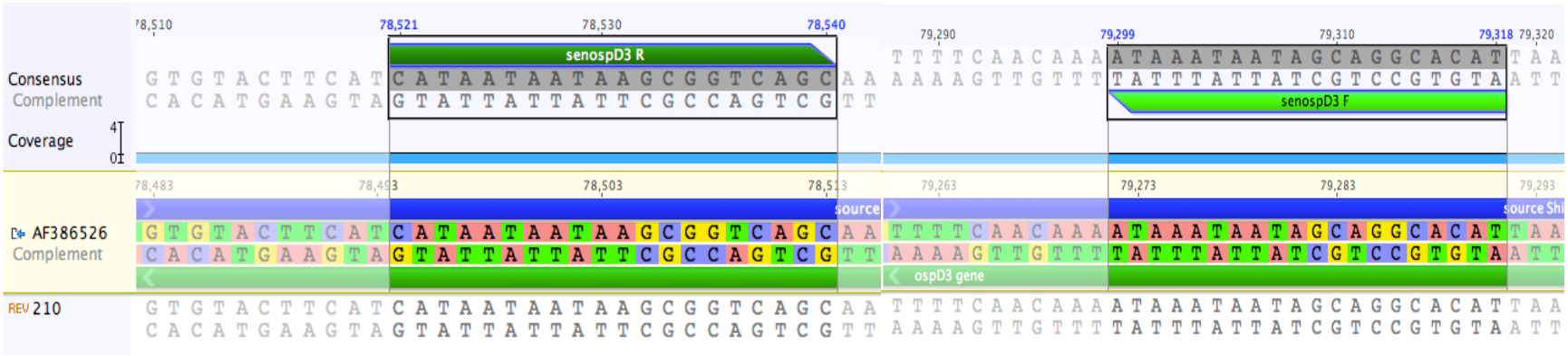


S. flexneri #52_PNG isolate

ipaBCD Forward and Reverse primers

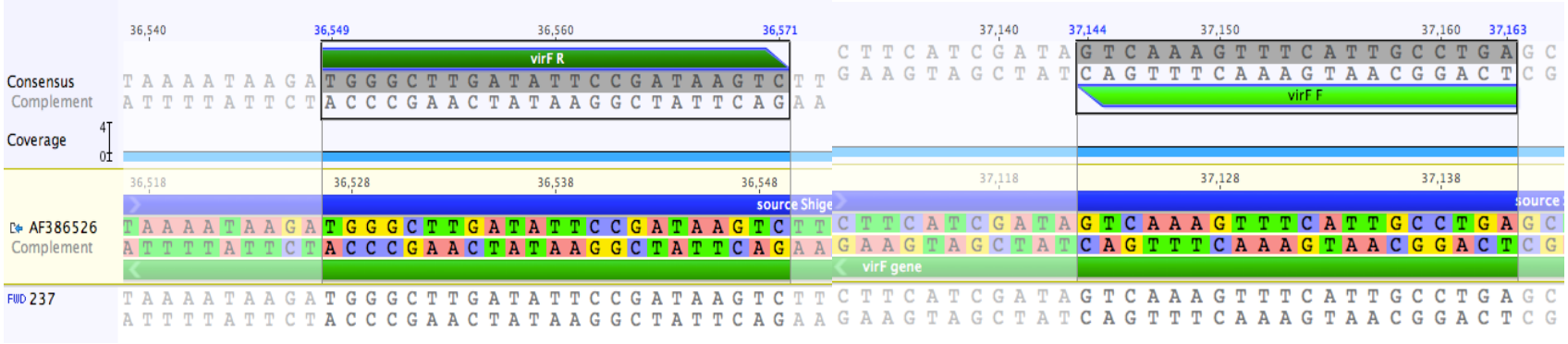


senospD3 Forward and Reverse primers

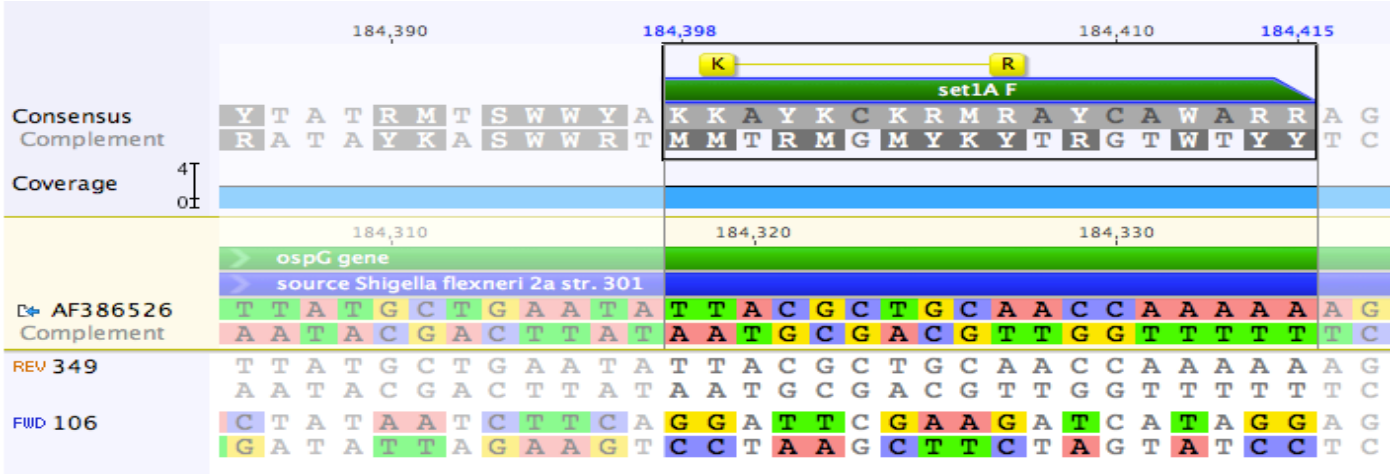


S. flexneri #52_PNG isolate

virF Forward and Reverse primers

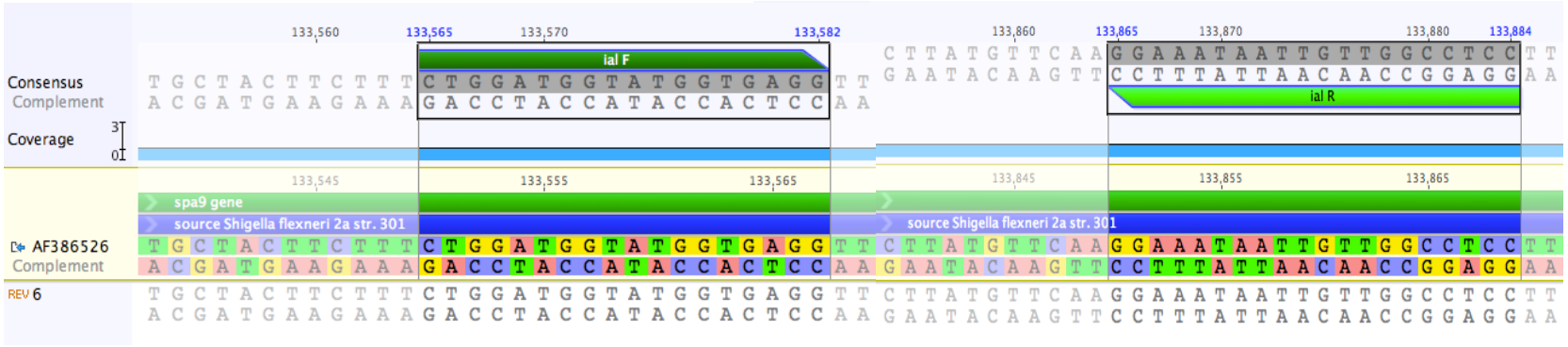


Set1A Forward primer with 2 miss-matches (Note: Only the Forward primer was detected in this isolate)

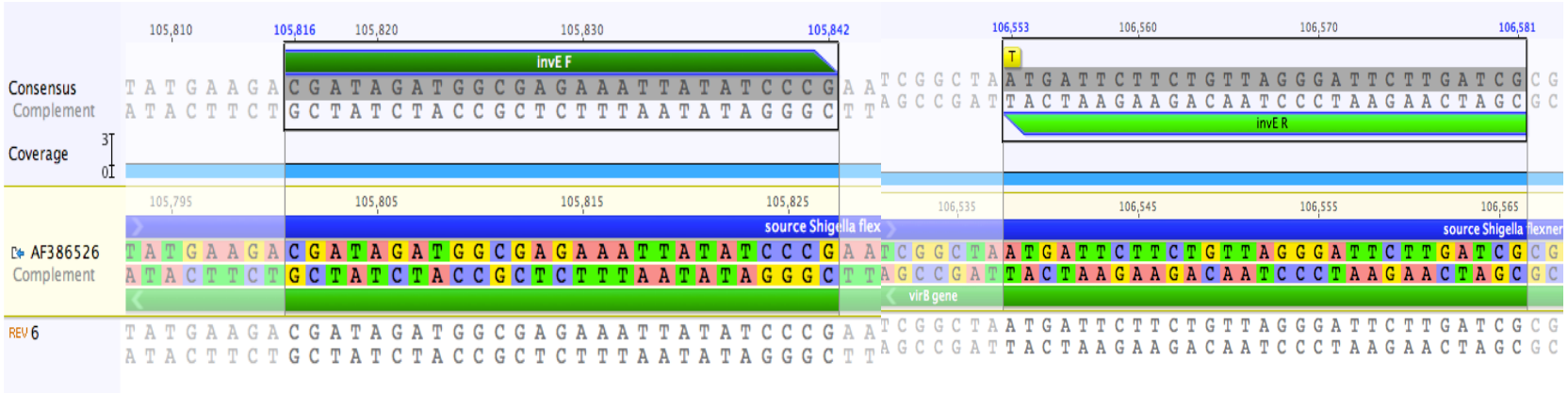


S. flexneri #53_PNG isolate

ial Forward and Reverse primers

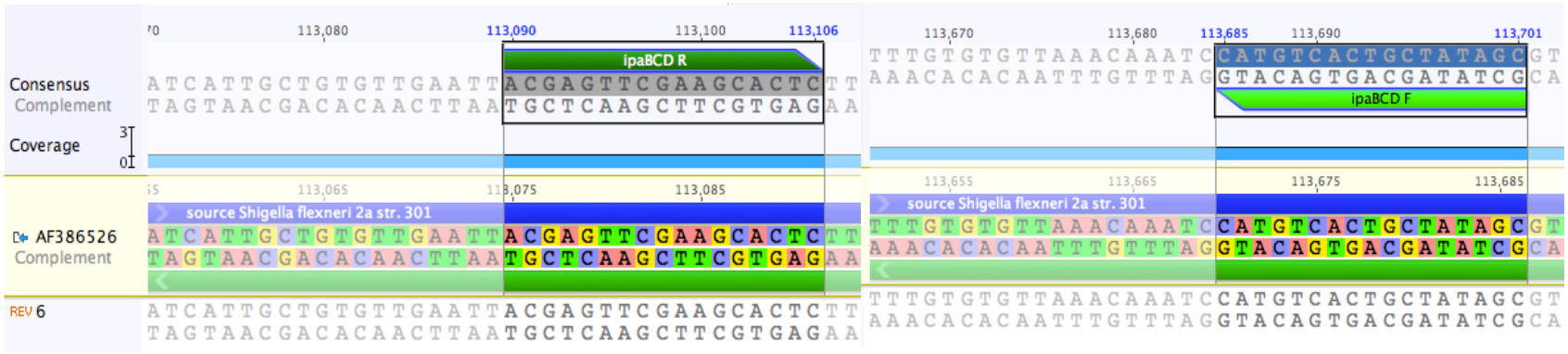


invE Forward and Reverse primers

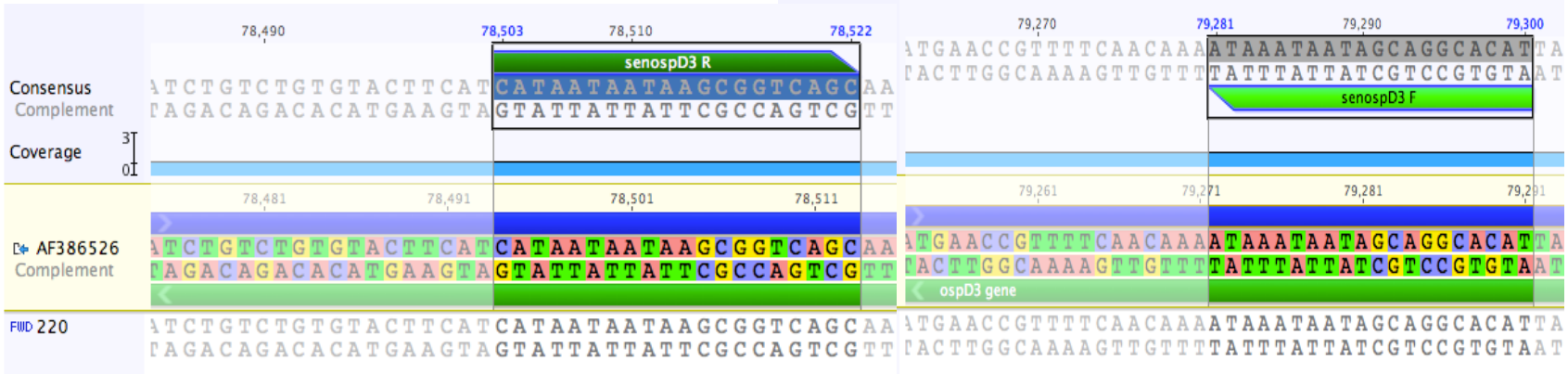


S. flexneri #53_PNG isolate

ipaBCD Forward and Reverse primers

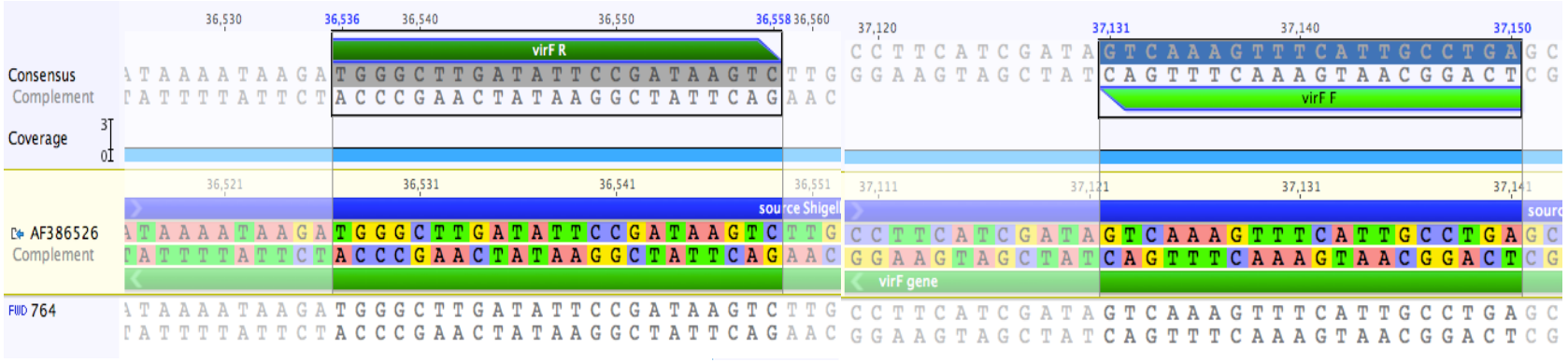


senospD3 Forward and Reverse primers

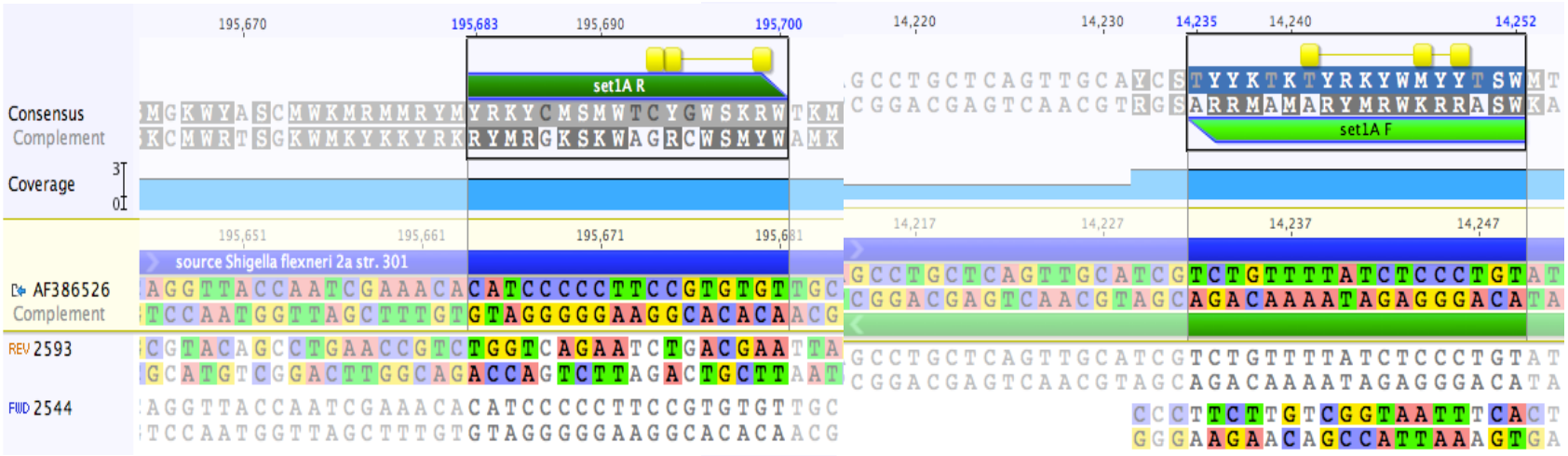


S. flexneri #53_PNG isolate

virF Forward and Reverse primers

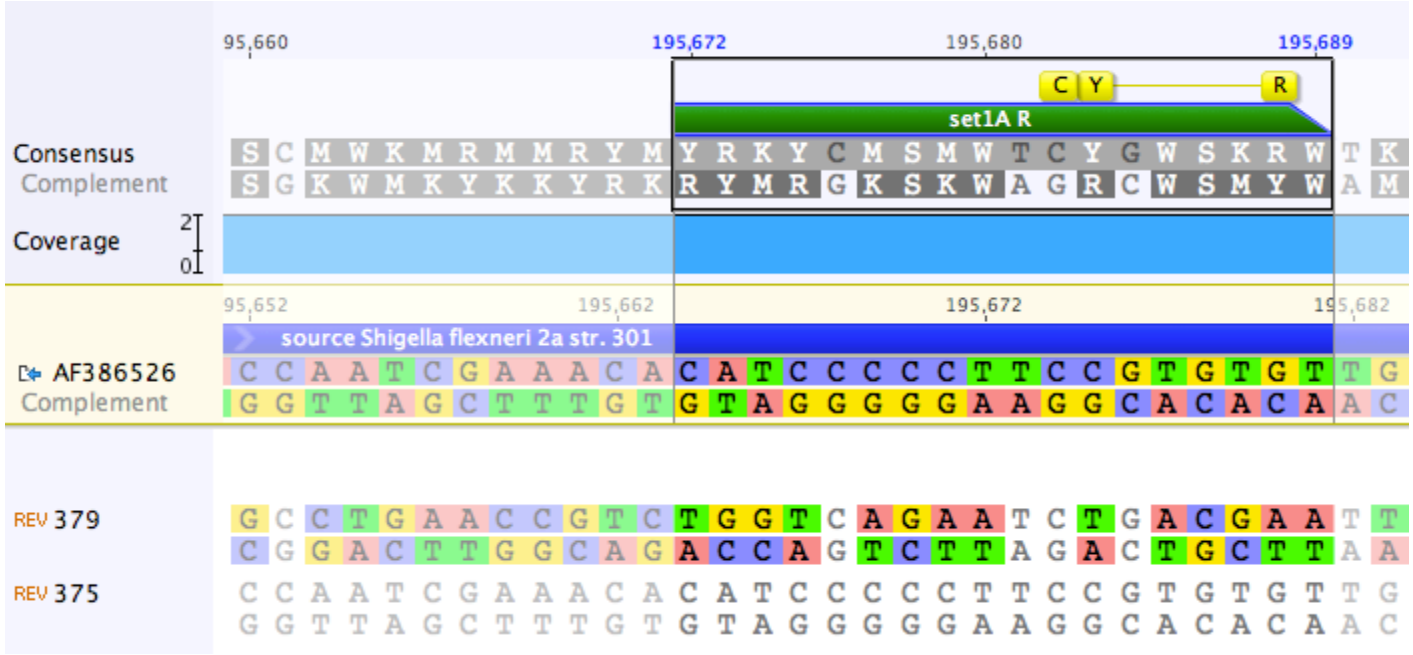


Set1A Forward and Reverse primers (Note: 3 miss-matches found in both primers)



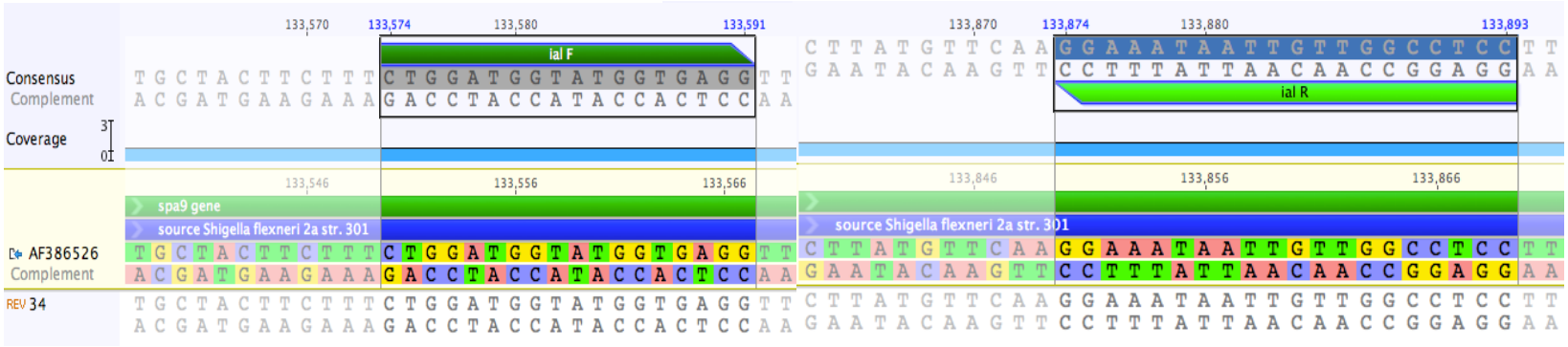
S. flexneri #61_PNG isolate

Set1A Reverse primer with 3 miss-matches (Note: Only the Reverse primer was detected in this isolate)

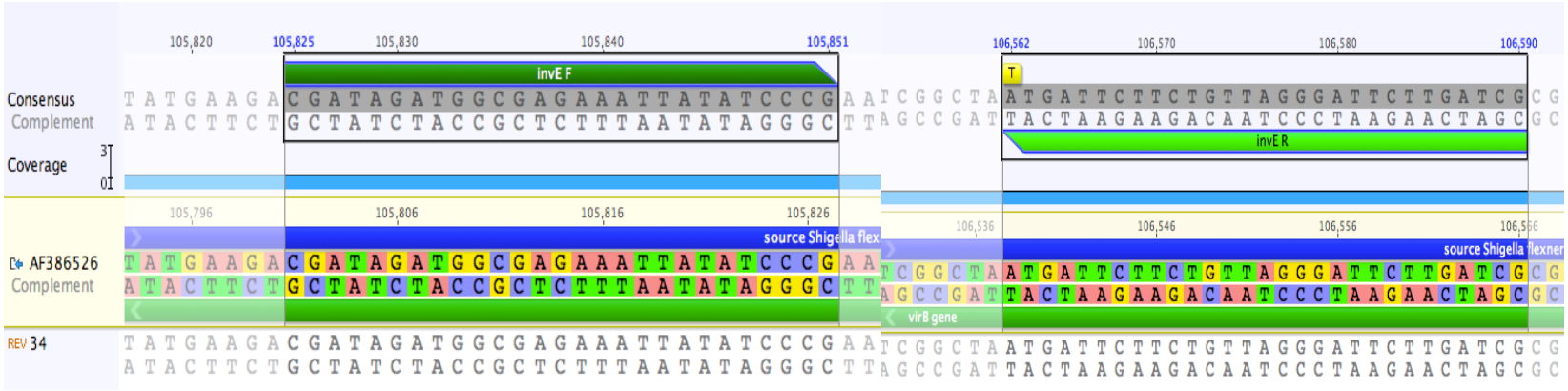


S. flexneri #62_PNG isolate

ial Forward and Reverse primers

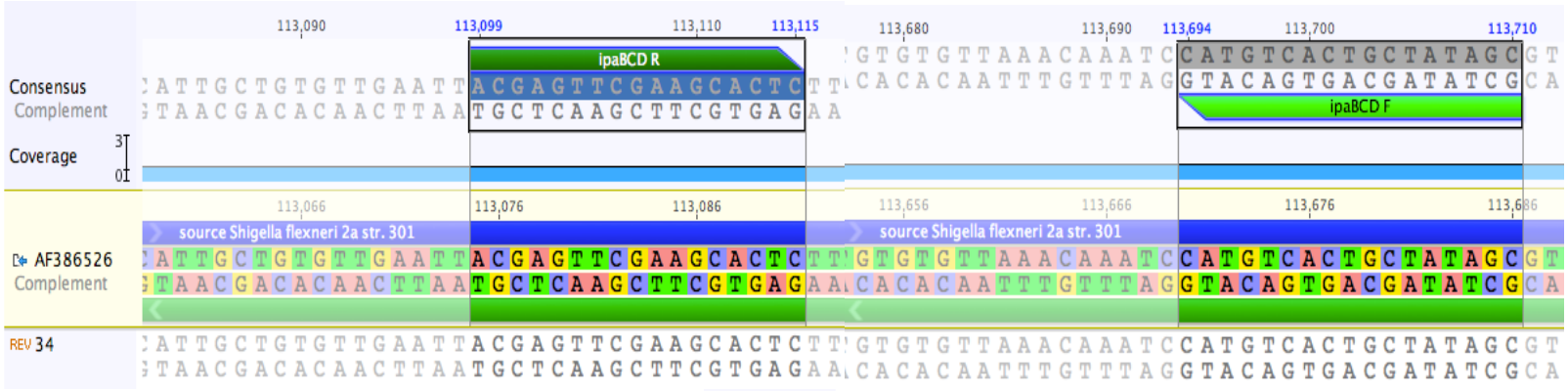


invE Forward and Reverse primers

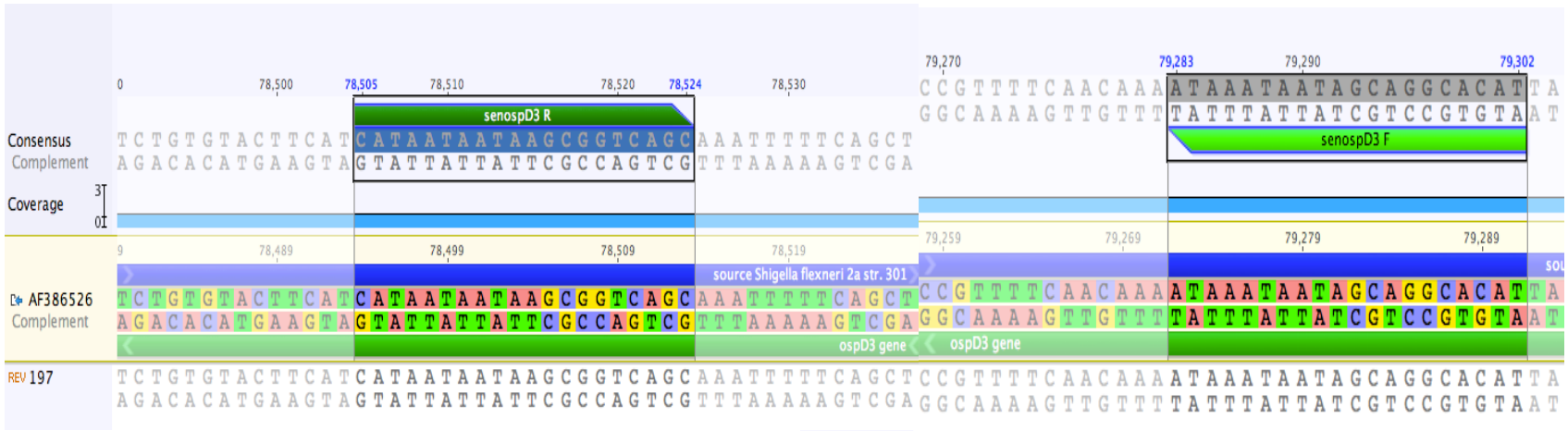


S. flexneri #62_PNG isolate

ipaBCD Forward and Reverse primers

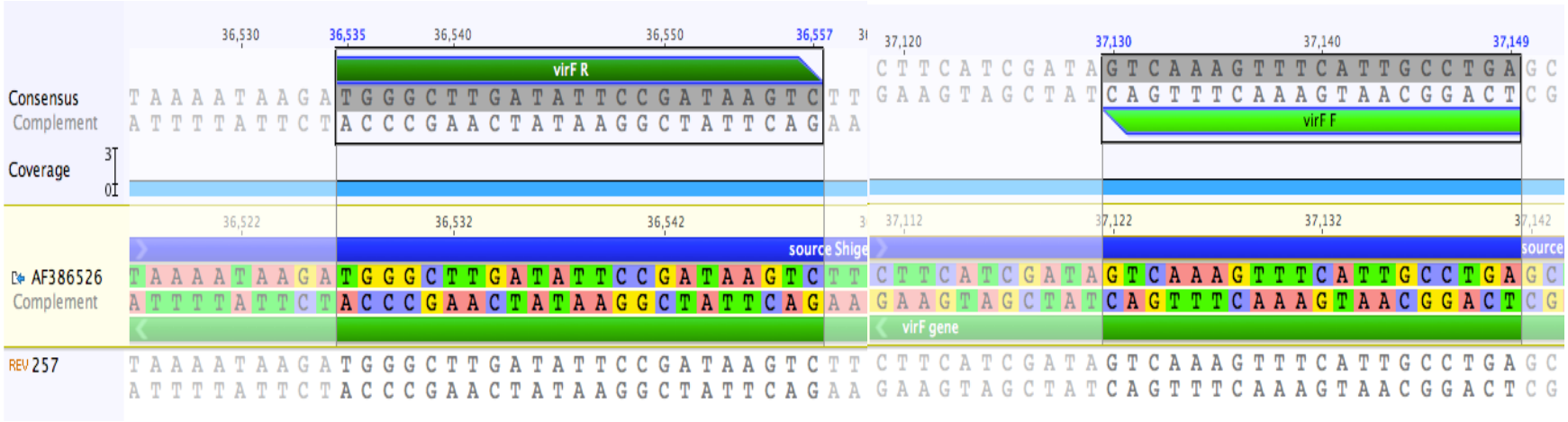


senospD3 Forward and Reverse primer



S. flexneri #62_PNG isolate

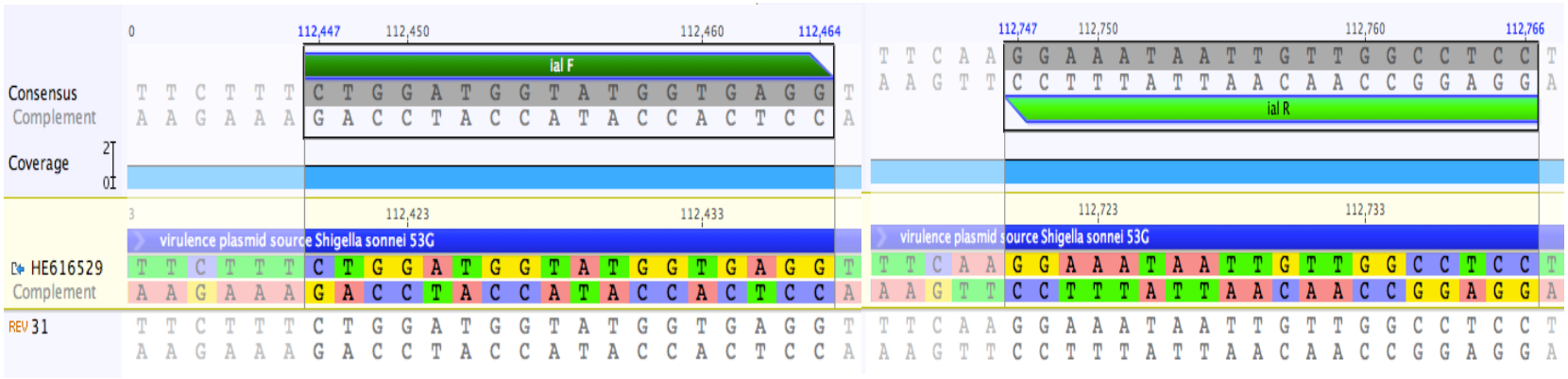
virF Forward and Reverse primer



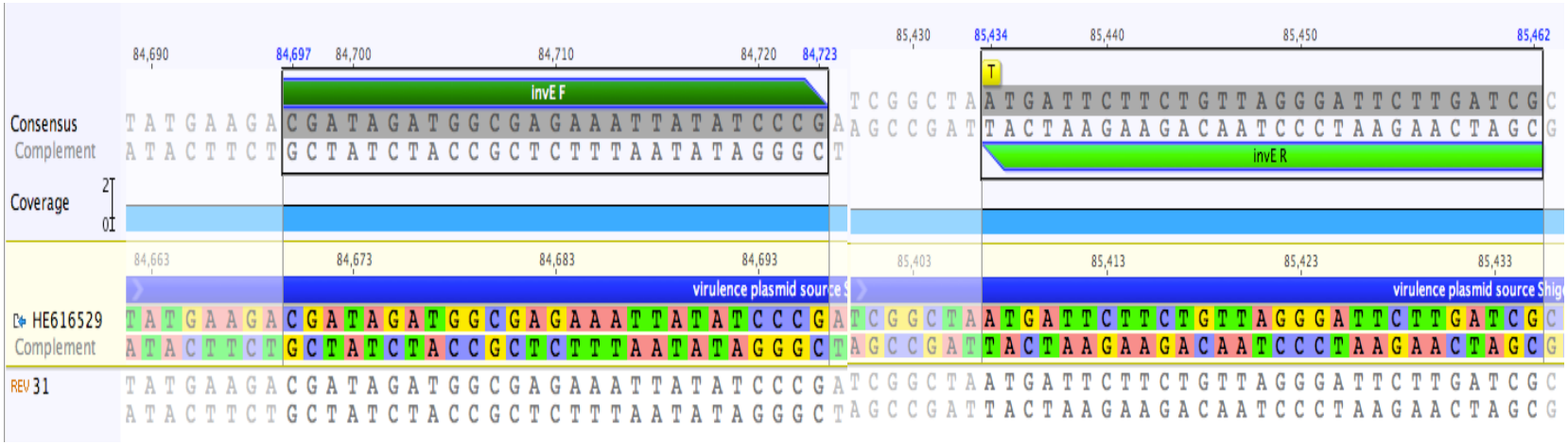
Appendix 4.16 ***Shigella sonnei*** **PCR primer checks using Geneious**

S. sonnei #3_PNG isolate

ial Forward and Reverse primers

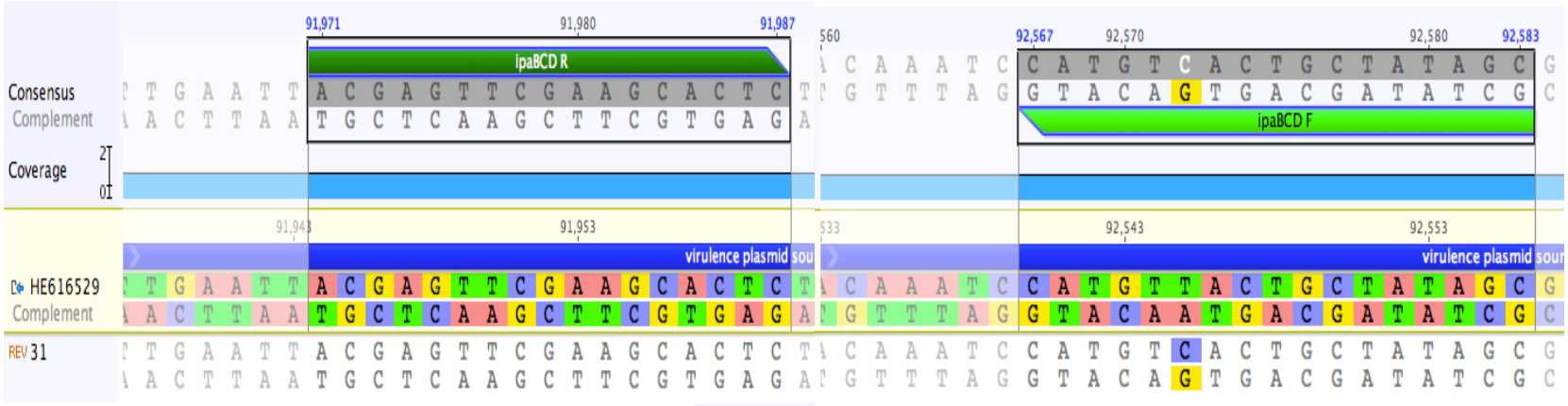


invE Forward and Reverse primers



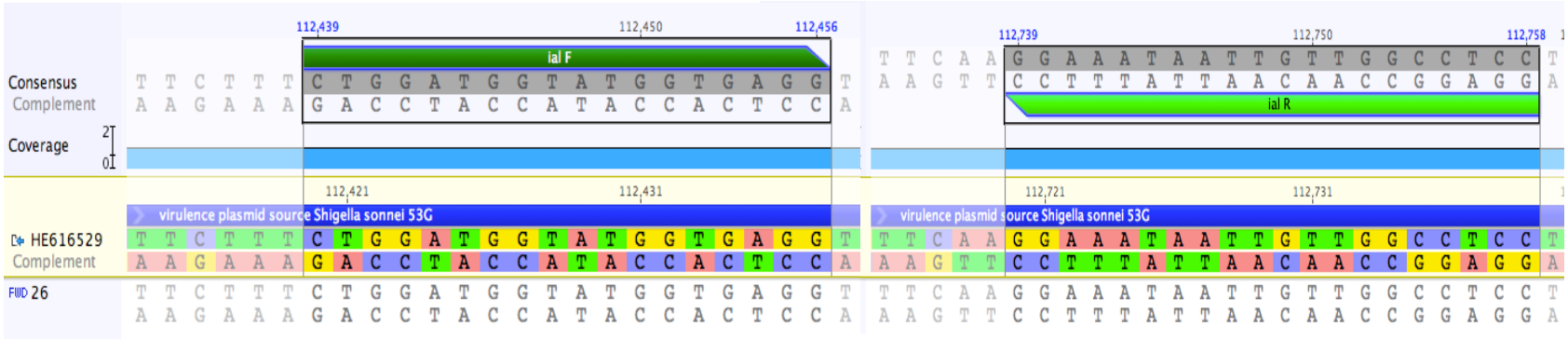
S. sonnei #3_PNG isolate

ipaBCD Forward and Reverse primers

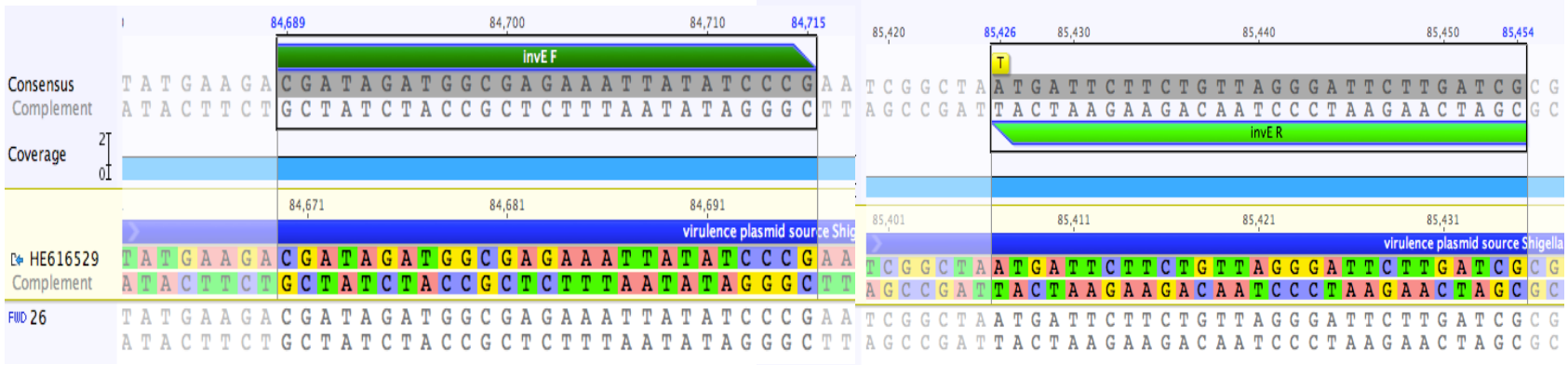


S. sonnei #18_VUT isolate

ial Forward and Reverse primers

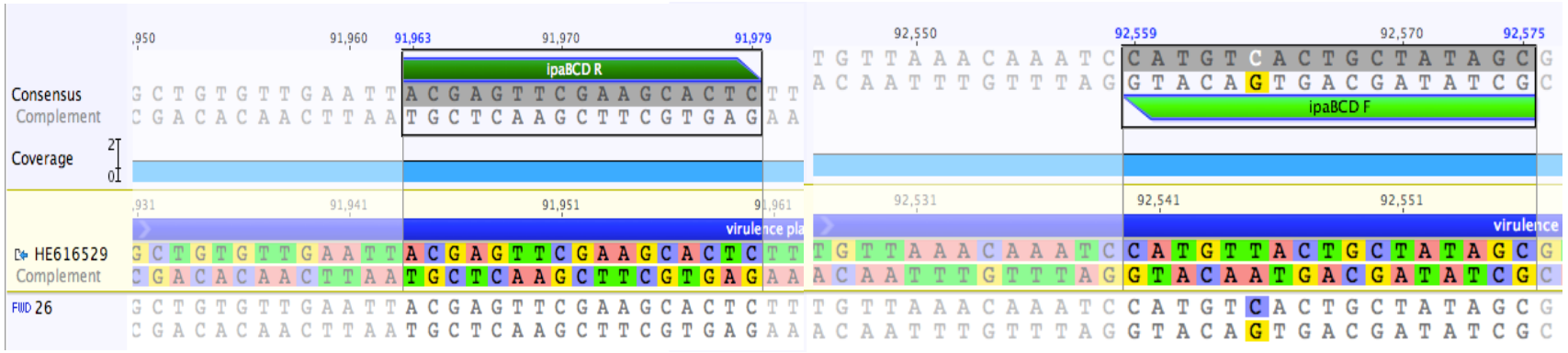


invE Forward and Reverse primers

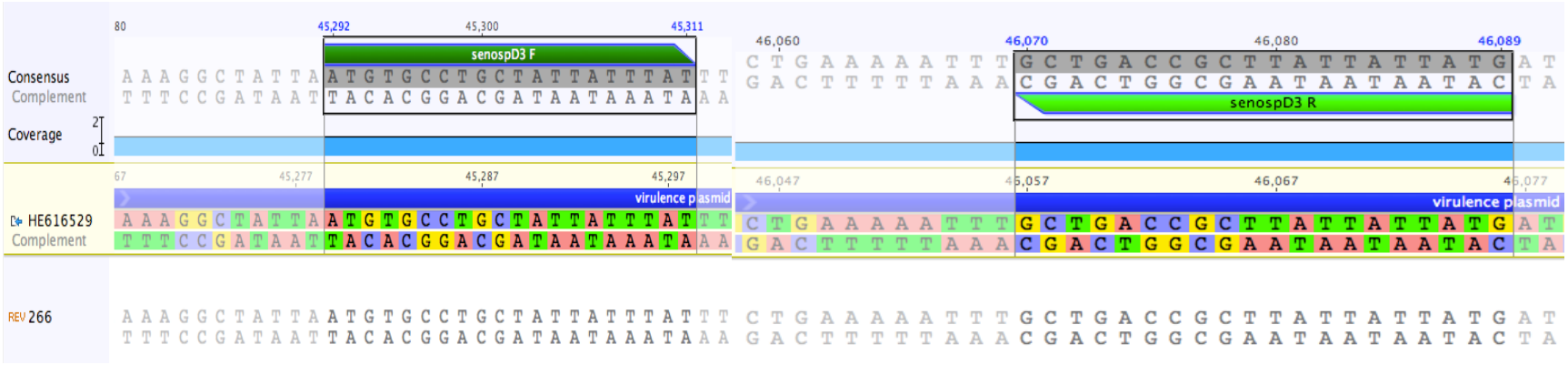


S. sonnei #18_VUT isolate

ipaBCD Forward and Reverse primers

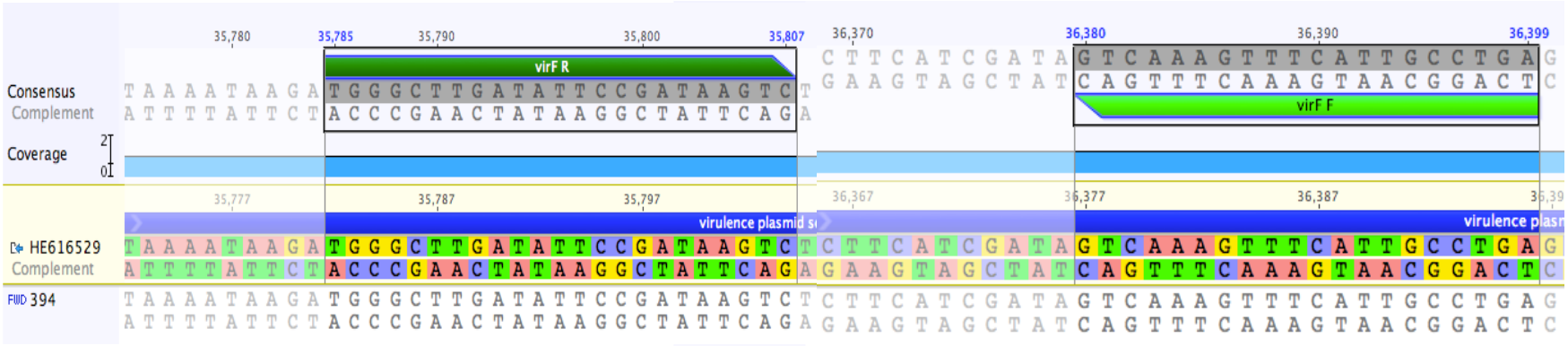


senospD3 Forward and Reverse primers



S. sonnei #18_VUT isolate

virF Forward and Reverse primers



Set1A Reverse primer with 3 miss-matches (Note: Only the Reverse primer was detected in this isolate)