Phenotypic and genotypic characteristics for *Escherichia coli* strains responsible for bacterial bloom events in Australia

Buddhie Samanmalie Nanayakkara

March 2019

A thesis submitted for the Degree of Doctor of Philosophy of
The Australian National University

Declaration

I declare that this thesis titled 'Phenotypic and genotypic characteristics for *Escherichia coli* strains responsible for bacterial bloom events in Australia' is the product of my original work. This thesis contains no material which has been accepted for the award of any other degree or diploma in any university. To the best of my knowledge, it contains no material previously published or written by another person, except where due reference is made in the text.

Buddhie Samanmalie Nanayakkara
March 2019

Acknowledgements

I express my sincere gratitude to my primary supervisor Prof. David Gordon, for the guidance and encouragement given throughout my research. The knowledge and skills I learnt from you are immeasurable. Your composure and understanding kept me going, always. David, you are an inspiration! I am deeply grateful to members of my supervisory panel Dr. Claire O'Brien and Prof. Gavin Huttley for the guidance. Thank you Claire for the constructive feedback on my work.

Thank you, all Gordon Lab members past and present; Sam, Angel, Belinda, Heli, Judith, Ojas, Mah, Mukta, Truc, Veronica, Jaki, and Madiha for the research and lab skills learnt, helping me whenever needed, and for making a beautiful work environment. Not only are we lab mates, we became good friends. My dear Angel, thank you for being so caring a friend.

I thank everyone in E&E and RSB for all the help. You all made my stay there one full of fond memories. My dear Kalya Nangi, you are one with so big a heart, and you sure did keep me confident and smiling. My dear friends Bo, Mas Arild, Fito, Shukhrat, James, and Hee-Jin, thank you for the immeasurable support. We sure did drink many a cup of tea sweetened not with sugar but with chats and laughter.

Thank you all my friends in Ursula Hall, ANU, and those outside of ANU, especially Aunty Savithri, Dulamanie Akki, Sean Aiya, Dulanjali, Buddhima, Upul Aiya, Adarshi, Dimanthi, and many others more. You all made the beautiful Canberra so much more beautiful, a home away from home.

I am deeply grateful to all my teachers past and present, especially Prof. Charmalie Abayasekara, Prof. Anoma Perera, and Dr. Jilushi Damunupola, for motivating me to do better.

Thank you my dearest Ammi and Thaththi, and Mahattaiya for your unconditional love. You are my strength, and if I have achieved anything in life, it is because of you.

Last but not least, thank you to The Australian National University and Australia for giving me this opportunity to pursue Science.

Thank you!

Abstract

Escherichia coli is widely used as an indicator of recent faecal contamination of drinking and recreational waters. However, evidence suggests that *E. coli* can proliferate in the environment outside a host, confounding its use as a faecal indicator. *E. coli* strains that produce significantly elevated counts of 10,000 - 100,000 cells/100 ml of water (bloom events) are reported from freshwater reservoirs and recreational lakes in Australia. Bloom strains are not faecal associated and may represent free-living *E. coli*. A limited number of strains are responsible for bloom events and all belong to *E. coli* phylogroups A and B1. Bloom strains have acquired a capsule originating from *Klebsiella*.

Diversity and distribution of *Klebsiella* capsules in *E. coli* overall and in bloom strains were investigated. A PCR-based protocol was developed to detect capsule-positive *E. coli* and discriminate strains that harbour bloom strain-associated capsule types. Bloom strain attributes that could lead to the elevated cell densities observed in bloom events were experimented. The B1 bloom strain from the east coast (termed B1-001) which is numerically dominant and always present in bloom events was further characterised, and why B1-001 strain has not been detected in recent bloom events assessed using the Colilert-18® system was investigated.

Frequency of *Klebsiella* capsules in *E. coli* was only 7% and 23 different *Klebsiella* capsule types were detected. All bloom strains were encapsulated and seven *Klebsiella* capsule types were detected among the eight bloom strains isolated to date. Capsules were observed only in strains from *E. coli* phylogroups A, B1, and C, and all encapsulated strains were of O-serogroups O8, O9, and O89. Capsule gene region and the adjacent O-antigen gene region in encapsulated *E. coli* are a result of a horizontal gene transfer event that occurred between *E. coli* and *Klebsiella*. The PCR accurately detected known bloom, non-bloom encapsulated, and capsule-negative strains.

A pan genome comparison of phylogroup A *E. coli* revealed that the iron uptake system encoded by *fecIRABCDE* operon was over-represented among bloom strains (100%) compared to non-bloom *E. coli* (<39%). Growth assays however showed that the *fec* operon is unlikely

to contribute to the elevated cell densities. In contrast, strains that were encapsulated had a growth rate advantage compared to capsule-negative strains.

B1-001 bloom strain was closely related to *Shigella* but its closest relatives were lactose-positive *E. coli*. Several features that are beneficial for a free-living lifestyle such as flagella and curli were disrupted in B1-001. In Colilert-18®, B1-001 was heavily outcompeted by the two phylogroup A bloom strains from the east coast, explaining why B1-001 was not detected in recent bloom events assessed using Colilert-18®.

The current study suggests that any *E. coli* strain that harbours a *Klebsiella* capsule may be able to produce elevated counts under conducive environmental conditions. The recurrence of bloom events across Australia confounds the use of *E. coli* as a water quality indicator and urges a shift to alternative indicators.

Table of Contents

Chapter 1. General Introduction

1.1 *Escherichia coli*

The genus *Escherichia* constitutes several species, namely, *Escherichia coli*, *E. fergusonii*, *E. albertii*, *E. blattae*, *E. vulneris* and *E. hermannii*. The latter three species however are not considered as valid members of *Escherichia* as they are distantly related to the other *Escherichia* species (Gordon, 2013). *E. coli* belongs to the Proteobacteria phylum of Bacteria, within the class Gammaproteobacteria, and family, *Enterobacteriaceae*. It is a Gram-negative, motile, non-spore forming, facultative anaerobe (Hartl and Dykhuizen, 1984; Leclerc et al., 2001). *E. coli* has long been recognised as a normal inhabitant of the gastrointestinal (GI) tract of warm-blooded animals (Alm et al., 2011). Inside a host, *E. coli* exists as a harmless commensal, or occasionally as a pathogen that causes intestinal or extra-intestinal infection (van Elsas et al., 2011). Regardless, *E. coli* does not require a host to reproduce, as it is capable of surviving in the outside environment (Ishii and Sadowsky, 2008; Alm et al., 2011). Due to its importance from a public health perspective, its use as a water quality indicator, and use in molecular biology, *E. coli* is one of the best-studied organisms on earth (Savageau, 1983; Hartl and Dykhuizen, 1984).

The life cycle of *E. coli* comprises two phases spent in two different habitats; the primary habitat and the secondary habitat (Savageau, 1983). The primary habitat is the lower intestine of warm-blooded animals while the secondary habitat includes environments outside the host, including soil, water, and sediment (Savageau, 1983; Ishii and Sadowsky, 2008). At the same time, ectothermic animals are known to harbour *E. coli* (Hansen et al, 2008; Frick et al., 2018), and *E. coli* can also survive in free-living protozoa (Barker et al., 1999; Brown et al., 2002). *E. coli* spends almost half of its life cycle in each of the primary and secondary habitats (Savageau, 1983). Transition from the primary to secondary habitat occurs when *E. coli* is

released to the external environment through faeces, which contains about 10^6 *E. coli* per gram (Hartl and Dykhuizen, 1984). The association of *E. coli* with animals, among several other criteria, has led to its widespread use as an indicator of recent faecal contamination of water (Leclerc et al., 2001; Ishii and Sadowsky, 2008). Important assumptions in using *E. coli* as an indicator organism are that it is unable to multiply outside a host, and all cells are identical in their ability to survive in the open environment (Bonde, 1966; Leclerc et al., 2001; Barnes and Gordon, 2004). However, a growing body of information suggests that *E. coli* can not only survive but also proliferate in the external environment, including water, soil, and sediment (Carrillo et al., 1985; Solo-Gabriele et al., 2000; Byappanahalli et al., 2003). This confounds the use of *E. coli* as a reliable water quality indicator. *E. coli* 'bloom' strains produce elevated counts in freshwater reservoirs and recreational lakes in Australia. These strains provide another example of where *E. coli* has gained the ability to proliferate in water (Power et al., 2005). Characterising these bloom-associated strains is important in terms of water quality monitoring and assessment.

The primary aim of the current research was to genotypically and phenotypically characterise the bloom-associated *E. coli* strains and investigate what attributes contribute to the elevated counts observed in bloom events.

1.2 Genetic diversity of *E. coli*

The species *E. coli* is characterised by a high level of genetic diversity and high genomic plasticity which are reflected by the commensal to intestinal and extra-intestinal pathogenic strains (Touchon et al., 2009). An average *E. coli* genome is 5 Mb in size and comprises

approximately 4700 genes. Out of these, about 2000 genes are shared among all strains, and are referred to as the core genome (Rasko et al., 2008; Touchon et al., 2009). Genes shared among some, but not all strains, is referred to as the variable genome, and other genes may be strain-specific. The full complement of genes in the *E. coli* species, including the core and variable genes, is referred to as the pan genome, and it comprises approximately 17,800 genes (Touchon et al., 2009). Genes involved in fundamental metabolic processes such as biosynthesis and degradation of compounds, energy metabolism, and regulation of osmolarity form the core genome (Rasko et al., 2008; Touchon et al., 2009; Vieira et al., 2011; Maddamsetti et al., 2017). Variable genes include for example those that are horizontally transferred and virulence factors, which enable strains to quickly adapt to a different environmental niche (van Elsas et al., 2011; Vieira et al., 2011). More than 90% of the pan genome is made up of variable genes, and contribute to ~80% of each genome. This means that about 80% of the genes found in a typical *E. coli* genome are not found in all strains (Lukjancenko et al., 2010). On average, the gene content of two *E. coli* genomes may differ by more than 30%, while genome size within the species varies by about 1 Mb, which equates to more than 1000 genes (Bergthorsson and Ochman, 1998; Touchon et al., 2009). Taken together, the variable 15,000 genes drive the high genomic plasticity observed in the species (Rasko et al., 2008; Touchon et al., 2009).

The overall genetic diversity of the species is driven by mutation and recombination in the core genome, including horizontal gene transfer of the variable genome (Médigue et al., 1991; LeClerc et al., 1996; Touchon et al., 2009). The gene content of *E. coli* is under constant flux, which is restricted to a few localities on the chromosome, uniform across all representatives of the species (Touchon et al., 2009). The capsule *cps* locus is a recombination hotspot in not

only *E. coli,* but also other species, including *Streptococcus pneumoniae* and *Klebsiella pneumoniae* (Milkman et al., 2003; Didelot et al., 2012; Alqasim et al., 2014; Wright et al., 2014; Mostowy et al., 2017). Genetic elements acquired via horizontal gene transfer undergo further divergence, mainly through subsequent homologous recombination events at the flanking regions (Touchon et al., 2009). Although *E. coli* undergoes recombination at a rate higher than that of mutation, their effect does not obscure the phylogenetic inference of the strains (Touchon et al., 2009).

Plasmids are a form of mobile genetic elements horizontally transferred among bacteria, and as such, they are a source for recombination and contribute to genetic variation (Sørensen et al., 2005). Plasmids play a key role in both inter-species and intra-species gene transfer and thereby the genetic diversity of *E. coli* (Hartl and Dykhuizen, 1984; Boyd and Hartl, 1997). Adaptive traits such as antibiotic resistance and virulence can be horizontally transferred among strains through plasmids (Hartl and Dykhuizen, 1984; de la Cruz and Davies, 2000; Sørensen et al., 2005).

1.3 Genetic structure and classification of *E. coli*

1.3.1 Phylogenetic groups and evolution

E. coli has a predominantly clonal genetic structure (Selander and Levin, 1980; Desjardins et al., 1995). Multi-locus enzyme electrophoresis (MLEE) (Ochman and Selander, 1984; Herzer et al., 1990) and other DNA marker techniques (Desjardins et al., 1995; Clermont et al., 2000; Escobar-Páramo et al., 2004; Gordon et al., 2008) have identified four major phylogenetic groups (phylogroups) of *E. coli*, namely A, B1, B2, and D. Four minor phylogroups designated

C, E, F (Tenaillon et al., 2010; Clermont et al., 2013) (Figure 1.1), and cryptic clade I (Luo et al., 2011) are also recognised. Phylogroup A is divided into two subcategories as A1 and A0 (Clermont et al., 2013; Clermont et al., 2015). Phylogroups A and B1 are sister clades while B2 strains form a monophyletic group which is thought to be the most basal group of *E. coli* (Lecointre et al., 1998). Phylogroup D strains form at least two distinct clades. Phylogroup C does not form a decidedly distinct group of its own, and is closely related to B1 (Gordon, 2013). Strains of phylogroup F are closely related to those of B2 and D (Gordon, 2013). The most prominent member of phylogroup E is the well-known enterohaemorrhagic strain of *E. coli* O157:H7 (Alm et al., 2011).

Members of the different phylogroups vary in their ability to cause infections (Boyd and Hartl, 1998; Johnson et al., 2001), their sugar utilisation profiles (Gordon, 2004), antibiotic resistance profiles (Gordon, 2004), genome size (Bergthorsson and Ochman, 1998), and growth rate-temperature relationship (Gordon, 2004). For example, members of phylogroups A and B1 are known to be smaller than those of phylogroups B2 and D (Bergthorsson and Ochman, 1998). As a result of these differences, strains of different phylogroups are associated with particular ecological niches and life history characteristics (Gordon and Cowling, 2003; Gordon et al., 2008; Touchon et al., 2009). Strains of phylogroups A and B1 are generalists and occupy a wide range of host and non-host habitats. These strains are over-represented in water (Power et al., 2005), while those of B2 and D are infrequent in the open environment (Power et al., 2005; Walk et al., 2007; Gordon, 2013). B2 and D strains mainly occur in mammals and birds, and B2 strains are reckoned to be the most host-adapted (Zhang et al., 2002; Nowrouzian et al., 2006; Gordon, 2013). B2 strains are far less common in water and cold-blooded vertebrates, compared to warm-blooded vertebrates (Gordon, 2013). Apart from animal hosts, evidence suggests an association of phylogroup B2 strains with free-living protozoa. For example, the

virulence factors of B2 strains are shown to have coincidentally evolved to resist grazing by free-living protozoa, rather than for virulence intrinsically (Adiba et al., 2010).

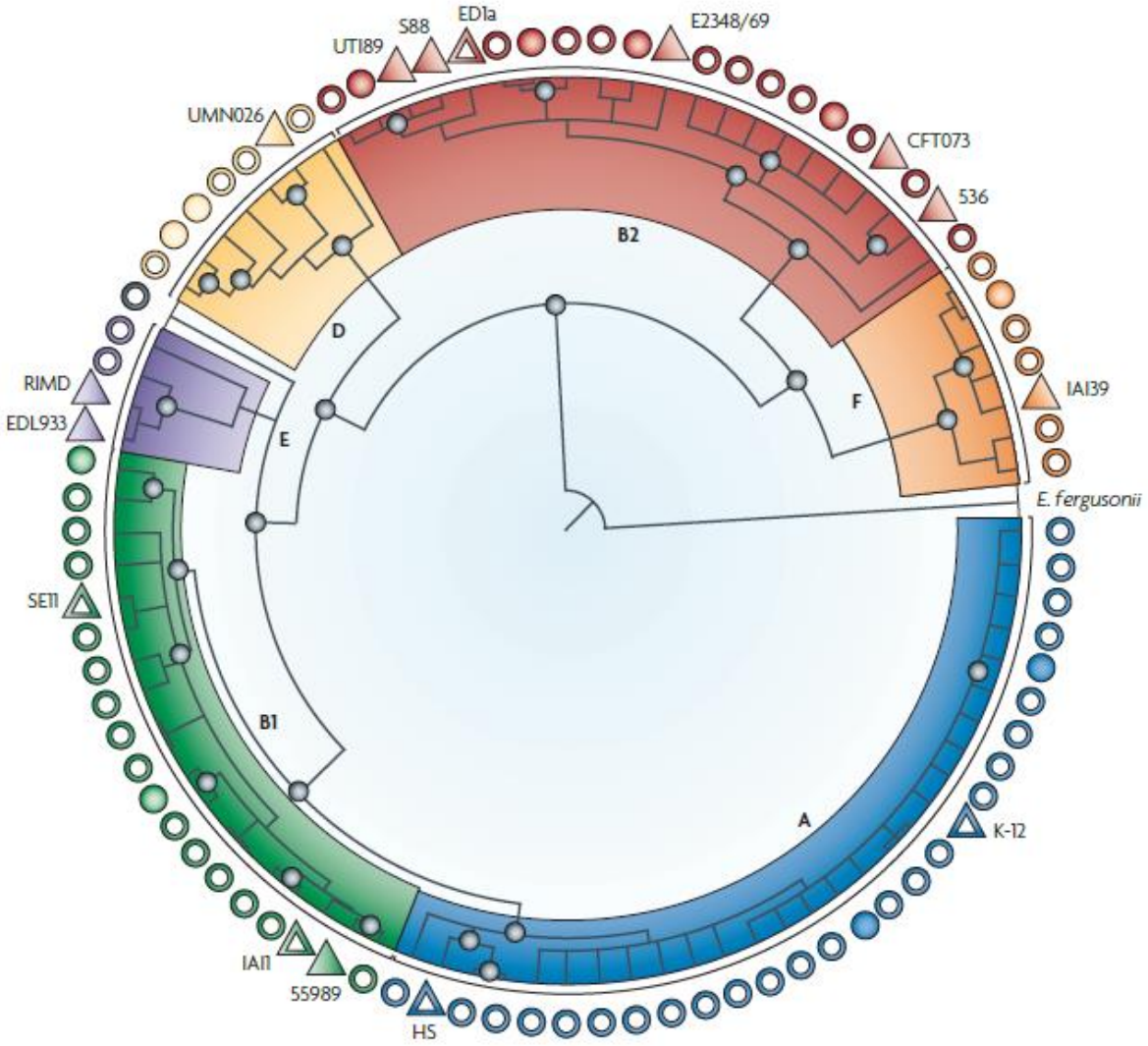


Figure 1.1. The phylogenetic tree of *E. coli* indicating the main phylogroups, constructed by Tenaillon and colleagues (2010) using the nucleotide sequences of eight housekeeping genes. The inner coloured ranges highlight the different phylogroups. The circles on the outer ring indicate strains from the ECOR (*E. coli* reference) collection, while the triangles represent genome reference strains. Symbols that are bold stand for pathogenic *E. coli*, while those that are open represent commensals. The tree has been rooted on *E. fergusonii*. Among other recognised species within the genus *Escherichia*, *E. fergusonii* is the species most closely related to *E. coli*. The use of *E. fergusonii* to root the tree has limited the long-branch attraction artefact and has enabled the construction of a robust phylogeny (Tenaillon et al., 2010).

1.3.2 Relationship between *E. coli* and *Shigella*

The genus *Shigella* constitutes four species, namely *Shigella boydii*, *S. sonnei*, *S. flexneri*, and *S. dysenteriae* (Lan and Reeves, 2002; Escobar-Paramo et al., 2003). There is considerable evidence to suggest that *Shigella* genotypes are interspersed with *E. coli* genotypes, and *Shigella* does not even represent a subspecies or a distinct group within *E. coli* (Ochman et al., 1983; Pupo et al., 1997; Pupo et al., 2000; Lan and Reeves, 2002; Lan et al., 2004). A key juncture in *Shigella* evolution is the acquisition by an *E. coli* ancestor, of the virulence plasmid which houses the virulence determinants (Lan and Reeves, 2002; Escobar-Paramo et al., 2003). *Shigella* differs from *E. coli* in having more Insertion Sequence (IS) elements, inversions, translocations, deletions, and acquisition of DNA segments, particularly pathogenicity islands transmitted by bacteriophages (Jin et al., 2002).

1.3.3 Serogroups and serotypes

The O-antigen forms the surface layer of the lipopolysaccharide outer membrane of Gram-negative bacteria (Iguchi et al., 2014; DebRoy et al., 2016). It is composed of Lipid A, an inner and outer core oligosaccharide, and the O-polysaccharide (O-antigen) (Kido et al., 1995; Willis and Whitfield, 2013; DebRoy et al., 2016). O-antigens are used for serogroup designation, and to date 184 different O-types (serogroups) have been described for *E. coli* by the World Health Organization Collaborating Centre for Reference and Research on *Escherichia* and *Klebsiella* (https://www.ssi.dk/English.aspx) (Iguchi et al., 2014), while DebRoy and colleagues (2016) report the existence of 196 O-serogroups. O-antigen biosynthesis in *E. coli* operates in two major pathways (Willis and Whitfield, 2013). One is *wzy*-dependent and involves Wzx (O-

antigen flippase) and Wzy (O-antigen polymerase). In this pathway the individual repeat units are formed in the cytoplasmic face of the cell membrane and flipped across to the periplasm by Wzx. Polymerization by Wzy takes place in the periplasm (Willis and Whitfield, 2013; DebRoy et al., 2016). The other pathway is ATP-binding cassette (ABC) transporter-dependent, which relies on Wzt (ABC transporter ATP-binding protein) and Wzm (ABC transporter permease). Here the O-antigen repeat units are extended through the action of glycosyltransferases in the cytoplasmic face of the inner membrane. The extended polymers are transported across the cell membrane by Wzt and Wzm (Greenfield and Whitfield, 2012). A third pathway exists but is not well characterised (DebRoy et al., 2016). A comprehensive study done by Iguchi and colleagues (2014) involving 184 serogroups reports that each serogroup carries either *wzx*/*wzy* or *wzt*/*wzm* genes. Out of 182 O-antigen gene clusters, the vast majority (n = 171) had the *wzx*/*wzy* genes, and are synthesized by the *wzy*-dependent pathway. The other 11 serogroups are associated with *wzt*/*wzm* genes, and are ABC transporter-dependent (Kido et al., 1995; Iguchi et al., 2014).

The serotype of a strain of *E. coli* is defined by the combination of the O-antigen type (O-type), capsular K-antigen type, and the flagellar H-antigen type; O:K:H (Kauffmann, 1947; Ørskov and Ørskov, 1991). In the current study, serogroups were assigned to whole genome sequences of strains using the web-based tool SerotypeFinder (https://cge.cbs.dtu.dk/services/SerotypeFinder/) of the Centre for Genomic Epidemiology (Joensen et al., 2015).

1.3.4 Multi-Locus Sequence Typing (MLST)

Multi-locus sequence typing is thought to be the 'gold standard' for typing *E. coli* (Larsen et al., 2012). MLST schemes use different combinations of genes to assign a sequence type (ST) to a strain (Clermont et al., 2015). Each different nucleotide sequence of an allele of a gene is assigned a unique number, and the resultant combination of the allele numbers defines a particular ST (Maiden et al., 1998). Traditional typing methods suffered limited portability between laboratories, hindering reliable comparisons, and were also expensive and time-consuming. The advent of whole genome sequencing and its continually reducing cost enable reliable multi-locus sequence typing using web servers (Larsen et al., 2012).

Currently, three MLST schemes are available to assign an ST to a strain of *E. coli* (Clermont et al., 2015); Whittam scheme, Acthman scheme, and Institut Pasteur scheme. The Whittam scheme is hosted by the Michigan State University, USA and this scheme originally focussed on enteropathogenic *E. coli* (Reid et al., 2000). The Acthman scheme is maintained by the Warwick Medical School, UK (Wirth et al., 2006), and the Institut Pasteur scheme by the Institut Pasteur, France (Jaureguy et al., 2008). The latter two schemes are general and do not focus on any group of *E. coli* in particular (Clermont et al., 2015). Each scheme is based on a different gene combination, while *icd* is common to all three schemes (Clermont et al., 2015). In the current study, STs were assigned to strains based on the *E. coli* MLST scheme #1, i.e., the Acthman scheme (Wirth et al., 2006) employing the seven genes *adk*, *fumC*, *gyrB*, *icd*, *mdh*, *purA*, and *recA* (Larsen et al., 2012), through the web-based tool MLST 2.0 (https://cge.cbs.dtu.dk/services/MLST/) (Larsen et al., 2012) available on the Centre for Genomic Epidemiology website. Compared to phylogenies inferred through the other two

MLST schemes, those inferred through the Acthman scheme are the most compatible with the whole genome phylogeny (Clermont et al., 2015).

1.4 Ecological structure of *E. coli*

1.4.1 *E. coli* in the primary habitat

As mentioned above, *E. coli* occurs in two habitats, the primary habitat (host) and secondary habitat (environment). *E. coli* is predominantly found in homeothermic vertebrates which include mammals and birds (Gordon, 2013). The prevalence of *E. coli* is the highest in mammals, followed by birds where it is less prevalent. In contrast, the likelihood of isolating *E. coli* from ectothermic vertebrates in Australia, i.e., fish, amphibians, and reptiles is low (Gordon and Cowling, 2003). Yet, various ectothermic animals are known to harbour *E. coli* (Gordon and Cowling, 2003; Hansen et al., 2008; Frick et al., 2018). The *E. coli* in fish is of faecal origin and it is argued that fish do not represent a distinct source of *E. coli*, but rather a vector of *E. coli* from other animal sources (Hansen et al., 2008). Among Australian mammals, *E. coli* is more likely to be isolated from omnivores and herbivores compared to carnivores, indicating that host diet influences the occurrence of *E. coli*. The likelihood of isolating *E. coli* is higher among mammals from temperate, semi-arid, and grassland habitats compared to those from tropical habitats, while *E. coli* is unlikely to occur in hosts from desert habitats (Gordon and Cowling, 2003). Hosts living close to human habitation have a higher likelihood to harbour *E. coli* than those that live afar (Gordon and Cowling, 2003).

The primary habitat is nutrient-rich, is more uniform, and better regulated compared to the secondary habitat, particularly in terms of temperature (Savageau, 1983). The low pH in the

stomach presents a highly stressful environment for *E. coli* but upon passage to the intestinal environment the conditions are more amenable for survival (van Elsas et al., 2011). Here, the host's diet, physiological state, immune system, and resident microbiota affect the survival and dynamics of *E. coli* (Savageau, 1974).

1.4.2 *E. coli* in the external environment

The primary and secondary habitats vary enormously in their physical properties and nutrient availability. In stark contrast are the conditions in the environment external to a host where *E. coli* faces wide fluctuations in both biotic and abiotic conditions (van Elsas et al., 2011). Several stress conditions including temperature fluctuations and extremes (Terzieva and McFeters, 1991; Solic and Krstulovic, 1992), desiccation (Desmarais et al., 2002), salinity (Solic and Krstulovic, 1992), osmotic stress (Record Jr et al., 1998), limited nutrient levels (Carrillo et al., 1985; Korhonen and Martikalnon, 1991; Desmarais et al., 2002), varying levels of oxygen (van Elsas et al., 2011), ultraviolet (UV) radiation (Pommepuy et al., 1992; Solic and Krstulovic, 1992; Alkan et al., 1995), and predation (Mezrioui et al., 1995; Desmarais et al., 2002) challenge the growth of *E. coli* in the environment (van Elsas et al., 2011). Hence, it is generally perceived that the intestinal environment is more conducible than the open environment for *E. coli* growth (Savageau, 1983).

The concentration of *E. coli* varies vastly among external environmental habitats; from 2 - 39 cfu (colony forming units)/100 ml in an irrigation reservoir (Jokinen et al., 2019), around 11 cfu/g of dry weight in surface sand of a freshwater recreational beach (Alm et al., 2003), between 5 – 1150 cfu/g in organic soil (Ishii et al., 2006), and 10^2 – 10^4 cfu/100 ml in estuarine

bed sediment (Schang et al., 2018) and coastal lagoon water (Perini et al., 2015), to over 10^6 cfu/100 ml in community wastewater (Paulshus et al., 2019). Most *E. coli* have an average half-life of around 1 day in water, between 0.5 - 2 days in sediment, and between 2 - 3 days in soil (Savageau, 1983; Vinten et al., 2002). Until fairly recently, it has been presumed that *E. coli* is transient in occurrence in the external environment and is unable to survive for long periods and proliferate (Winfield et al., 2003).

Although a 'typical' *E. coli* cell spends half of its life outside a host, multiple studies show that *E. coli* can become 'naturalised' to spend all its life in the external environment (Ishii and Sadowsky, 2008 and references therein). A growing body of data from the USA, Puerto Rico, Hawaii, and Australia, for instance, indicates that *E. coli* can not only survive for extended periods but also proliferate in water (Carrillo et al., 1985; Power et al., 2005; McCrary et al., 2013), wastewater pond sludge (Schwarz et al., 2019), soil (Solo-Gabriele et al., 2000; Byappanahalli et al., 2006; Ishii et al., 2006), algae (Whitman et al., 2003; Byappanahalli et al., 2007), sediments (Byappanahalli et al., 2003; Schang et al., 2018), and sand (Whitman and Nevers, 2003) in tropical, subtropical, and temperate environments. For example, *E. coli* have been shown to survive and regrow in tropical soils and rainforest freshwater habitats in Hawaii and Puerto Rico, respectively (Carrillo et al., 1985; Hardina and Fujioka, 1991; Byappanahalli and Fujioka, 1998; Byappanahalli and Fujioka, 2004). Other studies show that *E. coli* can regrow in tidally influenced riverbank surface soil in sub-tropical Florida (Solo-Gabriele et al., 2000; Desmarais et al., 2002). Environmental strains of *E. coli* that are independent of faecal deposits have also been reported from temperate streambank forest soils/sediment in a southern Lake Michigan watershed, in north-west Indiana, USA (Byappanahalli et al., 2003; Byappanahalli et al., 2006).

As opposed to environmental growth of *E. coli*, a sub-fraction of the population may persist for long in habitats including water, soil, and sediment (Davies et al., 1995; Korajkic et al., 2019). An over-representation of stress-response genes is observed in environmental isolates of *E. coli* and certain strains may exist for extended periods in a stress-induced viable but non-culturable (VBNC) state (Davies et al., 1995; Halliday and Gast, 2010; Zhi et al., 2019). Such a phenomenon, where the VBNC state is speculated to enable persistence of *E. coli* cells has been observed in fresh, estuarine and marine waters (Xu et al., 1982; Pommepuy et al., 1996; Na et al., 2006), and soil (Ishii et al., 2009).

Habitat factors including high nutrient/energy source availability, moisture content, warm optimum temperatures, and reduced predation are conducive to proliferation in the open environment (Carrillo et al., 1985; Solo-Gabriele et al., 2000; Desmarais et al., 2002; Byappanahalli and Fujioka, 2004; Ishii and Sadowsky, 2008). The efficacy in acquiring varied nutrients, broad temperature range (7.5 - 49 ^{o}C, with a 37 ^{o}C optimum), and being a facultative anaerobe that can survive in both aerobic and anaerobic conditions can drive growth and competitiveness of *E. coli* in diverse habitats (Ihssen and Egli, 2005; Ishii and Sadowsky, 2008). Being a chemoheterotroph, the acquisition of carbon (C) sources largely determines the fate of *E. coli* (Ihssen and Egli, 2005; Byanapahalli et al., 2006; van Elsas et al., 2011).

Overall, the vast environmental versatility of the species is a consequence of their genomic plasticity and differential regulation of gene expression (Savageau, 1983; Kudva et al., 1998; Touchon et al., 2009; van Elsas et al., 2011). Exactly how genomic diversity translates into environmental behaviour is not well understood (van Elsas et al., 2011). Properties encoded by the core genome (Touchon et al., 2009), to a great extent, determine the metabolic characters and resistance to stress, while the variable genes may complement survival, for

instance via siderophore mediated iron uptake (van Elsas et al., 2011). Further, horizontal gene transfer events contribute to genomic plasticity on a large scale (van Elsas et al., 2011).

The alternate sigma factor σ^S encoded by *rpoS* is induced for starvation and general stress resistance, including conditions of nutrient scarcity (Lange and Hengge-Aronis, 1991; Peterson et al., 2005), temperature extremes and fluctuations (Hengge-Aronis, 1996; Muffler et al., 1997), osmotic stress (Hengge-Aronis, 1996), and extreme acidity (Small et al., 1994; Hengge-Aronis, 1996). The starvation response of *E. coli* entails two strategies: the first is nutrient scavenging through the production of forage proteins such as CAP (catabolite activator protein)/Crp (Cyclic AMP receptor protein) that enable the use of alternative C sources; second, if scavenging is insufficient, entering an inactive stationary phase (Matin et al., 1989; Peterson et al., 2005). De-repression of alternative catabolic pathways, a higher catabolic flexibility, and simultaneous utilization of multiple sugars and amino acids have been observed in *E. coli* growing under low levels of C and energy sources (Ihssen and Egli, 2005), like those observed in open environments. Taken together, *E. coli* that either survive for extended periods or regrow in the external environment adapt to these conditions under the influence of selection (Alm et al., 2011).

1.5 *E. coli* as a water quality indicator

Contamination of water with human and animal faeces poses a significant health risk as faeces may harbour a range of harmful pathogens including the hepatitis A virus, *Cryptosporidium*, and *Campylobacter* (Barnes and Gordon, 2004; WHO, 2004; Ishii and Sadowsky, 2008). Waterborne diarrhoeal diseases alone account for 1.8 million deaths each year worldwide. It

is estimated that proper water treatment can bring down the incidence of diarrhoea by 35 - 39% (WHO, 2004). Therefore, water authorities perform routine testing of drinking, recreational, and ground waters to detect recent faecal contamination (Ishii and Sadowsky, 2008; U.S. EPA, 2012; WHO, 2017). The need for faecal indicator organisms arose because direct testing for pathogens is expensive, requires more highly trained staff, and is labour-intensive and time-consuming (Bitton, 2011). The presence of a faecal indicator organism in water reflects the potential presence of enteric pathogens (Leclerc et al., 2001; Bitton, 2011; U.S. EPA, 2012). Consequently, faecal associated bacteria have long been used as indicators of the level of faecal contamination and as an index of water quality (Leclerc et al., 2001).

An ideal faecal indicator organism should meet a list of criteria. Firstly, it should be restricted to the GI tract and be present in the faeces of humans and other warm-blooded animals. It should be present when pathogens are present and outnumber pathogens, and should not be pathogenic. An ideal indicator should also be more resistant to disinfection and adverse environmental conditions compared to pathogens, and should be easily detected by rapid, inexpensive laboratory methods. Further, an ideal indicator organism should not be present in the absence of faecal contamination, should have a short lifespan outside a host, and should be unable to multiply outside a host. Lastly, all cells of the species should be equal in their ability to survive and reproduce in the external environment (Leclerc et al., 2001; Barnes and Gordon, 2004; Ishii and Sadowsky, 2008; Bitton, 2011). *E. coli* is largely restricted to the GI tract of humans and mammals, accounting for 1% of the bacterial biomass (Leclerc et al., 2001; Gordon and Cowling, 2003). One gram of colon content of mammals contains on average 10^6 cells of *E. coli* (Hartl and Dykhuizen, 1984). Overall, *E. coli* meets most of these criteria to be a good indicator, and upon being first suggested as an indicator organism by Schardinger in

1892, has widely been used as an indicator in monitoring water safety (Leclerc et al., 2001; Ishii and Sadowsky, 2008).

The World Health Organization (WHO, 1997) and the Australian drinking water guidelines (Australian Government NHMRC, 2011) dictate that *E. coli* should not be detected in a drinking water sample of a minimum volume of 100 ml. In recreational freshwater, both *E. coli* and intestinal enterococci are employed as indicators (Australian Government NHMRC, 2008; U.S. EPA, 2012). According to the U.S. EPA (2012), recreational water bodies should be closed if a single water sample contains more than 235 cfu of *E. coli*/100 ml or if the geometric mean (GM) of *E. coli* in at least 5 samples uniformly distributed in a 30-day interval exceeds 126 cfu/100 ml. Within Australia, recreational water bodies are regulated primarily on a state level and the guidelines vary among states/territories. Depending on local conditions such as environmental factors and influence of human and animal populations, different states utilise different primary indicators; intestinal enterococci, *E. coli* or other, and different permissible levels of indicators (Australian Government NHMRC, 2008).

How *E. coli* copes with the transition from primary to secondary habitat carries significant implications for water quality assessment (Gordon, 2001). Regardless of the widespread use of *E. coli* as an indicator, a growing body of evidence suggests that *E. coli* can not only survive but can also multiply in environments external to a host, including water, soil, sediment, manure, algae, and plants (Carrillo et al., 1985; Solo-Gabriele et al., 2000; Desmarais et al., 2002; Ishii and Sadowsky, 2008). Therefore, the presence of *E. coli* in water may not necessarily be due to faecal contamination, but may represent 'naturalised' strains in water or strains that have re-suspended from sediment (Solo-Gabriele et al., 2000; Desmarais et al., 2002; Pachepsky and Shelton, 2011). This violates key assumptions for an ideal indicator and

confounds the use of *E. coli* in monitoring water quality (Carrillo et al., 1985; Solo-Gabriele et al., 2000; Desmarais et al., 2002; Power et al., 2005; Ishii and Sadowsky, 2008). Consequently, several alternative indicators including intestinal enterococci, *Clostridium perfringens* (Australian Government NHMRC, 2008), Bacteroidales (Savichtcheva et al., 2007), human enteric viruses, enteric bacteriophages, and coliphages are proposed for water quality monitoring (Leclerc et al., 2001; U.S. EPA, 2012; McMinn et al., 2017). In terms of enterococci, recreational water bodies should be closed if the enterococcal count exceeds 70 cfu/100 ml on a single-sample basis or if the GM for enterococci exceeds 35 cfu/100 ml of recreational water (U.S. EPA, 2012). However, regrowth of enterococci in the environment (Anderson et al., 1997; Desmarais et al., 2002) and the rarity of *C. perfringens* in faeces of ruminant herbivores (Vierheilig et al., 2013) for example, make designating an 'ideal' indicator an ongoing challenge (Desmarais et al., 2002).

1.6 Methods for testing water for faecal contamination

Recent times have seen a gradual transition in methods used for testing water for faecal contamination (Rompré et al., 2002). These methods target the detection of total coliforms and faecal coliforms/*E. coli* in water samples (Edberg and Edberg, 1988; Rompré et al., 2002). Total coliforms include members of the genera *Escherichia, Klebsiella, Enterobacter,* and *Citrobacter* (Edberg and Edberg, 1988). The multiple tube fermentation (MTF) method and the membrane filtration (MF) technique were traditionally at the forefront of water testing to detect total coliforms (Rompré et al., 2002). In the MF method, the water sample is filtered through a sterile filter membrane with a pore diameter of 0.45 µm ± 0.02 µm, which retains the bacteria (Goetz et al., 1951; Rompré et al., 2002). Enumeration of colonies is done upon

incubation of the filter membrane on selective media. On m-Endo-type media containing lactose, coliforms produce red colonies with a metallic sheen (Rompré et al., 2002). Faecal coliforms including *E. coli* can be enumerated on m-FC medium upon incubation at an elevated temperature of 44.5 °C (Geldreich et al., 1965; Ciebin et al., 1995; Rompré et al., 2002).

The current definition of *E. coli* used by the water industry is that *E. coli* expresses two enzymes, namely, β-D-galactosidase and β-D-glucuronidase (APHA, AWWA, WEF - Standard methods for the examination of water and wastewater, 2018). The use of the enzymes β-D-galactosidase and β-D-glucuronidase for the detection and enumeration of total coliforms and *E. coli*, respectively, has resulted in novel chromogenic and fluorogenic defined substrates (Rompré et al., 2002). The Colilert® system, patented by the IDEXX Laboratories Inc., USA (www.idexx.com/water), is one such defined substrate that detects both total coliforms and *E. coli* in water simultaneously (Edberg and Edberg, 1988). The detection relies on two nutrient indicator compounds present in the medium. The first is Ortho-Nitrophenyl-β-D-galactopyranoside (ONPG), which is used to detect coliforms, while the second, 4-Methyl-umbelliferyl-β-D-glucuronide (MUG) detects *E. coli*. The Colilert® medium is mixed with a water sample and the enzymes of any coliforms/*E. coli* in water react with the indicators. Coliforms have the enzyme β-D-galactosidase which metabolises ONPG, turning the medium from colourless to yellow. *E. coli* carry the enzyme β-D-glucuronidase which metabolises MUG to produce blue fluorescence under UV light of 365 nm (Rice et al., 1990; Manafi and Rotter, 1991; Kinzelman et al., 2005; IDEXX, 2019). The Colilert® system can detect these bacteria at concentrations as low as 1 cfu/100 ml of water (Edberg and Edberg, 1988). This system has high sensitivity and specificity, allows growth of injured coliforms, provides results within a maximum of 24 hours without the need for confirmatory tests, and does not feed non-target

organisms (Edberg and Edberg, 1988). The Colilert-18® medium which is used in the current study is an optimised Colilert® formulation (Budnick et al., 2001).

1.7 *E. coli* 'bloom' events in Australian lakes

E. coli strains that produce significantly elevated counts of 10,000 - 100,000 cells/100 ml of water, well above the 'safe' cut-off level of 235 cells/100 ml (Ishii and Sadowsky, 2008; U.S. EPA, 2012) have been reported from freshwater reservoirs and recreational lakes in Australia (Power et al., 2005). These elevated counts resemble bloom events, and strains responsible for bloom events are termed bloom strains. In recent times *E. coli* bloom events have been a regular occurrence in freshwater reservoirs and lakes across Australia, particularly in the Australian Capital Territory (ACT), Queensland (Qld), New South Wales (NSW), Western Australia (WA), and Victoria (VIC). Lake Burley Griffin and Lake Ginninderra, two recreational lakes in the ACT, Hinze Dam in South East Qld, Tallowa dam and dams in the Hunter Valley in NSW, and Lake Burragorang, the major drinking water supply reservoir for greater metropolitan Sydney, NSW, experienced recurring bloom events. A single bloom event occurred in a freshwater reservoir in WA in 2015.

A limited number of strains (n = 8) are associated with bloom events and all belong to the *E. coli* phylogenetic groups A and B1 (Power et al., 2005). Three strains have been isolated from the bloom events in freshwater bodies on the east coast (ACT, NSW, and Qld); two of these strains belong to phylogroup A, out of which one belongs to subgroup A1, while the other belongs to subgroup A0. The third bloom strain from the east coast belongs to phylogroup B1 (termed B1-001 strain) (Power et al., 2005). Five different strains genotypically distinct from

each other, have been isolated from the WA bloom event, and four of these strains belong to subgroup A1 of phylogroup A, while the other strain belongs to phylogroup B1.

The strains responsible for the elevated counts are present at low concentrations during non-bloom periods. Bloom strains have never been isolated from a vertebrate host and they are genotypically distinct from faecal *E. coli* isolates. Considering the population of the regions in question and the volume of the lakes, it is highly unlikely that the elevated counts are a result of recent faecal contamination (Power et al., 2005). Sanitary surveys also have revealed that these strains are not a result of faecal input. These observations suggest that the bloom stains may in fact represent environmentally free-living *E. coli* that undergo proliferation in water (Power et al., 2005). Elevated counts produced by free-living strains in water can lead to unnecessary panic and closure of water bodies including recreational lakes.

Something in common to all of the bloom strains is that they have all acquired a group 1 capsule originating from *Klebsiella* (Power et al., 2005; Nanayakkara et al., 2019). The capsule confers a highly mucoid phenotype when the strains are grown on MacConkey agar (Power et al., 2005; Nanayakkara et al., 2019).

1.8 The capsule

When present, the polysaccharide capsule is the outermost layer of the bacterial cell surface (Whitfield and Roberts, 1999). Therefore, the capsule masks the O-antigen used for serotyping (Kauffmann, 1947). *E. coli* capsules are classified into four major groups numbered 1 through 4 (Whitfield and Roberts, 1999). The group 1 capsule is highly similar structurally, genetically, and in terms of its expression characteristics, to that of *Klebsiella* (Amor and Whitfield, 1997;

Rahn et al., 1999; Whitfield, 2006). The biosynthesis of *Klebsiella* capsules and *E. coli* group 1 capsules occurs via the *wzy*-dependent polymerisation pathway (Whitfield, 2006). The *E. coli* group 1 capsule is known to co-express with a limited number of O-serogroups including O8, O9, O20, and O101 (Whitfield and Roberts, 1999). O8, O9, and O101 are ABC transporter-dependent serogroups (DebRoy et al., 2016). Strains with group 1 capsules do not co-express colanic acid, which is an extracellular polysaccharide (Jayaratne et al., 1993; Rahn et al., 1999; Whitfield and Paiment, 2003; Whitfield, 2006).

Klebsiella capsule gene cluster spans a 10-30 kbp region of the genome (Wyres et al., 2016). It is thought that the capsule is one of the most crucial virulence determinants of certain strains of *Klebsiella* (Simoons-Smit et al., 1986; Wyres et al., 2016). Inside a host, it protects the cell from phagocytosis, complement-mediated killing, and lethal effects of serum (Williams et al., 1983; Álvarez et al., 2000). Different capsule types are implicated in virulence to different extents; for instance, K1 and K2 have been associated with hyper-virulent *K. pneumoniae* strains (Shon et al., 2013), while certain capsule types have a minimal effect on virulence (Kabha et al., 1995). The capsule also plays a critical role in adapting to novel environments (Rendueles et al., 2017). In the external environment, the capsule protects the bacterial cell from adverse conditions encountered including desiccation, UV radiation, predation, and bacteriophage infection (Weiner et al., 1995; Rendueles et al., 2017). In a recent study carried out by Wyres and colleagues (2016), 134 *Klebsiella* capsule synthesis loci were identified, including 31 novel loci. These different loci encode different *Klebsiella* capsule variants/types.

1.9 Bacterial growth curve

Bacterial growth progresses through the succession of different phases of changing growth rate (Monod, 1949). The phases are, in sequential order, lag phase, acceleration phase, exponential growth phase, retardation phase, stationary phase, and phase of decline (Figure 1.2) (Monod, 1949). During lag phase, no discernible growth occurs as the cell synthesises enzymes required for the utilisation of compounds present in its new substrate (Monod, 1949). Starting from zero after the lag phase, the specific growth rate increases and reaches a maximum growth rate (Zwietering et al., 1990). Gradual cessation of growth occurs after the exponential phase, halting the growth rate. This leads to the stationary phase by which time the cells have exhausted most of the nutrients available in the medium. The onset of the stationary phase is highly regulated and is governed by RpoS (Navarro Llorens et al., 2010). Growth cessation results from nutrient exhaustion including iron limitation, accumulation of toxic metabolites during growth, and change in pH (Monod, 1949). Temperature, nutrient availability, competition, and predation all affect growth rate (White et al., 1991; Thingstad and Lignell, 1997).

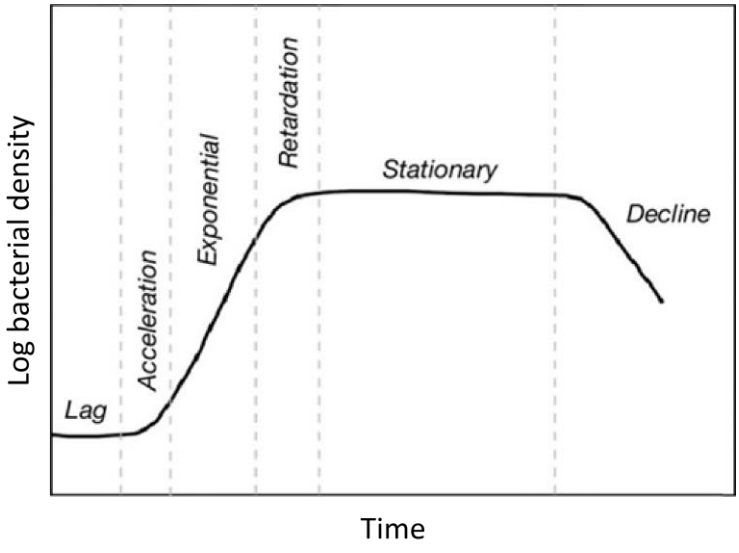


Figure 1.2. The growth curve of bacteria (Nordstrom and Campbell, 2014), as described by Monod (1949).

1.10 Research aims

E. coli bloom events are a regular occurrence in freshwater reservoirs and recreational lakes in Australia. The occurrence of these elevated counts confounds routine water quality monitoring and assessment carried out by water authorities, as bloom events can lead to unnecessary closures of water bodies. It is important that bloom-associated strains are characterised in detail.

The aim of the current study was to genotypically and phenotypically characterise bloom-associated *E. coli* strains. As the *Klebsiella* capsule is a key attribute of the bloom strains (Power et al., 2005), the diversity and distribution of *Klebsiella* capsules in a large collection of non-bloom and bloom-associated *E. coli* was investigated. A PCR (polymerase chain reaction) protocol was developed to detect *Klebsiella* capsule-positive *E. coli* and discriminate *E. coli* strains that harbour bloom strain-associated *Klebsiella* capsule types/variants.

While bloom strains belong to *E. coli* phylogroups A and B1, the phylogroup A strains represent the majority of bloom strains isolated to date. A collection of phylogroup A strains was investigated using comparative genomics and laboratory experiments to determine factors that could contribute to the elevated counts observed in bloom events.

As the initial investigations have provided a *Shigella sonnei* identification for the phylogroup B1 bloom strain isolated from the east coast (also termed B1-001 strain) (Power et al., 2005), its evolution and genome content were studied using a comparative genomics approach. Further, previous investigations of east coast bloom events have shown that the B1-001 bloom strain was numerically dominant while one or both of the phylogroup A bloom strains were

also present. However, the B1-001 bloom strain has failed to appear in recent bloom events assessed using Colilert-18®, and this phenomenon was investigated.

1.11 References

Adiba, S., Nizak, C., Van Baalen, M., Denamur, E., and Depaulis, F. (2010) From grazing resistance to pathogenesis: the coincidental evolution of virulence factors. *PloS One* 5(8): e11882.

Alkan, U., Elliott, D.J., and Evison, L.M. (1995) Survival of enteric bacteria in relation to simulated solar radiation and other environmental factors in marine waters. *Water Res* 29(9): 2071-2080.

Alm, E.W., Burke, J., and Spain, A. (2003) Fecal indicator bacteria are abundant in wet sand at freshwater beaches. *Water Res* 37(16): 3978-3982.

Alm, E.W., Walk, S.T., and Gordon, D.M. (2011) The Niche of *Escherichia coli*. In *Population Genetics of Bacteria: A Tribute to Thomas S. Whittam*. Walk, S.T., and Feng, P.C.H. (eds). Washington, DC: ASM Press, pp. 69–89.

Alqasim, A., Scheutz, F., Zong, Z., and McNally, A. (2014) Comparative genome analysis identifies few traits unique to the *Escherichia coli* ST131 H30Rx clade and extensive mosaicism at the capsule locus. *BMC Genomics*. 15: 830.

Álvarez, D., Merino, S., Tomás, J.M., Benedí, V.J., and Albertí, S. (2000) Capsular polysaccharide is a major complement resistance factor in lipopolysaccharide O side chain-deficient *Klebsiella pneumoniae* clinical isolates. *Infect Immun* 68(2): 953-955.

Amor, P.A., and Whitfield, C. (1997) Molecular and functional analysis of genes required for expression of group IB K antigens in *Escherichia coli*: characterization of the *his*-region containing gene clusters for multiple cell-surface polysaccharides. *Mol Microbiol* 26(1): 145-161.

Anderson, S.A., Turner, S.J., and Lewis, G.D. (1997) Enterococci in the New Zealand environment: implications for water quality monitoring. *Water Sci Technol* 35(11-12): 325-331.

APHA, AWWA, WEF – American Public Health Association, American Water Works Association, Water Environment Federation. (2018) 9223 Enzyme substrate coliform test

2017. In *Standard methods for the examination of water and wastewater,* 23rd Ed. https://www.standardmethods.org/doi/abs/10.2105/SMWW.2882.194

Australian Government NHMRC - National Health and Medical Research Council. (2008) Guidelines for managing risks in recreational water. https://www.nhmrc.gov.au/sites/default/files/images/guidelines-for-managing-risks-in-recreational-water.pdf

Australian Government NHMRC – National Health and Medical Research Council. (2011) Australian drinking water guidelines 6. Version 3.4. Updated October 2017. https://www.nhmrc.gov.au/sites/default/files/documents/reports/aust-drinking-water-guidelines.pdf

Barker, J., Humphrey, T.J., and Brown, M.W. (1999) Survival of *Escherichia coli* 0157 in a soil protozoan: implications for disease. *FEMS Microbiol Lett* 173(2): 291-295.

Barnes, B., and Gordon, D.M. (2004) Coliform dynamics and the implications for source tracking**.** *Environ Microbiol* 6(5): 501-509.

Bergthorsson, U., and Ochman, H. (1998) Distribution of chromosome length variation in natural isolates of *Escherichia coli*. *Mol Biol Evol* 15(1): 6-16.

Bitton, G. (2011) *Wastewater Microbiology*, 4th ed. Hoboken, NJ: Wiley-Blackwell.

Bonde, G.J. (1966) Bacteriological methods for estimation of water pollution. *Health laboratory science* 3(2): 124-128.

Boyd, E.F., and Hartl, D.L. (1997) Recent horizontal transmission of plasmids between natural populations of *Escherichia coli* and *Salmonella enterica*. *J Bacteriol* 179(5): 1622-1627.

Boyd, E.F., and Hartl, D.L. (1998) Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *J Bacteriol* 180(5): 1159-1165.

Brown, M.R., Smith, A.W., Barker, J., Humphrey, T.J., and Dixon, B. (2002) *E. coli* O157 persistence in the environment. *Microbiology* 148(1): 1-2.

Budnick, G.E., Howard, R.T., and Mayo, D. (2001) Comparison of Colilert-18 to the mTEC agar method for the enumeration of *Escherichia coli* in recreational waters. In *American Society of Microbiologists Annual Convention*, *May*.

Byappanahalli, M.N., and Fujioka, R.S. (1998) Evidence that tropical soil environment can support the growth of *Escherichia coli*. *Water Sci Technol* 38(12): 171.

Byappanahalli, M., and Fujioka, R. (2004) Indigenous soil bacteria and low moisture may limit but allow faecal bacteria to multiply and become a minor population in tropical soils. *Water Sci Technol* 50(1): 27-32.

Byappanahalli, M., Fowler, M., Shively, D., and Whitman, R. (2003) Ubiquity and persistence of *Escherichia coli* in a Midwestern coastal stream. *Appl Environ Microbiol* 69(8): 4549-4555.

Byappanahalli, M.N., Whitman, R.L., Shively, D.A., Sadowsky, M.J., and Ishii, S. (2006) Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. *Environ Microbiol* 8(3): 504-513.

Byappanahalli, M.N., Whitman, R.L., Shively, D.A., Ferguson, J., Ishii, S., and Sadowsky, M.J. (2007) Population structure of *Cladophora*-borne *Escherichia coli* in nearshore water of Lake Michigan. *Water Res* 41(16): 3649-3654.

Carrillo, M., Estrada, E., and Hazen, T.C. (1985) Survival and enumeration of the fecal indicators *Bifidobacterium adolescentis* and *Escherichia coli* in a tropical rain forest watershed. *Appl Environ Microbiol* 50(2): 468-476.

Ciebin, B.W., Brodsky, M.H., Eddington, R., Horsnell, G., Choney, A., Palmateer, G., and Shears, G. (1995) Comparative evaluation of modified m-FC and m-TEC media for membrane filter enumeration of *Escherichia coli* in water. *Appl Environ Microbiol* 61(11): 3940-3942.

Clermont, O., Bonacorsi, S., and Bingen, E. (2000) Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* 66(10): 4555-4558.

Clermont, O., Christenson, J.K., Denamur, E., and Gordon, D.M. (2013) The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 5(1): 58-65.

Clermont, O., Gordon, D., and Denamur, E. (2015) Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. *Microbiology* 161(5): 980-988.

Davies, C.M., Long, J.A., Donald, M., and Ashbolt, N.J. (1995) Survival of fecal microorganisms in marine and freshwater sediments. *Appl Environ Microbiol* 61(5): 1888-1896.

DebRoy, C., Fratamico, P.M., Yan, X., Baranzoni, G., Liu, Y., Needleman, D.S., and Mwangi, M. (2016) Comparison of O-antigen gene clusters of all O-serogroups of *Escherichia coli* and proposal for adopting a new nomenclature for O-typing. *PLoS One* 11(1): e0147434.

de la Cruz, F., and Davies, J. (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends Microbiol* 8(3): 128-133.

Desjardins, P., Picard, B., Kaltenböck, B., Elion, J., and Denamur, E. (1995) Sex in *Escherichia coli* does not disrupt the clonal structure of the population: evidence from random amplified polymorphic DNA and restriction-fragment-length polymorphism. *J Mol Evol* 41(4): 440-448.

Desmarais, T.R., Solo-Gabriele, H.M., and Palmer, C.J. (2002) Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. *Appl Environ Microbiol* 68(3): 1165-1172.

Didelot, X., Méric, G., Falush, D., and Darling, A.E. (2012) Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics* 13: 256.

Edberg, S.C., and Edberg, M.M. (1988) A defined substrate technology for the enumeration of microbial indicators of environmental pollution. *Yale J Biol Med* 61(5): 389-399.

Escobar-Paramo, P., Giudicelli, C., Parsot, C., and Denamur, E. (2003) The evolutionary history of *Shigella* and enteroinvasive *Escherichia coli* revised. *J Mol Evol* 57(2): 140-148.

Escobar-Páramo, P., Sabbagh, A., Darlu, P., Pradillon, O., Vaury, C., Denamur, E., and Lecointre, G. (2004) Decreasing the effects of horizontal gene transfer on bacterial phylogeny: the *Escherichia coli* case study. *Mol Phylogenet Evol* 30(1): 243-250.

Frick, C., Vierheilig, J., Linke, R., Savio, D., Zornig, H., Antensteiner, R., et al. (2018) Poikilothermic animals as a previously unrecognized source of fecal indicator bacteria in a backwater ecosystem of a large river. *Appl Environ Microbiol* 84(16): e00715-18.

Geldreich, E.E., Clark, H.F., Huh, C.B., and Best, L.C. (1965) Fecal-coliform-organism medium for the membrane filter technique. *J Am Water Works Assoc* 57(2): 208-214.

Goetz, A., Tsuneishi, N., Kabler, P.W., Streicher, L., and Neumann, H.G. (1951) Application of molecular filter membranes to the bacteriological analysis of water [with discussion]. J *Am* Water Works Assoc 43(12), 943-984.

Gordon, D.M. (2001) Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. *Microbiology* 147(5): 1079-1085.

Gordon, D.M. (2004) The influence of ecological factors on the distribution and the genetic structure of *Escherichia coli*. *EcoSal Plus* 1(1).

Gordon, D.M. (2013) The Ecology of *Escherichia coli*. In *Escherichia coli: Pathotypes and Principles of Pathogenesis.* Donnenberg, M.S. (ed). San Diego, CA: Academic Press, pp. 3-20.

Gordon, D.M., and Cowling, A. (2003) The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology* 149(12): 3575-3586.

Gordon, D.M., Clermont, O., Tolley, H., and Denamur, E. (2008) Assigning *Escherichia coli* strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method. *Environ Microbiol* 10(10): 2484-2496.

Greenfield, L.K., and Whitfield, C. (2012) Synthesis of lipopolysaccharide O-antigens by ABC transporter-dependent pathways. *Carbohydr Res* 356: 12-24.

Halliday, E., and Gast, R.J. (2010) Bacteria in beach sands: an emerging challenge in protecting coastal water quality and bather health. *Environ Sci Technol* 45(2): 370-379.

Hansen, D.L., Clark, J.J., Ishii, S., Sadowsky, M.J., and Hicks, R.E. (2008) Sources and sinks of *Escherichia coli* in benthic and pelagic fish. *J Great Lakes Res*, 34(2): 228-234.

Hardina, C.M., and Fujioka, R.S. (1991) Soil: the environmental source of *Escherichia coli* and enterococci in Hawaii's streams. *Environ Toxicol Water Qual* 6(2): 185-195.

Hartl, D.L., and Dykhuizen, D.E. (1984) The population genetics of *Escherichia coli*. *Annu Rev Genet* 18(1): 31-68.

Hengge-Aronis, R. (1996) Back to log phase: σS as a global regulator in the osmotic control of gene expression in *Escherichia coli*. *Mol Microbiol* 21(5): 887-893.

Herzer, P.J., Inouye, S, Inouye, M, and Whittam, T.S. (1990) Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli. J Bacteriol* 172(11): 6175-6181.

IDEXX. (2019) Colilert-18. https://www.idexx.com/en/water/water-products-services/colilert-18/

Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M. et al. (2014) A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res* 22(1): 101 – 107.

Ihssen, J., and Egli, T. (2005) Global physiological analysis of carbon-and energy-limited growing *Escherichia coli* confirms a high degree of catabolic flexibility and preparedness for mixed substrate utilization. *Environ Microbiol* 7(10): 1568-1581.

Ishii, S., and Sadowsky, M.J. (2008) *Escherichia coli* in the environment: Implications for water quality and human health. *Microbes Environ* 23(2): 101-108.

Ishii, S., Ksoll, W.B., Hicks, R.E., and Sadowsky, M.J. (2006) Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. *Appl Environ Microbiol* 72(1): 612-621.

Ishii, S., Yan, T., Vu, H., Hansen, D. L., Hicks, R.E., and Sadowsky, M.J. (2009) Factors controlling long-term survival and growth of naturalized *Escherichia coli* populations in temperate field soils. *Microbes Environ* doi:10.1264/jsme2.ME09172

Jaureguy, F., Landraud, L., Passet, V., Diancourt, L., Frapy, E., Guigon, G. et al. (2008) Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC genomics* 9(1): 560.

Jayaratne, P., Keenleyside, W.J., MacLachlan, P.R., Dodgson, C., and Whitfield, C. (1993) Characterization of *rcsB* and *rcsC* from *Escherichia coli* O9: K30: H12 and examination of the role of the *rcs* regulatory system in expression of group I capsular polysaccharides. *J Bacteriol* 175(17): 5384-5394.

Jin, Q., Yuan, Z., Xu, J., Wang, Y., Shen, Y., Lu, W., et al. (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.

Joensen, K.G., Tetzschner, A.M., Iguchi, A., Aarestrup, F.M., and Scheutz, F. (2015) Rapid and easy in silico serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. *J Clin Microbiol* 53(8): 2410-2426.

Johnson, J.R., Delavari, P., Kuskowski, M., and Stell, A.L. (2001) Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. *J Infect Dis* 183(1): 78-88.

Jokinen, C.C., Hillman, E., and Tymensen, L. (2019) Sources of generic *Escherichia coli* and factors impacting guideline exceedances for food safety in an irrigation reservoir outlet and two canals. *Water Res* 156: 148-158.

Kabha, K., Nissimov, L., Athamna, A., Keisari, Y., Parolis, H., Parolis, L.A. et al. (1995) Relationships among capsular structure, phagocytosis, and mouse virulence in *Klebsiella pneumoniae*. *Infect Immun* 63(3): 847-852.

Kauffmann, F. (1947) The serology of the coli group. *J Immunol* 57(1): 71-100.

Kido, N., Torgov, V.I., Sugiyama, T., Uchiya, K., Sugihara, H., Komatsu, T. et al. (1995) Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E. coli* O9 *rfb*

gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. *J Bacteriol* 177(8): 2178-2187.

Kinzelman, J.L., Singh, A., Ng, C., Pond, K.R., Bagley, R.C., and Gradus, S. (2005) Use of IDEXX Colilert-18® and Quanti-Tray/2000 as a rapid and simple enumeration method for the implementation of recreational water monitoring and notification programs. *Lake Reserv Manage* 21(1): 73-77.

Korajkic, A., McMinn, B.R., Ashbolt, N.J., Sivaganesan, M., Harwood, V.J., and Shanks, O.C. (2019) Extended persistence of general and cattle-associated fecal indicators in marine and freshwater environment. *Sci Total Environ* 650: 1292-1302.

Korhonen, L.K., and Martikalnon, P.J. (1991) Survival of *Escherichia coli* and *Campylobacter jejuni* in untreated and filtered lake water. *J Appl Bacteriol* 71(4): 379-382.

Kudva, I.T., Blanch, K., and Hovde, C.J. (1998) Analysis of *Escherichia coli* O157: H7 survival in ovine or bovine manure and manure slurry. *Appl Environ Microbiol* 64(9): 3166-3174.

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

Lan, R., Alles, M.C., Donohoe, K., Martinez, M.B., and Reeves, P.R. (2004) Molecular evolutionary relationships of enteroinvasive *Escherichia coli* and *Shigella* spp. *Infect Immun* 72(9): 5080-5088.

Lange, R., and Hengge-Aronis, R. (1991) Identification of a central regulator of stationary-phase gene expression in *Escherichia coli. Mol Microbiol* 5(1): 49-59.

Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., and Lund, O. (2012) Multilocus sequence typing of total genome sequenced bacteria. *J Clin Microbiol* 50(4): 1355-1361.

Leclerc, H.D.A.A., Mossel, D.A.A., Edberg, S.C., and Struijk, C.B. (2001) Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. *Annu Rev Microbiol* 55(1): 201-234.

LeClerc, J.E., Li, B., Payne, W.L., and Cebula, T.A. (1996) High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science*, 274(5290): 1208-1211.

Lecointre, G., Rachdi, L., Darlu, P., and Denamur, E. (1998) *Escherichia coli* molecular phylogeny using the incongruence length difference test. *Mol Biol Evol* 15(12): 1685-1695.

Lukjancenko, O., Wassenaar, T.M., and Ussery, D.W. (2010) Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol* 60(4): 708-720.

Luo, C., Walk, S.T., Gordon, D.M., Feldgarden, M., Tiedje, J.M., and Konstantinidis, K.T. (2011) Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proc Natl Acad Sci USA* 108(17): 7200-7205.

Maddamsetti, R., Hatcher, P.J., Green, A.G., Williams, B.L., Marks, D.S., and Lenski, R.E. (2017) Core genes evolve rapidly in the long-term evolution experiment with *Escherichia coli*. *Genome Biol Evol* 9(4): 1072-1083.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., and Feavers, I. M. (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 95(6): 3140-3145.

Manafi, M., and Rotter, M.L. (1991) A new plate medium for rapid presumptive identification and differentiation of *Enterobacteriaceae*. *Int J Food Microbiol* 14(2): 127-134.

Matin, A., Auger, E.A., Blum, P.H., and Schultz, J.E. (1989) Genetic basis of starvation survival in non-differentiating bacteria. *Annu Rev Microbiol* 43(1): 293-314.

McCrary, K.J., Case, C.L.H., Gentry, T.J., and Aitkenhead-Peterson, J.A. (2013) *Escherichia coli* regrowth in disinfected sewage effluent: Effect of DOC and nutrients on regrowth in laboratory incubations and urban streams. *Water Air Soil Poll* 224(2): 1412.

McMinn, B.R., Ashbolt, N.J., and Korajkic, A. (2017) Bacteriophages as indicators of faecal pollution and enteric virus removal. *Lett Appl Microbiol* 65(1): 11-26.

Médigue, C., Rouxel, T., Vigier, P., Hénaut, A., and Danchin, A. (1991) Evidence for horizontal gene transfer in *Escherichia coli* speciation. *J Mol Biol* 222(4): 851-856.

Mezrioui, N., Baleux, B., and Troussellier, M. (1995) A microcosm study of the survival of *Escherichia coli* and *Salmonella typhimurium* in brackish water. *Water Res* 29(2): 459-465.

Milkman, R., Jaeger, E., and McBride, R.D. (2003) Molecular evolution of the *Escherichia coli* chromosome. VI. Two regions of high effective recombination. *Genetics* 163(2): 475-483.

Monod, J. (1949) The growth of bacterial cultures. *Annu Rev Microbiol* 3(1): 371-394.

Mostowy, R.J., Croucher, N.J., De Maio, N., Chewapreecha, C., Salter, S.J., Turner, P. et al. (2017) Pneumococcal capsule synthesis locus *cps* as evolutionary hotspot with potential to generate novel serotypes by recombination. *Mol Biol Evol* 34(10): 2537–2554.

Muffler, A., Barth, M., Marschall, C., and Hengge-Aronis, R. (1997) Heat shock regulation of sigmaS turnover: a role for DnaK and relationship between stress responses mediated by sigmaS and sigma32 in *Escherichia coli*. *J Bacteriol* 179(2): 445-452.

Na, S.H., Miyanaga, K., Unno, H., and Tanji, Y. (2006) The survival response of *Escherichia coli* K12 in a natural environment. *Appl Microbiol Biotechnol* 72(2): 386-392.

Nanayakkara, B.S., O'Brien, C.L., and Gordon, D.M. (2019) Diversity and distribution of *Klebsiella* capsules in *Escherichia coli*. *Environ Microbiol Rep* 11(2): 107-117.

Navarro Llorens, J.M., Tormo, A., and Martínez-García, E. (2010) Stationary phase in Gram-negative bacteria. *FEMS Microbiol Rev* 34(4): 476-495.

Nordstrom D.K., and Campbell K.M. (2014) Modeling low-temperature geochemical processes. In *Treatise on Geochemistry*. Drever, J.I., Holland H.D., and Turekian K.K. (eds). Oxford: Elsevier pp. 27-68.

Nowrouzian, F.L., Adlerberth, I., and Wold, A.E. (2006) Enhanced persistence in the colonic microbiota of *Escherichia coli* strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. *Microbes Infect* 8(3): 834-840.

Ochman, H, and Selander, R.K. (1984) Standard reference strains of *Escherichia coli* from natural populations. *J Bacteriol* 157(2): 690-693.

Ochman, H., Whittam, T.S., Caugant, D.A., and Selander, R.K. (1983) Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. *Microbiology* 129(9): 2715-2726.

Ørskov, F., and Ørskov, I. (1992) *Escherichia coli* serotyping and disease in man and animals. *Can J Microbiol* 38(7): 699-704.

Pachepsky, Y.A., and Shelton, D.R. (2011) *Escherichia coli* and fecal coliforms in freshwater and estuarine sediments. *Crit Rev Env Sci Tec* 41(12): 1067-1110.

Paulshus, E., Kühn, I., Möllby, R., Colque, P., O'Sullivan, K., Midtvedt, T., et al. (2019) Diversity and antibiotic resistance among *Escherichia coli* populations in hospital and community wastewater compared to wastewater at the receiving urban treatment plant. *Water Res* 161: 232-241.

Perini, L., Quero, G.M., García, E.S., and Luna, G.M. (2015) Distribution of *Escherichia coli* in a coastal lagoon (Venice, Italy): temporal patterns, genetic diversity and the role of tidal forcing. *Water Res* 87: 155-165.

Peterson, C.N., Mandel, M.J., and Silhavy, T.J. (2005) *Escherichia coli* starvation diets: essential nutrients weigh in distinctly. *J Bacteriol* 187(22): 7549-7553.

Pommepuy, M., Guillaud, J.F., Dupray, E., Derrien, A., Le Guyader, F., and Cormier, M. (1992) Enteric bacteria survival factors. *Wat Sci Tech* 25(12): 93-103.

Pommepuy, M., Butin, M., Derrien, A., Gourmelon, M., Colwell, R.R., and Cormier, M. (1996) Retention of enteropathogenicity by viable but nonculturable *Escherichia coli* exposed to seawater and sunlight. *Appl Environ Microbiol* 62(12): 4621-4626.

Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., and Slade, M.B. (2005) Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. *Environ Microbiol* 7(5): 631-640.

Pupo, G.M., Karaolis, D.K., Lan, R., and Reeves, P.R. (1997) Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. *Infect Immun* 65(7): 2685-2692.

Pupo, G.M., Lan, R., and Reeves, P.R. (2000) Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci USA* 97(19): 10567-10572.

Rahn, A., Drummelsmith, J., and Whitfield, C. (1999) Conserved organization in the *cps* gene clusters for expression of *Escherichia coli* group 1 K antigens: relationship to the colanic acid biosynthesis locus and the *cps* genes from *Klebsiella pneumoniae*. *J Bacteriol* 181(7): 2307-2313.

Rasko, D.A., Rosovitz, M.J., Myers, G.S., Mongodin, E.F., Fricke, W.F., Gajer, P., and Henderson, I.R. (2008) The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J Bacteriol* 190(20): 6881-6893.

Record Jr, M.T., Courtenay, E.S., Cayley, D.S., and Guttman, H.J. (1998) Responses of *E. coli* to osmotic stress: large changes in amounts of cytoplasmic solutes and water. *Trends Biochem Sci* 23(4): 143-148.

Reid, S.D., Herbelin, C.J., Bumbaugh, A.C., Selander, R.K., and Whittam, T.S. (2000) Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* 406(6791): 64-67.

Rendueles, O., Garcia-Garcerà, M., Néron, B., Touchon, M., and Rocha, E.P.C. (2017) Abundance and co-occurrence of extracellular capsules increase environmental breadth: Implications for the emergence of pathogens. *PLOS Pathog* 13(7): e1006525.

Rice, E.W., Allen, M.J., and Edberg, S.C. (1990) Efficacy of beta-glucuronidase assay for identification of *Escherichia coli* by the defined-substrate technology. *Appl Environ Microbiol* 56(5): 1203-1205.

Rompre, A., Servais, P., Baudart, J., De-Roubin, M.R., and Laurent, P. (2002) Detection and enumeration of coliforms in drinking water: current methods and emerging approaches. *J Microbiol Methods* 49(1): 31-54.

Savageau, M.A. (1974) Genetic regulatory mechanisms and the ecological niche of *Escherichia coli*. *Proc Natl Acad Sci USA* 71(6): 2453-2455.

Savageau, M.A. (1983) *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. *Amer Nat* 122(6): 732-744.

Savichtcheva, O., Okayama, N., and Okabe, S. (2007) Relationships between *Bacteroides* 16S rRNA genetic markers and presence of bacterial enteric pathogens and conventional fecal indicators. *Water Res* 41(16): 3615-3628.

Selander, R.K., and Levin, B.R. (1980) Genetic diversity and structure in *Escherichia coli* populations. *Science* 210(4469): 545-547.

Schang, C., Lintern, A., Cook, P.L.M., Rooney, G., Coleman, R., Murphy, H.M., et al. (2018) *Escherichia coli* survival and transfer in estuarine bed sediments. *River Res Appl* 34(6): 606-614.

Schwarz, K.R., Sidhu, J.P.S., Toze, S., Li, Y., Lee, E., Gruchlik, Y., and Pritchard, D.L. (2019) Decay rates of *Escherichia coli*, *Enterococcus* spp., F-specific bacteriophage MS2, somatic coliphage and human adenovirus in facultative pond sludge. *Water Res* 154: 62-71.

Shon, A.S., Bajwa, R.P., and Russo, T.A. (2013) Hypervirulent (hypermucoviscous) *Klebsiella pneumoniae*: a new and dangerous breed. *Virulence* 4(2): 107-118.

Simoons-Smit, A.M., Verweij-van Vught, A.M.J.J., and MacLaren, D.M. (1986) The role of K antigens as virulence factors in *Klebsiella*. *J Med Microbiol* 21(2): 133-137.

Small, P., Blankenhorn, D., Welty, D., Zinser, E., and Slonczewski, J.L. (1994) Acid and base resistance in *Escherichia coli* and *Shigella flexneri*: role of *rpoS* and growth pH. *J Bacteriol* 176(6): 1729-1737.

Solic, M., and Krstulovic, N. (1992) Separate and combined effects of solar radiation, temperature, salinity, and pH on the survival of faecal coliforms in seawater. *Marine Poll Bull* 24(8): 411-416.

Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R., and Palmer, C.J. (2000) Sources of *Escherichia coli* in a coastal subtropical environment. *Appl Environ Microbiol* 66(1): 230-237.

Sørensen, S.J., Bailey, M., Hansen, L.H., Kroer, N., and Wuertz, S. (2005) Studying plasmid horizontal transfer in situ: a critical review. *Nat Rev Microbiol* 3(9): 700-710.

Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010) The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol* 8(3): 207-217.

Terzieva, S.I., and McFeters, G.A. (1991) Survival and injury of *Escherichia coli*, *Campylobacter jejuni*, and *Yersinia enterocolitica* in stream water. *Can J Microbiol* 37(10): 785-790.

Thingstad, T.F., and Lignell, R. (1997) Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat Microb Ecol* 13: 19-27.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P. et al. (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5(1): e1000344.

U.S. EPA - United States Environmental Protection Agency. (2012) Recreational water quality criteria. Office of Water. 820-F-12-058. https://www.epa.gov/sites/production/files/2015-10/documents/rwqc2012.pdf.

van Elsas, J.D., Semenov, A.V., Costa, R., and Trevors, J.T. (2011) Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J* 5(2): 173-183.

Vieira, G., Sabarly, V., Bourguignon, P.Y., Durot, M., Le Fèvre, F., Mornico, D. et al. (2011) Core and panmetabolism in *Escherichia coli*. *J Bacteriol* 193(6): 1461-1472.

Vierheilig, J., Frick, C., Mayer, R.E., Kirschner, A.K.T., Reischer, G.H., Derx, J. et al. (2013) *Clostridium perfringens* is not suitable for the indication of fecal pollution from ruminant wildlife but is associated with excreta from nonherbivorous animals and human sewage. *Appl Environ Microbiol* 79(16): 5089-5092.

Vinten, A.J.A., Lewis, D.R., Fenlon, D.R., Leach, K.A., Howard, R., Svoboda, I., and Ogden, I. (2002) Fate of *Escherichia coli* and *Escherichia coli* O157 in soils and drainage water following cattle slurry application at 3 sites in southern Scotland. *Soil Use Manage* 18(3): 223-231.

Walk, S.T., Alm, E.W., Calhoun, L.M., Mladonicky, J.M., and Whittam, T.S. (2007) Genetic diversity and population structure of *Escherichia coli* isolated from freshwater beaches. *Environ Microbiol* 9(9): 2274-2288.

Weiner, R., Langille, S., and Quintero, E. (1995) Structure, function and immunochemistry of bacterial exopolysaccharides. *J Ind Microbiol* 15(4): 339-346.

White, P.A., Kalff, J., Rasmussen, J.B., and Gasol, J.M. (1991) The effect of temperature and algal biomass on bacterial production and specific growth rate in freshwater and marine habitats. *Microb Ecol* 21(1): 99-118.

Whitfield, C. (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu Rev Biochem* 75: 39-68.

Whitfield, C., and Paiment, A. (2003) Biosynthesis and assembly of Group 1 capsular polysaccharides in *Escherichia coli* and related extracellular polysaccharides in other bacteria. *Carbohydr Res* 338(23): 2491-2502.

Whitfield, C., and Roberts, I.S. (1999) Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol Microbiol* 31(5): 1307-1319.

Whitman, R.L., and Nevers, M.B. (2003) Foreshore sand as a source of *Escherichia coli* in nearshore water of a Lake Michigan beach. *Appl Environ Microbiol* 69(9): 5555-5562.

Whitman, R.L., Shively, D.A., Pawlik, H., Nevers, M.B., and Byappanahalli, M.N. (2003) Occurrence of *Escherichia coli* and enterococci in *Cladophora* (Chlorophyta) in nearshore water and beach sand of Lake Michigan. *Appl Environ Microbiol* 69(8): 4714-4719.

WHO-World Health Organization. (1997) Guidelines for drinking-water quality, Volume 3 Surveillance and control of community supplies, 2nd edn. World Health Organization, Geneva.

WHO-World Health Organization. (2004) The global burden of disease - 2004 update. World Health Organization, Geneva.

WHO-World Health Organization. (2017) Guidelines for drinking-water quality, Microbial aspects, 4th edn. incorporating the 1st addendum. World Health Organization, Geneva.

Williams, P., Lambert, P.A., Brown, M.R., and Jones, R.J. (1983) The role of the O and K antigens in determining the resistance of *Klebsiella aerogenes* to serum killing and phagocytosis. *Microbiology* 129(7): 2181-2191.

Willis, L.M., and Whitfield, C. (2013) Capsule and lipopolysaccharide. In *Escherichia coli: pathotypes and principles of pathogenesis*. Donnenberg, M.S. (ed). San Diego, CA: Academic Press, pp. 533-556.

Winfield, M.D., and Groisman, E.A. (2003) Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. *Appl Environ Microbiol 69*(7): 3687-3694.

Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L.H., and Achtman, M. (2006) Sex and virulence in *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* 60(5): 1136-1151.

Wright, M.S., Perez, F., Brinkac, L., Jacobs, M.R., Kaye, K., Cober, E. et al. (2014) Population structure of KPC-producing *Klebsiella pneumoniae* isolates from midwestern US hospitals. *Antimicrob Agents Chemother* 58(8): 4961-4965.

Wyres, K.L., Wick, R.R., Gorrie, C., Jenney, A., Follador, R., Thomson, N.R., and Holt, K.E. (2016) Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom* 2(12): e000102.

Xu, H.S., Roberts, N., Singleton, F L., Attwell, R.W., Grimes, D.J., and Colwell, R. R. (1982) Survival and viability of nonculturable *Escherichia coli* and *Vibrio cholerae* in the estuarine and marine environment. *Microb Ecol* 8(4): 313-323.

Zhang, L., Foxman, B., and Marrs, C. (2002) Both urinary and rectal *Escherichia coli* isolates are dominated by strains of phylogenetic group B2. *J Clin Microbiol* 40(11): 3951-3955.

Zhi, S., Banting, G., Stothard, P., Ashbolt, N.J., Checkley, S., Meyer, K. et al. (2019) Evidence for the evolution, clonal expansion and global dissemination of water treatment-resistant naturalized strains of *Escherichia coli* in wastewater. *Water Res* 156: 208-222.

Zwietering, M.H., Jongenburger, I., Rombouts, F.M., and Van't Riet, K. (1990) Modeling of the bacterial growth curve. *Appl Environ Microbiol* 56(6): 1875-1881.

Chapter 2. Diversity and distribution of *Klebsiella* capsules in *Escherichia coli* [1]

[1] Nanayakkara, O'Brien, and Gordon. (2019) *Environ Microbiol Rep* 11(2): 107-117.

2.1 Abstract

Escherichia coli strains responsible for elevated counts (blooms) in freshwater reservoirs and lakes in Australia carry a capsule originating from *Klebsiella*. All bloom strains belong to *E. coli* phylogroups A and B1, while phylogroup A represents the majority. The diversity and distribution of *Klebsiella* capsules in known *E. coli* bloom strains, and in a large collection of *E. coli* isolated from a variety of sources were investigated.

The occurrence of *Klebsiella* capsules in *E. coli* was about 7% overall and 23 different capsule types were detected. All bloom strains were encapsulated, and seven different capsule types were detected among the eight bloom strains isolated to date. Capsules were observed in strains from phylogroups A, B1, and C, but were absent from phylogroups B2, D, E, and F. In general, few A, B1, or C lineages were capsule-positive, but when a lineage was encapsulated multiple different capsule types were present. All capsule-positive strains were of serogroups O8, O9, and O89. Regardless of the phylogroup, O9 strains were more likely to be capsule-positive than O8 strains. It appears that both the capsule region and the O-antigen gene region co-transfer to *E. coli* from *Klebsiella* via horizontal gene transfer events.

Pan genome analysis indicated that there were only modest differences between encapsulated and non-encapsulated strains of phylogroup A. The *fecIRABCDE* operon was over-represented among the phylogroup A bloom strains compared to non-bloom *E. coli*. Overall, the possession of a *Klebsiella* capsule, but not the type of capsule, is likely a key determinant of bloom ability.

2.2 Introduction

Contamination of drinking and recreational waters by animal faeces poses a significant human health risk, as faeces may serve as a source of harmful pathogens, such as the hepatitis A virus (Szewzyk et al., 2000; Gordon, 2001; Alm et al., 2011; U.S. EPA, 2012). Faecal-associated bacteria have long been used as faecal indicator species. An ideal faecal indicator, among other attributes, should not be present in a water body in the absence of faecal contamination and be unable to multiply outside a host (Bonde, 1966; Power et al., 2005; Bitton, 2011). *Escherichia coli* has been assumed to exhibit many of the attributes of an ideal indicator, and for many years the species has been widely used as an indicator of recent faecal contamination of drinking and recreational waters (Edberg et al., 2000; U.S. EPA, 2012; Gordon, 2013). However, there is a growing body of evidence suggesting that *E. coli* can not only survive for extended periods, but also proliferate in environments such as water (Carrillo et al., 1985), soil (Solo-Gabriele et al., 2000; Byappanahalli et al., 2006; Ishii et al., 2006), algae (Whitman et al., 2003; Byappanahalli et al., 2007), and sediments (Byappanahalli et al., 2003), thus confounding its use as a water quality indicator (Bonde, 1966; Alm et al., 2011; van Elsas et al., 2011).

In Australia, significantly elevated *E. coli* counts have been reported from freshwater reservoirs and recreational lakes (Power et al., 2005). These elevated count events have been termed *E. coli* 'bloom' events, as counts from 10,000 – 100,000 cells/100 ml of water have been reported; counts well above the maximum allowed level of 235 cfu (colony forming units)/100 ml in a single sample (Ishii and Sadowsky, 2008; U.S. EPA, 2012). Sanitary surveys indicate that these elevated count events cannot be attributed to faecal contamination, and indeed, achieving cell counts this high would require an unachievable level of faecal

contamination (Power et al., 2005). The strains responsible for the bloom events can be isolated from the water bodies at any time, and have not been detected in the faeces of humans or other animals. These outcomes suggest that the presence of these strains in a water body is largely independent of faecal inputs and indeed, the strains responsible may represent, free-living *E. coli* (Power et al., 2005; Alm et al., 2011).

Relatively few strains have been found to be responsible for *E. coli* bloom events, and all belong to *E. coli* phylogroups A and B1. All bloom strains have a mucoid phenotype and encode a group 1 capsule originating from *Klebsiella* (Power et al., 2005; Nanayakkara et al., 2019). Restriction fragment length polymorphism (RFLP) analysis of the capsule region has revealed that the group 1 capsules possessed by the strains are not identical (Power et al., 2005).

The polysaccharide capsule, when present, is the outermost layer that envelopes the bacterial cell surface (Whitfield and Roberts, 1999). In *E. coli*, capsules are classified into four major groups (Whitfield and Roberts, 1999). The group 1 capsule exhibits high similarity to those of *Klebsiella* in structural, genetic, and expression terms (Amor and Whitfield, 1997; Rahn et al., 1999; Whitfield, 2006). The biosynthesis of *Klebsiella* capsules and *E. coli* group 1 capsules occurs via a *wzy*-dependent polymerization pathway (Whitfield, 2006). The *Klebsiella* capsule spans a 10-30 kbp region in the genome and has been implicated in virulence for certain strains of *Klebsiella* (Simoons-Smit et al., 1986; Struve and Krogfelt, 2003; Lawlor et al., 2005; Wyres et al., 2016). It protects the cell from phagocytosis, complement-mediated killing, and lethal effects of serum (Williams et al., 1983; Álvarez et al., 2000; Lin et al., 2004). Similarly, in the external environment the capsule protects the bacterial cell from adverse conditions encountered including desiccation, UV radiation, predation, and bacteriophage infection (Weiner et al., 1995; Rendueles et al., 2017).

Given the strong association of the Australian 'bloom' status with the possession of a *Klebsiella* capsule, the objective of the present study was to examine the diversity and distribution of *Klebsiella* capsules in known *E. coli* bloom strains, and in a large collection of *E. coli* isolated from a variety of sources.

2.3 Materials and Methods

2.3.1 Diversity and distribution of capsule types

2.3.1.1 *Strain selection*

Whole genome sequence (WGS) data for 1194 Australian *E. coli* strains isolated from a wide range of sources were used to determine the frequency and diversity of *Klebsiella* capsules. The strains comprised 332 phylogroup A strains, 300 phylogroup B1 strains, 261 phylogroup B2 strains, 142 phylogroup D strains, 72 phylogroup E strains, 68 phylogroup F strains, and 19 phylogroup C strains. These strains were isolated from a variety of sources across Australia (Gordon and FitzGibbon, 1999; Gordon and Cowling, 2003; Power et al., 2005; Blyton et al., 2014; Blyton et al., 2015; Vangchhia et al., 2016). To investigate where other encapsulated strains occur in the *E. coli* phylogeny, the study was extended to include encapsulated strains from two publicly available databases. First, the *Klebsiella* capsule core gene *galF* was used to query the NCBI (National Centre for Biotechnology Information) *E. coli* database using BLAST and 135 strains with >90% sequence similarity were selected at random and the assemblies were downloaded. Second, group 1 capsules are known to be associated with the ATP-binding cassette (ABC) transporter-dependent serogroups O8, O9, and O89 in *E. coli* (Kido et al., 1995; Amor and Whitfield, 1997; Drummelsmith et al., 1997). Using the *E. coli*/*Shigella* database in EnteroBase (https://enterobase.warwick.ac.uk/species/index/ecoli) a subset of 99 O8, O9,

and O89 strains were selected representing one example of each O:H type combination present in the database. The phylogroups of the strains have been determined experimentally or *in silico* (Clermont et al., 2000; Clermont et al., 2013; Beghain et al., 2018).

2.3.1.2 *Determination of capsule status and type*

The WGS data of all 1194 Australian strains were screened for the *Klebsiella* capsule using Kaptive, a software tool used for the rapid identification of *Klebsiella* capsule loci in whole genome sequences by comparison to a reference database of known *Klebsiella* capsule types (Wyres et al., 2016). Strains predicted by Kaptive to possess a capsule were examined morphologically on MacConkey and Congo red plates. The strains' morphology was found to be consistent with the Kaptive predictions and showed that strains having ≥90% coverage and identity with the best match capsule locus could be considered capsule-positive. These were the criteria used to score the WGS data downloaded from NCBI and Enterobase.

The Harvest suite of tools (Treangen et al., 2014) was used to infer phylogeny of 869 strains belonging to phylogroups A, B1, and C. The tree was rooted on the phylogroup B2 strain ED1a.

2.3.2 Within-capsule screening

Kaptive extracts the nucleotide sequence of the capsule region from whole genome sequences. The capsule regions of the 237 capsule-positive strains were annotated using Prokka (Seemann, 2014) and a pan genome analysis was conducted using Roary (Page et al., 2015) and Scoary (Brynildsrud et al., 2016).

2.3.3 Capsule flanking region

Capsule flanking regions of strains representative of each capsule type were investigated by aligning the capsule-positive genomes in progressiveMauve (Darling et al., 2004) and determining the sequence similarity using the NCBI BLASTn. The draft genomes were reordered against the *E. coli* K-12 reference genome prior to alignment. *E. coli* K-12 has the colanic acid gene cluster and is negative for *Klebsiella* capsule. The arrangement of the *his* operon, O-antigen (*rfb*) gene region, and capsule region of *Klebsiella* strains identified using BLAST was compared against capsule-positive/negative *E. coli* using progressiveMauve genome alignments (Darling et al., 2004).

2.3.4 Association of the capsule with the O-antigen and other capsule types

Encapsulated strains are known to be associated with particular ABC transporter-dependent O-antigens (Kido et al., 1995; Amor and Whitfield, 1997; Drummelsmith et al., 1997). The serotype of all strains was determined *in silico* using SerotypeFinder 1.1 (Joensen et al., 2015) within the Centre for Genomic Epidemiology website (http://www.genomicepidemiology.org). The O-type data were incorporated in the phylogroups A, B1, and C *E. coli* phylogeny.

To infer the origin of the O-antigen genes, the *wzt*/*wzm* nucleotide sequences of encapsulated and capsule-negative *E. coli* strains representing O8, O9, and O89 were extracted from SerotypeFinder 1.1 (Joensen et al., 2015). These were queried against the Microbial Nucleotide BLAST database in NCBI.

Rendueles and colleagues (2017) reported the presence of multiple other capsule types in Bacteria and Archaea. Their programme CapsuleFinder (https://research.pasteur.fr/en/tool/capsulefinder/) is based on the use of HMM profiles to detect essential proteins involved in capsule biogenesis and uses computational models to describe the capsule components and their organisation. CapsuleFinder can detect eight capsule types: Wzx/Wzy-dependent, Group II and III or ABC-dependent, PGA capsule (Poly-γ-d-glutamate proteic capsule), synthase-dependent-HAS (hyaluronic acid capsules), synthase-dependent CPS3 (capsules of type cps3), Group IV_f (based on *Francisella tularensis* GroupIV capsule), Group IV_e (based on *Escherichia coli* GroupIV capsule), and Group IV_s (based on *Salmonella sp.* GroupIV capsule). The WGS data for all strains found to possess a *Klebsiella* capsule were screened using CapsuleFinder.

2.3.5 Pan genome comparison

Phylogroup A strains represent the majority of bloom strains isolated to date. Hence, the pan genome of a collection of phylogroup A strains was compared, based on whether they were bloom-associated or not (52 strains), or harboured a *Klebsiella* capsule or not (341 strains). Genomes were annotated using Prokka (Seemann, 2014) and pan genome analyses were conducted using Roary (Page et al., 2015). Scoary (Brynildsrud et al., 2016) was used to assess the association between the capsule and other genomic components. For a gene to be considered over-represented in one group (e.g. encapsulated strains) compared to a second group (e.g. non-encapsulated strains), the gene had to be present in the first group at a frequency >75% and present in the second group at a frequency ≤60%; and the difference in the frequencies of the gene in the two groups had to be >30%. For a gene to be considered

under-represented, then its frequency needed to be <50% in the first group and >75% in the second group; and the difference in the frequencies of the gene in the two groups had to be >30%. These criteria were determined based on prior background research, so as to avoid the exclusion of attributes that were already known to be specific for a group, to limit the inclusion of attributes that were marginal, and also to maintain consistency between comparisons.

2.4 Results

2.4.1 Diversity and distribution of capsule types

Of the 1194 Australian *E. coli* strains screened, 82 strains (6.9%) were *Klebsiella* capsule-positive. The encapsulated strains comprised 53 out of 332 phylogroup A (16%), 21 out of 300 phylogroup B1 (7%), and eight out of 19 phylogroup C (42%) strains. Capsules were not observed among the 261 phylogroup B2, 142 D, 72 E, and 68 F strains. Encapsulated strains detected in this study represented 23 of the 134 distinct capsule synthesis loci reported by Wyres and colleagues (2016). Forty different capsule types were found among the capsule-positive strains from the NCBI database. Among the 99 EnteroBase strains with a serogroup of O8, O9, O89, 27 strains (27%) were capsule-positive, representing 17 different capsule types. The distribution of capsules among the phylogroup A, B1, and C strains is depicted in Figure 2.1.

Among the Australian phylogroup A and B1 strains, the frequency of encapsulated strains varied with respect to the phylogroup membership of the strain and the source of the isolate (nominal logistic regression: phylogroup, $p = 0.727$; source, $p = 0.028$; phylogroup*source, $p < 0.0001$). Phylogroup C strains were excluded as the number of isolates was too small for

meaningful analysis (Table 2.1). Phylogroup A strains isolated from water (excluding bloom strains) were more likely to be encapsulated while strains from native birds and poultry meat were unlikely to be encapsulated (Table 2.1). By contrast, B1 strains from water were rarely encapsulated, while encapsulated strains were most likely to be detected among the poultry meat B1 isolates (Table 2.1). In general, encapsulated strains can be isolated from a variety of vertebrate hosts or from water samples. However, the bloom strains have never been isolated from a vertebrate host (Power et al., 2005). This suggests that the bloom strains represent a distinct subset of encapsulated and presumably free-living *E. coli*. Yet, as the capsule appears to be a key feature of the bloom strains, the possibility remains that any encapsulated strain, regardless of its isolation source, might be able to produce 'blooms'. There is no evidence to indicate if encapsulated strains from a particular source are more likely or less likely to become 'bloom' strains.

A number of phylogroup B2, D, and A strains were identified by Kaptive as encoding the *Klebsiella* capsule deletion variant KL156-D1. These strains were not considered to be true capsule-positive strains due to several reasons. Firstly, KL156-D1 co-occurred with the colanic acid gene cluster, which encapsulated strains do not harbour (Jayaratne et al., 1993; Whitfield and Paiment, 2003; Whitfield, 2006). In strains encoding KL156-D1, we also detected a flanking region having *wzzB*, which determines the length of O-antigen polysaccharide in *wzy*-dependent O-serogroups, and is associated with non-encapsulated strains (Whitfield and Roberts, 1999; Iguchi et al., 2014). Another reason was that the serogroups of these strains were typically O-types other than the capsule-associated O8, O9, and O89. Lastly, these strains lacked the characteristic colony morphology of encapsulated strains.

A total of 35 bloom isolates representing eight bloom strains obtained at different time periods from multiple localities in Australia were used in the study (Supplemental Table 2.1). Each of the strains known to have been responsible for 'bloom' events exhibited a different capsule type. The eastern Australian bloom strains belonged to phylogroups A and B1. The phylogroup A sequence type 10 (ST10) strain had either capsule type KL16 or KL113, which are almost identical, while the ST609 strain had capsule type KL49. The phylogroup B1 bloom strain (ST1494) had a KL53 capsule type. Five strains: four phylogroup A strains and one B1 strain, were isolated from a single bloom event in Western Australia and each strain had a different capsule type, i.e., KL31 (ST58), KL53, KL60 (ST10), KL63 and KL101 (ST227).

The encapsulated strains were present in a limited subset of all lineages, regardless of their phylogroup membership (Figure 2.1). When a clade was capsule-positive, it contained multiple capsule types. In other words, closely related lineages carried diverse capsule types.

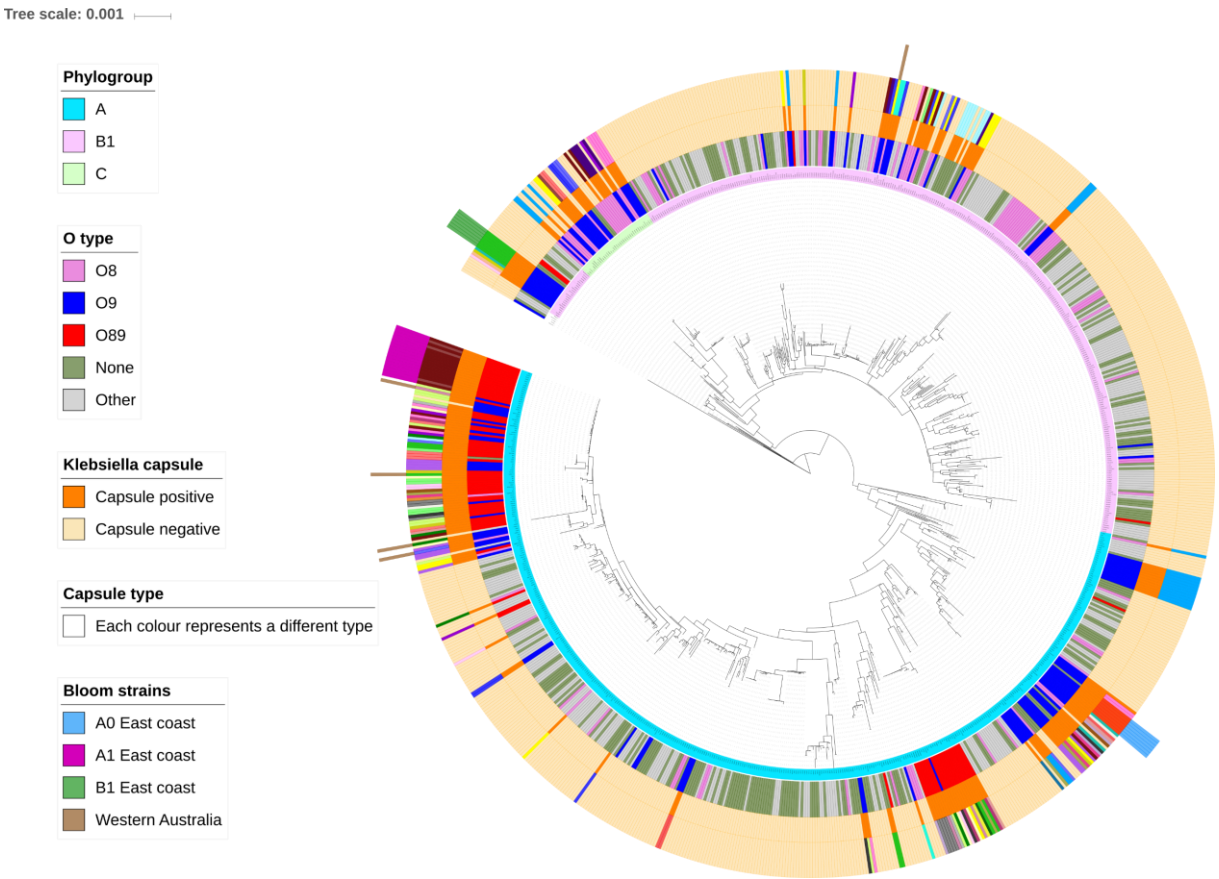


Figure 2.1. Distribution of *Klebsiella* capsules and O-types in *E. coli* phylogroups A, B1, and C. The phylogenetic tree was constructed using the whole genome sequences of 869 *E. coli* strains. These included environmental and host-associated strains from an Australian collection, and those downloaded from NCBI and EnteroBase. The innermost ring depicts the phylogroup membership of the strains; A, B1, or C. The second innermost ring denotes the O-type of the strains, while the third ring denotes the presence/absence of a *Klebsiella* capsule. The fourth ring indicates the capsule type, with each colour corresponding to a different capsule type. The outermost ring depicts the bloom strains. The tree was rooted on the phylogroup B2 strain *E. coli* ED1a and annotated using the web interface Interactive tree of life (iTOL) (Letunic and Bork, 2016).

Table 2.1. The occurrence of *Klebsiella* capsules in terms of the phylogroup and source of isolation of the Australian *E. coli* strains

Phylogroup	Source*	Number capsule-positive	Number capsule-negative	% capsule-positive
A	Bird	0	56	0
	Human	12	77	13.5
	Non-human mammal	3	27	10.0
	Poultry meat	2	89	2.2
	Water	9	29	23.7
B1	Bird	1	48	2.0
	Human	3	38	7.3
	Non-human mammal	2	32	5.9
	Poultry meat	7	41	14.6
	Water	1	118	0.8
C	Bird	0	0	0
	Human	6	5	54.5
	Non-human mammal	0	1	0
	Poultry meat	1	4	20.0
	Water	1	0	100

* The bloom strains and 4 strains isolated from fish or reptiles were excluded from the analysis.

2.4.2 Capsule Region

The capsule region was flanked by *galF* (UDP-glucose pyrophosphorylase) and *ugd* (UDP-glucose 6-dehydrogenase). The region between *galF* and *ugd* among encapsulated strains varied from 13.4 kbp – 30.3 kbp and encoded an average of 15 genes.

Previous studies have shown that the group 1 capsule has been inserted between the *his* (histidine biosynthesis) operon and *galF* in *E. coli* (Whitfield, 2006) (Figure 2.2). In *E. coli* K-12 the O-antigen gene cluster is flanked by the *his* operon and *galF*, while the colanic acid gene cluster is located upstream of *galF*. Encapsulated strains notably lacked genes of the *wzy*-dependent O-antigen clusters, including *wzzB* (Batchelor et al., 1991; Iguchi et al., 2014). The *wzx/wzy* genes that typically determine the O-antigen in *E. coli* have, in encapsulated strains,

been replaced by *wzt* and *wzm*. The best BLAST query hits (>95%identity) in NCBI for *wzt* from encapsulated *E. coli* strains were for *Klebsiella* and encapsulated *E. coli*, indicating a likely *Klebsiella* origin for *wzt*. Transposases/insertion sequence (IS) elements were present on either side of the capsule cluster in most, but not all encapsulated strains. The position of the capsule region insertion was not identical in all encapsulated strains, but occurred within 1-9 genes upstream of the *his* operon. The arrangement of *his*, O-antigen region (*rfb*) and the capsule region (*cps*) of *Klebsiella,* was comparable to that of encapsulated *E. coli* strains (Figure 2.2). The gene region for colanic acid biosynthesis did not co-occur with the *Klebsiella* capsule.

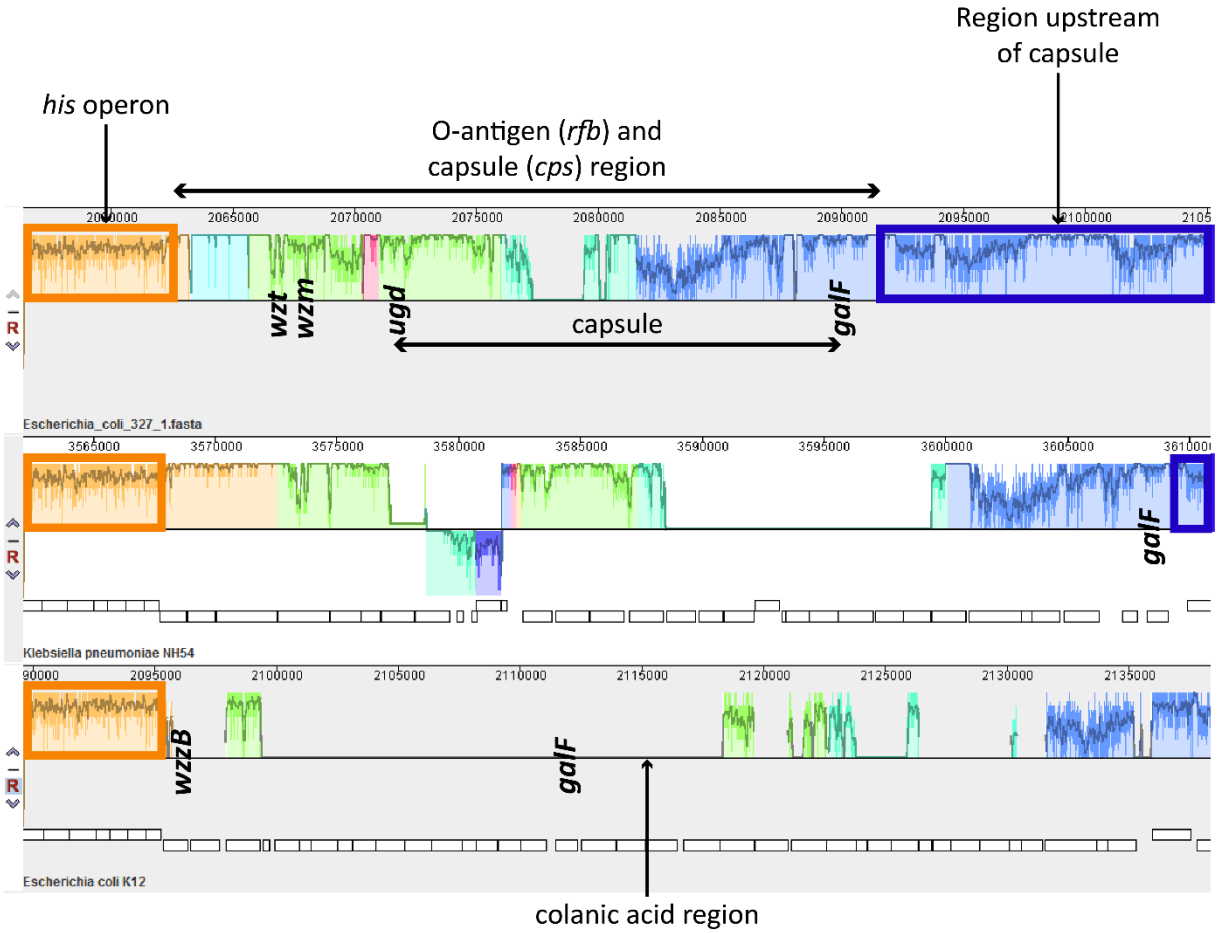


Figure 2.2. progressiveMauve alignment of the *his* operon (outlined in orange), O-antigen region, and the capsule locus of *E. coli* bloom strain 327_1, *Klebsiella pneumoniae* NH54, and capsule-negative strain *E. coli* K-12, in order from top to bottom. Coloured regions represent locally collinear blocks (LCBs). The upstream region of the capsule is outlined in blue and contains *yegH* and *asmA*. In *E. coli* K-12, the region with no homology to the other two strains is the colanic acid gene cluster.

2.4.3 Association of the capsule with the O-antigen and other capsule types

All capsule-positive strains were O8, O9, or O89 (220 out of 240 strains; 91.7%) or the serogroup could not be determined. However, not all of the O8, O9, and O89 strains were encapsulated (Table 2.2). Capsule-negative strains were of O-serogroups which were both *wzy*-dependent and ABC transporter-dependent. Strains that were capsule-positive and O8/O9/O89 clustered together in multiple clades whereas those that were O8/O9/O89 and capsule-negative occurred throughout phylogroups A, B1, and C (Figure 2.1). Phylogroup C strains were most likely to have one of either an O8/O9/O89 serogroup (17 out of 19 strains; 89%), while the frequency of the strains with these serogroups was 20% (67 out of 332 strains) for phylogroup A strains and 17% (50 out of 301 strains) for B1 strains. O8/O9/O89 strains were uncommon among phylogroups B2, D, E, and F (Table 2.2). Phylogroup A strains were more likely to be O9/O89, while O8 strains were more common among phylogroup B1 strains. However, regardless of a strain's phylogroup membership, those with an O9 serogroup were more likely to be capsule-positive than the O8 strains (Figure 2.1) (Table 2.2). The H-types of the encapsulated strains were variable. Strains with the same capsule type had different H-types, while similar H-types occurred with different capsule types.

Strains encoding a *Klebsiella* capsule (228 strains) were next investigated for the presence of other capsule types reported in Bacteria and Archaea. When a strain was positive for any of the eight capsule types reported by Rendueles and colleagues (2017), the capsule types identified were Wzx/Wzy-dependent, Group IV_e, or Group IV_s (Table 2.3). Group IV_e was the most abundant (214 out of 228 strains; 93.9%) followed by Group IV_s (174 out of 228 strains; 76.3%), and Wzx/Wzy-dependent (167 out of 228 strains; 73.2%). The presence of these three capsule types was variable among the bloom strains.

Table 2.2. Presence/absence of *Klebsiella* capsules in O8, O9, or O89 *E. coli* strains of different phylogroups

Phylogroup	Serogroup*	Capsule-positive n (%)
A (n=332)	O8 (n=16)	4 (25.0%)
	O9 (n=25)	22 (88.0%)
	O89 (n=26)	25 (96.2%)
B1 (n=301)	O8 (n=35)	2 (5.7%)
	O9 (n=14)	13 (92.9%)
	O89 (n=1)	0 (0%)
C (n=19)	O8 (n=10)	2 (20.0%)
	O9 (n=7)	6 (85.7%)
B2 (n=245)	O8 (n=13)	0 (0%)
D (n=142)	O8 (n=2)	0 (0%)
E (n=72)	O8 (n=1)	0 (0%)
	O9 (n=1)	0 (0%)
F (n=68)	O8 (n=2)	0 (0%)

* Strains with an O9 serogroup were not observed for phylogroups B2, D, and F, while strains with an O89 serogroup were absent from phylogroups C, B2, D, E, and F.

Table 2.3. Occurrence of Wzx/Wzy-dependent, GroupIV_s, and GroupIV_e capsule types in *Klebsiella* capsule-positive *E. coli*

Wzx/Wzy-dependent	GroupIV_s	GroupIV_e	Number of strains
-	-	-	3
-	-	+	16
-	+	+	45
+	-	-	6
+	-	+	29
+	+	-	5
+	+	+	124

2.4.4 Variable gene content of bloom and other encapsulated *E. coli*

Pan genome analysis of the capsule region of encapsulated strains showed that there were no capsule-specific genes unique to all bloom strains compared to other encapsulated strains.

The majority of the bloom strains were members of phylogroup A, as were the bulk of the strains found to possess a *Klebsiella* capsule. Consequently, to determine if variable gene

content varied between bloom strains and other encapsulated strains, or between encapsulated and non-encapsulated strains, the pan genome analysis was restricted to phylogroup A strains.

No genes were found to be unique to bloom strains compared to other encapsulated strains, and apart from capsule genes, no genes were unique to encapsulated strains compared to non-encapsulated strains. However, all phylogroup A bloom strains encoded the ferric citrate uptake system (*fecIRABCDE*), while ≤60% of non-bloom encapsulated strains, and fewer than 39% of non-bloom phylogroup A strains encoded the *fec* genes. Genes of the *fim* operon (96.3% versus ≤60.0%) and *cas* genes (≥85.2% versus ≤56.0%) were also over-represented among the bloom strains compared to non-bloom encapsulated strains (Supplemental Table 2.2A). Other genes were found to be over- or under-represented when bloom strains were compared to other encapsulated phylogroup A strains (Supplemental Tables 2.2A, 2.2B) or when encapsulated strains were compared to non-encapsulated phylogroup A strains (Supplemental Tables 2.3A, 2.3B). Overall, the variable gene content of bloom or encapsulated strains did not differ substantially from non-encapsulated strains (Figure 2.3). The phylogroup A1 bloom strains isolated from the east coast group somewhat separately from the rest of the bloom and encapsulated strains (Figure 2.3). This might be attributed to differences in sequence type (ST10 versus others), capsule type (KL16 and KL113 versus others), virulence profiles (bloom strains lack an array of virulence factors) etc.

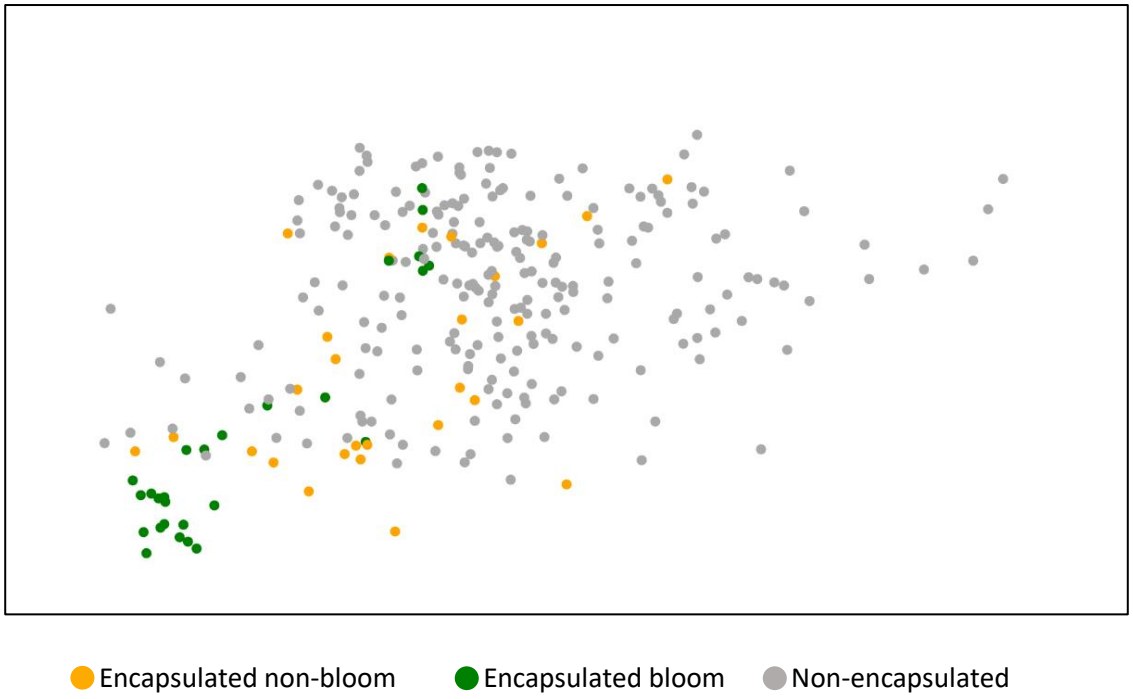


Figure 2.3. Principal coordinate (PCoA) plot of the variable gene content of phylogroup A encapsulated non-bloom (n = 27; orange), bloom (n = 27; green), and non-encapsulated (n = 276; grey) *E. coli* strains. Analysis was done using PAST3 (Hammer et al., 2001). Axis 1 captures 9.7% of the variation and axis 2 captures 4.2% of the variation. The group of green dots at the bottom left represents the phylogroup A1 bloom strains isolated from the east coast.

2.5 Discussion

Overall, about 7% of *E. coli* strains have acquired a *Klebsiella* capsule. However, the distribution of encapsulated strains is non-random and restricted to phylogroups A, B1, and C. *Klebsiella* capsules are very rare or absent among strains belonging to phylogroups B2, D, E, and F. The variable gene content of a strain varies with its phylogroup membership and, on average, each phylogroup has a distinct variable gene content (Touchon et al., 2009; Clermont et al., 2013). This outcome suggests that the acquisition and maintenance of a group 1 capsule depends on the genomic background of the strains. Further, it suggests that there is some

aspect of the genomic background of B2, D, E, and F strains that is incompatible with the possession of a *Klebsiella* capsule, although the nature of this incompatibility is unknown.

Another clearly non-random association between the possession of a group 1/*Klebsiella* capsule and the genomic background of a strain is the long recognized association between group 1 capsule and the serogroups O8, O9, and O89 (Kauffmann, 1947; Kido et al., 1995; Amor and Whitfield, 1997; Drummelsmith et al., 1997; Whitfield and Roberts, 1999). O-antigen biosynthesis in *E. coli* operates in two major pathways (Willis and Whitfield, 2013). One is *wzy*-dependent and carries Wzx (O-antigen flippase) and Wzy (O-antigen polymerase). The other pathway is ATP-binding cassette (ABC) transporter-dependent relying on Wzt (ABC transporter ATP-binding protein) and Wzm (ABC transporter permease). Among 182 O-antigen gene clusters, the vast majority (n = 171) had *wzx*/*wzy* genes, the remaining 11 had *wzt*/*wzm* genes and O8, O9, and O89 are among these ABC transporter-dependent serogroups (Kido et al., 1995; Iguchi et al., 2014).

BLAST comparisons of *wzt* from O8, O9, and O89 *E. coli* strains, regardless of their capsule status, most often resulted in a very close match with the *wzt* gene from either a strain of *Klebsiella* or an encapsulated *E. coli* strain. This outcome suggests that the entire region comprising the O-antigen and capsule gene clusters is the consequence of a single recombination event. The fact that O8 and O9 of *E. coli* are identical to O5 and O3 of *K. pneumoniae*, respectively (Jansson et al., 1985; Saeki et al., 1993) and that O8, O9 co-express with the *E. coli* group 1/*Klebsiella* capsule (Whitfield and Roberts, 1999) also supports the notion that both the O-antigen and capsule region co-transfer. This is further corroborated by *E. coli* O9a, which is suggested to have arisen due to a post-recombination mutation in an O-antigen gene upon transfer of *K. pneumoniae* O3 to *E. coli* (Sugiyama et al., 1998).

There are phylogroup B2, D, E, and F strains with an O8 or O9 serogroup, but these strains are capsule-negative. If it is assumed that the O-antigen and capsule region are transferred as a single block from *Klebsiella* to *E. coli*, then the absence of a capsule would suggest that these strains have subsequently lost the capsule region genes but maintained the O-antigen region from *Klebsiella*. In turn, this outcome also indicates that the genomic background of strains belonging to these phylogroups is incompatible with the maintenance of a *Klebsiella* capsule.

Acquisition of the *Klebsiella* capsule has occurred in a limited number of lineages, but when a lineage has acquired a *Klebsiella* capsule, there are often multiple *Klebsiella* capsule types in the lineage (Figure 2.1). A similar distribution can be observed in O-serogroups, where the occurrence of multiple O-serogroups within a single lineage is not unusual (Ingle et al., 2016). This outcome is unlikely to be the result of a single *Klebsiella* capsule acquisition event, followed by subsequent evolution of the capsule region in the lineage. Different capsule types are defined not only by their sequence similarity but also by the presence of a particular suite of capsule region genes (Wyres et al., 2016). Therefore, while capsule types within a lineage might diverge due to the loss of genes within the capsule region, the gain of genes can only be the consequence of horizontal gene transfer events. Further, the presence of transposase/IS elements on either side of the capsule cluster is also indicative of a horizontal gene transfer event, as reported previously (Rahn et al., 1999). Once a lineage has acquired the *Klebsiella* capsule region and likely the associated O-antigen region as well, then it may be that this lineage has an increased likelihood of experiencing subsequent lateral gene transfer events involving this region. The *Klebsiella* capsule region is typically 15 - 30 kbp and the region would be larger if the O-antigen region is also considered. This region would represent a large target for homologous recombination with capsule regions from other encapsulated

E. coli strains or from *Klebsiella*. Consequently, subsequent homologous recombination of the capsule region would be expected to occur at a higher rate than *de novo* acquisition of the capsule region through non-homologous recombination. The capsule (*cps*) region around the *rfb* locus has been identified as a recombination hotspot in multiple bacterial species including *Streptococcus pneumoniae*, *K. pneumoniae*, and *E. coli* (Milkman et al., 2003; Didelot et al., 2012; Alqasim et al., 2014; Wright et al., 2014; Mostowy et al., 2017).

The encapsulated strains do not carry the colanic acid gene cluster, indicating that the acquisition of the capsule has been at the cost of colanic acid, as has been previously reported (Rahn et al., 1999). The great majority of non-encapsulated *E. coli* strains are positive for colanic acid. Colanic acid appears to be the default exopolysaccharide, particularly in phylogroups B2, D, E, and F. Given the different temperatures of expression and the genetic localization, the expression of the *Klebsiella* capsule/*E. coli* group 1 capsule and colanic acid are mutually exclusive events (Rahn et al., 1999; Whitfield, 2006).

Three strains have been associated with bloom events in eastern Australia (Power et al., 2005) and five distinct strains were isolated during a recent bloom event in Western Australia. All bloom strains are encapsulated, and given that only about 7% of *E. coli* are encapsulated, then this indicates that the possession of a *Klebsiella* capsule is required if a strain is to be capable of causing bloom events. Among the eight bloom strains, there are seven distinct capsule types, again suggesting that it is the possession of the capsule *per se* that is important rather than the possession of a particular *Klebsiella* capsule variant. Of the two strains with the same capsule type, one is a phylogroup B1 strain while the other belongs to phylogroup A, further suggesting that it is the possession of a capsule that is important rather than the genomic background of the strain. The pan genome analysis did not detect any genes either uniquely

present or absent in bloom strains or other encapsulated *E. coli*. However, the pan genome analysis indicated that the iron uptake system encoded by the *fecIRABCDE* operon was over-represented in the phylogroup A bloom strains, and its importance for the bloom status remains to be investigated. Genes of the *fim* operon that code for type 1 pilus synthesis, assembly, and regulation (Orndorff and Falkow, 1984; Klemm, 1986) are more prevalent in bloom strains compared to non-bloom encapsulated strains. The type 1 pilus is required for the attachment of *E. coli* cells to abiotic surfaces (Pratt and Kolter, 1998; Cookson et al., 2002) and might enhance the free-living lifestyle of the bloom strains. Further experimentation with bloom water samples is needed to infer if bloom strains prefer to reside as part of biofilms. The east coast B1 bloom strain, however, appears to be incapable of forming biofilms and is likely planktonic (Chapter 5). CRISPR (clustered regularly interspaced short palindromic repeats)-associated *cas* genes are also over-represented in the bloom strains. The CRISPR/Cas system provides resistance against viral infection (Horvath and Barrangou, 2010) and would enhance the survival of the strains.

Rendueles and colleagues (2017) revealed that capsules are more likely to occur in free-living species than in pathogens, reversing the long-held belief that capsules are associated more with virulence (Simoons-Smit et al., 1986; Williams et al., 1990; Lawlor et al., 2005). Possession of a capsule is believed to enhance strain survival and persistence through overcoming predation and adverse environmental conditions including desiccation, osmotic stress, and UV radiation encountered in the external environment. The capsule may facilitate a strain's ecological transitions, thereby increasing its environmental range (Weiner et al., 1995; Rendueles et al., 2017). Thus, being encapsulated may enhance a strains' persistence in the external environment, relative to non-encapsulated *E. coli*. Indeed, studies indicate that in

eastern Australian reservoirs and recreational lakes, bloom strains can be isolated at any time, even when *E. coli* counts are low (non-bloom periods).

Although encapsulation appears to be essential for conferring bloom status on an *E. coli* strain, it seems unlikely that it can be the sole trait required. At any given point in time, multiple *E. coli* genotypes are present in a water body (Casarez et al., 2007; Higgins et al., 2007). Bloom events are associated with nutrient influx events such as dust storms or the autumn die-off of aquatic vegetation (Water Research Australia, 2019). Encapsulated strains might be more likely to persist in water bodies compared to other *E. coli* strains, due to their potentially enhanced survival, and can exploit these nutrient influx events and achieve high cell densities. However, this does not explain why other strains of *E. coli* present in the water body are also not capable of doing this. Further studies are required to identify traits that confer a growth rate advantage on the bloom strains relative to non-encapsulated *E. coli*.

2.6 Acknowledgements

This study was funded in part by an Australian Research Council Linkage Grant (LP120100327). Additional support was provided by Melbourne Water, Water Corporation of Western Australia, Queensland Bulk Water Authority, Hunter Water Corporation, South East Water, Yarra Valley Water, City West Water, Sydney Water Corporation, South Australian Water Corporation, Sydney Catchment Authority, Fitzroy River Water, and Victorian Department of Health. Portions of the research were also funded by Water Research Australia (2018) Management of Environmental *E. coli*, Project 1101.

2.7 References

Alm, E.W., Walk, S.T., and Gordon, D.M. (2011) The niche of *Escherichia coli*. In *Population genetics of bacteria: a tribute to Thomas S. Whittam.* Walk, S.T., and Feng, P.C.H. (eds). Washington, DC: ASM Press, pp. 69-89.

Alqasim, A., Scheutz, F., Zong, Z., and McNally, A. (2014) Comparative genome analysis identifies few traits unique to the *Escherichia coli* ST131 H30Rx clade and extensive mosaicism at the capsule locus. *BMC Genomics* 15: 830.

Álvarez, D., Merino, S., Tomás, J. M., Benedí, V. J., and Albertí, S. (2000) Capsular polysaccharide is a major complement resistance factor in lipopolysaccharide O side chain-deficient *Klebsiella pneumoniae* clinical isolates. *Infect Immun* 68(2): 953-955.

Amor, P.A., and Whitfield, C. (1997) Molecular and functional analysis of genes required for expression of group IB K antigens in *Escherichia coli*: characterization of the *his*-region containing gene clusters for multiple cell-surface polysaccharides. *Mol Microbiol* 26(1): 145-161.

Batchelor, R.A., Haraguchi, G.E., Hull, R.A., and Hull, S.I. (1991) Regulation by a novel protein of the bimodal distribution of lipopolysaccharide in the outer membrane of *Escherichia coli*. *J Bacteriol* 173(18): 5699-5704.

Beghain, J., Bridier-Nahmias, A., Le Nagard, H., Denamur, E., and Clermont, O. (2018) ClermonTyping: an easy-to-use and accurate *in silico* method for *Escherichia* genus strain phylotyping. *Microb Genom* 4: e000192.

Bitton, G. (2011) *Wastewater Microbiology*, 4th ed. Hoboken, NJ: Wiley-Blackwell.

Blyton, M.D.J., Cornall, S.J., Kennedy, K., Colligon, P., and Gordon, D.M. (2014) Sex-dependent competitive dominance of phylogenetic group B2 *Escherichia coli* strains within human hosts. *Environ Microbiol Rep* 6(6): 605-610.

Blyton, M.D.J., Pi, H., Vangchhia, B., Abraham, S., Trott, D.J., Johnson, J.R., and Gordon, D.M. (2015) Genetic structure and antimicrobial resistance of *Escherichia coli* and cryptic clades in birds with diverse human associations. *Appl Environ Microbiol* 81(15): 5123-5133.

Bonde, G.J. (1966) Bacteriological methods for estimation of water pollution. *Health Lab Sci* 3(2): 124-128.

Brynildsrud, O., Bohlin, J., Scheffer, L., and Eldholm, V. (2016) Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* 17: 238.

Byappanahalli, M., Fowler, M., Shively, D., and Whitman, R. (2003) Ubiquity and persistence of *Escherichia coli* in a Midwestern coastal stream. *Appl Environ Microbiol* 69(8): 4549-4555.

Byappanahalli, M.N., Whitman, R.L., Shively, D.A., Sadowsky, M.J., and Ishii, S. (2006) Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. *Environ Microbiol* 8(3): 504-513.

Byappanahalli, M.N., Whitman, R.L., Shively, D.A., Ferguson, J., Ishii, S., and Sadowsky, M.J. (2007) Population structure of *Cladophora*-borne *Escherichia coli* in nearshore water of Lake Michigan. *Water Res* 41(16): 3649-3654.

Carrillo, M., Estrada, E., and Hazen, T.C. (1985) Survival and enumeration of the fecal indicators *Bifidobacterium adolescentis* and *Escherichia coli* in a tropical rain forest watershed. *Appl Environ Microbiol* 50(2): 468-476.

Casarez, E.A., Pillai, S.D., and Di Giovanni, G.D. (2007) Genotype diversity of *Escherichia coli* isolates in natural waters determined by PFGE and ERIC-PCR. *Water Res* 41(16): 3643-3648.

Clermont, O., Bonacorsi, S., and Bingen, E. (2000) Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* 66(10): 4555-4558.

Clermont, O., Christenson, J.K., Denamur, E., and Gordon, D.M. (2013) The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 5(1): 58-65.

Cookson, A.L., Cooley, W.A., and Woodward, M.J. (2002) The role of type 1 and curli fimbriae of Shiga toxin-producing *Escherichia coli* in adherence in abiotic surfaces. *Int J Med Microbiol* 292(3-4): 195-205.

Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14(7): 1394-1403.

Didelot, X., Méric, G., Falush, D., and Darling, A.E. (2012) Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics* 13: 256.

Drummelsmith, J., Amor, P.A., and Whitfield, C. (1997) Polymorphism, duplication, and IS1-mediated rearrangement in the chromosomal *his-rfb-gnd* region of *Escherichia coli* strains with group IA and capsular K antigens. *J Bacteriol* 179(10): 3232-3238.

Edberg, S.C., Rice, E.W., Karlin, R.J., and Allen M.J. (2000) *Escherichia coli*: the best biological drinking water indicator for public health protection. *J Appl Microbiol* 88(S1): 106S-116S.

Gordon, D.M. (2001) Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. *Microbiology* 147: 1079-1085.

Gordon, D.M. (2013) The ecology of *Escherichia coli*. In *Escherichia coli: Pathotypes and principles of pathogenesis*. Donnenberg, M.S. (ed). San Diego, CA: Academic Press, pp. 3-20.

Gordon, D.M., and Cowling, A. (2003) The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology* 149: 3575-3586.

Gordon, D.M., and FitzGibbon, F. (1999) The distribution of enteric bacteria from Australian mammals: host and geographical effects. *Microbiology* 145: 2663-2671.

Hammer, Ø., Harper, D.A.T., and Ryan, P.D. (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontol Electronica* 4: 1-9.

Higgins, J., Hohn, C., Hornor, S., Frana, M., Denver, M., and Joerger, R. (2007) Genotyping of *Escherichia coli* from environmental and animal samples. *J Microbiol Methods* 70(2): 227-235.

Horvath, P., and Barrangou, R. (2010) CRISPR/Cas, the immune system of Bacteria and Archaea. *Science* 327(5962): 167-170.

Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M. et al. (2014) A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res* 22(1): 101-107.

Ingle, D.J., Valcanis, M., Kuzevski, A., Tauschek, M., Inouye, M., Stinear, T. et al. (2016) *In silico* serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and between pathogenic lineages. *Microb Genom* 2: 0.000064.

Ishii, S., and Sadowsky, M.J. (2008) *Escherichia coli* in the environment: Implications for water quality and human health. *Microbes Environ* 23(2): 101-108.

Ishii, S., Ksoll, W.B., Hicks, R.E., and Sadowsky, M.J. (2006) Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. *Appl Environ Microbiol* 72(1): 612-621.

Jansson, P-E., Lönngren, J., Widmalm, G., Leontein, K., Slettengren, K., Svenson, S.B. et al. (1985) Structural studies of the O-antigen polysaccharides of *Klebsiella* O5 and *Escherichia coli* O8. *Carbohydr Res* 145(1): 59-66.

Jayaratne, P., Keenleyside, W.J., MacLachlan, P.R., Dodgson, C., and Whitfield, C. (1993) Characterization of *rcsB* and *rcsC* from *Escherichia coli* O9: K30: H12 and examination of the role of the *rcs* regulatory system in expression of group I capsular polysaccharides. *J Bacteriol* 175(17): 5384-5394.

Joensen, K.G., Tetzschner, A.M., Iguchi, A., Aarestrup, F.M., and Scheutz, F. (2015) Rapid and easy *in silico* serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. *J Clin Microbiol* 53(8): 2410-2426.

Kauffmann, F. (1947) The serology of the coli group. *J Immunol* 57(1): 71-100.

Kido, N., Torgov, V.I., Sugiyama, T., Uchiya, K., Sugihara, H., Komatsu, T. et al. (1995) Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E. coli* O9 *rfb* gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. *J Bacteriol* 177(8): 2178-2187.

Klemm, P. (1986) Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *Escherichia coli*. *EMBO J* 5(6): 1389-1393.

Lawlor, M.S., Hsu, J., Rick, P.D., and Miller, V.L. (2005) Identification of *Klebsiella pneumoniae* virulence determinants using an intranasal infection model. *Mol Microbiol* 58(4): 1054-1073.

Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44(W1): W242-W245.

Lin, J.C., Chang, F.Y., Fung, C.P., Xu, J.Z., Cheng, H.P., Wang, J.J. et al. (2004). High prevalence of phagocytic-resistant capsular serotypes of *Klebsiella pneumoniae* in liver abscess. *Microbes and infection* 6(13): 1191-1198.

Milkman, R., Jaeger, E., and McBride, R.D. (2003) Molecular evolution of the *Escherichia coli* chromosome. VI. Two regions of high effective recombination. *Genetics* 163(2): 475-483.

Mostowy, R.J., Croucher, N.J., De Maio, N., Chewapreecha, C., Salter, S.J., Turner, P. et al. (2017) Pneumococcal capsule synthesis locus *cps* as evolutionary hotspot with potential to generate novel serotypes by recombination. *Mol Biol Evol* 34(10): 2537–2554.

Nanayakkara, B.S., O'Brien, C.L., and Gordon, D.M. (2019) Diversity and distribution of *Klebsiella* capsules in *Escherichia coli*. *Environ Microbiol Rep* 11(2): 107-117.

Orndorff, P.E., and Falkow, S. (1984) Organization and expression of genes responsible for type 1 piliation in *Escherichia coli*. *J Bacteriol* 159(2): 736-744.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G. et al. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22): 3691-3693.

Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., and Slade, M.B. (2005) Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. *Environ Microbiol* 7(5): 631-640.

Pratt, L.A., and Kolter, R. (1998) Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Mol Microbiol* 30(2): 285-293.

Rahn, A., Drummelsmith, J., and Whitfield, C. (1999) Conserved organization in the *cps* gene clusters for expression of *Escherichia coli* group 1 K antigens: relationship to the colanic acid biosynthesis locus and the *cps* genes from *Klebsiella pneumoniae*. *J Bacteriol* 181(7): 2307-2313.

Rendueles, O., Garcia-Garcerà, M., Néron, B., Touchon, M., and Rocha, E.P.C. (2017) Abundance and co-occurrence of extracellular capsules increase environmental breadth: Implications for the emergence of pathogens. *PLoS Pathog* 13: 1-30.

Saeki, A., Kido, N., Sugiyama, T., Ohta, M., Iwashita, T., Uchiya, K., and Kato, N. (1993) Isolation of *rfb* gene clusters directing the synthesis of O polysaccharides consisting of mannose homopolymers and serological analysis of lipopolysaccharides. *Microbiol Immunol* 37(8): 601-606.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14): 2068 – 2069.

Simoons-Smit, A.M., Verweij-van Vught, A.M.J.J., and MacLaren, D.M. (1986) The role of K antigens as virulence factors in *Klebsiella*. *J Med Microbiol* 21(2): 133-137.

Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R., and Palmer, C.J. (2000) Sources of *Escherichia coli* in a coastal subtropical environment. *Appl Environ Microbiol* 66(1): 230-237.

Struve, C., and Krogfelt, K.A. (2003) Role of capsule in *Klebsiella pneumoniae* virulence: lack of correlation between in vitro and in vivo studies. *FEMS Microbiol Lett* 218(1): 149-154.

Sugiyama, T., Kido, N., Kato, Y., Koide, N., Yoshida, T., and Yokochi, T. (1998) Generation of *Escherichia coli* O9a serotype, a subtype of *E. coli* O9, by transfer of the *wb** gene cluster of *Klebsiella* O3 into *E. coli* via recombination. *J Bacteriol* 180(10): 2775-2778.

Szewzyk, U., Szewzyk, R., Manz, W., and Schleifer, K.-H. (2000) Microbiological safety of drinking water. *Annu Rev Microbiol* 54: 81-127.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P. et al. (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5: e1000344.

Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014) The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 15: 524.

U.S. EPA - United States Environmental Protection Agency. (2012) Recreational water quality criteria. Office of Water 820-F-12-058. https://www.epa.gov/sites/production/files/2015-10/documents/rwqc2012.pdf.

van Elsas, J.D., Semenov, A.V., Costa, R., and Trevors, J.T. (2011) Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J* 5(2): 173-183.

Vangchhia, B., Abraham, S., Bell, J.M., Collignon, P., Gibson, J.S., Ingram, P.R. et al. (2016) Phylogenetic diversity, antimicrobial susceptibility and virulence characteristics of phylogroup F *Escherichia coli* in Australia. *Microbiology* 162(11): 1904-1912.

Water Research Australia. (2019) Management of environmental *E. coli* blooms. Factsheet Project 1101. https://www.waterra.com.au › publications › document-search

Weiner, R., Langille, S., and Quintero, E. (1995) Structure, function and immunochemistry of bacterial exopolysaccharides. *J Ind Microbiol* 15(4): 339-346.

Whitfield, C. (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu Rev Biochem* 75: 39-68.

Whitfield, C., and Paiment, A. (2003) Biosynthesis and assembly of Group 1 capsular polysaccharides in *Escherichia coli* and related extracellular polysaccharides in other bacteria. *Carbohydr Res* 338(23): 2491-2502.

Whitfield, C., and Roberts, I.S. (1999) Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol Microbiol* 31(5): 1307-1319.

Whitman, R.L., Shively, D.A., Pawlik, H., Nevers, M.B., and Byappanahalli, M.N. (2003) Occurrence of *Escherichia coli* and enterococci in *Cladophora* (Chlorophyta) in nearshore water and beach sand of Lake Michigan. *Appl Environ Microbiol* 69(8): 4714-4719.

Williams, P., Lambert, P.A., Brown, M.R., and Jones, R.J. (1983) The role of the O and K antigens in determining the resistance of *Klebsiella aerogenes* to serum killing and phagocytosis. *Microbiology* 129(7): 2181-2191.

Williams, P., Ciurana, B., Camprubi, S., and Tomas, J.M. (1990) Influence of lipopolysaccharide chemotype on the interaction between *Klebsiella pneumoniae* and human polymorphonuclear leucocytes. *FEMS Microbiol Lett* 69: 305-310.

Willis, L.M., and Whitfield, C. (2013) Capsule and lipopolysaccharide. In *Escherichia coli: Pathotypes and principles of pathogenesis.* Donnenberg, M.S. (ed). San Diego, CA: Academic Press, pp. 533-556.

Wright, M.S., Perez, F., Brinkac, L., Jacobs, M.R., Kaye, K., Cober, E. et al. (2014) Population structure of KPC-producing *Klebsiella pneumoniae* isolates from midwestern US hospitals. *Antimicrob Agents Chemother* 58(8): 4961-4965.

Wyres, K.L., Wick, R.R., Gorrie, C., Jenney, A., Follador, R., Thomson, N.R., and Holt, K.E. (2016) Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom* 2(12): 0.000102.

2.8 Supplemental Material

Supplemental Table 2.1. Isolation details of the bloom strains

Bloom strain group	Number of isolates	Date of isolation	Geographic origin*
East coast A0	6	2002-2014	Hinze Dam, Qld; Hunter valley, NSW
East coast A1	17	2002-2014	Hinze Dam, Qld; Hunter valley, NSW; Lake Ginninderra, ACT; Shoalhaven, NSW; Sullivans Creek, ACT
East coast B1	7	2002-2015	Googong Dam, NSW; Hinze Dam, Qld; Shoalhaven, NSW; Wyaralong Dam, Qld
Western Australia A1-4 strains	1 each	2015	WA
Western Australia B1	1	2015	WA

* ACT = Australian Capital Territory; NSW = New South Wales; Qld = Queensland; WA = Western Australia

Supplemental Table 2.2A. Genes over-represented among phylogroup A bloom strains compared to phylogroup A non-bloom encapsulated *E. coli*

Gene	Gene product/function	% occurrence in bloom strains (n = 27) (>75 %)	% occurrence in non-bloom encapsulated strains (n = 25) (≤60%)
yjhV	Conserved hypothetical protein	100	56.0
fecE	Ferric citrate transport	100	56.0
ydaM	Sensor domain-containing diguanylate cyclase	100	56.0
fecB	Ferric citrate transport system	100	56.0
fecD	Ferric citrate transport system	100	56.0
ybjL	Putative transmembrane potassium ion transporter	100	60.0
fecC	Ferric citrate transport system	100	60.0
fecI	Ferric citrate transport system	100	60.0
fecA	Ferric citrate transport system	100	60.0
fecR	Ferric citrate transport system	100	60.0
fimA	Major type 1 subunit fimbrin (pilin)	96.3	40.0
group_9550	Hcp1 family type VI secretion system	96.3	56.0
fimE	Controls phase variation of type 1 fimbriae	96.3	56.0
fimB	Type 1 fimbriae regulatory protein FimB	96.3	56.0
fimI	Type 1 pilus biosynthesis	96.3	56.0
fimC	Chaperone type 1 pilus biosynthesis	96.3	60.0
tamB	Putative outer membrane protein	96.3	60.0
group_6187	Hypothetical protein	92.6	28.0
group_4967	Hypothetical protein	92.6	32.0
casB	type I-E CRISPR-associated protein Cse2/CasB	92.6	48.0
group_1154	DUF4132 domain-containing protein	92.6	52.0
casE	Encodes CRISPR system Cascade subunit CasE	92.6	52.0
yehQ	Hypothetical protein	92.6	56.0
cstA	Carbon starvation protein A	92.6	56.0
yfaW	L-rhamnonate dehydratase	92.6	56.0
cas2	CRISPR-associated protein	92.6	56.0
group_13527	Conserved hypothetical protein	92.6	56.0
yaiW	Putative lipoprotein	92.6	56.0
group_10425	Conserved protein of unknown function	92.6	60.0
group_12337	Hypothetical protein	88.9	32.0
group_10439	Hypothetical protein	88.9	32.0
gspD_1	Type II secretion system protein GspD	88.9	36.0
ydeS	Fimbrial-like adhesin protein	88.9	36.0
group_12403	Hypothetical protein	88.9	40.0
yieH	6-phosphogluconate phosphatase	88.9	40.0
group_6978	Pyruvate:ferredoxin (flavodoxin) oxidoreductase	88.9	40.0
yagP	Hypothetical protein	88.9	48.0
gnsB	Protein GnsB	88.9	48.0
cas1	Type I-E CRISPR-associated endonuclease Cas1	88.9	48.0

Gene	Gene product/function	% occurrence in bloom strains (n = 27) (>75 %)	% occurrence in non-bloom encapsulated strains (n = 25) (≤60%)
casA	Type I-E CRISPR-associated protein Cse1/CasA	88.9	48.0
yegJ	DUF2314 domain-containing protein	88.9	48.0
gspE_1	Type II secretion system protein GspE	88.9	52.0
casC	Type I-E CRISPR-associated protein Cas7/Cse4/CasC	88.9	52.0
yghF	Type II secretion system protein GspC	88.9	56.0
yodB	Cytochrome b561	88.9	56.0
yghG	Hypothetical protein	88.9	56.0
pppA	Prepilin peptidase	88.9	56.0
gspF_1	Type II secretion system protein GspF	88.9	56.0
gspG_1	Type II secretion system protein GspG	88.9	56.0
group_1600	Lipopolysaccharide heptosyltransferase family protein	88.9	56.0
casD	Type I-E CRISPR-associated protein Cas5/CasD	85.2	48.0
bglF	PTS beta-glucoside transporter subunit EIIBCA	85.2	52.0
group_8165	DNA ligase B	81.5	16.0
dicB	Division inhibition protein DicB	81.5	40.0
nohB	Host specificity protein J	81.5	44.0
ydfD	DUF1482 family protein	81.5	44.0
yieL	Xylanase	81.5	48.0
group_3309	Sensor domain-containing phosphodiesterase	77.8	32.0

Supplemental Table 2.2B. Genes under-represented among phylogroup A bloom strains compared to phylogroup A non-bloom encapsulated *E. coli*

Gene	Gene product/function	% occurrence in bloom strains (n = 27) (<50%)	% occurrence in non-bloom encapsulated strains (n = 25) (>75%)
ligB	DNA ligase B	14.8	84.0
yebB	YebB family permuted papain-like enzyme	29.6	84.0
erfK	L,D-transpeptidase	29.6	84.0
group_7642	YjjI family glycine radical enzyme	29.6	76.0
ypdF	Aminopeptidase	29.6	76.0
potG	Putrescine transport ATP-binding protein PotG	33.3	96.0
betT	Choline transporter	33.3	92.0
betB	Betaine-aldehyde dehydrogenase	33.3	92.0
ykgF	Ferredoxin-like LutB family protein	33.3	92.0
group_11760	FUSC family protein	33.3	88.0
safA	Two-component-system connector protein SafA	33.3	80.0
group_5813	Capsule assembly Wzi family protein	37.0	100
hyfR_2	Hydrogenase-4 transcriptional activator	37.0	100
cheZ	Protein phosphatase CheZ	37.0	92.0
yeeE	YeeE/YedE family protein	37.0	88.0
yfdE	Acetyl-CoA--oxalate CoA-transferase	37.0	80.0
ydcM_1	Transposase	37.0	80.0
aspS	Aspartate--tRNA ligase	37.0	80.0
yfaL	AIDA-I family autotransporter YfaL	37.0	76.0
group_14522	DUF2057 family protein	40.7	100
yfgO	AI-2E family transporter	40.7	96.0
ykgH	Hypothetical protein	40.7	88.0
dadX	Alanine racemase catabolic	40.7	84.0
ykgC	Pyridine nucleotide-disulfide oxidoreductase	40.7	80.0
yahA	Cyclic di-GMP phosphodiesterase YahA	40.7	76.0
fliY	Cystine ABC transporter substrate-binding protein	48.1	96.0
elaD	Deubiquitinase	48.1	80.0

Supplemental Table 2.3A. Genes over-represented among phylogroup A encapsulated strains compared to phylogroup A capsule-negative *E. coli*

Gene	Gene product/function	% occurrence in encapsulated strains (n = 111) (>75%)	% occurrence in capsule-negative strains (n = 230) (≤60%)
galF	Putative uridylyltransferase subunit with GalU	99.1	0
kanE	Alpha-D-kanosaminyltransferase	95.5	8.7
ugd	UDP-glucose 6-dehydrogenase	93.7	0
hisG	ATP phosphoribosyltransferase	93.7	59.6
oppA	Peptide ABC transporter	92.8	55.7
rfaD	ADP-L-glycero-D-mannoheptose-6-epimerase	91.9	50.4
frmA	Formaldehyde dehydrogenase	91.0	55.7
yigG	Inner membrane protein	90.1	52.2
mfd	Transcription-repair coupling factor	90.1	57.8
arnB	UDP-L-Ara4O C-4 transaminase	90.1	59.6
arpA_1	Regulator of acetyl CoA synthetase	89.2	49.6
acrD	AcrAD-TolC multidrug efflux transport system	89.2	57.8
alsB	D-allose ABC transporter	88.3	57.4
yfaX	Putative DNA-binding transcriptional regulator	88.3	57.8
nepI	Purine ribonucleoside efflux transporter	88.3	57.8
iraD	Inhibitor of sS proteolysis	87.4	39.6
nirB	Nitrite reductase, large subunit	87.4	55.7
group_13623	Hypothetical protein	86.5	48.3
rhaD	Rhamnulose-1-phosphate aldolase monomer	86.5	53.5
tolQ	Colicin A import System	86.5	54.8
recQ	ATP-dependent DNA helicase	86.5	55.2
melB	Melibiose:H^+/Na^+/Li^+ symporter	86.5	55.7
group_11014	Hypothetical protein	86.5	56.5
insG	IS1 predicted transposase	85.6	49.1
ariR	Regulator of acid resistance	85.6	53.5
alsK	D-allose kinase	84.7	43.0
insG_2	Hypothetical protein	84.7	46.5
citF	Citrate lyase	84.7	51.7
bglH	Carbohydrate-specific outer membrane porin	84.7	52.6
yfgH	Putative lipoprotein	83.8	34.3
yjbL	Putative protein	83.8	43.9
leuA	2-isopropylmalate synthase	83.8	49.6
group_14921	Hypothetical protein	83.8	51.3
actP	Acetate/glycolate transporter	82.9	49.1
dtpB	Dipeptide/tripeptide:H^+ symporter DtpB	82.9	49.1
yneJ	DNA-binding transcriptional regulator	82.9	50.4
rdoA	Serine/threonine protein kinase	82.9	51.3
rsmC	16S rRNA methyltransferase	82.9	51.7
rarD	Putative chloramphenicol resistance permease	82.0	40.9
group_9715	Hypothetical protein	82.0	40.9

Gene	Gene product/function	% occurrence in encapsulated strains (n = 111) (>75%)	% occurrence in capsule-negative strains (n = 230) (≤60%)
ilvH	Acetolactate synthase	82.0	45.2
ynbD	Putative phosphatase	82.0	47.4
gapA_2	Glyceraldehyde 3-phosphate dehydrogenase A	82.0	47.4
yfgI	Putative membrane protein	81.1	38.3
ydeE	Putative transport protein YdeE	81.1	40.9
chbR	DNA-binding transcriptional dual regulator	81.1	44.8
rhsB	RhsB protein in *rhs* element	81.1	46.1
aldA	Aldehyde dehydrogenase A, NAD-linked	81.1	47.0
aphA	Acid phosphatase monomer	81.1	48.7
sdsP	SdsRQP multidrug efflux transport system	81.1	50.4
leuC	Isopropylmalate isomerase	80.2	45.2
leuO	LeuO DNA-binding transcriptional activator	80.2	46.1
trg	Methyl accepting chemotaxis protein	80.2	47.8
group_15111	Hypothetical protein	80.2	48.3
sgbH	3-keto-L-gulonate 6-phosphate decarboxylase	80.2	48.7
yiaK	2,3-diketo-L-gulonate reductase monomer	80.2	49.6
bcsQ	Putative cellulose biosynthesis protein	79.3	28.7
yhiS_1	Hypothetical protein	79.3	33.5
arpA_1	Hypothetical protein	79.3	43.9
group_25103	Hypothetical protein	78.4	33.9
ydjO	Putative protein	78.4	34.8
lexA	LexA DNA-binding transcriptional repressor	77.5	33.0
yjiA	P-loop guanosine triphosphatase	77.5	38.7
yafT	Putative aminopeptidase	77.5	40.0
group_14981	Hypothetical protein	77.5	40.4
ykfM	Hypothetical protein	77.5	43.9
sohB	Putative inner membrane peptidase	76.6	8.7
ygcU	Putative FAD-containing dehydrogenase	76.6	29.1
fecR	Ferric citrate transport system	76.6	40.9
group_2783	Hypothetical protein	76.6	42.2
group_8730	Hypothetical protein	75.7	34.3
yhaB	Putative protein	75.7	36.1
elfC	Putative outer membrane usher protein	75.7	39.6
fecI	Ferric citrate transport system	75.7	40.9
yjbS	Hypothetical protein	75.7	43.5

Supplemental Table 2.3B. Genes under-represented among phylogroup A encapsulated strains compared to phylogroup A capsule-negative *E. coli*

Gene	Gene product/function	% occurrence in encapsulated strains (n = 111) (<50%)	% occurrence in capsule-negative strains (n = 230) (>75%)
sohB	Putative inner membrane peptidase	22.5	91.7
caiB	CaiB monomer	35.1	81.7
ybeF	DNA-binding transcriptional regulator	35.1	78.3
ygcR	Putative flavoprotein	36.0	75.7
topA	DNA topoisomerase I	36.9	85.7
cysB	CysB-O-acetyl-L-serine transcriptional regulator	37.8	80.0
ligB	DNA ligase	40.5	81.7
ybdG	Mechanosensitive YbdG monomer	41.4	96.1
kch	K^+ channel Kch monomer	42.3	93.0
nfsB	Dihydropteridine reductase monomer	44.1	94.8
ybdK	Carboxylate-amine ligase	45.9	99.1
gltD	Glutamate synthase, small subunit	45.9	84.3
ybdJ	Putative inner membrane protein	48.6	99.6
group_16489	Hypothetical protein	48.6	99.1
yeeO	YeeO MATE transporter	49.5	80.0

Chapter 3. PCR-based method to detect *Klebsiella* capsules in *Escherichia coli* and discriminate encapsulated strains harbouring bloom strain-associated capsule types

3.1 Abstract

Elevated counts of *Escherichia coli* in recreational water is of public health concern and causes the closure of recreational water bodies for public use. As *E. coli* bloom strains are not faecal-associated and are likely free-living, closure of water bodies due to bloom events is unnecessary. All bloom strains have acquired a capsule from *Klebsiella* and seven variants of the capsule, termed capsule types, are found among bloom strains.

A PCR protocol was developed to detect capsule-positive *E. coli* and discriminate strains that harbour bloom strain-associated *Klebsiella* capsule types. The PCR accurately detected all known bloom strains and other encapsulated strains, and it accurately identified capsule-negative strains. The PCR-based capsule status and type determinations were consistent with the *in silico* results obtained using Kaptive. Applying our PCR method for 1632 *E. coli* strains isolated from water, we found that 4% of the strains screened were capsule-positive, while 1.7% carried bloom strain-associated capsule types.

3.2 Introduction

Water authorities routinely test drinking and recreational waters for recent faecal contamination (Ishii and Sadowsky, 2008; U.S. EPA, 2012; WHO, 2017). *Escherichia coli* is used as a faecal indicator and its presence in water indicates the possible presence of enteric-associated pathogens (Leclerc et al., 2001; Bitton, 2011). Elevated counts of *E. coli*, i.e.; counts exceeding 235 cfu/100 ml sample of recreational water, are considered unsafe and lead to the closure of water bodies for public access (Ishii and Sadowsky, 2008; U.S. EPA, 2012). During bloom events, *E. coli* cell counts increase from 10,000 to 100,000 cfu/100 ml of water. As

bloom strains are not faecal-associated and are likely free-living, closure of water bodies due to these elevated counts is unnecessary. A protocol that can be used by the water authorities for the detection and discrimination of *E. coli* bloom strains, would be useful.

Possession of a *Klebsiella* capsule is a key attribute of bloom strains (Power et al., 2005; Nanayakkara et al., 2019). Wyres and colleagues (2016) report 134 distinct *Klebsiella* capsule variants, which are termed capsule types. Seven different *Klebsiella* capsule types are detected among the eight *E. coli* bloom strains isolated to date (Nanayakkara et al., 2019); i.e., three capsule types are detected among the three bloom strains from the east coast, while five capsule types are identified among the five bloom strains isolated from Western Australia (Table 3.1). One bloom strain from the east coast and another bloom strain from Western Australia (WA) share the same capsule type of KL53 (Table 3.1).

The current study was carried out with the objective of designing a polymerase chain reaction (PCR) protocol to detect *Klebsiella* capsule-positive *E. coli* and discriminate strains that harbour bloom strain-associated *Klebsiella* capsule types, to differentiate bloom strains from non-bloom strains.

3.3 Materials and Methods

3.3.1 Primer design

As described in Chapter 2, 1194 bloom and non-bloom *E. coli* strains were previously screened for the *Klebsiella* capsule using Kaptive, a software tool for the rapid identification of *Klebsiella* capsule loci and their types in whole genome sequence (WGS) data (Wyres et al., 2016). The nucleotide sequences of the capsule regions of 82 (out of 1194) capsule-positive *E. coli* strains,

were extracted using Kaptive (Wyres et al., 2016), annotated using Prokka (Seemann, 2014), and the pan genome analysed using Roary (Page et al., 2015). Capsule core gene *galF* which is present in all capsule gene clusters regardless of the capsule type, was identified. Different genes are unique to different capsule types, and accordingly, genes unique to each bloom strain-associated capsule type were identified (Table 3.1). A consensus sequence for *galF* was obtained using CLC Genomics workbench v.9.5.3 (https://www.qiagenbioinformatics.com). Primers were designed targeting regions of these genes, using the web-based PrimerQuest tool in Integrated DNA Technologies (IDT; https://sg.idtdna.com/PrimerQuest/Home/Index) (Table 3.1). Overall, two primer pools were designed; the first pool targeting the capsule types of the east coast bloom strains, and the second pool targeting the capsule types of the Western Australian bloom strains. Each pool had a primer pair targeting the capsule core gene *galF* (Table 3.1).

3.3.2 PCR conditions

All PCR reactions were carried out in a volume of 20 µl. Pool 1 master mix contained 4.0 µl of 5× MyTaq Red reaction buffer (supplied with *Taq* polymerase), 0.2 µl of MyTaq HS DNA Polymerase (hot-start, 5u/µl), 0.8 µl of each primer (at 10 µM), 8.4 µl of UltraPure distilled water, and 1.0 µl of template DNA. The amounts were the same for pool 2, except that the volume of UltraPure water was 5.2 µl. The PCR reaction conditions were as follows: initialisation for 10 minutes at 94 °C; 25 cycles of 20 seconds at 94 °C for denaturation, 30 seconds at 57 °C for annealing, and 2 minutes at 68 °C for extension; followed by a final extension of 10 minutes at 72 °C, and a final hold at 10 °C. PCR cycling was carried out in a Veriti 96 well Thermal Cycler (Applied Biosystems). PCR products were loaded on 1.5% agarose

gels prepared with 1×TBE (Tris/Borate/EDTA) buffer and Ethidium bromide (2 µl/100 ml) as the nucleic acid stain, and electrophoresed in TBE buffer for 45 minutes at 150 V. Gels were imaged under ultraviolet (UV) light in a BIO-RAD Gel Doc XR+ Imaging System.

Table 3.1. Primer sequences and PCR product sizes of the *Klebsiella* capsule PCR

Bloom strain phylogroup	Capsule type	Gene target	Primer*	Primer sequences	PCR product size (bp)
Pool 1 - East coast bloom strains					
-	all	*galF*	galF.F	5'-GTGACGCACTCCTCGAA-3'	668
			galF.R	5'-CCGTAGGTGACGAAGGC-3'	
A0	KL49	*wcuI*	wcuI.F	5'-GATGTCATTTACAATGCCGTGAG-3'	542
			wcuI.R	5'-GCAGAATTTGCCCATGAGTATC-3'	
B1	KL53	*wcuE*	wcuE.F	5'-ATTGTTGAGTGGTCAGGAAGAA-3'	437
			wcuE.R	5'-GCTACTTCTGGAAGCGATGTAA-3'	
A1	KL16	*wcsT*	wcsT.F	5'-ATTGGGCCTAATTGCTTGATTG-3'	293
			wcsT.R	5'-GCCGGTACACCACCATAAATA-3'	
Pool 2 – Western Australian bloom strains					
-	all	*galF*	galF.F	5'-GTGACGCACTCCTCGAA-3'	668
			galF.R	5'-CCGTAGGTGACGAAGGC-3'	
A1	KL60	*wcqX*	wcqX.F	5'-CCTTGCCAAGCTATAGAGTTTCAT-3'	834
			wcqX.R	5'-CCCAAGCCTCATTAACGTCATC-3'	
A1	KL101	*wcuN*	wcuN.F	5'-GTTGTAGGTGGTGAGGTATTAGG-3'	558
			wcuN.R	5'-ACCACTCGCGTAACCAATAG-3'	
A1	KL53	*wcuE*	wcuE.F	5'-ATTGTTGAGTGGTCAGGAAGAA-3'	437
			wcuE.R	5'-GCTACTTCTGGAAGCGATGTAA-3'	
B1	KL31	*wctG*	wctG.F	5'-GTGTTGGGTGAAGGTGATGATA-3'	339
			wctG.R	5'-CGCTTCCTTTGCGTCTAAATC-3'	
A1	KL63	*wcsD*	wcsD.F	5'-ATACGCGGGTTGGAGAATATG-3'	231
			wcsD.R	5'-CGACTACTACCTCGTATTGGTTTAC-3'	

Seven different *Klebsiella* capsule types were present among the eight bloom strains. The capsule type KL53 occurs in two strains: one east coast strain and one Western Australian strain. Primer pairs were designed to target a region of the capsule core gene *galF* which is present in all encapsulated strains, and to target regions of genes that are unique for each capsule type of bloom strains.
* F - forward primer; R - reverse primer

3.3.3 PCR-screening of *E. coli* isolated from water

A collection of 1632 *E. coli* strains from water, comprising 998 strains isolated from water bodies in Sydney and 634 strains isolated from water bodies in Brisbane were screened for *Klebsiella* capsule and bloom strain-associated capsule types using the PCR protocol designed. DNA extracted from *E. coli* strains known to be positive for the target *Klebsiella* capsule types of each pool (bloom strains) (Table 3.1) and DNA extracted from a capsule-positive strain having a non-target capsule type (e.g. KL123 – positive for *galF*) was used as positive controls, while DNA extracted from a known capsule-negative strain (similar to *E. coli* K-12) was used as the negative control (Figure 3.1). The PCR master mix without any template DNA was used as the blank (Figure 3.1).

3.4 Results and Discussion

The PCR-based capsule status and type determinations were consistent with the *in silico* results obtained using Kaptive (Figure 3.1). The PCR yielded a band for *galF* in all known *Klebsiella* capsule-positive *E. coli* strains, regardless of the capsule type (Figure 3.1). Further, a *galF* PCR product was not observed for known capsule-negative *E. coli*, including those having capsule deletion variant KL156-D1 (Wyres et al., 2016), which is not considered a true capsule type (Nanayakkara et al., 2019). Although capsule-negative *E. coli* such as *E. coli* K-12 harbour a gene *galF* in the colonic acid gene region (Chapter 2), sequence comparisons indicate that it is variable from the *galF* of *Klebsiella* capsule. *E. coli* strains similar to *E. coli* K-12 in terms of capsule status and *galF* (based on sequence comparisons) were verified using the PCR. The PCR primers did not amplify their *galF* and confirmed that these strains were

negative for the *Klebsiella* capsule. The capsule type-specific primer pairs accurately detected the target capsule types, and did not produce bands with non-target capsule types, indicating high specificity (Figure 3.1). Overall, the PCR accurately detected all known bloom strains and other encapsulated strains, and accurately identified capsule-negative strains. The primers were highly specific and a problem of false positives did not arise within *E. coli*. Yet, the primers would amplify capsule genes of *Klebsiella* strains if present. Therefore, prior identification of the strains as *E. coli* is necessary.

galF of encapsulated strains is in the terminal part of the capsule cluster and codes for core sugar nucleotide precursors (Rahn and Whitfield, 2003; Wyres et al., 2016). The genes that are unique to particular capsule types are located in the variable central region of the *Klebsiella* capsule gene cluster and code for sugar synthesis, processing, and export proteins (Wyres et al., 2016).

Of the 1632 *E. coli* strains from water screened using PCR, a total of 65 strains (4.0%) were capsule-positive; 48 out of 998 strains (4.8%) from Sydney and 17 out of 634 strains (2.7%) from Brisbane. In contrast, the frequency of capsules in host-associated and water *E. coli* isolates collectively was approximately 7% (Chapter 2). Of the strains from Sydney, 20 capsule-positive strains (2.0%) had bloom-associated capsule types; KL63 (7 strains), KL16 (6 strains), KL53 (6 strains), and KL60 (1 strain). Eight strains (1.3%) from Brisbane harboured bloom-associated capsule types; i.e., KL53 (4 strains), KL16 (2 strains), KL31 (1 strain), and KL63 (1 strain). The overall frequency of bloom associated-capsule types among the water isolates was 1.7%. Twenty-eight strains from Sydney and nine from Brisbane had capsule types other than the bloom strain-associated capsule types targeted in the PCR protocol. As speculated previously in Chapter 2, any *E. coli* strain with a *Klebsiella* capsule, regardless of the capsule

type, may be able to produce blooms. Taken together, the PCR method detected a total of 65 (4.0%) capsule-positive *E. coli* strains that might have the ability to produce elevated counts in water, and 28 of these 65 strains (43%) had capsule types that were detected among the bloom strains isolated to date.

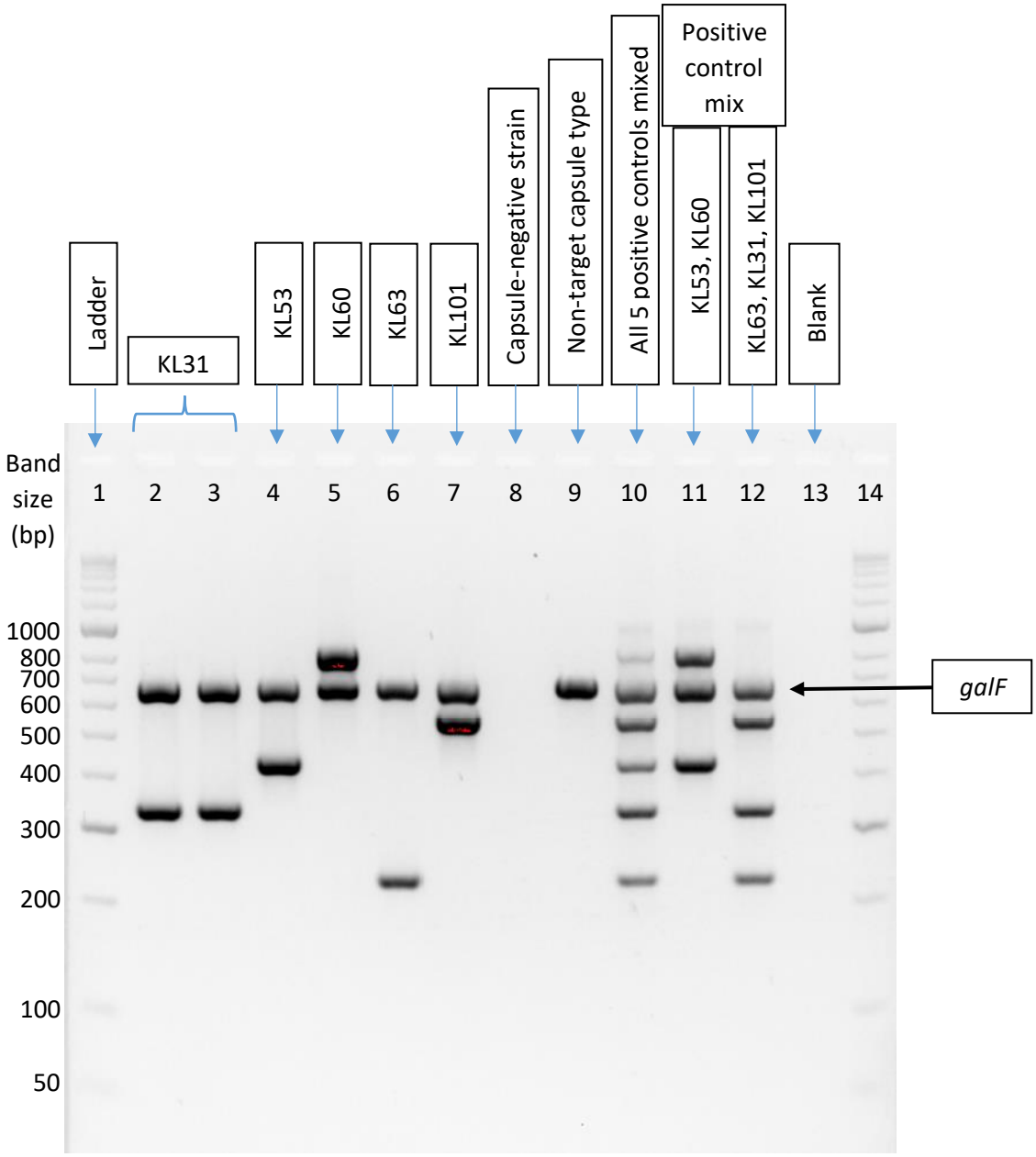


Figure 3.1. Capsule PCR products of pool 2 run on a 1.5 % agarose gel for 45 minutes at 150 V. Known test strains were run separately (lanes 2-9) and mixed together (lanes 10-12). The capsule status and type of the test strains have previously been determined using Kaptive and are consistent with the results of the PCR. For screening purposes, positive controls were mixed as shown in lanes 11 and 12. For pool 1, positive controls of KL16 and KL49 were mixed together while KL53 was used separately, considering the band sizes.

The current PCR-based protocol can be used to detect *E. coli* strains that harbour a *Klebsiella* capsule and specifically those with bloom strain-associated capsule types. The assay however was not designed for crude DNA extracts from the wells of a Colilert® tray, as it would be difficult to design primers that would not amplify capsule genes of *Klebsiella* strains, if present. The PCR protocol would be a convenient and rapid means for the water authorities to detect probable *E. coli* bloom strains. Upon enumeration by Colilert Quanti-Tray®, *E. coli* strains responsible for elevated counts can be PCR-tested for encapsulation. When elevated counts of *E. coli* are reported from recreational lakes, if the strains responsible can be rapidly identified as harmless bloom strains that are not faecal-associated, unnecessary panic and closure of water bodies can be avoided. All bloom strains isolated to date do not harbour an array of genes implicated in virulence in *E. coli* (Power et al., 2005; Chapter 5; unpublished data). However, capsules do occur in *E. coli* from a variety of vertebrates, and some of these strains may be enteric or extraintestinal pathogens. Therefore, the PCR protocol may be combined with a protocol to screen strains for virulence (Jagals et al., 2013), especially when monitoring water quality. Yet this could have limitations and complexities, for example, deciding on the range of virulence factors that must be screened to ensure validity and safety, and warrants future work to clarify this. Bloom strains may express particular virulence factors against the different organisms that are present in the bloom water microbiome, and these might have a human health significance. Hence, a thorough prior study is needed before deciding on the range of virulence factors to be screened for, in terms of human safety. This PCR protocol can be easily extended to detect more capsule types, for example, if future capsule types are found from bloom strains. It can also be extended to identify the capsule types of the 57% of *galF*-positive strains that were not given a capsule type using the current method.

3.5 References

Bitton, G. (2011) *Wastewater Microbiology*, 4th ed. Hoboken, NJ: Wiley-Blackwell.

Ishii, S., and Sadowsky, M.J. (2008) *Escherichia coli* in the environment: Implications for water quality and human health. *Microbes Environ* 23(2): 101-108.

Jagals, P., Barnard, T.G., Mokoena, M.M., Ashbolt, N., and Roser, D.J. (2013) Pathogenic *Escherichia coli* in rural household container waters. *Water Sci Technol* 67(6): 1230-1237.

Leclerc, H.D.A.A., Mossel, D.A.A., Edberg, S.C., and Struijk, C.B. (2001) Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. *Annu Rev Microbiol* 55(1): 201-234.

Nanayakkara, B.S., O'Brien, C.L., and Gordon, D.M. (2019) Diversity and distribution of *Klebsiella* capsules in *Escherichia coli*. *Environ Microbiol Rep* 11(2): 107-117.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G. et al. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22): 3691-3693.

Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., and Slade, M.B. (2005) Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. *Environ Microbiol* 7(5): 631-640.

Rahn, A., and Whitfield, C. (2003) Transcriptional organization and regulation of the *Escherichia coli* K30 group 1 capsule biosynthesis (*cps*) gene cluster. *Mol Microbiol* 47(4): 1045-1060.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14): 2068 – 2069.

U.S. EPA - United States Environmental Protection Agency. (2012) Recreational water quality criteria. Office of Water 820-F-12-058. https://www.epa.gov/sites/production/files/2015-10/documents/rwqc2012.pdf.

WHO-World Health Organization. (2017) Guidelines for drinking-water quality, Microbial aspects, 4th edn. incorporating the 1st addendum. World Health Organization, Geneva.

Wyres, K.L., Wick, R.R., Gorrie, C., Jenney, A., Follador, R., Thomson, N.R., and Holt, K.E. (2016) Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom* 2(12): 0.000102.

Chapter 4. Phenotypic characteristics contributing to the enhanced growth of *Escherichia coli* bloom strains

4.1 Abstract

During bloom events, *Escherichia coli* cell counts increase to 10,000 – 100,000 cells/100 ml of water. The strains responsible for bloom events belong to *E. coli* phylogenetic groups A and B1, and all have acquired a capsule from *Klebsiella*. A pan-genome comparison of phylogroup A *E. coli* revealed that the ferric citrate uptake system (*fecIRABCDE*) was over-represented in phylogroup A bloom strains compared to non-bloom *E. coli*.

A series of experiments was carried out to investigate if the capsule together with ferric citrate uptake system could confer a growth rate advantage on *E. coli*. *E. coli* strains representing four 'genotypes' with respect to capsule and *fec* operon; i.e. cap+fec+, cap+fec-, cap-fec+, and cap-fec- were selected. Their growth rates were compared in the presence/absence of iron (Fe) and/or citrate in media, and in four different carbon sources. Growth was also compared at low (0.01 mM) and high (10 mM) glucose concentrations and the lag phase of the strains was determined.

Encapsulated strains had a growth rate advantage regardless of the media composition and the presence/absence of the *fec* operon, and they had a shorter lag phase compared to capsule-negative strains. The results suggest that the *Klebsiella* capsule may enhance nutrient uptake or utilisation by a strain, thereby conferring a growth rate advantage on *E. coli*. Also, due to its protective role, the capsule makes it more likely that bloom strains will be present during a nutrient input event, which precedes a bloom. These may explain why it is only capsule-positive strains that produce elevated cell counts and cause bloom events.

4.2 Introduction

Bloom events are a regular occurrence in freshwater reservoirs and recreational lakes across Australia (Power et al., 2005). During bloom events, *Escherichia coli* cell counts increase to 10,000 – 100,000 cells/100 ml of water, well above the 'safe' cut-off level of 235 cfu/100 ml in a single sample of recreational water (Ishii and Sadowsky, 2008; U.S. EPA, 2012). A limited number of strains are responsible for causing blooms, and all belong to the *E. coli* phylogenetic groups A and B1. Among these, the phylogroup A strains represent the majority of bloom strains isolated to date. Strains of groups A and B1 are recognized as generalists, occur in a wide range of host and non-host environments, and are over-represented in water (Power et al., 2005; Gordon, 2013).

Bloom strains can be isolated from the water bodies even during non-bloom periods (Power et al., 2005). The highly elevated counts of bloom strains during bloom events suggests that they outcompete all other co-occurring *E. coli*. Something in common to the bloom strains is that they have all acquired a capsule originating from *Klebsiella* (Power et al., 2005; Nanayakkara et al., 2019). The *Klebsiella* capsule is an outermost layer of polysaccharide, approximately 160 nm in thickness, which covers the bacterial cell surface (Amako et al., 1988; Whitfield and Roberts, 1999). In the external environment, the capsule is believed to enhance cell persistence by providing protection from adverse conditions, including ultra violet (UV) radiation, desiccation, osmotic stress, predation by protozoa, and bacteriophage infection (Weiner et al., 1995; Rendueles et al., 2017). Curli, an amyloid fibre produced by some enterobacteria, and cellulose production, also enhance cell survival in external environments (Prigent-Combaret et al., 2000; Römling and Galperin, 2015). Bloom events are triggered by nutrient influxes to water bodies such as dust storms and the autumn die-off of aquatic

vegetation (Littlefield-Wyer, 2006). Given that the capsule enhances cell persistence, encapsulated strains are more likely to be present in water than non-encapsulated strains when a nutrient influx event occurs. A typical water sample however contains multiple genotypes of *E. coli* (Casarez et al., 2007; Higgins et al., 2007). As the capsule is not known to be metabolically active, the presence of a capsule alone does not seem to explain why only encapsulated strains, and not all co-occurring *E. coli*, increase in number in response to nutrient influxes.

Iron is a limited nutrient and is fundamentally important for bacterial growth (Neilands, 1981; Braun, 2003; Leiby et al., 2012; Braun and Hantke, 2013; Westrich et al., 2016). Iron received via Saharan dust has been implicated in causing up to a 30-fold increase in culturable *Vibrio* in the Caribbean and sub-tropical Atlantic waters (Westrich et al., 2016). A pan genome analysis of phylogroup A *E. coli* strains revealed that genes of the ferric citrate uptake system; *fecIRABCDE* (Staudenmaier et al., 1989) were present in 100% of the phylogroup A bloom strains compared to the non-bloom phylogroup A *E. coli* (<39%) (Nanayakkara et al., 2019). The ferric citrate uptake system operates aerobically for the uptake of iron via a citrate chelator, as dinuclear ferric citrate (Frost and Rosenberg, 1973; Braun, 2003; Grass, 2006). The contribution of this citrate mediated iron uptake system and the co-occurring *Klebsiella* capsule to the overall growth rate of *E. coli* has not been investigated. The combination of this iron uptake system and capsule may confer bloom strains a growth rate advantage, thereby leading to the elevated cell counts observed during bloom events.

Given that the capsule alone does not seem to explain the growth advantage of bloom strains, experiments were undertaken with the objective of exploring other phenotypic attributes and growth characteristics that may contribute to the elevated cell counts observed in bloom

events. Curli and cellulose production was also investigated, as these traits are thought to enhance cell survival in external environments.

4.3 Materials and Methods

4.3.1 Detection of curli and cellulose production using Congo red

Strain selection: Eighty-eight *E. coli* strains belonging to phylogroups A, B1, and C were used to investigate the curli/cellulose production phenotype. Among these strains, 82 were positive for the *Klebsiella* capsule and included 30 bloom strains, while six strains were negative for the capsule. The capsule status of the strains was previously determined using the *Klebsiella* capsule detection software tool Kaptive (Wyres et al., 2016).

Congo red is an amyloid-binding dye that binds to curli and cellulose. Colonies that co-express curli and cellulose deplete Congo red in the media and consequently stain red (Westermark et al., 1999; Römling, 2005; Reichhardt et al., 2015). Congo red indicator plates were made using lysogeny agar (tryptone 10 g/l, yeast extract 5 g/l, agar 13 g/l) without NaCl, followed by the addition of Congo red (40 μg/ml) (Römling et al., 1998). Overnight cultures of the strains in lysogeny broth were diluted in sterile normal saline (NaCl 0.85% w/v) and aliquots were spread in duplicate on Congo red agar plates. One set of plates was incubated at 37 ^{o}C, while the other set was incubated at room temperature. Colony morphology was observed regularly for eight days.

Calcofluor (fluorescent brightener 28) is a fluorescent dye that stains cellulose. Cellulose-producing cells grown in the presence of calcofluor, fluoresce under UV light of 366 nm (Zogaj et al., 2001). The cellulose detection media was prepared by adding Calcofluor (0.02%) (Da Re

and Ghigo, 2006; Uhlich et al., 2006) and 1 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (Da Re and Ghigo, 2006) to sterile lysogeny broth (LB) agar. Overnight cultures of LB were diluted in sterile normal saline (0.85% w/v), and a 10 µl aliquot of the 10^{-6} dilution together with 100 µl of sterile normal saline was spread on Calcofluor agar plates, then incubated at 37 ^{o}C for 96 h. The colonies were observed for cellulose production by exposure to UV light in a UVP Ultraviolet Transilluminator, using a constant exposure time for every strain. Fluorescence indicates cellulose production by a strain.

4.3.2 Growth rate in the presence/absence of iron and/or citrate

Strain selection: Assuming that the *fecIRABCDE* operon in conjunction with the *Klebsiella* capsule might confer a growth advantage on *E. coli*, a collection of 24 phylogroup A *E. coli* strains which were either *Klebsiella* capsule-positive or -negative (cap+/cap-) and either did or did not carry genes for the *fec* operon (fec+/fec-) were used for the experiments described below. The strains represented four 'genotypes': cap+fec+, cap+fec-, cap-fec+, and cap-fec-, and each genotype was represented by six strains. The selected strains have a diversity of sequence type (ST)-capsule type combinations, and were isolated from a variety of sources (Table 4.1).

Davis minimal media without citrate was prepared using potassium phosphate (dibasic trihydrate) (7 g/l), potassium phosphate (monobasic anhydrous) (2 g/l), ammonium sulfate (1 g/l), magnesium sulfate (0.1 g/l), and thiamine (0.002 g/l) (Leiby et al., 2012). Glucose (1 mM) was used as the carbon source. Sodium citrate tribasic dihydrate equivalent to 1.7 mM disodium citrate was added to prepare minimal media containing citrate (cit+). Fe(II) was

supplemented immediately before the experiment, from ammonium iron(II) sulfate to achieve a final Fe(II) concentration of 10 µM (iron+). Four different media; iron+cit+, iron+cit-, iron-cit+, and iron-cit- were prepared.

The strains were dilution streaked on MacConkey agar from -80 ^{o}C freezer cultures and incubated overnight at 37 ^{o}C. An isolated colony was inoculated into lysogeny broth, incubated overnight at 37 ^{o}C with shaking at 170 r.p.m., then diluted in sterile normal saline (0.85% w/v). For acclimation, an aliquot (100 µl) was inoculated into Davis minimal media without citrate and incubated for 24 h at 37 ^{o}C with shaking. A 10 µl aliquot of the diluted acclimation culture (10^4 cells/ml) was added to the wells of a microtitre plate, each containing 190 µl of one of the four media, so as to achieve a starting cell density of approximately 500 cells/ml. The cell density was confirmed by spread plating. These cultures were incubated at 37 ^{o}C in a PowerWave$_x$ 340 (BIO-TEK INSTRUMENTS, INC.) plate reader, which automatically records the optical density (OD) at 600 nm every 10 minutes, with the plate being shaken at high intensity for one minute prior to every read. Incubation was carried out until stationary phase was reached. In a microtitre plate, each media sample contained one example of each strain and the experiment was repeated four times. The maximum growth rate was estimated by regressing the natural log of the OD readings against time. The 5-6 consecutive data points that gave the maximum slope were used to calculate the maximum growth rate.

The same experiment described above was carried out after removing iron from the media, by adding the iron chelator 2,2'-Dipyridyl (250 µM) (Massé and Gottesman, 2002). This was to determine if any growth rate advantage would be mitigated in the absence of iron, or persist.

Table 4.1. Characteristics of the *E. coli* strains used for the growth experiments

Strain	Capsule status	Capsule type	*fec* operon	Source of isolate	Sequence Type	Serogroup
Hunter	Positive	KL49	Positive	Bloom	609	O9
H288	Positive	KL31	Positive	Human	10	O8
E258	Positive	KL16	Positive	Bloom	10	O89
H645	Positive	KL123	Positive	Human	227	O9
2099_1.1	Positive	KL101	Positive	Bloom	227	O9
H331	Positive	KL127	Positive	Human	4238	O9
64_3_AC10	Positive	KL31	Negative	Human	10	O8
H446	Positive	KL39	Negative	Human	46	O9
64_6_TC7	Positive	KL31	Negative	Human	6881	O8
H474	Positive	KL127	Negative	Human	10	O9
E7112	Positive	KL25	Negative	Water	361	O9
A1_25	Positive	KL57	Negative	Poultry meat	6050	None
2H_157_1	Negative	-	Positive	Human	10	O12
45_2_HU10	Negative	-	Positive	Human	398	O11
B1893	Negative	-	Positive	Bird	6927	O109
C1_10	Negative	-	Positive	Poultry meat	10	O49
B1867	Negative	-	Positive	Bird	1303	None
E4173	Negative	-	Positive	Water	10	O131
2H_255_4	Negative	-	Negative	Human	6929	O132
H034	Negative	-	Negative	Human	409	O75
H656	Negative	-	Negative	Human	398	None
B1820	Negative	-	Negative	Bird	48	None
67_4_Ti7	Negative	-	Negative	Human	216	None
E7285	Negative	-	Negative	Water	48	None

4.3.3 Presence/absence of genes involved in other major iron uptake systems

E. coli strains carry several other iron uptake systems (Grass, 2006) including enterobactin (Langman et al., 1972; Ozenberger et al., 1987; Braun, 2003; Grass, 2006), aerobactin (de Lorenzo et al., 1986; Braun, 2003), salmochelin (Hantke et al., 2003), yersiniabactin (Braun, 2003; Bultreys et al., 2006), and systems for Fe(II) uptake (Braun, 2003; Lau et al., 2015). The presence/absence of these iron uptake systems in whole genomes sequences of the strains used for the iron/citrate assays was investigated using progressiveMauve genome alignments

(Darling et al., 2004) and the MicroScope platform (https://www.genoscope.cns.fr/agc/microscope/home/).

4.3.4 Growth rate in carbon sources that differ in their uptake mechanism

The results of the growth experiments with or without iron indicated a growth rate advantage for the encapsulated strains compared to capsule-negative strains. An experiment using four sugars that differ from each other in their uptake mechanism through the bacterial outer membrane (porins OmpF and LamB) and inner membrane (phosphoenolpyruvate: carbohydrate phosphotransferase system-PTS and non-PTS), was performed to determine if the encapsulated strains had an enhanced nutrient uptake system (Travisano and Lenski, 1996). Mannitol and glycerol diffuse through the outer membrane porin OmpF (Travisano and Lenski, 1996), while maltose and trehalose enter through the porin LamB (Nikaido and Vaara, 1985; Benz et al., 1986; Klein and Boos, 1993). Subsequently, mannitol and trehalose cross the inner membrane and enter the cytoplasm via the PTS system (Erni, 1989; Boos et al., 1990; Deutscher et al., 2006), while glycerol and maltose enter the cytoplasm via the non-PTS system (Postma et al., 1993; Deutscher et al., 2006).

A single colony was taken from MacConkey agar and inoculated into Davis minimal media (10.6 g/l) having glucose (1 mM) and thiamine (0.002 g/l). These acclimation cultures were incubated for 24 h at 37 ^{o}C with shaking at 170 r.p.m. Davis minimal media each containing one of the four sugars; mannitol, glycerol, maltose, or trehalose (180 mg/l), was added to wells of a microtitre plate in 190 µl aliquots. The acclimation cultures were diluted to a cell density of 10^4 cells/ml in sterile normal saline (0.85 % w/v) and 10 µl was used to inoculate the media

in the microtitre plate, so as to have a starting cell density of approximately 500 cells/ml. Incubation was done in an automated PowerWave$_x$ 340 plate reader at 37 ^oC where the OD of the cultures was measured at 600 nm every 10 minutes. The plate was shaken at high intensity before each read and incubation continued until stationary phase was reached. The maximum growth rates were calculated as explained previously.

4.3.5 Growth at low and high glucose concentrations

The following experiment was conducted to determine if encapsulated strains have a growth advantage in either low (0.01 mM) or high (10 mM) glucose concentration. Strains were grown for a prescribed amount of time and then cell counts were determined.

Minimal media containing a low glucose concentration (0.01 mM) was used for inoculating the strains and the cell densities at time (t) = 0 were determined. The cultures were incubated for 5.5 h at 37 ^oC with shaking, then the final cell densities were calculated. A lag phase of 0.5 h was assumed based on prior background research, when determining the incubation time. The incubation time was adjusted to 3 h for high glucose concentration (10 mM) inoculum, taking the faster growth rate into account. The final cell density (N_f) was compared between genotypes, using the starting cell density (N_0) as a covariate.

4.3.6 Length of lag phase in 10 mM glucose

The lag time of the strains in 10 mM glucose was determined as follows: Davis minimal media with glucose (1 mM) was inoculated with a single colony followed by incubation for 24 h at 37

^{o}C with shaking. A 10 µl aliquot of the 24 h culture was used to inoculate each well of a microtitre plate containing 190 µl of Davis minimal media having glucose (10 mM). The microtitre plate was incubated in an automated PowerWave$_x$ 340 plate reader at 37 ^{o}C and OD readings at 600 nm were taken every 5 minutes. The plate was shaken at high intensity for one minute before every read. The logarithm of the OD readings was plotted against time to obtain the growth curves. The lag time was the time interval before the first two consecutive increases in OD, as obtained from the growth curve. The experiment was replicated four times.

4.3.7 Statistical analysis

Statistical analyses were performed using JMP 12.2.0 software (2015 SAS Institute, Inc.). Analysis of variance (ANOVA) was conducted with a full factorial model, with genotype, media, and interaction as model effects, and technical replicates were incorporated as a random effect. The final cell density (N_f) at low (0.01 mM) and high (10 mM) glucose concentrations was analysed by genotype, using the starting cell density (N_0) as a covariate.

4.4 Results

4.4.1 Curli and cellulose production

The following colony morphotypes are characteristic of curli/cellulose producers when incubated on agar containing Congo red; production of both curli and cellulose – red, dry, and rough (rdar); curli alone – brown, dry, and rough (bdar); cellulose alone – pink, dry, and rough

(pdar); and neither curli nor cellulose – smooth and white (saw) (Zogaj et al., 2001; Solano et al., 2002; Römling, 2005; Uhlich et al., 2006).

We observed that the capsule hindered staining of curli/cellulose by Congo red, and masked the morphotype of the colonies. Therefore, the Congo red assay could not be used to determine curli or cellulose production by encapsulated strains. Among the strains incubated at 37 ^{o}C, 57% had a highly mucoid phenotype and included bloom and non-bloom encapsulated strains. Fewer strains (35%) were dry and included both encapsulated and capsule-negative strains. When the same strains were incubated at room temperature, a majority (64%) exhibited a highly mucoid phenotype. Most, but not all, bloom strains were in this group, which also included non-bloom encapsulated strains. A considerable number of strains (32%), regardless of being encapsulated or not, had a dry colony morphology. On Congo red agar, all strains that produced a mucoid phenotype were encapsulated, but not all encapsulated strains produced a mucoid phenotype. Overall, the bloom strains could not be distinguished from encapsulated non-bloom strains using colony morphology. In the Calcofluor assay, all strains but two produced fluorescence under UV light, indicating that cellulose production was not distinctive of the bloom strains.

4.4.2 The effect of capsule, *fec* operon, iron, and citrate on the growth rate

The encapsulated strains had a mean growth rate that was significantly higher than that of both capsule-negative genotypes, and grew better than the capsule-negative strains in all media (genotype, $p<0.0001$; media, $p<0.0001$; interaction, $p = 0.002$) (Figure 4.1). Overall, the mean growth rate was significantly higher in media with both iron and citrate [iron+cit+], and

neither iron nor citrate [iron-cit-], compared to media having only one component. However, the growth rate of the cap+fec+ strains was more or less similar independent of media, and this caused the significant interaction effect.

When dipyridyl was added to the media an overall decline in the growth rate was observed, with the exception of cap+fec- strains (genotype, $p<0.0001$; media, $p = 0.01$; interaction, $p = 0.60$) (Figure 4.2). Across media, the cap+fec- strains had a significantly higher growth rate compared to the other genotypes. On average, growth was better in media with iron present than when it was absent, indicating that dipyridyl had not completely eliminated iron availability. This may have been caused by an insufficiency in the amount of 2,2'-Dipyridyl used, or due to its slow binding of iron, or due to an unpredicted instability of the iron-dipyridyl complex. As was observed in the first experiment, the growth rate of the cap+fec+ strains did not vary with media.

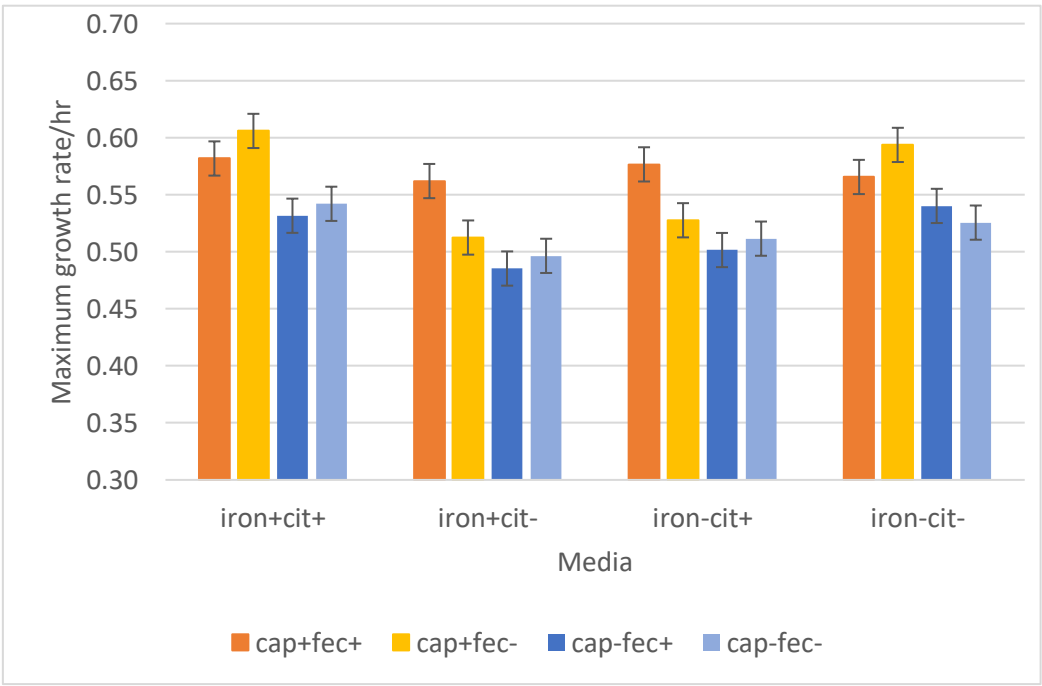


Figure 4.1. Growth rate of the strains of the four capsule/*fec* genotypes in four different media with or without iron and/or citrate. n = 24; Mean ± SE.

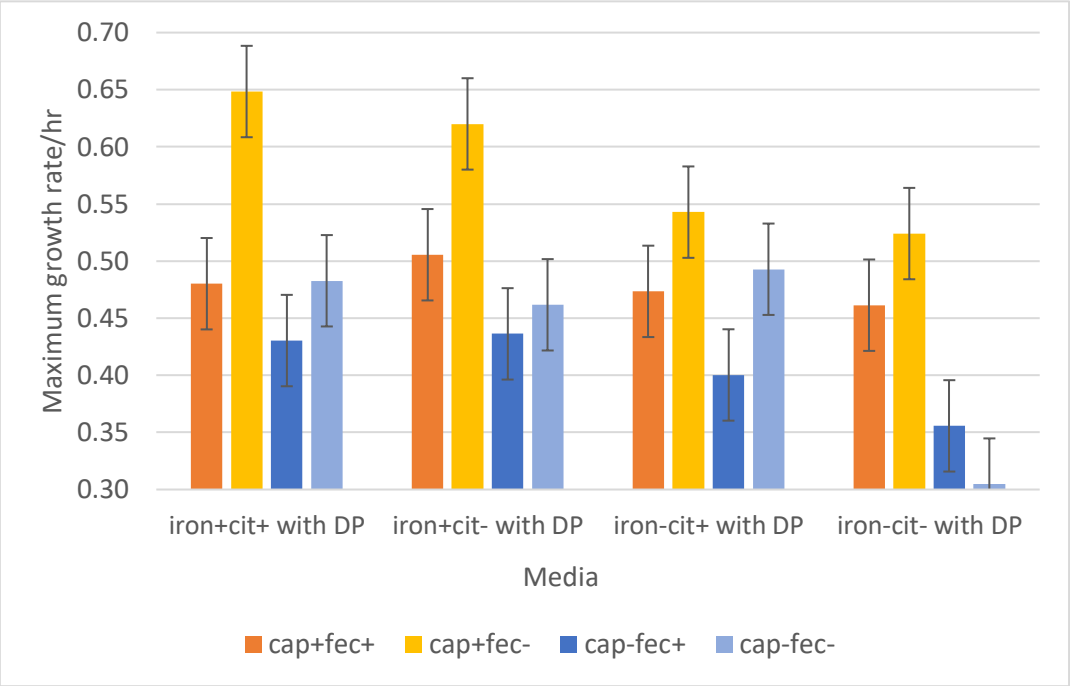


Figure 4.2. Growth rate of the strains of the four capsule/*fec* genotypes in media with reduced iron. DP stands for 2,2'-Dipyridyl. n = 6; Mean ± SE.

4.4.3 Presence/absence of other major iron uptake systems

All strains were positive for the Fe(II) uptake systems encoded by *feoABC* and *efeUOB*. The enterobactin biosynthesis and uptake system was present in all encapsulated strains, while it was present in 92% (11 out of 12 strains) of the capsule-negative strains. Aerobactin and salmochelin were present in 17%, and yersiniabactin in 25% of the encapsulated strains. Among capsule-negative strains, 17% had aerobactin, while 33% carried genes for yersiniabactin. Salmochelin was absent among the capsule-negative strains (Supplemental Table 4.1).

4.4.4 Growth rate in different carbon sources

The growth rate of the strains was highest in trehalose, followed by maltose, with the exception of cap-fec- strains. The strains exhibited the lowest growth rates in mannitol and glycerol (p<0.0001). Except for mannitol, the encapsulated strains grew better compared to the capsule-negative strains (Figure 4.3). In trehalose, the maximum growth rate of the capsule-positive strains was markedly higher compared to the capsule-negative strains, suggesting that harbouring the capsule increases the growth rate in trehalose.

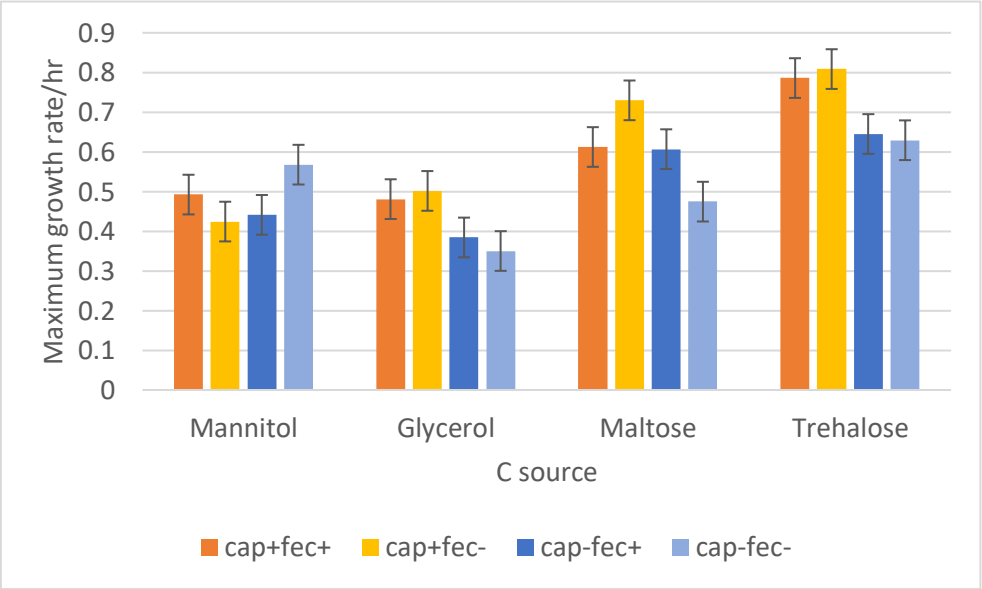


Figure 4.3. Growth rate of the strains of the four capsule/*fec* genotypes in four different C sources. n = 6; Mean ± SE.

4.4.5 Growth at low and high glucose concentrations

At both low (0.01 mM) and high (10 mM) glucose concentrations, the cap+fec+ strains grew better than the other genotypes (Figures 4.4 and 4.5). At both glucose concentrations, the genotype effect was significant, while the initial cell density N_0 did not have a significant effect on the outcomes (0.01 mM glucose: genotype, p = 0.03; N_0, p = 0.2; interaction, p = 0.3; 10 mM glucose: genotype, p = 0.03; N_0, p = 0.4; interaction, p = 0.4). In 0.01 mM glucose, the cap+fec+ strains grew significantly better than the cap-fec- strains, but the growth of the cap+fec- and cap-fec+ strains did not differ significantly from the other genotypes. In 10 mM glucose, the cap+fec+ strains grew significantly better than the cap-fec+ strains, yet the growth of the cap+fec- and cap-fec- strains did not differ significantly from the others. Overall, the cap+fec+ strains grew better than the others, regardless of the glucose concentration.

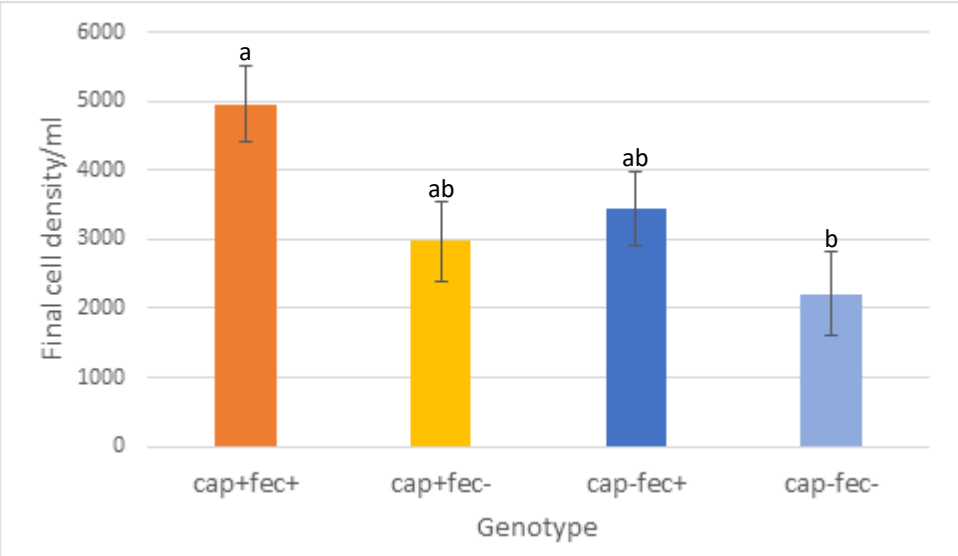


Figure 4.4. Final cell density of the four capsule/*fec* genotypes in minimal media with 0.01 mM glucose. n = 6; Mean ± SE. Values not connected by the same letter are significantly different.

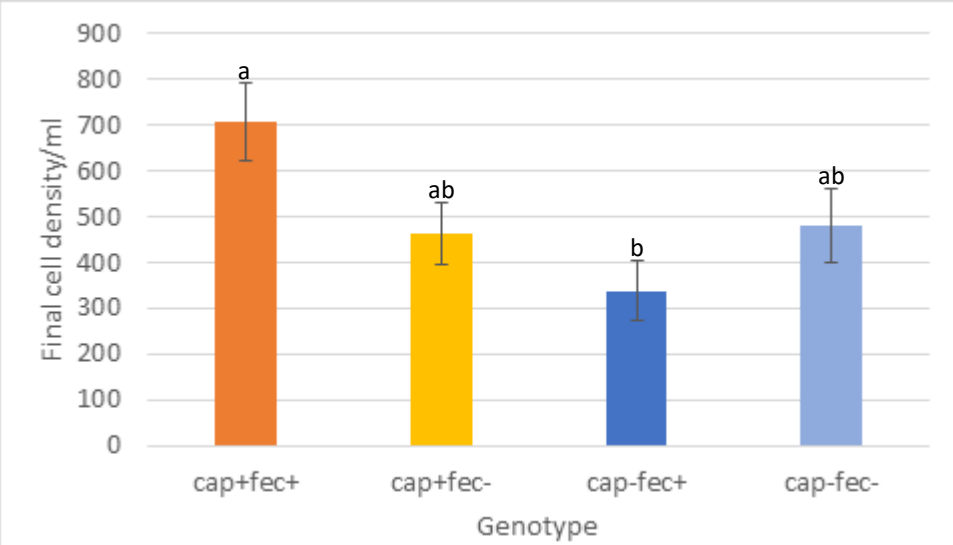


Figure 4.5. Final cell density of the four capsule/*fec* genotypes in minimal media with 10 mM glucose. n = 6; Mean ± SE. Values not connected by the same letter are significantly different.

4.4.6 Length of lag phase in 10 mM glucose

The encapsulated strains had a lag phase that was shorter than that of the capsule-negative strains (Figure 4.6). The genotype effect was significant (p<0.0001). The cap+fec+ genotype had the shortest lag phase followed by the cap+fec- strains, while the cap-fec- genotype had the longest lag phase.

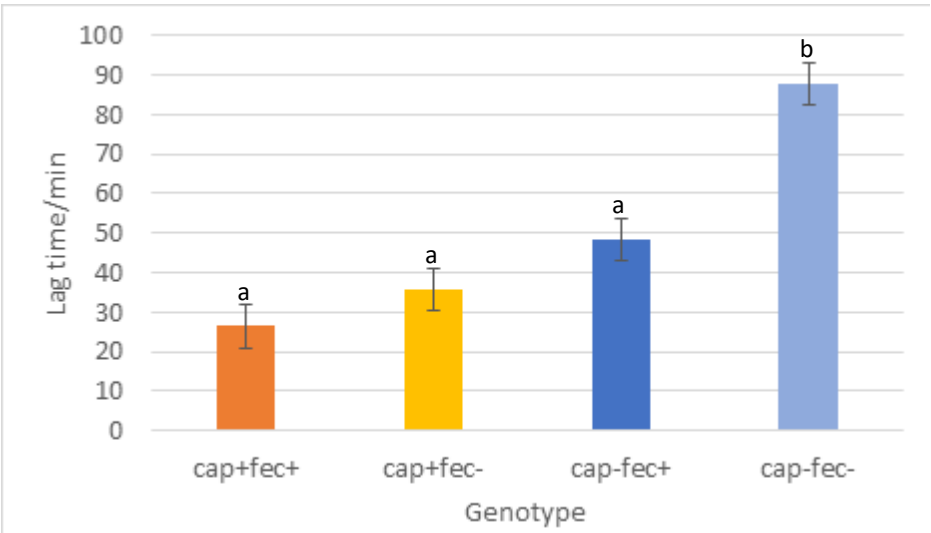


Figure 4.6. Lag time of the four capsule/*fec* genotypes in minimal media with 10 mM glucose. n = 24; Mean ± SE. Values not connected by the same letter are significantly different.

4.5 Discussion

We show that encapsulated strains are at a growth rate advantage relative to non-encapsulated strains regardless of the media composition, and that they have a shorter lag phase.

Whole genome sequence comparisons revealed that the ferric citrate uptake system *fecIRABCDE* is over-represented in the phylogroup A bloom strains (100%) compared to phylogroup A non-bloom *E. coli* (<39%) (Nanayakkara et al., 2019). The ferric citrate system acquires Fe(III) that is bound to a citrate chelator, which shuttles iron to the FecA receptor without citrate entering the cell (Staudenmaier et al., 1989; Braun, 2003; Grass, 2006). Hence, the functioning of the *fec* operon requires both Fe(III) and citrate to be present in the culture media. In the absence of iron, citrate is presumed to have no effect on growth. This is because most *E. coli* are unable to utilise citrate aerobically as a sole carbon source due to the absence of an aerobically-functioning transport system (Koser, 1924; Lara and Stokes, 1952; Hall, 1982).

In both experiments involving iron and citrate, on average the encapsulated strains grew better compared to the capsule-negative strains, regardless of the presence or absence of the *fec* operon. The growth rate of the cap+fec+ strains did not vary with the media (Figures 4.1 and 4.2). The cap+fec- strains and the capsule-negative genotypes grew better when both iron and citrate were present and when neither iron nor citrate was present compared to when one component was absent (Figure 4.1). The results of these experiments show that the subtraction of iron using dipyridyl reduces the growth rate across genotypes except for cap+fec-, proving that iron enhances bacterial growth, as reported previously (Leiby et al., 2012; Braun and Hantke, 2013; Westrich et al., 2016); however, the results do not suggest

that the presence of the *fec* operon plays a key role in determining the growth rate. Rather, the dominant growth response appears to have been caused by the capsule and not by the *fec* operon, iron or citrate. The over-representation of the *fecIRABCDE* genes in the phylogroup A bloom strains might merely be due to chance, and the absence of the *fec* operon in the phylogroup B1 bloom strain further validates this.

The strains used in the iron/citrate assays carry genes for several iron uptake systems including the siderophore enterobactin. Multiple iron uptake systems can operate simultaneously (Frost and Rosenberg, 1973; Braun and Hantke, 2013), and certain proteins are shared among different iron uptake systems, for instance, degraded iron-salmochelin can be transported via enterobactin-related FepB, FepD, FepG, and FepC (Hantke et al., 2003). The variable presence of multiple other iron uptake systems in the strains complicates this study, which focuses solely on the ferric citrate system. Despite the presence of other iron uptake systems, the capsule-positive strains were shown to have a growth rate advantage relative to the capsule-negative strains, and that was largely independent of the presence or absence of iron.

Encapsulated strains had an enhanced growth rate in trehalose compared to non-encapsulated strains. The induction of the outer membrane porin LamB (λ receptor) is crucial for the efficient uptake of trehalose into the cell (Klein and Boos, 1993), which subsequently passes through the PTS system (Boos et al., 1990; Travisano and Lenski, 1996). In the non-PTS sugar maltose, which is also taken through LamB (Nikaido and Vaara, 1985; Benz et al., 1986), the encapsulated strains do not out-perform the capsule-negative strains to the same extent as in trehalose. We speculate that the enhanced growth in trehalose may be due to the interplay of LamB and the PTS system. Overall, our data suggest that encapsulated strains have

an enhanced trehalose-related nutrient uptake and/or utilisation system, but further studies are required to confirm this.

Glucose concentrations of 1 mM resulted in cell densities of around 10^8 cells/ml whereas *E. coli* counts during bloom events are 10^2 – 10^3 cells/ml of water. Therefore, the nutrient concentrations needed to obtain the cell densities observed in bloom events would be very low. This suggests that bloom strains grow better at low nutrient concentrations. Experimentally, the growth advantage of the cap+fec+ strains was observed at both low (0.01 mM) and high (10 mM) glucose concentrations. However, of the two encapsulated genotypes, only the cap+fec+ strains but not the cap+fec- strains grew well at these glucose concentrations, which could imply an effect of the *fec* operon. However, as iron was not added to the minimal media used in this experiment, the *fec* operon cannot be expected to affect growth.

All *E. coli* strains associated with bloom events are *Klebsiella* capsule-positive, while the frequency of *Klebsiella* capsules in *E. coli* overall is only 7% (Nanayakkara et al., 2019). This means that the *Klebsiella* capsule is a vital attribute of bloom strains. Rendueles and colleagues (2017) also revealed that capsules are more likely to occur in free-living species than in pathogens, suggesting a significant role of capsules in bacterial survival outside a host. From an ecological perspective, the possession of a capsule is believed to enhance strain persistence by providing protection from desiccation, osmotic stress, UV radiation, and protozoal predation (Weiner et al., 1995; Rendueles et al., 2017). Due to these protective features, it is likely that encapsulated strains are more likely to be present in water when nutrient influx events, such as dust storms, occur. Nutrient influx events are known to precede bloom events. However, a typical water sample contains multiple genotypes of *E. coli* (Casarez

et al., 2007; Higgins et al., 2007), and it might be expected that all strains present should respond to a nutrient influx event. However, it is only the capsule-positive strains that produce elevated counts in response to nutrient influxes. The results of the present study indicate that the capsule-positive strains are better able to take advantage of nutrient influxes because they have shorter lag phases which means that they start dividing earlier, and likely an enhanced nutrient uptake system, leading to higher growth rates than non-encapsulated strains.

The *Klebsiella* capsule region is made up of approximately 14-25 genes, but none of these genes are thought to be actively involved in metabolism. Why then does the presence of the capsule region confer a growth advantage to these strains? Perhaps capsule-associated genes are involved in increasing the rate of cell division. In studies by Gervais and colleagues (1992) and Carballès and colleagues (1999), the capsule synthesis regulator *rcsB* was implicated in the activation of *ftsZ* which controls cell division in *E. coli*. In another study, the capsule core gene product GalF via its interaction with GalU, led to increased cellular levels of UDP-Glucose (Marolda and Valvano, 1996). UDP-Glucose serves as a precursor for carbohydrate biosynthesis and hence may lead to the production of energy (Marolda and Valvano, 1996; Berbís et al., 2015; Ebrecht et al., 2015). Amako and colleagues (1988) showed that the *Klebsiella* capsule is thicker (160 nm) than the *E. coli* K1 capsule which is less than 10 nm in thickness. The same study demonstrated that the *Klebsiella* capsule has a denser arrangement of fine fibres compared to the *E. coli* K1 capsule (Amako et al., 1988). Could the structure of the *Klebsiella* capsule be key to the growth advantage, by being more conducive to nutrient uptake? Is it possible that nutrients are stored in the capsule? Further studies are required to elucidate the role of the *Klebsiella* capsule, to explore its involvement in nutrient uptake,

cellular metabolism, and overall growth promotion. Investigation of the transcriptional responses under similar growth conditions would help understand what genes confer the decreased lag time and enhanced growth of encapsulated strains.

4.6 References

Amako, K., Meno, Y., and Takade, A. (1988) Fine structures of the capsules of *Klebsiella pneumoniae* and *Escherichia coli* K1. *J Bacteriol* 170(10): 4960-4962.

Benz, R., Schmid, A., Nakae, T., and Vos-Scheperkeuter, G.H. (1986) Pore formation by LamB of *Escherichia coli* in lipid bilayer membranes. *J Bacteriol* 165(3): 978-986.

Berbís, M.A., Sánchez-Puelles, J.M, Cañada, F.J., and Jiménez-Barbero, J. (2015) Structure and function of prokaryotic UDP-glucose pyrophosphorylase, a drug target candidate. *Curr Med Chem* 22(14): 1687-1697.

Boos, W., Ehmann, U., Forkl, H., Klein, W., Rimmele, M., and Postma, P. (1990) Trehalose transport and metabolism in *Escherichia coli*. *J Bacteriol* 172(6): 3450-3461.

Braun, V. (2003) Iron uptake by *Escherichia coli*. *Front Biosci* 8: s1409-1421.

Braun, V., and Hantke, K. (2013) The tricky ways bacteria cope with Iron limitation. In *Iron uptake in bacteria with emphasis on E. coli and Pseudomonas*. Chakraborty, R., Braun, V., Hantke, K., and Cornelis, P. (eds). Dordrecht: Springer, pp. 31-66.

Bultreys, A., Gheysen, I., and de Hoffmann, E. (2006) Yersiniabactin production by *Pseudomonas syringae* and *Escherichia coli*, and description of a second yersiniabactin locus evolutionary group. *Appl Environ Microbiol* 72(6): 3814-3825.

Carballès, F., Bertrand, C., Bouché, J-P., and Cam, K. (1999) Regulation of *Escherichia coli* cell division genes *ftsA* and *ftsZ* by the two-component system *rcsC–rcsB*. *Mol Microbiol* 34(3): 442 – 450.

Casarez, E.A., Pillai, S.D., and Di Giovanni, G.D. (2007) Genotype diversity of *Escherichia coli* isolates in natural waters determined by PFGE and ERIC-PCR. *Water Res* 41(16): 3643-3648.

Da Re, S., and Ghigo, J-M. (2006) A CsgD-independent pathway for cellulose production and biofilm formation in *Escherichia coli. J Bacteriol* 188(8): 3073-3087.

Darling, A.C.E., Mau. B., Blattner, F.R., and Perna. N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14(7): 1394-1403.

de Lorenzo, V., Bindereif, A., Paw, B.H., and Neilands, J.B. (1986) Aerobactin biosynthesis and transport genes of plasmid ColV-K30 in *Escherichia coli* K-12. *J Bacteriol* 165(2): 570-578.

Deutscher, J., Francke, C., and Postma, P.W. (2006) How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol Mol Biol Rev* 70(4): 939-1031.

Ebrecht, A.C., Orlof, A.M., Sasoni, N., Figueroa, C.M., Iglesias, A.A., and Ballicora, M.A. (2015) On the ancestral UDP-Glucose pyrophosphorylase activity of *galF* from *Escherichia coli*. *Front Microbiol* 6: 1253.

Erni, B. (1989) Glucose transport in *Escherichia coli*. *FEMS Microbiol Lett* 63: 13-24.

Frost, G.E., and Rosenberg, H. (1973) The inducible citrate-dependent iron transport system in *Escherichia coli* K12. *Biochim Biophys Acta Biomembr* 330 (1): 90-101.

Gervais, F.G., Phoenix, P., and Drapeau, G.R. (1992) The *rcsB* gene, a positive regulator of colanic acid biosynthesis in *Escherichia coli*, is also an activator of *ftsZ* expression. *J Bacteriol* 174(12): 3964-3971.

Gordon, D.M. (2013) The ecology of *Escherichia coli*. In *Escherichia coli: Pathotypes and principles of pathogenesis.* Donnenberg, M.S. (ed). San Diego, CA: Academic Press, pp. 3-20.

Grass, G. (2006) Iron transport in *Escherichia coli*: all has not been said and done. *Biometals* 19(2): 159-172.

Hall, B.G. (1982) Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. *J Bacteriol* 151(1): 269-273.

Hantke, K., Nicholson, G., Rabsch, W., and Winkelmann, G. (2003) Salmochelins, siderophores of *Salmonella enterica* and uropathogenic *Escherichia coli* strains, are recognized by the outer membrane receptor IroN. *Proc Natl Acad Sci USA* 100(7): 3677-3682.

Higgins, J., Hohn, C., Hornor, S., Frana, M., Denver, M., and Joerger, R. (2007) Genotyping of *Escherichia coli* from environmental and animal samples. *J Microbiol Methods* 70(2): 227-235.

Ishii, S., and Sadowsky, M.J. (2008) *Escherichia coli* in the environment: Implications for water quality and human health. *Microbes Environ* 23(2): 101-108.

Klein, W., and Boos, W. (1993) Induction of the lambda receptor is essential for effective uptake of trehalose in *Escherichia coli. J Bacteriol* 175(6): 1682-1686.

Koser, S.A. (1924) Correlation of citrate utilization by members of the colon-aerogenes group with other differential characteristics and with habitat. *J Bacteriol* 9(1): 59-77.

Langman, L., Young, I.G., Frost, G.E., Rosenberg, H., and Gibson, F. (1972) Enterochelin system of iron transport in *Escherichia coli*: mutations affecting ferric-enterochelin esterase. *J Bacteriol* 112(3): 1142-1149.

Lara, F.J.S., and Stokes, J.L. (1952) Oxidation of citrate by *Escherichia coli*. *J Bacteriol* 63(3): 415-420.

Lau, C.K., Krewulak, K.D., and Vogel, H.J. (2015) Bacterial ferrous iron transport: the Feo system. *FEMS Microbiol Rev* 40(2): 273-298.

Leiby, N., Harcombe, W.R., and Marx, C.J. (2012) Multiple long-term, experimentally-evolved populations of *Escherichia coli* acquire dependence upon citrate as an iron chelator for optimal growth on glucose. *BMC Evol Biol* 12(151): 1-9.

Littlefield-Wyer, J.G. (2006) Genetic and phenotypic characterisation of *Escherichia coli* isolated from the primary and secondary habitat (PhD Thesis). Canberra, ACT, Australia: The Australian National University.

Marolda, C.L., and Valvano, M.A. (1996) The *GalF* protein of *Escherichia coli* is not a UDP-glucose pyrophosphorylase but interacts with the *GalU* protein possibly to regulate cellular levels of UDP-glucose. *Mol Microbiol* 22(5): 827-840.

Massé, E., and Gottesman, S. (2002) A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci USA* 99(7): 4620-4625.

Nanayakkara, B.S., O'Brien, C.L., and Gordon, D.M. (2019) Diversity and distribution of *Klebsiella* capsules in *Escherichia coli*. *Environ Microbiol Rep* 11(2): 107-117.

Neilands, J.B. (1981) Iron absorption and transport in microorganisms. *Annu Rev Nutr* 1(1): 27-46.

Nikaido, H., and Vaara. M. (1985) Molecular basis of bacterial outer membrane permeability. *Microbiol Rev* 49(1): 1-32.

Ozenberger, B.A., Nahlik, M.S., and McIntosh, M.A. (1987) Genetic organization of multiple *fep* genes encoding ferric enterobactin transport functions in *Escherichia coli*. *J Bacteriol* 169(8): 3638-3646.

Postma, P.W., Lengeler, J.W., and Jacobson, G.R. (1993) Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev* 57(3): 543-594.

Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., and Slade, M.B. (2005) Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. *Environ Microbiol* 7 (5): 631-640.

Prigent-Combaret, C., Prensier, G., Le Thi, T.T., Vidal, O., Lejeune, P., and Dorel, C. (2000) Developmental pathway for biofilm formation in curli-producing *Escherichia coli* strains: role of flagella, curli and colanic acid. *Environ Microbiol* 2(4): 450-464.

Reichhardt, C., Jacobson, A.N., Maher, M.C., Uang, J., McCrate, O.A., Eckart, M., and Cegelski, L. (2015) Congo red interactions with curli-producing *E. coli* and native curli amyloid fibers. *PloS One* 10(10): e0140388.

Rendueles, O., Garcia-Garcerà, M., Néron, B., Touchon, M., and Rocha, E.P.C. (2017) Abundance and co-occurrence of extracellular capsules increase environmental breadth: Implications for the emergence of pathogens. *PLoS Pathog* 13(7): 1-30.

Römling, U. (2005) Characterization of the rdar morphotype, a multicellular behaviour in *Enterobacteriaceae*. *Cell Mol Life Sci* 62(11): 1234-1246.

Römling, U., and Galperin, M.Y. (2015) Bacterial cellulose biosynthesis: diversity of operons, subunits, products, and functions. *Trends Microbiol* 23(9): 545-557.

Römling, U., Sierralta, W.D., Eriksson, K., and Normark, S. (1998) Multicellular and aggregative behaviour of *Salmonella typhimurium* strains is controlled by mutations in the *agfD* promoter. Mol Microbiol 28(2): 249-264.

Solano, C., García, B., Valle, J., Berasain, C., Ghigo, J.-M., Gamazo, C., and Lasa, I. (2002) Genetic analysis of *Salmonella enteritidis* biofilm formation: critical role of cellulose. *Mol Microbiol* 43(3): 793-808.

Staudenmaier, H., Van Hove, B., Yaraghi, Z., and Braun, V. (1989) Nucleotide sequences of the *fecBCDE* genes and locations of the proteins suggest a periplasmic-binding-protein-dependent transport mechanism for iron(III) dicitrate in *Escherichia coli*. *J Bacteriol* 171(5): 2626-2633.

Travisano, M., and Lenski, R.E. (1996) Long-term experimental evolution in *Escherichia coli*. IV. Targets of selection and the specificity of adaptation. *Genetics*. 143(1): 15-26.

Uhlich, G.A., Cooke, P.H., and Solomon, E.B. (2006) Analyses of the red-dry-rough phenotype of an *Escherichia coli* O157: H7 strain and its role in biofilm formation and resistance to antibacterial agents. *Appl Environ Microbiol* 72(4): 2564-2572.

U.S. EPA – United States Environmental Protection Agency. (2012) Recreational water quality criteria. Office of Water 820-F-12-058. http://water.epa.gov/scitech/swguidance/standards/criteria/health/recreation/upload/RWQC2012.pdf

Weiner, R., Langille, S., and Quintero, E. (1995) Structure, function and immunochemistry of bacterial exopolysaccharides. *J Ind Microbiol* 15(4): 339-346.

Westermark, G.T., Johnson, K.H., and Westermark, P. (1999) Staining methods for identification of amyloid in tissue. *Methods Enzymol* 309: 3-25.

Westrich, J.R., Ebling, A.M., Landing, W.M., Joyner, J.L., Kemp, K.M., Griffin, D.W., and Lipp, E.K. (2016) Saharan dust nutrients promote *Vibrio* bloom formation in marine surface waters. *Proc Natl Acad Sci USA* 113(21): 5964-5969.

Whitfield, C., and Roberts, I.S. (1999) Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol Microbiol* 31(5): 1307-1319.

Wyres, K.L., Wick, R.R., Gorrie, C., Jenney, A., Follador, R., Thomson, N.R., and Holt, K.E. (2016) Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom* 2(12): e000102.

Zogaj, X., Nimtz, M., Rohde, M., Bokranz, W., and Römling, U. (2001) The multicellular morphotypes of *Salmonella typhimurium* and *Escherichia coli* produce cellulose as the second component of the extracellular matrix. *Mol Microbiol* 39(6): 1452-1463.

4.7 Supplemental Material

Supplemental Table 4.1. Presence/absence of genes of other iron uptake systems in the *E. coli* strains used for the iron/citrate assays

Strain	*Klebsiella* capsule status*	*fec* operon *	Iron uptake systems†					
			Enterobactin	Aerobactin	Salmochelin	Yersinia bactin	Fe(II) uptake *feoABC*	Fe(II) uptake *efeUOB*
Hunter	1	1	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*
H288	1	1	*entABCDEF fepABCDEG ybdA fes*	-	-	*irp1234 56789 fyuA ybtA*	*feoABC*	*efeUOB*
E258	1	1	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*
H645	1	1	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*
2099_1.1	1	1	*entABCDEF fepABCDEG ybdA fes*	-	*iroBCDEN*	-	*feoABC*	*efeUOB*
H331	1	1	*entABCDEF fepABCDEG ybdA fes*	*iucABCD iutA*	[*iroB*] [*iroC*]	*irp1234 6789*[*irp 5*] *ybtA*	*feoABC*	*efeUOB*
64_3_AC10	1	0	*entABCDEF fepABCDEG ybdA fes*	-	-	*irp1345 6789*[*irp 2*] *fyuA ybtA*	*feoABC*	*efeUOB*
H446	1	0	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*
64_6_TC7	1	0	*entABCDEF fepABCDEG ybdA fes*	-	-	*irp1345 6789*[*irp 2*] *fyuA ybtA*	*feoABC*	*efeUOB*
H474	1	0	*entABCDEF fepABCDEG ybdA fes*	*iucABCD iutA*	-	-	*feoABC*	*efeUOB*
E7112	1	0	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*
A1_25	1	0	*entABCDEF fepABCDEG ybdA fes*	-	*iroBCDEN*	-	*feoABC*	*efeUOB*

Strain	*Klebsiella* capsule status*	*fec* operon *	Iron uptake systems†					
			Enterobactin	Aerobactin	Salmochelin	Yersinia bactin	Fe(II) uptake *feoABC*	Fe(II) uptake *efeUOB*
2H_157_1	0	1	*entABCDEF fepABCDEG ybdA fes*	*iucABCD iutA*	-	*irp1234 56789 fyuA ybtA*	*feoABC*	*efeUOB*
45_2_HU10	0	1	*entABCDEF fepABCDEG ybdA fes*	-	-	*irp1234 56789 fyuA ybtA*	*feoABC*	*efeUOB*
B1893	0	1	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*
C1_10	0	1	*entABCDEF fepABCDEG ybdA fes*	*iucABCD iutA*	-	*irp1234 56789 fyuA ybtA*	*feoABC*	*efeUOB*
B1867	0	1	*entABCEF*[*entD*] *fepABCDEG ybdA*	-	-	-	*feoABC*	*efeUOB*
E4173	0	1	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*
2H_255_4	0	0	*entABCDEF fepABCDEG ybdA fes*	-	-	*irp1234 56789 fyuA ybtA*	*feoABC*	*efeUOB*
H034	0	0	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*
H656	0	0	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*
B1820	0	0	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*
67_4_Ti7	0	0	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*
E7285	0	0	*entABCDEF fepABCDEG ybdA fes*	-	-	-	*feoABC*	*efeUOB*

* 1, present; 0, absent.
† A gene name within a square bracket indicates a fragment of the gene.

Chapter 5. Genotypic and phenotypic characteristics of a free-living strain of *Escherichia coli* responsible for bloom events

5.1 Abstract

Three *Escherichia coli* strains have been responsible for bloom events in the east coast of Australia, and all three have acquired a capsule from *Klebsiella*. One of these three strains belongs to *E. coli* phylogroup B1, is termed strain B1-001, and appears to be free-living in the environment.

Characteristics of the B1-001 bloom strain were investigated, including its phylogenetic relationship to *E. coli* and *Shigella*. The lactose (*lac*) operon, used for lactose transport and metabolism, and its flanking region, were studied in B1-001, other *E. coli*, and *Shigella*. A variable genome comparison was carried out using bloom isolates and a collection of *E. coli*.

The B1-001 bloom strain was closely related to *S. sonnei* and *S. boydii* and resembled *Shigella* by having a reduced genome size, being non-motile, and auxotrophic. It did not carry the virulence plasmid pINV of *Shigella* and its closest relatives were lactose-positive *E. coli*. Like typical *E. coli*, the bloom strain had a complete *lac* operon and a comparable flanking region. The bloom strain lacked genes required for flagella and curli biosynthesis, galactitol and carnitine metabolism, an arginine uptake system, putrescine transport, and 48 *E. coli* virulence genes. It has acquired a *Salmonella* bacteriophage and genes from *Klebsiella* outside of the capsule gene cluster. Genes unique to the B1-001 bloom strain compared to closely related non-bloom encapsulated *E. coli* do not seem to exclusively benefit its free-living lifestyle. Overall, it is intriguing to have a non-motile auxotroph that does not produce curli, presumably behaving as a free-living strain.

5.2 Introduction

Escherichia coli is used extensively as an indicator of recent faecal contamination of drinking and recreational waters, based on the assumption that it is unable to proliferate outside a host, and hence its presence in water indicates recent faecal contamination (Bonde, 1966; Edberg et al., 2000; U.S. EPA, 2012; Gordon, 2013). However, a growing body of evidence suggests that *E. coli* can not only survive for extended periods, but also proliferate in the environment outside a host. Studies carried out in the 1980s have indicated the extended survival and regrowth of *E. coli* in marine and fresh water habitats in Puerto Rico (Carrillo et al., 1985; Santo Domingo et al., 1989). Several subsequent studies from tropical and temperate regions also have indicated the extended survival and proliferation of *E. coli* in environments such as water, soil, algae, and manure (Hardina and Fujioka, 1991; Kudva et al., 1998; Fujioka et al., 1999; Solo-Gabriele et al., 2000; Whitman et al., 2003; Power et al., 2005; Ishii and Sadowsky, 2008). This confounds the use of *E. coli* as an ideal water quality indicator (Alm et al., 2011).

Perhaps the most compelling evidence against the use of *E. coli* as a water quality indicator comes from Australia, where significantly elevated *E. coli* counts have been reported for freshwater reservoirs and recreational lakes across the country (Power et al., 2005). These elevated count events have been termed *E. coli* 'bloom' events, as counts from 10,000 – 100,000 cells/100 ml of water have been reported, which are well above the 'safe' cut-off level of 235 cfu/100 ml of recreational water (Ishii and Sadowsky, 2008; U.S. EPA, 2012). Sanitary surveys indicate that these elevated count events cannot be attributed to faecal contamination, and indeed, achieving cell counts this high would require an unachievable level of faecal contamination (Power et al., 2005). The strains responsible for bloom events can be

isolated from the water bodies at any time, and have not been detected in the faeces of humans or other vertebrates (Power et al., 2005). These outcomes suggest that the presence of these strains in a water body is largely independent of faecal inputs and indeed, the strains responsible may represent, free-living *E. coli* (Power et al., 2005; Alm et al., 2011).

Relatively few strains have been found to be responsible for *E. coli* bloom events, and all of these are members of *E. coli* phylogenetic groups A and B1 (Power et al., 2005). Of the three strains isolated from bloom events in the east coast of Australia, the phylogroup B1 bloom strain, which is termed B1-001 strain, has always been found to be numerically dominant, while one or both of the phylogroup A strains may also be present. The current study concerns the B1-001 bloom strain. In addition to the possession of a capsule originating from *Klebsiella* which is present in all bloom strains (Power et al., 2005; Nanayakkara et al., 2019), the B-001 bloom strain has several other distinctive features. Unlike the majority of *E. coli* strains, it fails to produce acid and gas from lactose at 44.5 °C, and its optimum temperature for growth in lysogeny broth is significantly lower than most other strains of *E. coli* (Littlefield-Wyer, 2006). The B1-001 bloom strain also has an atypical phenotypic profile as it is auxotrophic, as well as indole, melibiose, and lysine negative. The API 20E identification system yields a *Citrobacter youngii* identification, while the BD BBL™ Crystal™ Enteric/Nonfermenter ID kit results in a *Shigella sonnei* identification (Power et al., 2005). Preliminary whole genome sequence analysis revealed that the B1-001 bloom strain is closely related to certain *Shigella* lineages and that it has an unusually small genome for a strain of *E. coli*. Yet, the B1-001 bloom strain expresses the two enzymes β-D-galactosidase and β-D-glucuronidase, which are used by the water industry to define *E. coli* (APHA, AWWA, WEF - Standard methods for the examination of water and wastewater, 2018).

The objective of this study was to further investigate the evolutionary relationship, and genotypic and phenotypic characteristics of this apparently free-living strain of *E. coli*.

5.3 Materials and Methods

Strain selection: Seven representatives of the B1-001 bloom strain were available: isolate E267 recovered from Lake Burley Griffin, Australian Capital Territory in 2003; isolate Hinze_Dam1 from a water sample collected from Hinze dam, Queensland (Qld) in 2009; E2095, E2711, and E9432 isolated from Shoalhaven dam, New South Wales (NSW) in 2011, 2012, and 2013, respectively; E8272 isolated from Wyaralong dam, Qld in 2013; and GOOG6 isolated from Googong dam, NSW in 2015. Subsequent genome comparisons revealed that all the B1-001 bloom isolates were highly similar and the isolate E267 was used as a representative of the B1-001 bloom strain.

Given that the B1-001 bloom strain has an unusually small genome and phenotypic characteristics suggestive of *Shigella*, the genomes of 53 *S. boydii*, *S. sonnei*, and *S. flexneri* and a number of *E. coli* strains were downloaded from the database of the National Centre for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/genome) for comparison. Whole genome sequence (WGS) data from a collection of Australian *E. coli* were also included (Nanayakkara et al., 2019).

5.3.1 Phylogenetic relationship of the B1-001 bloom strain to *E. coli* and *Shigella*

Draft genome sequences of the *E. coli* and *Shigella* strains were first reordered against a reference sequence using Mauve genome alignment visualization software (Darling et al., 2004), prior to alignment. The strains used as references for reordering the draft genomes were: *E. coli* phylogroup B1 strain IAI1 for reordering phylogroup B1 *E. coli* draft genomes, *S. boydii* strain Sb227 for reordering *S. boydii* draft genomes, and *S. sonnei* strain Ss046 for *S. sonnei* draft genomes. All of the *S. flexneri* strains used were reference genomes and did not require reordering.

Phylogenies were generated eliminating the sites of recombination using TOPALi (Milne et al., 2008) or Gubbins (Croucher et al., 2015). *Shigella* and *E. coli* strains together with E267 were aligned and the SNPs (single nucleotide polymorphisms) were obtained from Harvest v1.1.2 (Treangen et al., 2014), using the phylogroup B2 *E. coli* strain S88 as the outgroup, and the extracted SNP data were analysed using Gubbins (Croucher et al., 2015). Gubbins detects regions with high densities of base substitutions and eliminates these as they are considered to represent sites of recombination. Hence, tree inference is based on the putative point mutations outside of the regions of recombination.

5.3.2 *lac* operon, capsule, and genome size

Given the differences in the lactose operon of *Shigella* and *E. coli*, the lactose operon and its flanking region of representative *E. coli*, *Shigella* and bloom strains were investigated using MicroScope (https://www.genoscope.cns.fr/agc/microscope/home/) and genome alignments conducted using Mauve (Darling et al., 2004). As all bloom strains are

encapsulated, non-bloom strains closely related to the B1-001 bloom strain were screened for the presence/absence of the *Klebsiella* capsule using Kaptive (Wyres et al., 2016). The genome size of E267 was compared against *E. coli* strains representing the major *E. coli* phylogroups, and *Shigella* strains representing the three 'species': *S. boydii*, *S. sonnei*, and *S. flexneri*.

5.3.3 Variable genome comparison

The variable genome of B1-001 bloom isolate E267 and non-bloom *E. coli* strains was compared using the tools available in the MicroScope platform (https://www.genoscope.cns.fr/agc/microscope/home/). Phylogroup B1 *E. coli* strain IAI1 was set as the reference and a further 24 *E. coli* strains (NCBI RefSeq) representing the phylogroup diversity of the species were used for the comparison (Clermont et al., 2013). Genes that were present in at least 90% of the 25 *E. coli* strains but lacking in E267, and those unique to E267, were investigated. As all bloom strains have acquired a capsule originating from *Klebsiella* (Power et al., 2005; Nanayakkara et al., 2019), the variable genome of a collection of B1-001 bloom isolates (n = 3) and closely related non-bloom *Klebsiella*-capsule-positive *E. coli* strains (n = 4), was compared. The genomes were annotated using Prokka (Seemann, 2014) and the pan genome was analysed using Roary (Page et al., 2015).

5.3.4 Virulence screening

The B1-001 bloom isolates were screened for the presence of 48 genes that are implicated in virulence in *E. coli* using the CLC Genomics Workbench version 9.5.3 (https://www.qiagenbioinformatics.com). The virulence genes were *afaD*, *clbB*, *cnf1*,

*ColE*1_toxin, *colIa*_toxin, *colIb*_toxin, *etsC*, *fimH*, *focG*, *fyuA*, *hlyD*, *hra*, *ibeA*, *ireA*, *iroN*, *iucC*, *iutA*, *kpsE*, *lpfA*_LF82, microcin_V_toxin, microcin B17, *neuC*, *omp*_chromo, *ompT*_plasmid, *papG*, *sfaA*, *papC*, *sitA*, *tcpC*, *terC*, *traT*, *usp*, *vat*, *cah*_antigen_43, *cdiA*, *ybtS*, cloacin, *iha*, *eaeH*, *tia*, *upaG_eha*, microcin_47, *cdtB*, colicin B, colicin M, *senB*, *tsh*, and *arcA*_B6H13. The web-based tool, VirulenceFinder 2.0 (Joensen et al., 2014), was used to screen for the presence of the shiga toxin genes *stx1* and *stx2*. The presence/absence of *Shigella* virulence plasmids (pINV) was investigated using Mauve genome alignments (Darling et al., 2004).

5.4 Results

5.4.1 Relationship of the B1-001 bloom strain to *E. coli* and *Shigella*

Preliminary phylogenetic analysis revealed that the B1-001 bloom isolate E267 was closely related to *E. coli* strains of phylogroups B1 and C, but was not a true member of these phylogroups (Figure 5.1). Rather, E267 was most closely related to *S. sonnei* and *S. boydii* as well as to several other *E. coli* strains that were also lactose-positive and isolated from water and mammal faeces (Figure 5.1, Figure 5.2, and Figure 5.3).

An investigation of the distribution of *Klebsiella* capsules in *E. coli* (Nanayakkara et al., 2019) revealed four other non-bloom *E. coli* strains to be very closely related to the B1-001 bloom strain (Figure 5.3). All four strains were encapsulated and while the *Klebsiella* capsule type of the B1-001 bloom strain was KL53, the capsule types of the four non-bloom *E. coli* strains were: KL58 in strain KTE102, KL124 in strain M617, KL13 in strain 16009246, and KL107 in strain FSIS11704612. Although these strains had an intact *lac* operon and were encapsulated like the B1-001 bloom strain, the genome of these strains was significantly larger than the B1-

001 bloom isolates; 4.78 Mbp on average, as compared to 4.32 Mbp for the B1-001 bloom strain.

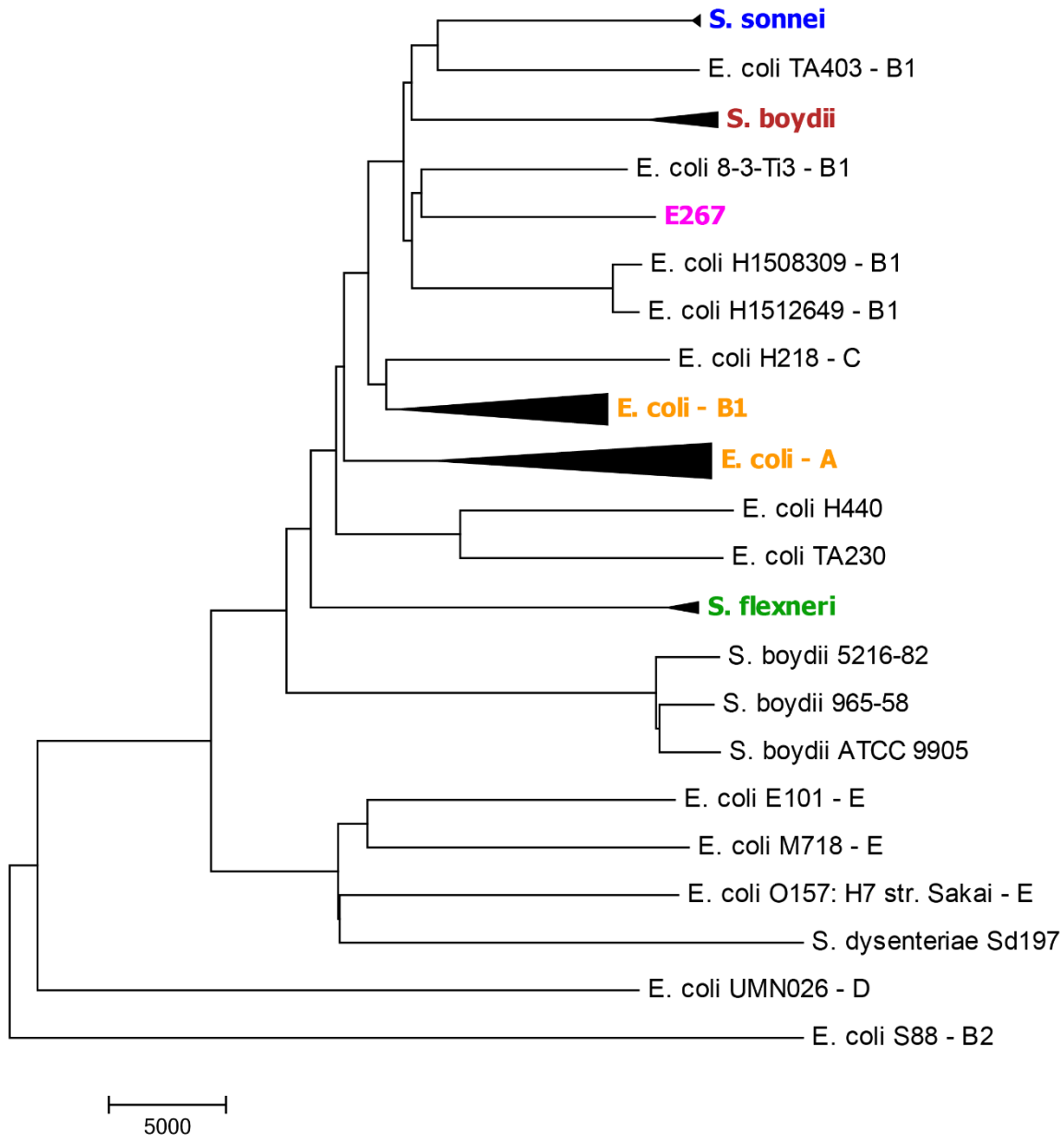


Figure 5.1. Neighbour-joining tree of the core genome data of B1-001 bloom strain E267, *E. coli*, and *Shigella*. Phylogroup of *E. coli* strains is indicated at the end of the strain name.

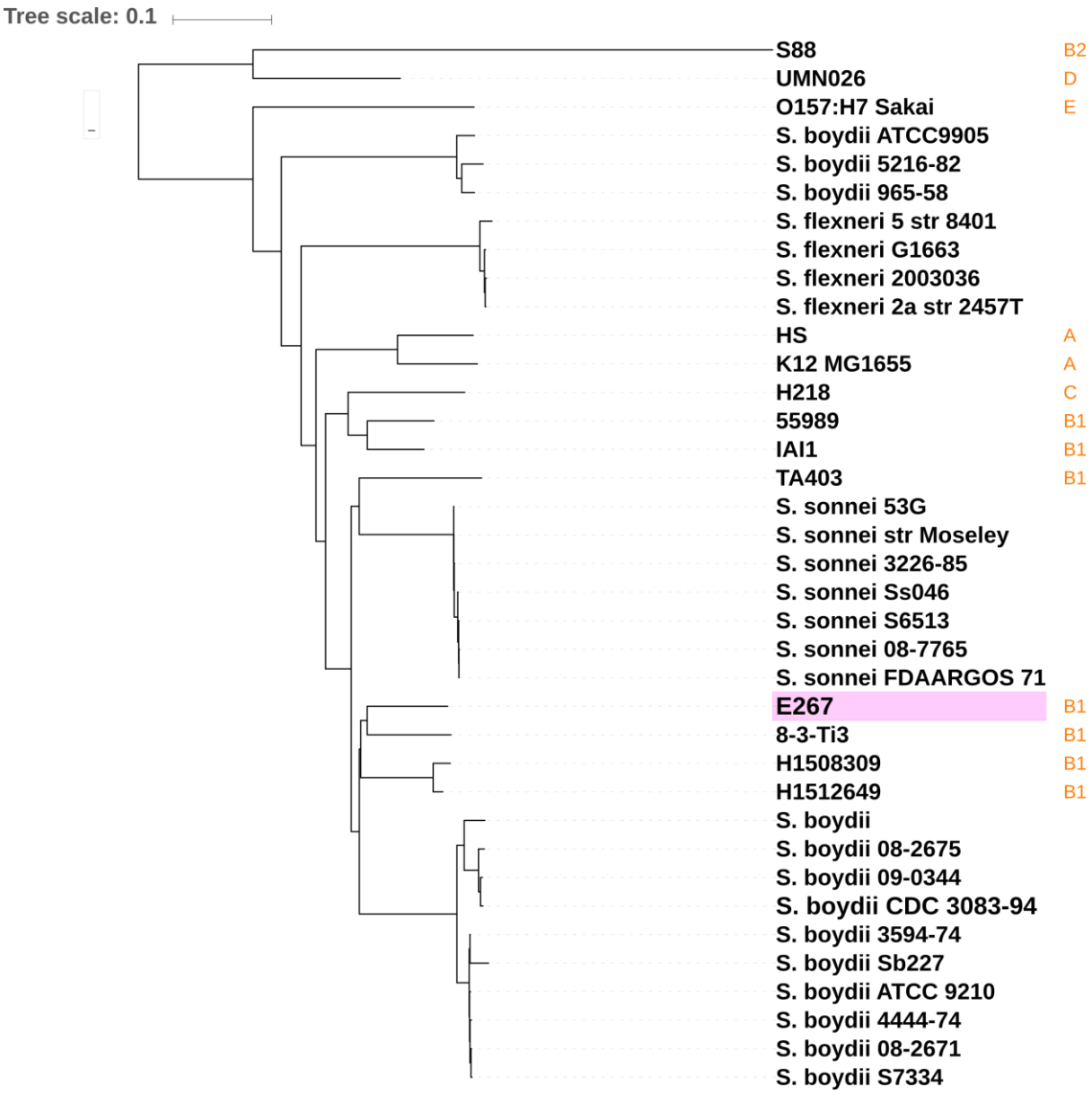


Figure 5.2. RAxML (randomised axelerated maximum likelihood) phylogeny generated by TOPALi (Milne et al., 2008) after eliminating the sites of recombination. The labels in orange indicated at the end of *E. coli* strain names depict the phylogroups of the *E. coli* strains. The tree was annotated using the web-based tool Interactive tree of life (iTOL) v3 (Letunic and Bork, 2016).

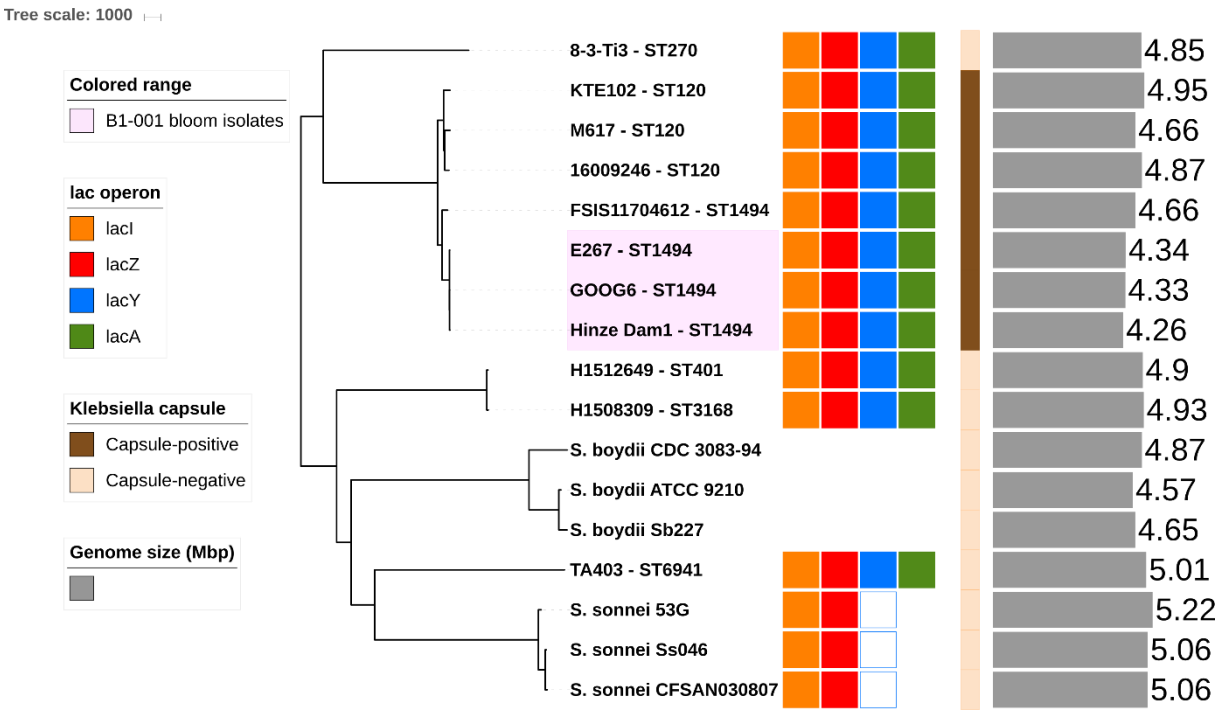


Figure 5.3. Phylogeny of B1-001 bloom isolates, *E. coli*, *S. sonnei*, and *S. boydii* with sites of recombination eliminated using Gubbins (Croucher et al., 2015). The tree was rooted on the phylogroup B2 strain *E. coli* S88, and annotated using Interactive tree of life (iTOL) v3 (Letunic and Bork, 2016). In terms of the *lac* operon, a coloured square indicates the presence of a *lac* gene while a white square indicates an impaired gene. The absence of a square indicates the absence of a gene.

5.4.2 *lac* operon and flanking region

The B1-001 bloom strain had the complete *lac* operon (*lacIZYA*) as do most *E. coli* (Figure 5.3). The *lac* operon had experienced multiple different deletion events in *Shigella*. *S. boydii* had lost the entire operon, while *S. sonnei* had lost *lacA* with *lacY* being completely or partially lacking (Figure 5.3), as observed previously (Ito et al., 1991).

The *lac* operon and its flanking region were compared for the same strains in Figure 5.3, using *E. coli* strain IAI1 as the reference (Figure 5.4). The *lac* operon flanking region of E267 was comparable to that of strain IAI1 except for the presence of bacteriophage-associated genes adjacent to *proAB* in E267. In *S. boydii* Sb227 a complete deletion event had occurred involving

the flanking region. In *S. sonnei* Ss046, a transversion event involving *mhp*, *yai*, and *proAB* had taken place in the region flanking its impaired *lac* operon (Figure 5.4).

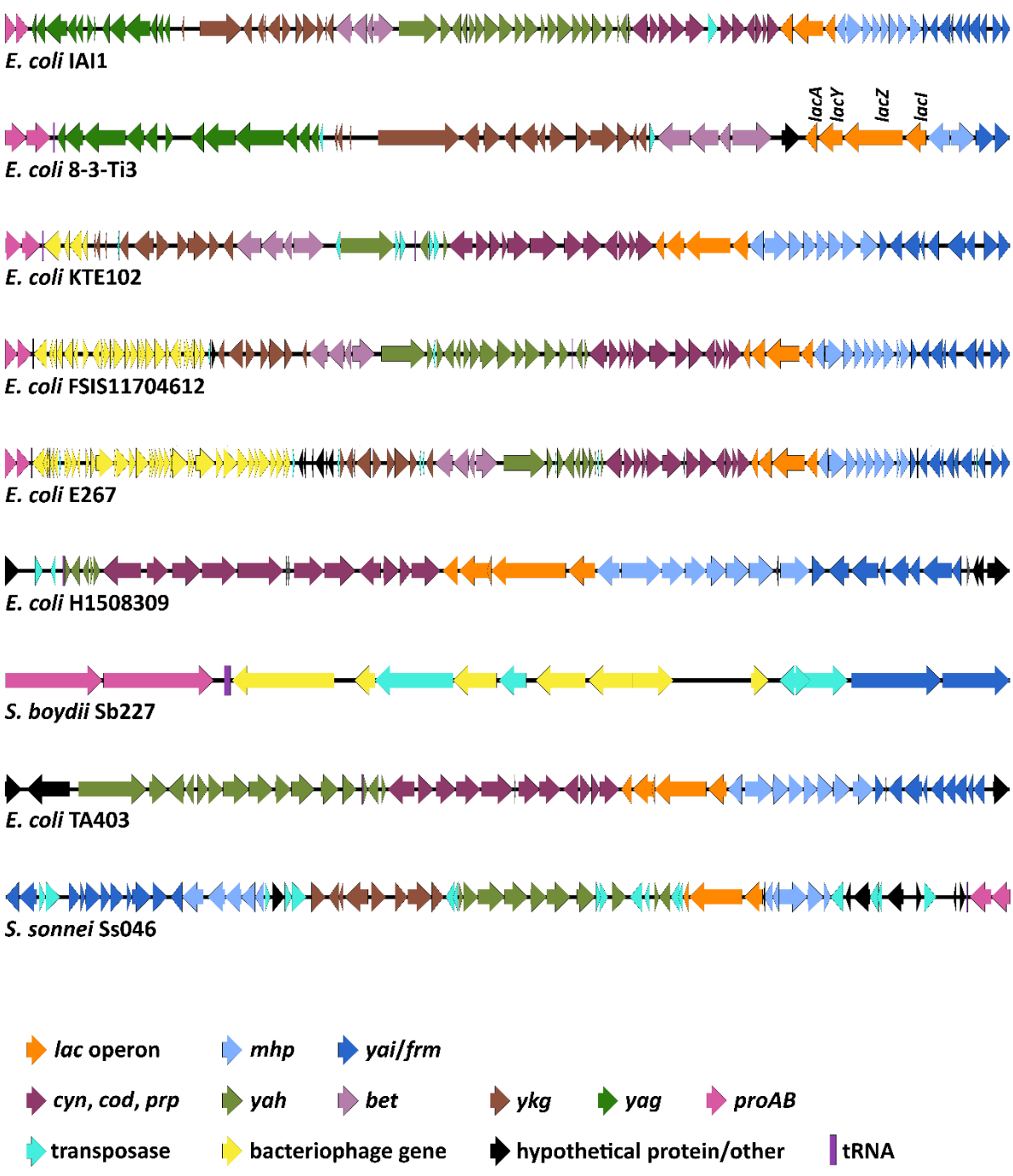


Figure 5.4. *lac* operon (orange) and the flanking region of *E. coli* IAI1, *E. coli* 8-3-Ti3, *E. coli* KTE102, *E. coli* FSIS11704612, *E. coli* E267, *E. coli* H1508309, *S. boydii* Sb227, *E. coli* TA403, and *S. sonnei* Ss046, in order from top to bottom. *E. coli* IAI1 was used as the reference and carries *lacIZY* and lacks *lacA*. The loci comparison figure was generated using Easyfig (Sullivan et al., 2011).

5.4.3 Variable gene content

The reduced genome size of the B1-001 bloom strain at 4.32 Mbp reflected a considerable loss of genome content. Of the genes shared by all 25 reference *E. coli* strains, 57 were absent from E267. From the genes present in at least 90% of the reference strains, 141 were absent in E267. Genes involved in colanic acid, curli, and flagella biosynthesis, carnitine and galactitol metabolism, and Fe^{2+} transport were among the major components lacking in E267 (Table 5.1). In addition to these genes, when the pan genome analysis was restricted to the encapsulated strains depicted in Figure 5.3, the following operons/genes were absent in the B1-001 bloom isolates and present in all four non-bloom encapsulated strains; arginine uptake system encoded by *artPIQMJ*, arginine decarboxylase *adiA*, putrescine transport genes *potFGHI*, nickel and cobalt homeostasis encoded by *rcnABR*, genes *yehABDE* for fimbrial biogenesis, and *atoDABE* for short-chain fatty acid degradation. In addition, several other genes including those coding for transcriptional regulators and uncharacterised proteins were also absent in the B1-001 bloom isolates compared to the four non-bloom encapsulated strains (Supplemental Table 5.1).

Approximately 136 genes were unique to E267 compared to the 25 reference *E. coli* strains, with the majority (95) being hypothetical proteins while the rest included bacteriophage-associated genes and those of the capsule cluster.

Table 5.1. Genes absent in E267 and present in >90% of the 25 *E. coli* strains

Function	Gene/s
Adhesion	*yadKN*, *ydeR*
Carnitine metabolism	*caiABCFT*
Colanic acid biosynthesis	*fcl*, *gmd*, *nudD*, *wcaACDEFIJKLM*, *wzxC*
Curli biogenesis	*csgABC*
Flagellar biosynthesis	*fliEFGHIJKLMNOPQR*
Galactitol degradation	*gatABCYZ*, *kbaYZ*
Melibiose catabolism	*melAR*
Glycogen biosynthesis	*glgS*
Iron transport	*efeUOB*
Glycolate utilisation	*glcBDG*
O-antigen biosynthesis	*cld*
Purine metabolism	*xdhBCD*
Uncharacterised proteins	*yegUVWX*
Uracil transport	*ybbW*

BLAST comparisons of genes unique to the B1-001 bloom isolates and absent in all four non-bloom encapsulated *E. coli* strains (Figure 5.3), revealed that about 45 genes were of *Klebsiella* origin, and these genes represented two groups in terms of genome location in *Klebsiella* and E267. Among these, approximately 15 genes appeared to have originated from a *Klebsiella* plasmid and comprised those encoding conserved hypothetical proteins and bacteriophage proteins. In the E267 genome this group of genes was located approximately 1,786,400 bp away from the capsule gene cluster. The second group of genes had a *Klebsiella* chromosomal origin and contained genes for hypothetical proteins, adenine-specific methyltransferase, RecE and RecT family proteins, and other bacteriophage related proteins. This cluster of genes was flanked by GMP synthase (glutamine-hydrolysing) (*guaA*) and exopolyphosphatase (*ppx*) in both *Klebsiella* and E267, and was not adjacent to the capsule gene cluster in either *Klebsiella* or E267. In E267 this block of genes was located approximately 458,000 bp away from the capsule gene region.

BLAST comparisons and screening for bacteriophages using PHASTER (Zhou et al., 2011; Arndt et al., 2016) revealed that the B1-001 bloom strain carries an extra-chromosomal element, a bacteriophage exhibiting close similarity to bacteriophage SSU5 of *Salmonella enterica* serovar Typhimurium reported by Kim and colleagues (2012). This phage was unique to the B1-001 bloom isolates compared to the four non-bloom encapsulated *E. coli*. The vast majority of the genes in the bacteriophage encoded hypothetical proteins while the rest included DNA polymerases, ligases, integrases, plasmid stability protein, several other bacteriophage-associated proteins, tellurite/colicin resistance protein, porphyrin biosynthesis protein, and cobalamin biosynthesis protein CobT.

No other *E. coli* genes of particular significance were present among the genes unique to the B1-001 bloom isolates compared to the four non-bloom encapsulated strains. Overall, none of the known genes specific to the bloom isolates seemed to be particularly significant for a free-living lifestyle.

The B1-001 bloom isolates did not carry any of the 48 *E. coli* genes implicated in virulence, or the virulence plasmid pINV and shiga toxin encoding *stx1* and *stx2* of *Shigella*. Investigation of the WGS data using the web-based tool ISsaga in ISfinder (http://issaga.biotoul.fr/) showed the B1-001 bloom strain to carry relatively few IS (Insertion Sequence) elements (25 IS elements in isolate E267, 21 in isolate Hinze_Dam1) compared to *Shigella* (412 IS elements in *S. boydii* Sb227, 460 in *S. boydii* CDC 3083-94, 399 in *S. sonnei* 53G, and 418 in *S. sonne*i Ss046).

5.5 Discussion

The B1-001 bloom strain is closely related to *S. boydii* and *S. sonnei* but is quite distinct. Certain superficial similarities exist between the B1-001 bloom strain and *Shigella* such as the reduced genome size, loss of motility, disrupted curli biogenesis, and auxotrophy. The lack of motility in *Shigella* is attributed to defective *fliF*, *fliD*, or *flhD* flagellar operons (Al Mamun et al., 1997), and most *Shigella* have disruptions in the curli-encoding *csg* locus caused by deletions or IS element insertions (Sakellaris et al., 2000). However, regardless of these superficial similarities to *Shigella*, the B1-001 bloom strain does not appear to be a non-pathogenic version of *Shigella*. *Shigella* are obligate intracellular pathogens (Ochman and Groisman, 1995) and are defined by the acquisition of the virulence plasmid pINV (Lan et al., 2001). The B1-001 bloom strain clearly lacks pINV, in addition to shiga toxin encoding *stx1* and *stx2*. Overall, the closest relatives of the B1-001 bloom strain are typical *E. coli* strains that are lactose-positive, as is the bloom strain itself.

Compared to its closest relatives, the unique feature of the B1-001 bloom strain is the bacteriophage SSU5, of *Salmonella enterica* serovar Typhimurium reported by Kim and colleagues (2012). However, BLAST comparisons reveal that this bacteriophage is present in a host of other *E. coli* strains, particularly those of the serotype O104:H4, which means that this bacteriophage is not unique to the B1-001 bloom strain. The annotations indicate that most of the proteins produced by the bacteriophage of the bloom strain are hypothetical proteins and those associated with structural and functional roles in the phage. Hence it is rather unlikely that this bacteriophage confers a fitness advantage to the apparently free-living B1-001 bloom strain.

The *Klebsiella* chromosomal and plasmid genes unique to the B1-001 bloom strain compared to the non-bloom encapsulated *E. coli* comprise mostly those encoding hypothetical proteins and bacteriophage-associated genes. In terms of function, none of these genes seems to be of particular significance for the likely free-living lifestyle of the bloom strain. These gene clusters are not located adjacent to the capsule gene cluster in either *Klebsiella* or the bloom strain, and hence they have likely been acquired independent of the capsule cluster. Although these hypothetical proteins have no apparent role in conferring the 'bloom' status, further analyses, especially at a transcriptional level, are needed to elucidate their role. It can be that they do play a role in environmental survival of the strains, given the genome reduction that has occurred over time.

The B1-001 bloom strain is an auxotroph for cysteine and glutamate (Chapter 6), which means that it is unable to grow unless these amino acids are present in its growth medium. Ideally, being a prototroph rather than an auxotroph is more beneficial for an essentially free-living strain, as a prototroph can synthesize amino acids on its own without having to depend on an external supply (McIver and Tapsall, 1993), thereby being able to grow even under limited nutrient conditions such as those represented by lakes and reservoirs. The bloom strains can grow at low densities during non-bloom periods when nutrients in water are in limited supply, and even in bloom waters, nutrients are known to be at relatively low concentrations (D. Gordon, personal communication). Hence, the ability of the B1-001 bloom strain to be presumably free-living and produce elevated counts while being auxotrophic, is indeed fascinating.

E. coli strains lose the colanic acid gene cluster as they acquire the capsule gene cluster from *Klebsiella*, and this explains the absence of the genes encoding colanic acid in the B1-001

bloom strain (Nanayakkara et al., 2019). The B1-001 bloom strain lacks operons encoding flagella and curli, which if present would have been beneficial for a free-living lifestyle. Flagella by providing motility aid cells in their search for nutrients (Ottemann and Miller, 1997). In addition, flagella aid biofilm development by allowing cells to come into contact with an inert surface and also by helping cells to move along the inert surface to spread the biofilm (Pratt and Kolter, 1998). Curli are bacterial extracellular adhesion structures which are also implicated in biofilm formation, and about 60% of environmental *E. coli* isolates are curli producers (Vidal et al., 1998; Prigent-Combaret et al., 2000; Sakellaris et al., 2000). The apparent loss of flagella and curli suggests that the B1-001 bloom strain has evolved to become 'planktonic' and not form aggregates or biofilms.

In water bodies where the B1-001 bloom strain is known to occur it can be detected at almost any time and not just during bloom events. However, its presence is unlikely to be due to vertebrate faecal inputs as it has never been detected in a vertebrate host, and its physiological characteristics make it unlikely that it could establish and persist in a bird or mammal. The B1-001 bloom strain is non-motile and an auxotroph, yet when a nutrient influx occurs, it is capable of out-competing other co-occurring strains of *E. coli* (Water Research Australia, 2019). Is it possible that the capsule, a distinctive feature of the bloom strain, traps nutrients, ensuring an ample nutrient supply for the cell? The results of Chapter 4 suggest that the capsule might enhance nutrient uptake by the cell. This in turn reiterates the fact that the capsule may be a key attribute in producing the elevated cell counts observed in bloom events (Nanayakkara et al., 2019). As a future study, the capsule region can be knocked out and survival of the strains under specific nutrient conditions can be observed, to infer what effect the capsule has on growth.

All of the evidence suggests that the B1-001 bloom strain is presumably a free-living strain whose presence and persistence in water bodies is completely independent of faecal inputs. The available data suggest that this bloom strain is restricted to Australia. Although there are other examples of *E. coli* strains that are ST1494 (the same sequence type as B1-001) in EnteroBase, those with whole genome data indicate that they have genome sizes that are significantly larger (4.5 – 4.7 Mbp) than the B1-001 bloom strain.

In Australia, the B1-001 bloom strain appears to be restricted to artificially impounded bodies of water. In these systems outflows tend to be relatively constant (water use), but inflows are highly variable and are often less than outflows, resulting in declining water volumes. For example, Warragamba Dam creates Lake Burragorang, the primary water supply reservoir for Sydney, NSW and in early 2012 the dam spilled for the first time in 12 years and during this period the storage capacity of the dam fell to less than a third. Thus, the dynamics of these storage reservoirs are very different from 'typical' lake and river systems and may have provided the conditions that encouraged the evolution of the B1-001 bloom strain.

Taken together, the B1-001 bloom strain stands out from the rest of the bloom strains especially in terms of its resemblance to *Shigella*, an intracellular pathogen. Yet, it is still a strain of *E. coli* which apparently has lost its ability to establish in a host but has evolved in such a way to become highly competent in its presumably free-living habit. This is exemplified by its sustained dominance and persistence in almost all bloom events reported from the east coast of Australia, compared to the likely more versatile phylogroup A bloom strains.

5.6 References

Alm, E.W., Walk, S.T., and Gordon, D.M. (2011) The niche of *Escherichia coli*. In *Population genetics of bacteria: A tribute to Thomas S. Whittam*. Walk, S.T., and Feng, P.C.H. (eds). Washington, DC: ASM Press, pp. 69–89.

Al Mamun, A.A., Tominaga, A., and Enomoto, M. (1997) Cloning and characterization of the region III flagellar operons of the four *Shigella* subgroups: genetic defects that cause loss of flagella of *Shigella boydii* and *Shigella sonnei*. *J Bacteriol* 179(14): 4493-4500.

APHA, AWWA, WEF – American Public Health Association, American Water Works Association, Water Environment Federation. (2018) 9223 Enzyme substrate coliform test 2017. In *Standard methods for the examination of water and wastewater,* 23rd Ed. https://www.standardmethods.org/doi/abs/10.2105/SMWW.2882.194

Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y., and Wishart, D.S. (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res* 44(W1): W16-W21.

Bonde, G.J. (1966) Bacteriological methods for estimation of water pollution. *Health laboratory science* 3(2): 124-128.

Carrillo, M., Estrada, E., and Hazen, T.C. (1985) Survival and enumeration of the fecal indicators *Bifidobacterium adolescentis* and *Escherichia coli* in a tropical rain forest watershed. *Appl Environ Microbiol* 50(2): 468-476.

Clermont, O., Christenson, J.K., Denamur, E., and Gordon, D.M. (2013) The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 5(1): 58-65.

Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D. et al. (2015) Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43(3): e15.

Darling, A.C.E., Mau. B., Blattner, F.R., and Perna. N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14(7): 1394-1403.

Edberg, S.C., Rice, E.W., Karlin, R.J., and Allen M.J. (2000) *Escherichia coli*: the best biological drinking water indicator for public health protection. *J Appl Microbiol* 88(S1): 106S-116S.

Fujioka, R., Sian-Denton, C., Borja, M., Castro, J., and Morphew, K. (1999) Soil: the environmental source of *Escherichia coli* and enterococci in Guam's streams. *J Appl Microbiol* 85(S1): 83S-89S.

Gordon, D.M. (2013) The ecology of *Escherichia coli*. In *Escherichia coli: Pathotypes and principles of pathogenesis.* Donnenberg, M.S. (ed). San Diego, CA: Academic Press, pp. 3-20.

Hardina, C.M., and Fujioka, R.S. (1991) Soil: the environmental source of *Escherichia coli* and enterococci in Hawaii's streams. *Environ Toxicol Water Qual* 6(2): 185-195.

Ishii, S., and Sadowsky, M.J. (2008) *Escherichia coli* in the environment: Implications for water quality and human health. *Microbes Environ* 23(2): 101-108.

Ito, H., Kido, N., Arakawa, Y., Ohta, M., Sugiyama, T., and Kato, N. (1991) Possible mechanisms underlying the slow lactose fermentation phenotype in *Shigella* spp. *Appl Environ Microbiol* 57(10): 2912-2917.

Joensen, K.G., Scheutz, F., Lund, O., Hasman, H., Kaas, R.S., Nielsen, E.M., and Aarestrup, F.M. (2014) Real-time whole-genome sequencing for routine typing, surveillance, and outbreak detection of verotoxigenic *Escherichia coli*. *J Clin Microbiol* 52(5): 1501-1510.

Kim, M., Kim, S., and Ryu, S. (2012) Complete genome sequence of bacteriophage SSU5 specific for *Salmonella enterica* serovar Typhimurium rough strains. *J Virol* 86(19): 10894-10894.

Kudva, I.T., Blanch, K., and Hovde, C.J. (1998) Analysis of *Escherichia coli* O157: H7 survival in ovine or bovine manure and manure slurry. *Appl Environ Microbiol* 64(9): 3166-3174.

Lan, R., Lumb, B., Ryan, D., and Reeves, P.R. (2001) Molecular evolution of large virulence plasmid in *Shigella* clones and enteroinvasive *Escherichia coli*. *Infect Immun* 69(10): 6303-6309.

Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44(W1): W242-W245.

Littlefield-Wyer, J.G. (2006) Genetic and phenotypic characterisation of *Escherichia coli* isolated from the primary and secondary habitat (PhD Thesis). Canberra, ACT, Australia: The Australian National University.

McIver, C.J., and Tapsall, J.W. (1993) Study of growth requirements other than cysteine of naturally occurring *Escherichia coli* and *Klebsiella* spp. auxotrophic for cysteine. *J Clin Microbiol* 31(10): 2790-2793.

Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D.F., and Wright, F. (2008) TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics* 25(1): 126-127.

Nanayakkara, B.S., O'Brien, C.L., and Gordon, D.M. (2019) Diversity and distribution of *Klebsiella* capsules in *Escherichia coli*. *Environ Microbiol Rep* 11(2): 107-117.

Ochman, H., and Groisman, E.A. (1995) The evolution of invasion by enteric bacteria. *Can J Microbiol* 41(7): 555-561.

Ottemann, K.M., and Miller, J.F. (1997) Roles for motility in bacterial–host interactions. *Mol Microbiol* 24(6): 1109-1117.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G. et al. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31(22): 3691-3693.

Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., and Slade, M.B. (2005) Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. *Environ Microbiol* 7(5): 631-640.

Pratt, L.A., and Kolter, R. (1998) Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Mol Microbiol* 30(2): 285-293.

Prigent-Combaret, C., Prensier, G., Le Thi, T.T., Vidal, O., Lejeune, P., and Dorel, C. (2000) Developmental pathway for biofilm formation in curli-producing *Escherichia coli* strains: role of flagella, curli and colanic acid. *Environ Microbiol* 2(4): 450-464.

Sakellaris, H., Hannink, N.K., Rajakumar, K., Bulach, D., Hunt, M., Sasakawa, C., and Adler, B. (2000) Curli loci of *Shigella* spp. *Infect Immun* 68(6): 3780-3783.

Santo Domingo, J.W., Fuentes, F.A., and Hazen, T.C. (1989) Survival and activity of *Streptococcus faecalis* and *Escherichia coli* in petroleum-contaminated tropical marine waters. *Environ Pollut* 56(4): 263-281.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14): 2068 – 2069.

Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R., and Palmer, C.J. (2000) Sources of *Escherichia coli* in a coastal subtropical environment. *Appl Environ Microbiol* 66(1): 230-237.

Sullivan, M.J., Petty, N.K., and Beatson, S.A. (2011) Easyfig: a genome comparison visualizer. *Bioinformatics* 27(7): 1009-1010.

Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014) The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 15(11): 524.

U.S. EPA - United States Environmental Protection Agency. (2012) Recreational water quality criteria. Office of Water. 820-F-12-058. https://www.epa.gov/sites/production/files/2015-10/documents/rwqc2012.pdf.

Vidal, O., Longin, R., Prigent-Combaret, C., Dorel, C., Hooreman, M., and Lejeune, P. (1998) Isolation of an *Escherichia coli* K-12 mutant strain able to form biofilms on inert surfaces: involvement of a new *ompR* allele that increases curli expression. *J Bacteriol* 180(9): 2442-2449.

Water Research Australia. (2019) Management of environmental *E. coli* blooms. Factsheet Project 1101. https://www.waterra.com.au › publications › document-search

Whitman, R.L., Shively, D.A., Pawlik, H., Nevers, M.B., and Byappanahalli, M.N. (2003) Occurrence of *Escherichia coli* and enterococci in *Cladophora* (Chlorophyta) in nearshore water and beach sand of Lake Michigan. *Appl Environ Microbiol* 69(8): 4714-4719.

Wyres, K.L., Wick, R.R., Gorrie, C., Jenney, A., Follador, R., Thomson, N.R., and Holt, K.E. (2016) Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom* 2(12): e000102.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., and Wishart, D.S. (2011) PHAST: a fast phage search tool. *Nucleic Acids Res* 39(suppl_2): W347-W352.

5.7 Supplemental Material

Supplemental Table 5.1. Genes absent in the B1-001 bloom isolates and present in all four non-bloom encapsulated *E. coli* strains closely related to the B1-001 bloom isolates

Product	Gene/s
23S rRNA (uracil(747)-C(5))-methyltransferase	*rlmC*
30S ribosomal protein S6--L-glutamate ligase	*rimK*
4Fe-4S mono-cluster protein	*yjdI*
Acetoacetate metabolism regulator	*hyfR_3*
Acetolactate synthase isozyme 1 small subunit	*ilvN*
Adenine deaminase	*ade*
Adhesins	*pgaCD, elfG_2*
Anaerobic fumarate respiratory system and regulator	*dcuBR*
Antitoxin	*prlF*
Cryptic beta-glucoside bgl operon antiterminator	*bglG*
DMT family transporter	*yicL*
DNA-binding response regulator	*fimZ_3,* group_1563
DUF1175 domain-containing protein	*yfaT*
DUF1198 domain-containing protein	group_1558
DUF1449 family protein	*yqiJ*
DUF2135 domain-containing protein	group_1779
DUF2238 domain-containing protein	group_1599
DUF2254 domain-containing protein	group_1591
DUF2300 domain-containing protein	*yfaQ*
DUF4132 domain-containing protein	group_548
DUF977 family protein	group_274
FAD-binding protein	group_420
Fructose-bisphosphate aldolase	*fbaB*
Fumarate hydratase class I, anaerobic	*fumB*
Galactarate dehydratase	*garD*
GntR family transcriptional regulator	*mngR*
Hexose phosphate transporter	*uhpT*
Hydroxymethylpyrimidine kinase, hydroxyethylthiazole kinase	*thiDM*
Hypothetical proteins	*yicS, yqiI*
Ketose-bisphosphate aldolase	*ydjI*
Lipoprotein	*ynfC*
Membrane protein	*yahN*
MFS transporter family glucose-6-phosphate receptor	*glpT_1*
Phosphate starvation protein	*phoH*
PTS system proteins	*manX_2, chbB, manX_1*
Purine ribonucleoside efflux pump	*nepI*
Rac prophage; site-specific recombinase	*pinR*
RHS repeat protein	*rhsD_1*
SDR family oxidoreductase	*fabG_1*

Product	Gene/s
Sensor protein	*uhpB*
TetR family transcriptional regulator	nemR_2
Transcriptional regulator	*adiY*
Two-component system sensor histidine kinase	*glnL_2*
Type 1 fimbrial protein	*ymdA*
Type IV conjugative transfer system protein	group_1502
Uncharacterised proteins	*ybjNOP*

Chapter 6. Growth characteristics of *Escherichia coli* bloom strains in Colilert-18® medium

6.1 Abstract

Escherichia coli strains that produce significantly elevated counts in water (blooms) have been reported from Australia. Three strains from phylogroups A and B1; A-000, A-010, and B1-001 have been isolated from the east coast bloom events. The B1-001 and A-010 strains fluoresce in Colilert-18® medium, while the A-000 strain does not. Although the B1-001 strain is known to be numerically dominant and always present in bloom water samples, the B1-001 strain has not been detected during assessment of the strains present, after growth in Colilert-18®.

Growth experiments were carried out to investigate the minimum frequency of B1-001 strain required to observe fluorescence in Colilert-18® when B1-001 co-occurs with the non-fluorescing A-000 strain, and also to explain why B1-001 strain is not detected when strains are assessed after growth in Colilert-18®.

The B1-001 strain was at a growth rate disadvantage and was heavily outcompeted by A-000 and A-010 strains in Colilert-18®. When B1-001 strain co-occurs with the non-fluorescing A-000 strain in a bloom event, fluorescence would be detected if B1-001 strain represents at least 50% of the cell population in the bloom water sample. As B1-001 strain is numerically dominant in bloom events, it is unlikely that fluorescence would be missed.

Assessing the strains present during a bloom event typically involves subsampling one or more positive wells of a Colilert® tray and plating to obtain single colonies. Sampling a single colony means that the most likely outcome will be that the out-competed and hence numerically less abundant B1-001 bloom strain is not sampled, while the single colony is more likely to be of the more abundant A-000 or A-010 strain. This explains why B1-001 is not detected during assessment of the strains present in a bloom water sample, after growth in Colilert-18®.

6.2 Introduction

Drinking and recreational water is subjected to routine microbiological testing for faecal contamination as faeces is a source of harmful pathogens (Szewzyk et al., 2000; Gordon, 2001; Alm et al., 2011; U.S. EPA, 2012). Routine water quality assessment protocols involve tests for the detection and quantification of total coliforms and faecal coliforms/*Escherichia coli* in water. The coliform group is defined as family *Enterobacteriaceae* members that metabolise Ortho-Nitrophenyl-β-D-galactopyranoside (Leclerc et al., 2001). Multiple tube fermentation and the membrane filtration technique are two protocols that were at the forefront of testing for total coliforms (Edberg and Edberg, 1988). Due to long turnaround times, subjectivity in result interpretation, and inadequacies in providing conclusive information from a public health perspective, these methods have mostly been replaced by a defined substrate (DS) technology (Edberg and Edberg, 1988). The Colilert-18® system is based on the DS technology patented by the IDEXX Laboratories Inc., USA (www.idexx.com/water). This system detects both total coliforms and *Escherichia coli* in water, simultaneously (Edberg and Edberg, 1988). The detection relies on two nutrient indicator compounds: Ortho-Nitrophenyl-β-D-galactopyranoside (ONPG), which detects coliforms; and 4-Methyl-umbelliferyl-β-D-glucuronide (MUG) which detects *E. coli*. Coliforms have the enzyme β-D-galactosidase which metabolises ONPG turning the medium from colourless to yellow. Most *E. coli* carry the enzyme β-D-glucuronidase which metabolises MUG to produce blue fluorescence under ultraviolet (UV) light of 365 nm (Figure 6.1) (Manafi and Rotter, 1991; IDEXX, 2019). The current definition of *E. coli* used by the water industry is that *E. coli* expresses β-D-galactosidase and β-D-glucuronidase (APHA, AWWA, WEF - Standard methods for the examination of water and wastewater, 2018).

E. coli bloom events have been observed regularly, particularly in eastern Australia where three strains are known to be responsible for the elevated counts: two phylogroup A strains, termed A-000 and A-010 (also referred to as A0 and A1, respectively, in previous chapters), and one phylogroup B1 strain, termed B1-001 (Power et al., 2005). Studies of bloom events using the membrane filtration protocol, and centred on bloom events in Warragamba Dam in New South Wales (NSW), Hinze Dam in Queensland, and Lake Burley Griffin in Canberra, Australian Capital Territory (ACT), showed that the B1-001 strain is numerically dominant while one or both of the phylogroup A bloom strains may also be present. However, more recent characterisations of bloom events using Colilert-18® have failed to detect the presence of the B1-001 bloom strain. All three bloom strains express β-D-galactosidase. However, the A-000 bloom strain does not express β-D-glucuronidase, is MUG-negative, and therefore does not produce fluorescence in Colilert® (Power et al., 2005). The A-010 and B1-001 bloom strains express β-D-glucuronidase, are MUG-positive and should be detectable using the Colilert-18® system.

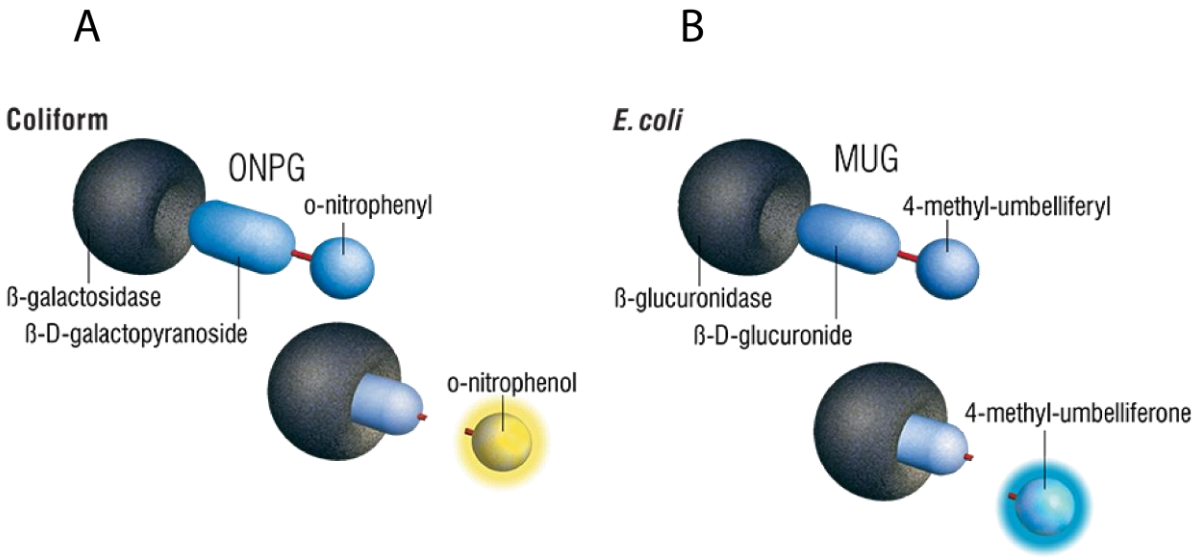


Figure 6.1. Colilert-18® system detection of coliforms based on ortho-Nitrophenyl-β-D-galactopyranoside (ONPG) metabolism by β-galactosidase, resulting in yellow media (A); and, *E. coli* based on the 4-Methyl-umbelliferyl-β-D-glucuronide (MUG) metabolism by β-glucuronidase resulting in blue fluorescence under UV light (B) (IDEXX, 2019).

Given the use of the Colilert® system for routine investigation of water samples and the apparent disappearance of the B1-001 bloom strain from recent bloom events assessed using the Colilert-18® system, a series of experiments was carried out to explore how the B1-001 bloom strain, and also the A-000 and A-010 bloom strains, behave in Colilert-18® medium.

6.3 Materials and Methods

6.3.1 Growth of the bloom strains in Colilert-18®

Strain selection: Eighteen *E. coli* bloom isolates representing the bloom strains A-000 (n = 6), A-010 (n = 6), and B1-001 (n = 6) were selected to determine the growth rate of the strains.

Strains were streaked on MacConkey agar from freezer cultures and an isolated colony was used to inoculate fresh Colilert-18® medium (5 ml) in flasks, which were incubated overnight at 35 ^{o}C with shaking at 170 r.p.m. Aliquots (100 µl) of the diluted cultures were inoculated into fresh Colilert-18® medium (4.9 ml) in flasks so as to have similar starting cell densities of approximately 10^4 cells/ml. The starting cultures (t = 0) were serially diluted in sterile normal saline (NaCl 0.85% w/v) and aliquots were plated on MacConkey agar to determine starting cells densities. The flasks were incubated at 35 ^{o}C with shaking, and sampling and plating conducted every hour for seven hours. The colony counts on MacConkey agar were used to obtain the cell density of the cultures at each sampling time. The maximum growth rate was estimated by regressing the natural log of the cell density against time. The 4-5 consecutive data points that gave the maximum slope were used to calculate the maximum growth rate.

To determine stationary phase cell densities of the bloom strains, the isolates were grown in Colilert-18® media as explained above and incubated overnight (18 hours) at 35 ^{o}C. The

overnight cultures were diluted in sterile normal saline (0.85% w/v) and an aliquot was plated on MacConkey agar followed by incubation overnight at 37 ^{o}C to estimate stationary phase cell densities (cfu/ml). The experiment was carried out in duplicate.

To investigate how the growth rate of the bloom strains varied with the concentration of the Colilert-18® medium, strains were dilution streaked on MacConkey agar from -80 ^{o}C freezer cultures and were incubated overnight at 37 ^{o}C. An isolated colony was inoculated into Colilert-18® medium and incubated overnight at 35 ^{o}C with shaking. Subsequently these cultures were diluted in sterile normal saline (0.85% w/v) to achieve a cell density of approximately 10^4 cells/ml. A series of two-fold dilutions of fresh Colilert-18® medium was prepared and these were inoculated with aliquots (100 µl) of the diluted overnight cultures above. The cultures were incubated at 35 ^{o}C with shaking, and at one, two, and three hours, the cultures were serially diluted, plated on MacConkey agar, and the cell densities determined. The growth rates were determined by regressing the natural log of cell density against time. For each bloom strain group, six biological replicates were conducted per Colilert-18® concentration.

6.3.2 Competition among bloom strains in Colilert-18®

Strain selection: Twelve *E. coli* bloom isolates representing the bloom strains A-000 (n = 4), A-010 (n = 4), and B1-001 (n = 4) were selected for the competition experiment.

Strains from -80 ^{o}C freezer cultures were dilution streaked on MacConkey agar and incubated overnight at 37 ^{o}C. An isolated colony was used to inoculate Colilert-18® medium in flasks and incubated overnight at 35 ^{o}C with shaking at 170 r.p.m. The cultures were diluted in sterile

normal saline (0.85% w/v) and 50 µl each of a B1-001 isolate and either an A-000 or A-010 isolate was co-inoculated into fresh Colilert-18® medium in flasks at equal starting cell densities of approximately 10^4 cells/ml. Aliquots were plated at time t = 0 to obtain the starting cell count of each of the two strains in each flask. The cultures were incubated for 24 hours at 35 ^oC with shaking at 170 r.p.m., after which the cultures were diluted and aliquots were plated on MacConkey agar to obtain the cell density of each competitor (Wiser and Lenski, 2015). The B1-001 isolates have a characteristic (small) colony morphology on MacConkey agar, which allows their easy differentiation from the A-000 and A-010 bloom colonies (Figure 6.2). Each of the four B1-001 isolates was competed against each of the four A-000 and A-010 isolates.

Figure 6.2. Colony morphology of the B1-001, A-000, and A-010 bloom strains on MacConkey agar.

The following equation (Wiser and Lenski, 2015) defines the fitness of A-000 and A-010 bloom isolates relative to the B1-001 isolates.

$$w = \frac{\ln\left(\frac{Af}{Ai}\right)}{\ln\left(\frac{Bf}{Bi}\right)}$$

Where; *w* = fitness, *A* = cell density of A bloom isolates and *B* = cell density of B1-001 isolates at the start (*i*) and at the end (*f*) of the experiment.

6.3.3 Fluorescence by B1-001 bloom strain in the presence of A-000 bloom strain

The outcomes of the experiments described above clearly showed that in Colilert-18®, the B1-001 isolates grew more slowly, produced lower stationary phase cell densities, and were clearly outcompeted by the A-000 and A-010 isolates. Further, the A-000 strain does not produce fluorescence in Colilert-18®. Consequently, these outcomes lead to the possibility of 'undetected bloom events' when B1-001 and A-000 strains co-occur. Therefore, an experiment was carried out to determine the minimum cell density of the B1-001 bloom strain required to produce fluorescence when the A-000 and B1-001 strains are both present.

A single A-000 isolate and two B1-001 isolates were used for this experiment. An isolated colony of the strains from MacConkey agar was inoculated into fresh Colilert-18® medium and flasks were incubated for 18 hours at 35 ^{o}C with shaking at 170 r.p.m. The stationary phase cell densities of the strains were determined. A 10 µl aliquot of the A-000 overnight culture was added to the Colilert-18® medium (1 ml) in Eppendorf tubes. Ten two-fold serial dilutions were made from the overnight cultures of the B1-001 isolates. A 10 µl aliquot from the serially diluted and undiluted cultures of the two B1-001 isolates was added separately to each

Eppendorf tube containing the A-000 isolate. The tubes were incubated for 18 hours at 35 °C and were observed for fluorescence in a BIO-RAD Gel Doc XR+ Imaging System. When fluorescence was not observed, the tubes were incubated for a further four hours and observed again (IDEXX, 2019). The experiment was carried out in duplicate.

6.3.4 Model to predict competition between phylogroup A and B1-001 bloom strains in Colilert-18®

A model of stationary phase kinetics was formulated for a phylogroup A bloom strain growing together with a B1-001 bloom isolate in a Colilert-18® batch culture over a period of 18 hours, as a modification of the model presented by Gordon and Riley (1999). The model was used to predict the final frequencies for given starting frequencies of the A and B1-001 strains.

As explained by Monod (1949) specific growth rate (φ) of the strains can be expressed as;

Equation (1)

$$\varphi = \left(\frac{\varphi_{max} R}{Q + R}\right)$$

Where; $\varphi\, max$ is the maximum growth rate, *R* is the resource concentration, and *Q* is the value of *R* when $\varphi/\varphi_{max} = 0.5$.

Derived from equation (1), the rate of change of the A and B1-001 populations for a single growth cycle can be expressed by the equations (2) and (3). The total cell count will increase as growth proceeds.

Equation (2)

$$\frac{dA}{dt} = \left(\frac{\varphi_{Amax} R}{Q_A + R}\right) A$$

Equation (3)

$$\frac{dB1}{dt} = \left(\frac{\varphi_{B1max} R}{Q_{B1} + R}\right) B1$$

Taking into account the number of cells available to divide (*A*, *B1*) at a given growth rate (φ) and the efficiency with which resources are converted to biomass (ϵ), the rate of change of resources can be expressed by equation (4). ϵ, $\varphi_{max,}$ and Q were experimental estimates.

Equation (4)

$$\frac{dR}{dt} = \epsilon_A \left(\frac{\varphi_{Amax} R}{Q_A + R}\right) A + \epsilon_{B1} \left(\frac{\varphi_{B1max} R}{Q_{B1} + R}\right) B1$$

6.3.5 Stationary phase cell densities of the bloom strains in different media

As the B1-001 strain grew more poorly compared to the phylogroup A bloom strains in Colilert-18®, their stationary phase cell densities were investigated in rich and minimal media to infer if the growth differences were the same.

Strain selection: The same set of 18 bloom isolates comprising A-000 (n = 6), A-010 (n = 6), and B1-001 (n = 6) strains used in the growth experiments above were selected.

Lysogeny broth (LB) was prepared using tryptone (10 g/l), yeast extract (5 g/l), and NaCl (10 g/l). Minimal media were prepared using the Davis minimal medium (10.6 g/l) and glucose/lactose (3 g/l). Each medium was inoculated with the isolates and was incubated overnight at 37 ^{o}C with shaking. At stationary phase, cultures were diluted in sterile normal saline (0.85% w/v) and an aliquot was plated on MacConkey agar followed by incubation at 37 ^{o}C overnight. Stationary phase cell densities were calculated from the plate counts. The experiment was carried out in duplicate.

6.3.6 Amino acid deficiency of the B1-001 bloom strain

The B1-001 bloom strain failed to produce growth in minimal media. However, growth was restored as the minimal glucose and lactose media were supplemented with yeast extract (1 mg/ml) or casamino acids (≥0.07% w/v), indicating that the B1-001 strain is likely auxotrophic for one or more amino acids. Therefore, minimal glucose media containing different cocktails of amino acids and thiamine (0.01%) were used to investigate if growth could be restored.

The first cocktail contained 12 different amino acids; tryptophan, phenylalanine, leucine, proline, aspartate, asparagine, lysine, glycine, arginine, methionine, putrescine, and ornithine (200 µg/ml) (Modified from Ahmed et al., 1988). The second cocktail contained 16 amino acids; 12 amino acids of the first cocktail and cysteine, serine, uracil, and glutamate (200 µg/ml). The third cocktail contained only four amino acids; cysteine (C), serine (S), uracil (U), and glutamate (G) at different combinations; CSUG, CSG, CUG, CG, and G alone. As growth was retarded without glutamate, it was included in all combinations.

6.3.7 Statistical analysis

Statistical analyses were performed using JMP 12.2.0 software (2015 SAS Institute, Inc.). Analysis of variance (ANOVA) was conducted with a full factorial model, with the bloom strain group as a model effect, and technical replicates incorporated as a random effect. The effect of the Colilert-18® concentration on the growth rate was analysed by non-linear modelling, using the Michaelis Menten model as a predictor formula.

6.4 Results

6.4.1 Growth rate and stationary phase cell density of bloom strains in Colilert-18®

The results indicated a significant effect of the bloom strain group on the maximum growth rate (p<0.0001) and the stationary phase cell density (p = 0.04) of the strains in Colilert-18®. The maximum growth rate of the B1-001 bloom strain was significantly lower than that of the A-000 and A-010 bloom strains (Table 6.1). The A-000 strain grew at a rate that was slightly lower than that of A-010 strain but the difference was not significant. The B1-001 strain had

the lowest stationary phase cell density in Colilert-18®, and this was significantly lower than that of the A-010 strain (Table 6.1).

Table 6.1. Maximum growth rate and stationary phase cell density of bloom strains in Colilert-18®

Bloom strain	Growth rate/hr ± SE	Stationary phase cell density/ml ± SE *
A-000	1.68 ± 0.07	$9.99 \times 10^8 \pm 2.10 \times 10^8$
A-010	1.72 ± 0.07	$1.42 \times 10^9 \pm 2.10 \times 10^8$
B1-001	1.15 ± 0.07	$5.56 \times 10^8 \pm 2.10 \times 10^8$

*n = 6

6.4.2 Growth rate with changing Colilert-18® concentration

The B1-001 bloom strain grew slower than the A-000 and A-010 bloom strains at all Colilert-18® concentrations tested (Figure 6.3). The Colilert-18® concentration when growth rate is half the maximum growth rate is termed the half saturation constant. The half saturation constant for the B1-001 strain was 0.08 ± 0.02 SE, while it was 0.03 ± 0.01 SE for the A-000 strain and 0.06 ± 0.02 SE for the A-010 strain. The values were not significantly different (p = 0.1).

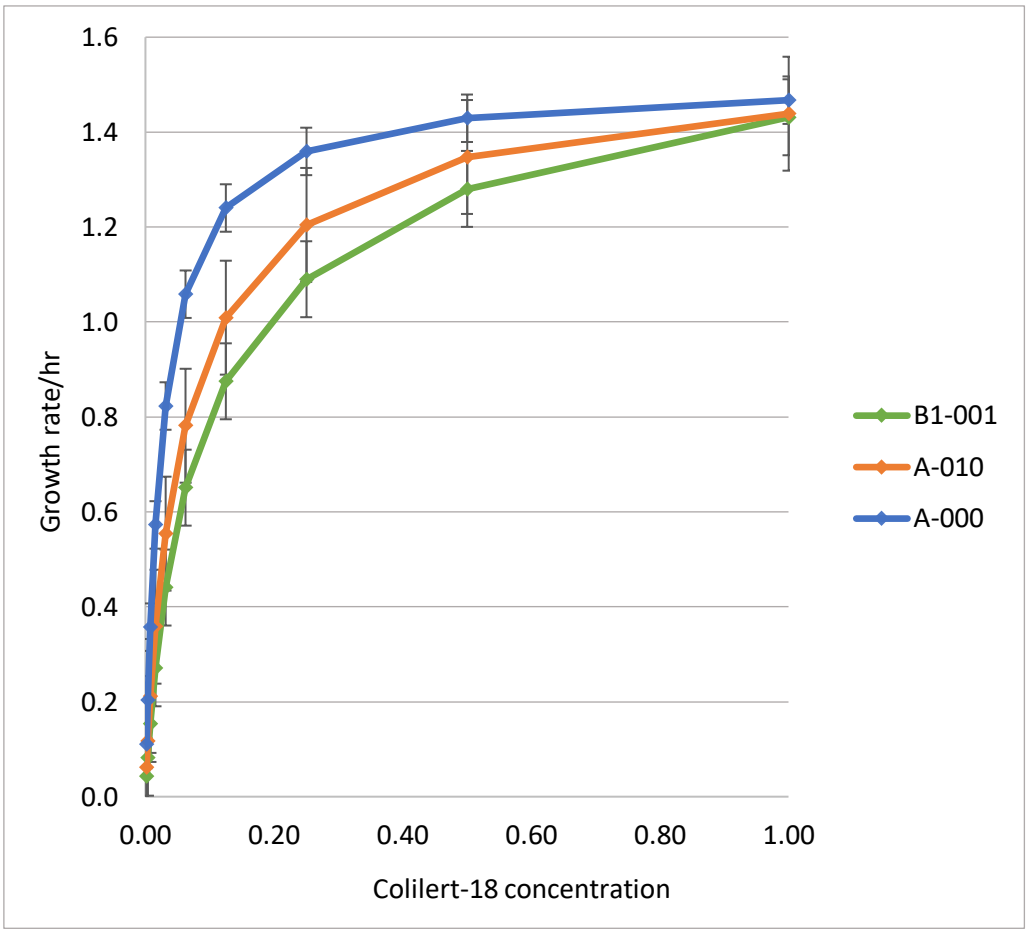


Figure 6.3. Growth rate of phylogroup A-000, A-010, and B1-001 bloom strains with changing Colilert-18® concentration. Six biological replicates were conducted for each bloom strain group, per Colilert-18® concentration. Mean ± SE.

6.4.3 Competition among bloom strains in Colilert-18®

Upon incubation, the B1-001 strain represented only 15.9% and 18.1% of the total cell count in the presence of A-000 and A-010 bloom strains, respectively. The relative fitness of both A-000 and A-010 strains was higher than that of the B1-001 strain, and the A-010 strain was significantly more fit than the A-000 strain (p = 0.04). Overall, the B1-001 strain was clearly outcompeted by the A-000 and A-010 strains in Colilert-18® when they were added at equal starting concentrations.

6.4.4 Model predictions and fluorescence by B1-001 in the presence of A-000

The B1-001 strain was outcompeted by the A-000 strain, and the A-000 strain is MUG-negative and does not fluoresce in Colilert-18®. Hence, it is possible that a bloom event consisting of the B1-001 and A-000 bloom strains could be undetected using the Colilert-18® system. Co-inoculation of the Colilert-18® medium with the A-000 strain and varying starting frequencies of the B1-001 bloom strain revealed that fluorescence was detected when the B1-001 bloom strain represented approximately 1% or more of the final cell population after incubation in the Colilert-18® medium (Figure 6.4).

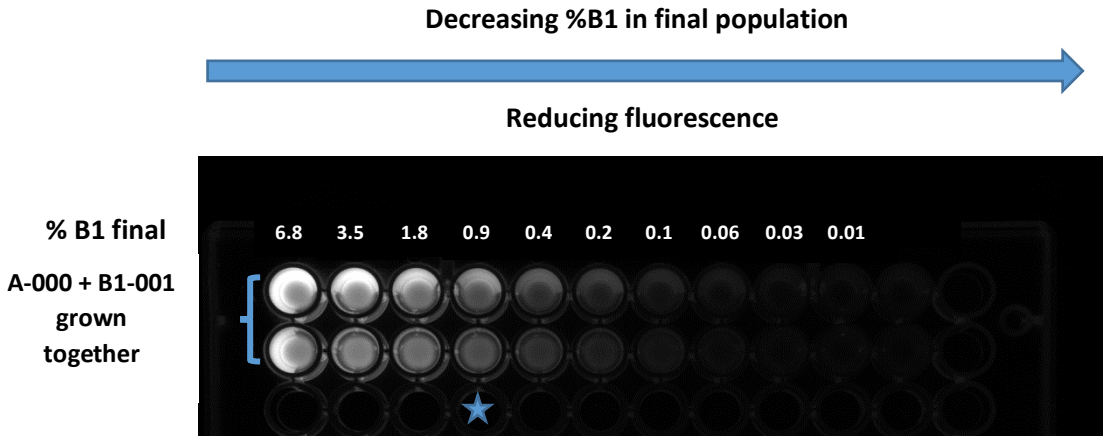


Figure 6.4. Fluorescence of a Colilert-18® mixed culture of A-000 and B1-001 bloom strains at 18 hours of incubation. The intensity of fluorescence decreased with decreasing cell density of the B1-001 strain. The star indicates the minimum level with clear detectable fluorescence.

6.4.5 Stationary phase cell densities in different media, and auxotrophy of B1-001

There was a significant effect of the bloom strain group on the stationary phase cell density in LB (p<0.0001). The B1-001 strain reached a cell density that was significantly lower than that of the A-010 and A-000 strains (Table 6.2). In contrast to LB, the B1-001 strain failed to grow

in minimal media. In minimal glucose, the cell density of the B1-001 strain was significantly lower than that of the A-010 strain (p = 0.002), while in minimal lactose, it was significantly lower than that of both A strains (p<0.0001) (Table 6.2).

The four amino acid combinations; cysteine, serine, uracil, and glutamate (CSUG); cysteine, serine, and glutamate (CSG); cysteine, uracil, and glutamate (CUG); and cysteine and glutamate (CG) restored growth of the B1-001 bloom strain in minimal media, but glutamate on its own did not produce growth. The stationary phase cell densities achieved with the four amino acid combinations were not significantly different (p = 0.2). Taken together, the presence of cysteine and glutamate was sufficient to restore growth, while CSG produced the best growth (Table 6.3).

Table 6.2. Stationary phase cell densities of the bloom strains in different media

Medium*	**Stationary phase cell density/ml ± SE**			***P*-value**
	A-000	**A-010**	**B1-001**	
LB	$6.02 \times 10^9 \pm 4.20 \times 10^8$	$6.29 \times 10^9 \pm 4.20 \times 10^8$	$4.31 \times 10^9 \pm 4.20 \times 10^8$	<0.0001
Minimal glucose	$1.84 \times 10^9 \pm 5.15 \times 10^8$	$3.10 \times 10^9 \pm 5.15 \times 10^8$	$4.65 \times 10^5 \pm 5.15 \times 10^8$	0.002
Minimal lactose	$2.67 \times 10^9 \pm 2.25 \times 10^8$	$2.87 \times 10^9 \pm 2.25 \times 10^8$	$8.58 \times 10^5 \pm 2.25 \times 10^8$	<0.0001

*LB – Lysogeny broth, n = 6

Table 6.3. Stationary phase cell densities achieved by the B1-001 bloom strain when minimal glucose medium was supplemented with different combinations of amino acids

Amino acid combination	**Stationary phase cell density/ml ± SE ***
cysteine + glutamate	$1.74 \times 10^9 \pm 2.11 \times 10^8$
cysteine + serine + glutamate	$2.31 \times 10^9 \pm 2.11 \times 10^8$
cysteine + uracil + glutamate	$1.79 \times 10^9 \pm 2.11 \times 10^8$
cysteine + serine + uracil + glutamate	$1.86 \times 10^9 \pm 2.11 \times 10^8$

*n = 6

6.5 Discussion

The B1-001 bloom strain has a slower growth rate and reaches a lower stationary phase cell density compared to the phylogroup A bloom strains in Colilert-18®. The growth disadvantage was observed in all media tested. The competitive superiority of the phylogroup A bloom strains does not appear to be due to bacteriocin production, as no bacteriocin related genes could be detected (unpublished data). It is also unlikely that bacteriophages are responsible for the competitive differences among strains, as all the bloom strains host one or more bacteriophages. Given that the B1-001 bloom strain is auxotrophic, the most likely explanation for the competitive differences of the bloom strains is their varying resource requirements.

The B1-001 bloom strain is auxotrophic for cysteine and glutamate, while the addition of serine further enhances growth. Cysteine biosynthesis serves as a major mode of sulfur assimilation by the cell (Neuwald et al., 1992), and its biosynthesis in *E. coli* proceeds through a cascade of genes including *cysB*, *cysTWA*, *cysPZ*, *cysDN*, *cysC*, *cysH*, *cysGIJ*, *cysE*, *cysMK*, and *cysQ*, grouped by function (Neuwald et al., 1992). According to whole genome sequence (WGS) data, the representative B1-001 isolate E267 carries all genes except *cysT* (sulfate permease). A *cysT* deletion or a non-functional gene in the cascade could have resulted in cysteine-dependence. Although serine is a precursor for the cysteine biosynthesis pathway (Abelson, 1954), serine in the absence of cysteine could not restore growth. Glutamate contributes about 88% of cellular nitrogen (Goss et al., 2001) and plays a key role in cellular osmoregulation (Yan et al., 1996). In enteric bacteria, the assimilation of ammonia into glutamate occurs through two pathways; the first involves glutamate dehydrogenase (*gdhA*), while the second involves glutamine synthetase (*glnA*) and glutamate synthase (*gltBD*) (Castaño et al., 1988; Goss et al., 2001). Strains lacking both *gdhA* and *gltBD* are auxotrophic

for glutamate (Brenchley et al., 1973). The B1-001 strain carries *gdhA*, *glnA*, and *gltBD* genes. Since ammonium salts are supplied in minimal media in the form of $(NH_4)_2SO_4$, if this genetic machinery were functional, the B1-001 strain would be able to synthesize its own glutamate and not be a glutamate auxotroph. Therefore, the glutamate-dependence likely indicates a functionally defective genetic machinery. Although the B1-001 strain lacks *gltF* (regulatory peptide) (Castaño et al., 1988), its impairment in *E. coli* does not seem to hinder nitrogen metabolism (Goss et al., 2001).

With the Colilert® system the presence of *E. coli* is demonstrated by the production of blue fluorescence (Edberg and Edberg, 1988), and *E. coli* is enumerated by counting the number of fluorescent wells in a Quanti-Tray® followed by referring to a most probable number (MPN) table (IDEXX, 2019). The A-000 bloom strain is MUG-negative and does not produce fluorescence in Colilert-18®, while the B1-001 strain does produce fluorescence. The B1-001 strain is heavily outcompeted by the A-000 strain in Colilert-18®. Therefore, when these two strains co-occur in a bloom event, fluorescence, and thus elevated counts, will only be detected if there is a sufficient concentration of B1-001 cells present in the water sample. According to the experimental results, when B1-001 strain co-occurs with the A-000 strain, fluorescence would be observed if the B1-001 strain represents 1% or more of the cell population after incubation in Colilert-18®. Hence, the B1-001 strain should represent at least 50% of the cell population in the bloom water sample for fluorescence to be detected (Figure 6.5). The fact that the B1-001 strain is usually numerically dominant in bloom events means that fluorescence is unlikely to be missed. Unless the B1-001 strain is present at very low frequencies compared to the A-000 strain, or the A-000 strain occurs on its own, it is unlikely that an entire bloom event will be missed with Colilert-18®.

Blue fluorescence is a result of the metabolism of MUG in Colilert-18® by β-glucuronidase (GUD) which is encoded by *uidA* (Novel and Novel, 1976). A clear majority of *E. coli* is GUD-positive though a minor proportion is phenotypically GUD-negative (Kämpfer et al., 1991; Manafi and Rotter, 1991; Frampton and Restaino, 1993). The exploration of WGS data using the MicroScope platform (https://www.genoscope.cns.fr/agc/microscope/home/) revealed that the A-000 bloom strain carries a fragment/s of *uidA* while *uidB* (proton-dependent transporter), *uidC* (membrane protein), and *uidR* (regulator) (Novel and Novel, 1976; Liang et al., 2005) are mostly intact. The inability of A-000 to metabolize MUG could be due to a deletion in *uidA* or an expression level defect.

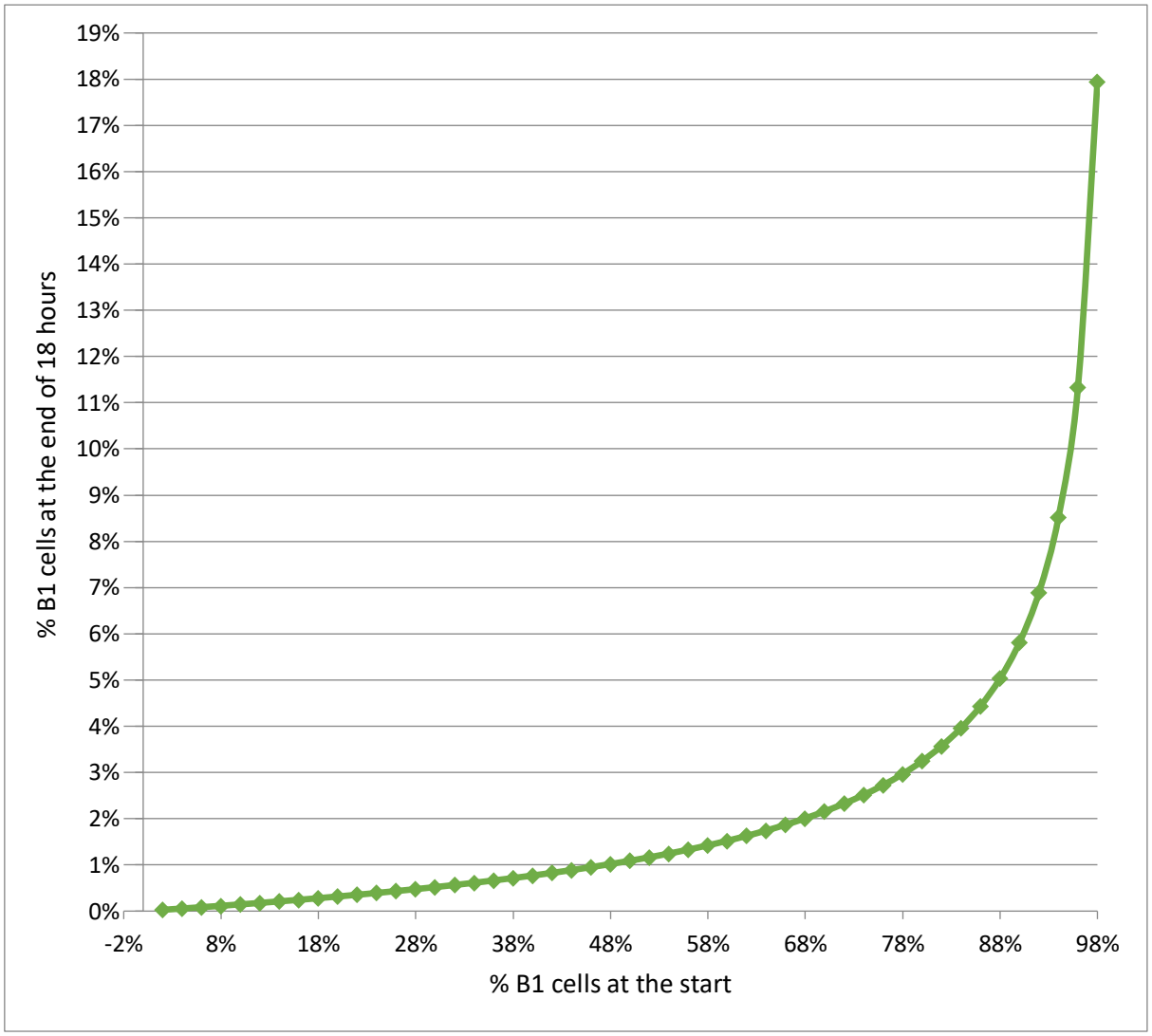


Figure 6.5. The percent B1-001 cells at the end versus the percent B1-001 cells at the start relative to the total cell population. The model assumed B1-001 and A-000 bloom strains growing together in Colilert-18® for 18 hours.

The characterisation of bloom events using membrane filtration in early work has revealed that the B1-001 bloom strain is always present and is numerically dominant, while one or both of the phylogroup A strains may also be present. The results of the present study reveal the reason for the apparent absence of the B1-001 bloom strain when bloom events are assessed using the Colilert® system. Assessing the strains present during a bloom event typically involves subsampling one or more positive wells of a Colilert® tray and plating onto an agar

plate to obtain single colonies for subsequent DNA extraction and phylotyping, as this has been shown to be the method that maximises strain diversity. Given the large growth rate differences between the B1-001 bloom strain and the phylogroup A bloom strains in Colilert-18® medium, even when B1-001 represents more than 95% of the cells present in the water sample, it may represent less than 20% of the cells present after the Quanti-Tray® has been incubated (Figure 6.5). Consequently, sampling a single colony means that the most likely outcome will be that the B1-001 bloom strain is not sampled and hence not identified, while the single colony is more likely to be of the more abundant A-000 or A-010 strain. This explains why the B1-001 strain has apparently 'disappeared' from recent bloom events. Even if direct PCR was applied to positive wells of a Colilert® tray, if the B1-001 bloom strain is present together with an A strain, PCR would be unlikely to detect B1-001 as it would be at a very low frequency given the growth rate advantage of the A strain.

Although the Colilert-18® defined substrate technology has increased specificity and applications in regulatory and public health aspects compared to the membrane filtration technique (Edberg and Edberg, 1988), caution should be taken when bloom water samples are assessed. Subsampling multiple colonies during assessment of strains might lead to the detection of the B1-001 strain. Further experimentation is needed to determine if changing the composition of Colilert-18® medium would be a solution, but any compositional change must ensure that the overall specificity and sensitivity of Colilert-18® is retained. Next generation versions of the Colilert® reagent are currently being applied in rapid, on-site detection systems. ColiMinder®, invented by the Viennese company Vienna Water Monitoring (VWM) Solutions, is one such fully automated version (https://www.vienna-water-monitoring.com/application-fields/). Like Colilert®, ColiMinder® detects *E. coli* and total

coliforms based on β-glucuronidase and β-galactosidase activity, respectively, and total bacteria based on alkaline phosphatase activity. It provides a rapid, efficient and economical assessment of microbiological water quality, and results can be obtained within 15 minutes (https://www.vienna-water-monitoring.com/application-fields/; Koschelnik et al., 2015; Koschelnik, 2016). As future work, it is worthwhile to investigate the growth characteristics of the bloom strains using ColiMinder®. Also as future work, environmental waters may be used to supply the large diversity of species expected, and spiked with the bloom strains that are pre-grown under different conditions, to better evaluate their responses, and to validate the current experimental results. Overall, an ecological study of the bloom water microbiome is warranted to understand the diversity of species present and their influence on the dynamics of *E. coli* bloom strains in Colilert-18®.

6.6 References

Abelson, P.H. (1954) Amino acid biosynthesis in *Escherichia coli*: isotopic competition with C14-glucose. *J Biol Chem* 206(1): 335-343.

Ahmed, Z.U., Sarker, M.R., and Sack, D.A. (1988) Nutritional requirements of shigellae for growth in a minimal medium. *Infect Immun* 56(4): 1007-1009.

Alm, E.W., Walk, S.T., and Gordon, D.M. (2011) The niche of *Escherichia coli*. In *Population genetics of bacteria: A tribute to Thomas S. Whittam*. Walk, S.T., and Feng, P.C.H. (eds). Washington, DC: ASM Press, pp. 69–89.

APHA, AWWA, WEF – American Public Health Association, American Water Works Association, Water Environment Federation. (2018) 9223 Enzyme substrate coliform test 2017. In *Standard methods for the examination of water and wastewater,* 23rd Ed. https://www.standardmethods.org/doi/abs/10.2105/SMWW.2882.194

Brenchley, J.E., Prival, M.J., and Magasanik, B. (1973) Regulation of the synthesis of enzymes responsible for glutamate formation in *Klebsiella aerogenes*. *J Biol Chem* 248(17): 6122-6128.

Castaño, I., Bastarrachea, F., and Covarrubias, A.A. (1988) *gltBDF* operon of *Escherichia coli*. *J Bacteriol* 170(2): 821-827.

Edberg, S.C., and Edberg, M.M. (1988) A defined substrate technology for the enumeration of microbial indicators of environmental pollution. *Yale J Biol Med* 61(5): 389-399.

Frampton, E.W., and Restaino, L. (1993) Methods for *Escherichia coli* identification in food, water and clinical samples based on beta-glucuronidase detection. *J Appl Microbiol* 74(3): 223-233.

Gordon, D.M. (2001) Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. *Microbiology* 147(5): 1079-1085.

Gordon, D.M., and Riley, M.A. (1999) A theoretical and empirical investigation of the invasion dynamics of colicinogeny. *Microbiology* 145(3): 655-661.

Goss, T.J., Perez-Matos, A., and Bender, R.A. (2001) Roles of glutamate synthase, *gltBD*, and *gltF* in nitrogen metabolism of *Escherichia coli* and *Klebsiella aerogenes*. *J Bacteriol* 183(22): 6607-6619.

IDEXX. (2019) Colilert-18. https://www.idexx.com/en/water/water-products-services/colilert-18/

Kämpfer, P., Rauhoff, O., and Dott, W. (1991) Glycosidase profiles of members of the family *Enterobacteriaceae*. *J Clin Microbiol* 29(12): 2877-2879.

Koschelnik, J. (2016) Application of Coliminder® technology for rapid detection of bacterial contamination in water: Practical examples and case studies. *International Conference and Expo on Water Microbiology and Novel Technologies – Water Microbiology 2016*. Chicago, USA, 18-19 July. https://www.longdom.org/conference-abstracts-files/1948-5948.C1.017-016.pdf

Koschelnik, J., Vogl, W., Epp, M., and Lackner, M. (2015) Rapid analysis of β-D-glucuronidase activity in water using fully automated technology. *Water Resour Manag* 8: 471.

Leclerc, H.D.A.A., Mossel, D.A.A., Edberg, S.C., and Struijk, C.B. (2001) Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. *Annu Rev Microbiol* 55(1): 201-234.

Liang, W.J., Wilson, K.J., Xie, H., Knol, J., Suzuki, S.I., Rutherford, N.G. et al. (2005) The *gusBC* genes of *Escherichia coli* encode a glucuronide transport system. *J Bacteriol* 187(7): 2377-2385.

Manafi, M., and Rotter, M.L. (1991) A new plate medium for rapid presumptive identification and differentiation of *Enterobacteriaceae*. *Int J Food Microbiol* 14(2): 127-134.

Monod, J. (1949) The growth of bacterial cultures. *Annu Rev Microbiol* 3(1): 371-394.

Neuwald, A.F., Krishnan, B.R., Brikun, I., Kulakauskas, S., Suziedelis, K., Tomcsanyi, T. et al. (1992) *cysQ*, a gene needed for cysteine synthesis in *Escherichia coli* K-12 only during aerobic growth. *J Bacteriol* 174(2): 415-425.

Novel, M., and Novel, G. (1976) Regulation of beta-glucuronidase synthesis in *Escherichia coli* K-12: constitutive mutants specifically derepressed for *uidA* expression. *J Bacteriol* 127(1): 406-417.

Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., and Slade, M.B. (2005) Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. *Environ Microbiol* 7(5): 631-640.

Szewzyk, U., Szewzyk, R., Manz, W., and Schleifer, K.-H. (2000) Microbiological safety of drinking water. *Annu Rev Microbiol* 54: 81-127.

U.S. EPA - United States Environmental Protection Agency. (2012) Recreational water quality criteria. Office of Water. 820-F-12-058. https://www.epa.gov/sites/production/files/2015-10/documents/rwqc2012.pdf

Wiser, M.J., and Lenski, R.E. (2015) A comparison of methods to measure fitness in *Escherichia coli*. *PLoS One* 10(5): e0126210.

Yan, D., Ikeda, T.P., Shauger, A.E., and Kustu, S. (1996) Glutamate is required to maintain the steady-state potassium pool in *Salmonella typhimurium*. *Proc Natl Acad Sci USA* 93(13): 6527-6531.

Chapter 7. General Discussion

7.1 *Escherichia coli* bloom events

Escherichia coli is generally perceived to be a normal inhabitant of the lower intestinal tract of warm-blooded animals (Savageau, 1983; Gordon and Cowling, 2003). *E. coli* released from a host through faeces survives for varying periods in aquatic and terrestrial environments, for instance, its half-life in water is only about a day, while in soil its about 1-5 days (Savageau, 1983). However, certain strains of *E. coli* are completely independent of a host for survival and are able to proliferate in the open environment (Carrillo et al., 1985; Hardina and Fujioka, 1991; Byappanahalli and Fujioka, 1998; Solo-Gabriele et al., 2000; Byappanahalli et al., 2003; Ishii et al., 2006).

Elevated counts of environmental *E. coli* reported from developed countries including the USA and Canada suggest that the elevated counts in these countries are typically less than 1000 cfu/100 ml. For example, the mean elevated *E. coli* concentrations in a stream that empties into a swimming beach along Lake Michigan, USA is 921 cfu/100 ml (Byappanahalli et al., 2003). In Lake Winnipeg in Canada, the elevated *E. coli* densities in bathing water is often less than 1000 cfu/100 ml (Williamson et al., 2004). In Australian bloom events by contrast, even the Colilert®-bias elevated counts are typically around 10,000 cfu/100 ml and often much higher, reaching levels as high as 100,000 cfu/100 ml. As defined by Power and colleagues (2005), a coliform 'bloom' is when a presumptive coliform count surpasses 10,000 cfu/100 ml, and so far it seems that *E. coli* 'bloom' events are restricted to Australia. Geographically, a major part of the Australian continent is desert and even the habitable regions are usually battered by water scarcity, where extended droughts are not uncommon. Bloom events occur when water temperatures are 18 ^{o}C or above, around autumn (Power et al., 2005), following the decay of aquatic vegetation and cyanobacterial blooms that release nutrients to water.

Further, the east coast of Australia where bloom events are frequent, experience short spells of strong winds that blow huge masses of dust from nearby desert and dry areas. In large lakes like Lake Winnipeg in Canada which extends over a massive area of 24,514 km², water rarely stays still; wind-induced wave action and fluctuations in the water level are frequent (Williamson et al., 2004). Most Australian lakes where bloom events occur are artificially impounded and are relatively small; Lake Burragorang spreads over an area of only 75 km^2, while the area of Lake Burley Griffin is only 6.64 km^2; they hardly ever spill and the water is mostly non-flowing. The absence of substantial wave action and water flow would cause nutrients to stay in much of the time. An interplay of these factors may be a reason behind the unprecedented *E. coli* densities observed in Australian water bodies compared to larger natural lakes in other developed countries.

Although the *E. coli* cell densities rise by two to three orders of magnitude at a bloom event, the final cell densities they reach is only 10^2 –10^3 cells/ml. In minimal media having a glucose concentration of 1 mM, *E. coli* were able to achieve a stationary phase cell density of approximately 10^8 cells/ml. Therefore, whatever the nutrients that the bloom strains are exploiting to produce the elevated counts would be present at relatively low concentrations. The elevated cell densities are known to drop to baseline levels within 2-3 days of a bloom event, indicating that the bloom strains exploit and exhaust the surplus nutrients in a short time period (Littlefield-Wyer, 2006). Of note, the nutrients that support bloom events are not yet defined.

The bloom strains are presumably free-living strains whose presence and persistence in water are beyond a doubt completely independent of faecal inputs. For instance, the elevated cell densities observed in Lake Burley Griffin in Canberra, ACT (Australian Capital Territory) during

bloom events would not be achieved even if every Canberran defecated directly to the lake for one whole week. If the elevated counts were to be caused by faecal input, it would also require that every person on average harbours the same dominant strain. This is highly impossible given that the likelihood of the same genotype being shared among two native Australian mammals is very rare (Gordon and Lee, 1999). Also of note, among 435 vertebrate *E. coli* isolates, none resembled any of the east coast bloom strains in terms of the genotypic or phenotypic profile (Power et al., 2005).

With regards to versatility, the bloom strains comprise two variants; the phylogroup A strains and the B1-001 bloom strain. Given the genotypic and phenotypic characteristics, the phylogroup A strains can establish in a vertebrate, while it is highly unlikely that the B1-001 bloom strain can. While the B1-001 strain is an auxotroph, the phylogroup A strains are prototrophic given their ability to grow profusely in minimal media (Chapter 6). Due to the absence of curli and flagella (Chapter 5), the B1-001 bloom strain is probably unable to produce aggregates or biofilms (Pratt and Kolter, 1998; Vidal et al., 1998; Prigent-Combaret et al., 2000), while the phylogroup A bloom strains are likely able to form biofilms. Among the genes over-represented in the phylogroup A bloom strains are *fim* genes encoding the type 1 pilus (Chapter 2) (Orndorff and Falkow, 1984; Klemm, 1986) which is a requirement for *E. coli* cells to attach to abiotic surfaces (Pratt and Kolter, 1998; Cookson et al., 2002). Taken together, the phylogroup A bloom strains may not be strictly free-living while all evidence suggests that the B1-001 bloom strain in fact is.

The B1-001 bloom strain has not only been present always but also been the dominant member in bloom water samples. Analyses performed in Lake Burley Griffin in Canberra, Tallowa Dam in New South Wales (Power et al., 2005), and several other reservoirs in the east

coast have shown that the B1-001 bloom strain persists even during non-bloom periods. The B1-001 bloom strain has also been isolated from the most recent bloom event reported from Victoria in January 2019. The occurrence of the B1-001 strain in water bodies several hundred kilometres apart may mean that this strain is widely distributed not only within but also among lakes and reservoirs. The B1-001 bloom strain is remarkable in that it is impaired in several traits that would be beneficial for its presumably free-living lifestyle (Chapter 5). Trait loss and resultant tiny genomes are typically observed in bacteria that engage in a symbiotic relationship with a host (Moran and Wernegreen, 2000; McCutcheon and Moran, 2012), and this is speculated to reflect host-dependence of the symbiont (McCutcheon and Moran, 2012). Loss of function such as motility, amino acid biosynthesis, and curli biogenesis is also observed in *Shigella*, which has evolved to become an obligate intracellular pathogen (Ahmed et al., 1988; Al Mamun et al., 1997; Sakellaris et al., 2000; Jin et al., 2002; Lan and Reeves, 2002). Yet, the B1-001 bloom strain is neither a symbiont, nor a *Shigella*, especially with the lack of pINV and shiga toxin genes. Evidence suggests that it is most likely a strain of *E. coli* (Chapter 5). *Shigella* are nevertheless interspersed with *E. coli* and are reckoned as clones of the latter (Ochman et al., 1983; Pupo et al., 2000). While the B1-001 bloom strain carries a bacteriophage exhibiting close similarity to bacteriophage SSU5 of *Salmonella enterica* serovar Typhimurium (Kim et al., 2012) (Chapter 5), it seems improbable that the phage confers a survival benefit. On the contrary, it may be that the growth retardation of the bloom strain in culture media is a result of the bacteriophage, if not solely due to auxotrophy. Overall, the numerical dominance and persistence of the B1-001 bloom strain over the likely more versatile phylogroup A bloom strains in bloom events, remain a mystery.

7.2 Assessment of bloom events

The Colilert-18® system, an optimised Colilert® formulation (Budnick et al., 2001), is routinely used by the Australian water authorities for the assessment of microbiological water quality (Hallas et al., 2008). Although it is unlikely that Colilert-18® will miss an entire bloom event, of note is how it affects strain diversity; Colilert-18® was shown to undermine strain diversity especially when bloom waters are assessed (Chapter 6). Subsampling multiple colonies, rather than a single colony, during assessment of strains might be a direct solution to detect B1-001 strain, yet subsequent phylotyping will be tedious. Also, given how heavily the B1-001 strain is outcompeted, it is rather doubtful that such a measure would ensure that the B1-001 strain is detected always.

Although the Colilert-18® system does not require confirmatory tests for detecting faecal indicators (Hallas et al., 2008), has higher specificity (Edberg and Edberg, 1988), and demands less time and expertise (Budnick et al., 2001) compared to the traditional membrane filtration (MF) techniques, the limitations of Colilert-18® observed in the current study place preference on MF technique over Colilert-18® particularly in an Australian perspective. The MF technique has sensitivity equal to Colilert-18® (Edberg and Edberg, 1988) and is known to provide accurate results with regards to capturing a wide spectrum of strain diversity when *E. coli* bloom events are assessed. The MF technique however is not without limitations such as subjectivity in result interpretation (Edberg and Edberg, 1988). In this respect, more alternative assessment methods need to be evaluated for suitability.

7.3 Attributes of bloom strains that likely contribute to elevated counts

As part of the current study, we attempted to investigate what cellular attributes of the bloom strains could have contributed to the elevated cell counts. All of our evidence suggests that the *Klebsiella* capsule which remarkably stands out in the Australian bloom strains, is likely the most influential factor. While frequency of capsule occurrence in bloom strains is 100%, it is only about 7% in *E. coli* overall (Nanayakkara et al., 2019) and 4% among water isolates (Chapter 3). Results show that neither the type of capsule acquired, nor the genomic background of a strain is necessarily important for *E. coli* to produce elevated counts. The *fecIRABCDE* operon which is the most pronounced genomic attribute over-represented in the phylogroup A bloom strains (Nanayakkara et al., 2019) did not have a significant effect on the growth rate, while the dominant growth response was caused by the capsule (Chapter 4). The capsule is the most distinctive feature that is harboured by the apparently 'impaired' B1-001 bloom strain as well. Overall, these observations suggest that the capsule alone is an essential attribute of bloom strains and, any *E. coli* strain that has acquired a capsule from *Klebsiella* might have the ability to produce bloom events. Capsules do occur in a variety of host-associated and environmental isolates (Nanayakkara et al., 2019), and should any of these isolates find its way to water, it might be able to produce bloom events provided that the environmental conditions are conducive. However, further investigation of environmental transcriptional regulation is needed to confirm if the capsule is essential to produce bloom events, or for some other aspect of the free-living survival of the strains.

During non-bloom periods bloom strains persist in water at densities less than 150 cells/100 ml, and following nutrient influx events, they rapidly increase in number to produce blooms (Power et al., 2005). The current study shows that the capsule may augment a strain's ability

to take advantage of nutrient influxes, leading to the production of elevated cell counts. It may be that the LamB mediated outer-membrane nutrient uptake system and/or the inner membrane-associated PTS system (phosphoenolpyruvate: carbohydrate phosphotransferase system) are better improved in encapsulated strains (Chapter 4). Or more simply, can the *Klebsiella* capsule act like a sponge to store nutrients, thereby enhancing nutrient uptake by the cell? Such a role may be determined by the functional structure of the capsule. The *Klebsiella* capsule is approximately 160 nm in thickness; more than ten times thicker than the *E. coli* K1 capsule which is less than 10 nm thick. The capsular structure of *Klebsiella* is formed by the arrangement of fine fibres in two distinct layers, where the innermost layer is dense while the outermost layer is loose and net-like (Amako et al., 1988). Taken together, the fact that the capsule can somehow contribute to an increased growth rate explains why it is only encapsulated strains, and not all co-occurring *E. coli*, that respond to nutrient influxes and produce elevated counts.

Both *E. coli* and *K. pneumoniae* colonise the intestinal tract, are faecal coliforms, and both can persist in water (Leclerc et al., 2001). This habitat sharing would enable the horizontal transfer of the capsule gene cluster between the two species. Considering the survival benefits the mucoid phenotype of the capsule confers on a cell (Weiner et al., 1995), it is likely that the bloom strains have acquired the capsule from *E. coli* strains that already possess the capsule, rather than *de novo* from *Klebsiella*.

Moreira and colleagues (2012) who investigated *E. coli* that persist on freshwater lake periphyton in Ontario, Canada, suggest that the ability to form biofilms leads to cell persistence in aquatic habitats. This is unlikely an explanation for the Australian bloom strains given that the B1-001 bloom strain is probably biofilm-deficient. Of note, most studies have

attempted to explain environmental re-growth and elevated counts of *E. coli* chiefly in terms of physical and environmental factors (Whitman et al., 2003; Williamson et al., 2004; Byappanahalli et al., 2007), and the cellular-level attributes, if any, that contribute to those elevated counts remain poorly understood.

7.4 Distribution of *Klebsiella* capsules in *E. coli*

Klebsiella capsules are found only in *E. coli* phylogenetic groups A, B1, and C and are absent from phylogroups B2, D, E, and F (Nanayakkara et al., 2019). The reason behind the absence of *Klebsiella* capsules in phylogroup B2, D, E, and F strains is not clear, but it is tempting to speculate that the genomic background of these phylogroups may be incompatible to maintain a capsule. The occurrence, though rare, of capsule-associated O8 and O9 among phylogroup B2, D, E, and F strains and the fact that the capsule gene cluster horizontally co-transfers with the O-antigen gene cluster from *Klebsiella* to *E. coli* propose that strains of these phylogroups may acquire a capsule but would be unable to retain it. Or else, the rarity of these O-types in B2, D, E, and F strains may suggest that the genomic background of these strains restricts the acquisition of the entire region of capsule and O-antigen from *Klebsiella*. Strains of phylogroups A and B1 are over-represented in environmental samples (Power et al., 2005; Gordon, 2013), while strains of phylogroups B2 and D are relatively rare in the external environment (Littlefield-Wyer, 2006; Gordon, 2013). The fact that the capsule may enhance strain persistence outside a host (Weiner et al., 1995; Rendueles et al., 2017) would explain the incidence of capsules in phylogroup A and B1 strains. Yet, the capsule is a major virulence factor for certain strains of *Klebsiella* (Simoons-Smit et al., 1986) especially for strains causing urinary tract infections (Struve and Krogfelt, 2003). It protects the bacterial cell from

phagocytosis, complement-mediated killing, and lethal effects of serum encountered inside a host (Williams et al., 1983; Álvarez et al., 2000; Lin et al., 2004). Given that strains of phylogroups B2, D, E, and F are frequently associated with hosts (Gordon and Cowling, 2003; Gordon, 2004; Nowrouzian et al., 2006; Jaureguy et al., 2008; Alm et al., 2011; Smati et al., 2013) and particularly B2 and to a lesser extent D strains are implicated in extraintestinal infections (Picard et al., 1999; Johnson et al., 2001a; Johnson et al., 2001b), the complete absence of *Klebsiella* capsules in these phylogroups is surprising and tempts further investigation.

7.5 The use of *E. coli* as a water quality indicator

The primary response by the water authorities in the event of elevated levels of *E. coli* in water is closing the water body for public use, as *E. coli* is of public health concern (Ishii and Sadowsky, 2008; U.S. EPA, 2012). Lake closures result in cancellation of sporting events, and considerable financial losses; according to the National Capital Authority of Australia, a bloom event has led to a loss of 80,000 AUD (Littlefield-Wyer, 2006). Yet, closure of water bodies due to bloom events would be an unnecessary action as the strains that contribute to elevated counts are most likely free-living *E. coli* whose presence is entirely independent of faecal input. The PCR-based protocol developed (Chapter 3) can be used by the water authorities for the detection and discrimination of bloom strains based on the presence of the *Klebsiella* capsule. This will aid prevent prolonged and unnecessary closures of water bodies due to future bloom events. The PCR protocol is an inexpensive and rapid tool in this regard, while Kaptive (Wyres et al., 2016), the software tool that detects *Klebsiella* capsules relies on whole genome

sequence data, the generation of which is expensive and demands time and expertise (Loman et al., 2012).

E. coli is widely used as an indicator of recent faecal contamination of drinking and recreational waters (Leclerc et al., 2001; Ishii and Sadowsky, 2008; WHO, 2017). A significant yet traditional assumption underlying its use as an indicator is that *E. coli* is incapable of cell division outside a host, and hence its presence in water indicates recent faecal contamination (Bonde, 1966; Leclerc et al., 2001; Alm et al., 2011). Adding to the wealth of evidence that confirms the ability of *E. coli* to proliferate in open environments (Carrillo et al., 1985; Solo-Gabriele et al., 2000; Whitman and Nevers, 2003; Whitman et al., 2003; Power et al., 2005; Byappanahalli et al., 2006; Ishii et al., 2006; Byappanahalli et al., 2007), *E. coli* bloom events are perhaps the most compelling and long-term evidence against the use of *E. coli* as a water quality indicator. As far as the recorded data goes, bloom events have occurred frequently since the 1970s in Lake Burragorang, the major water supply reservoir for metropolitan Sydney. Lake Burley Griffin, a recreational lake in the ACT has had bloom events from early 1990s. Similar events have been reported from several geographically distant localities across Australia including Queensland, Western Australia, and Victoria, which witnessed the most recent bloom event in January, 2019. The recurrence of *E. coli* bloom events in Australia calls for the evaluation and use of alternative indicators of water quality.

Intestinal enterococci are considered the best indicator for recreational fresh and marine waters in Australia (Australian Government NHMRC, 2008). The use of enterococci is also recommended by the World Health Organisation, United States, and the European Union (Boehm and Sassoubre, 2014). Considering the recurrence of *E. coli* bloom events, the Australian water authorities have shifted to use enterococci as the indicator organism in Lake

Burley Griffin. According to the recreational water quality guidelines defined by the NHMRC of the Australian Government (2008), recreational water is considered safe if the enterococcal count does not exceed 40 cfu/100 ml (95^{th} percentile). Of note, these guidelines differ among states. As per the U.S. EPA (2012), for recreational freshwater to be considered safe, the enterococcal count should not exceed 70 cfu/100 ml on a single-sample basis or the geometric mean of the enterococcal count should not exceed 35 cfu/100 ml in a 30-day interval.

If elevated counts of any usual indicator organism occur, it is recommended that water be assessed using an additional indicator to confirm if elevated counts are a result of proliferation in water. If the origin of the indicator cannot be traced, the Australian recreational water quality criteria recommend that *Clostridium perfringens* be enumerated (Australian Government NHMRC, 2008). Coliphages, intestinal enterococci, *Bacteroides* (Australian Government NHMRC, 2011), and thermotolerant coliforms (Australian Government NHMRC, 2011; WHO, 2017) are among the alternative indicators that can be used for drinking water in the event that the use of *E. coli* becomes confounding. Of note, no single indicator is universally perfect for all situations (Leclerc et al., 2001). Hence, selecting an indicator depending on the situation or the use of a combination of indicators would contribute to accurate water quality assessment.

7.6 Conclusions

E. coli bloom strains are represented by mainly two variants in terms of versatility; the B1-001 strain that is presumably free-living and unable to establish in a host, and the phylogroup A strains that are also free-living but may be able to establish in a host. The key feature that is

in common with all bloom strains is the *Klebsiella* capsule, which as per evidence so far, seems to be the trait that is most essential for conferring a bloom status on a strain of *E. coli*. We conclude that the possession of a capsule confers a growth rate advantage on *E. coli*.

We also conclude that the growth of different bloom strains in Colilert-18® is not uniform and leads to undermine the actual strain diversity. Care is needed when Colilert-18® is used for assessment of *E. coli* bloom waters.

Currently, a wealth of information suggests the ability of *E. coli* to proliferate in environments outside a host, confounding its use as a water quality indicator (Ishii and Sadowsky, 2008 and references therein; Alm et al., 2011). The recurrence of *E. coli* bloom events in freshwater reservoirs and recreational lakes across Australia urges a countrywide shift to alternative water quality indicators.

The current study also focussed on the occurrence of *Klebsiella* capsules in the species *E. coli* overall. *Klebsiella* capsules are not common in *E. coli*, yet capsules can occur in strains from a variety of hosts and water. Capsules occur only in strains from *E. coli* phylogroups A, B1, and C and are absent from phylogroups B2, D, E, and F. Capsule-positive strains exhibit a very non-random association with the O-antigen, where a highly limited subset of O-antigens occur in encapsulated *E. coli*. Both the *Klebsiella* capsule gene cluster and the adjacent O-antigen gene region in encapsulated *E. coli* are a result of a horizontal gene transfer event that occurred between *E. coli* and *Klebsiella* or encapsulated *E. coli*.

7.7 Future directions

In terms of explaining what exactly causes the *E. coli* bloom strains to produce elevated counts, the current study is limited in several aspects, including the lack of gene expression studies. qPCR can be used to quantify the expression level of target genes. Such an approach can be used to determine if there are gene expression level differences between bloom and non-bloom *E. coli* (Fitzmaurice et al., 2004; Carey et al., 2009). However, qPCR requires a set target gene and achieving this in terms of bloom strains requires a lot of further understanding. Analysis of the transcriptome of bloom strains using next-generation sequencing would be a more feasible and broad scale approach to understand gene regulation (Mutz et al., 2013). Transcriptome analyses can aid identify what genes impact on the bloom behaviour (growth rate/lag phase).

Further studies are required to elucidate the role of the *Klebsiella* capsule, to explore its involvement in nutrient uptake, cellular metabolism, and overall growth promotion. Functional analysis of genes can be carried out using knockout mutants (Ito et al., 2005). After knocking out the capsule gene region, the strains can be observed for survival under specific nutrient conditions. Growth rates of encapsulated strains and capsule-deficient mutants (Lawlor et al., 2005) of the same strains can be compared to determine if the loss of capsule genes leads to a growth rate difference. Also, the capsule gene cluster may be introduced to capsule-negative strains and the resultant growth effects can be studied. Knockout mutations of capsule regulatory genes *rcsB*, *rcsC* for example, or nutrient uptake-related genes *ompF*, *lamB* can be used to infer the effect of these genes on growth rate. Over-expression of capsule synthesis regulator *rcsB* is speculated to cause cells to divide earlier (Carballès et al., 1999),

and this is a lead worth investigating as the current study also shows that capsule-positive strains have shorter lag phases compared to capsule-negative strains (Chapter 4).

To determine if sugars accumulate in the capsule, radiolabelled sugars, including trehalose, can be provided to cells followed by measuring the radioactivity in the capsule (Liu et al., 1999), preferably upon a short incubation time to ensure that sugars are not transported into the cell interior. The functional structure of the capsule can be studied to determine if it augments nutrient uptake by the cell, either by acting as a sponge to store nutrients or through any other means. Advanced microscopic techniques would be useful in exploring the fine structure of the capsule (Amako et al., 1988). Independently looking at trehalose-effect(s) would be useful to elucidate if encapsulated strains possess an enhanced trehalose-related nutrient uptake/utilisation system, as speculated in Chapter 4.

The ecological aspects of the bloom microbiome need to be studied further, i.e., the diversity of bacteria, protozoa and coliphages that may be present and their influences on bloom events. The environmental factors that contribute to *E. coli* bloom events need to be explored in detail. Acquisition of a statistically amenable dataset would require that data is continually collected through several years. Constant monitoring of water bodies would be needed as bloom events do not last long; the cell densities drop to baseline levels within 2-3 days of a bloom event. A long-term chemical analysis of water from bloom events, across multiple geographical regions may help determine what exact nutrients the bloom strains exploit as substrates.

7.8 References

Ahmed, Z.U., Sarker, M.R., and Sack, D.A. (1988) Nutritional requirements of shigellae for growth in a minimal medium. *Infect Immun* 56(4): 1007-1009.

Alm, E.W., Walk, S.T., and Gordon, D.M. (2011) The niche of *Escherichia coli*. In *Population genetics of bacteria: A tribute to Thomas S. Whittam*. Walk, S.T., and Feng, P.C.H. (eds). Washington, DC: ASM Press, pp. 69–89.

Al Mamun, A.A., Tominaga, A., and Enomoto, M. (1997) Cloning and characterization of the region III flagellar operons of the four *Shigella* subgroups: genetic defects that cause loss of flagella of *Shigella boydii* and *Shigella sonnei*. *J Bacteriol* 179(14): 4493-4500.

Álvarez, D., Merino, S., Tomás, J.M., Benedí, V.J., and Albertí, S. (2000) Capsular polysaccharide is a major complement resistance factor in lipopolysaccharide O side chain-deficient *Klebsiella pneumoniae* clinical isolates. *Infect Immun* 68(2): 953-955.

Amako, K., Meno, Y., and Takade, A. (1988) Fine structures of the capsules of *Klebsiella pneumoniae* and *Escherichia coli* K1. *J Bacteriol* 170(10): 4960-4962.

Australian Government NHMRC - National Health and Medical Research Council. (2008) Guidelines for managing risks in recreational water. https://www.nhmrc.gov.au/sites/default/files/images/guidelines-for-managing-risks-in-recreational-water.pdf

Australian Government NHMRC - National Health and Medical Research Council. (2011) Australian drinking water guidelines 6. Version 3.5. Updated August 2018. https://www.nhmrc.gov.au/sites/default/files/documents/NHMRC%20ADWG%206%20-%20Version%203.5%20-%20Proof%203.pdf

Boehm, A.B., and Sassoubre, L.M. (2014) Enterococci as indicators of environmental fecal contamination. In *Enterococci: From commensals to leading causes of drug resistant infection.* Gilmore, M.S., Clewell, D.B., Ike, Y., and Shankar, N. (eds). Boston: Massachusetts Eye and Ear Infirmary.

Bonde, G.J. (1966) Bacteriological methods for estimation of water pollution. *Health laboratory science* 3(2): 124-128.

Budnick, G.E., Howard, R.T., and Mayo, D. (2001) Comparison of Colilert-18 to the mTEC agar method for the enumeration of *Escherichia coli* in recreational waters. In *American Society of Microbiologists Annual Convention*, *May*.

Byappanahalli, M.N., and Fujioka, R.S. (1998) Evidence that tropical soil environment can support the growth of *Escherichia coli*. *Water Sci Technol* 38(12): 171.

Byappanahalli, M., Fowler, M., Shively, D., and Whitman, R. (2003) Ubiquity and persistence of *Escherichia coli* in a Midwestern coastal stream. *Appl Environ Microbiol* 69(8): 4549-4555.

Byappanahalli, M.N., Whitman, R.L., Shively, D.A., Sadowsky, M.J., and Ishii, S. (2006) Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. *Environ Microbiol* 8(3): 504-513.

Byappanahalli, M.N., Whitman, R.L., Shively, D.A., Ferguson, J., Ishii, S., and Sadowsky, M.J. (2007) Population structure of *Cladophora*-borne *Escherichia coli* in nearshore water of Lake Michigan. *Water Res* 41(16): 3649-3654.

Carballès, F., Bertrand, C., Bouché, J-P., and Cam, K. (1999) Regulation of *Escherichia coli* cell division genes *ftsA* and *ftsZ* by the two-component system *rcsC–rcsB*. *Mol Microbiol* 34(3): 442 – 450.

Carey, C.M., Kostrzynska, M., and Thompson, S. (2009) *Escherichia coli* O157: H7 stress and virulence gene expression on Romaine lettuce using comparative real-time PCR. *J Microbiol Methods* 77(2): 235-242.

Carrillo, M., Estrada, E., and Hazen, T.C. (1985) Survival and enumeration of the fecal indicators *Bifidobacterium adolescentis* and *Escherichia coli* in a tropical rain forest watershed. *Appl Environ Microbiol* 50(2): 468-476.

Cookson, A.L., Cooley, W.A., and Woodward, M.J. (2002) The role of type 1 and curli fimbriae of Shiga toxin-producing *Escherichia coli* in adherence in abiotic surfaces. *Int J Med Microbiol* 292(3-4): 195-205.

Edberg, S.C., and Edberg, M.M. (1988) A defined substrate technology for the enumeration of microbial indicators of environmental pollution. *Yale J Biol Med* 61(5): 389-399.

Fitzmaurice, J., Glennon, M., Duffy, G., Sheridan, J.J., Carroll, C., and Maher, M. (2004) Application of real-time PCR and RT-PCR assays for the detection and quantitation of VT 1 and VT 2 toxin genes in *E. coli* O157: H7. *Mol Cell Probes* 18(2): 123-132.

Gordon, D.M. (2004) The influence of ecological factors on the distribution and the genetic structure of *Escherichia coli*. *EcoSal Plus* 1(1).

Gordon, D.M. (2013) The ecology of *Escherichia coli*. In *Escherichia coli: Pathotypes and principles of pathogenesis.* Donnenberg, M.S. (ed). San Diego, CA: Academic Press, pp. 3-20.

Gordon, D.M., and Cowling, A. (2003) The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology* 149(12): 3575-3586.

Gordon, D.M., and Lee, J. (1999) The genetic structure of enteric bacteria from Australian mammals. *Microbiology* 145(10): 2673-2682.

Hallas, G., Giglio, S., Capurso, V., Monis, P.T., and Grooby, W.L. (2008) Evaluation of chromogenic technologies for use in Australian potable water. *J Appl Microbiol* 105(4): 1138-1149.

Hardina, C.M., and Fujioka, R.S. (1991) Soil: the environmental source of *Escherichia coli* and enterococci in Hawaii's streams. *Environ Toxicol Water Qual* 6(2): 185-195.

Ishii, S., and Sadowsky, M.J. (2008) *Escherichia coli* in the environment: Implications for water quality and human health. *Microbes Environ* 23(2): 101-108.

Ishii, S., Ksoll, W.B., Hicks, R.E., and Sadowsky, M.J. (2006) Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. *Appl Environ Microbiol* 72(1): 612-621.

Ito, M., Baba, T., Mori, H., and Mori, H. (2005) Functional analysis of 1440 *Escherichia coli* genes using the combination of knock-out library and phenotype microarrays. *Metab Eng* 7(4): 318-327.

Jaureguy, F., Landraud, L., Passet, V., Diancourt, L., Frapy, E., Guigon, G. et al. (2008) Phylogenetic and genomic diversity of human bacteremic *Escherichia coli* strains. *BMC genomics* 9(1): 560.

Jin, Q., Yuan, Z., Xu, J., Wang, Y., Shen, Y., Lu, W. et al. (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.

Johnson, J.R., Delavari, P., Kuskowski, M., and Stell, A.L. (2001a) Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. *J Infect Dis* 183(1): 78-88.

Johnson, J.R., Stell, A.L., Delavari, P., Murray, A.C., Kuskowski, M., and Gaastra, W. (2001b) Phylogenetic and pathotypic similarities between *Escherichia coli* isolates from urinary tract infections in dogs and extraintestinal infections in humans. *J Infect Dis* 183(6): 897-906.

Kim, M., Kim, S., and Ryu, S. (2012) Complete genome sequence of bacteriophage SSU5 specific for *Salmonella enterica* serovar Typhimurium rough strains. *J Virol* 86(19): 10894-10894.

Klemm, P. (1986) Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *Escherichia coli*. *EMBO J* 5(6): 1389-1393.

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

Lawlor, M.S., Hsu, J., Rick, P.D., and Miller, V.L. (2005) Identification of *Klebsiella pneumoniae* virulence determinants using an intranasal infection model. *Mol Microbiol* 58(4): 1054-1073.

Leclerc, H.D.A.A., Mossel, D.A.A., Edberg, S.C., and Struijk, C.B. (2001) Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. *Annu Rev Microbiol* 55(1): 201-234.

Lin, J.C., Chang, F.Y., Fung, C.P., Xu, J.Z., Cheng, H.P., Wang, J.J. et al. (2004) High prevalence of phagocytic-resistant capsular serotypes of *Klebsiella pneumoniae* in liver abscess. *Microbes infect* 6(13): 1191-1198.

Littlefield-Wyer, J.G. (2006) Genetic and phenotypic characterisation of *Escherichia coli* isolated from the primary and secondary habitat (PhD Thesis). Canberra, ACT, Australia: The Australian National University.

Liu, J.Y., Miller, P.F., Gosink, M., and Olson, E.R. (1999) The identification of a new family of sugar efflux pumps in *Escherichia coli*. *Mol Microbiol* 31(6): 1845-1851.

Loman, N.J., Constantinidou, C., Chan, J.Z., Halachev, M., Sergeant, M., Penn, C.W. et al. (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol* 10(9): 599-606.

McCutcheon, J.P., and Moran, N.A. (2012) Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol* 10(1): 13-26.

Moran, N.A., and Wernegreen, J.J. (2000) Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol* 15(8): 321-326.

Moreira, S., Brown, A., Ha, R., Iserhoff, K., Yim, M., Yang, J. et al. (2012). Persistence of *Escherichia coli* in freshwater periphyton: biofilm-forming capacity as a selective advantage. *FEMS Microbiol Ecol* 79(3): 608-618.

Mutz, K.O., Heilkenbrinker, A., Lönne, M., Walter, J.G., and Stahl, F. (2013) Transcriptome analysis using next-generation sequencing. *Curr Opin Biotechnol* 24(1): 22-30.

Nanayakkara, B.S., O'Brien, C.L., and Gordon, D.M. (2019) Diversity and distribution of *Klebsiella* capsules in *Escherichia coli*. *Environ Microbiol Rep* 11(2): 107-117.

Nowrouzian, F.L., Adlerberth, I., and Wold, A.E. (2006) Enhanced persistence in the colonic microbiota of *Escherichia coli* strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. *Microbes Infect* 8(3): 834-840.

Ochman, H., Whittam, T.S., Caugant, D.A., and Selander, R.K. (1983) Enzyme polymorphism and genetic population structure in *Escherichia coli* and *Shigella*. *J Gen Microbiol* 129: 2715–2726.

Orndorff, P.E., and Falkow, S. (1984) Organization and expression of genes responsible for type 1 piliation in *Escherichia coli*. *J Bacteriol* 159(2): 736-744.

Picard, B., Garcia, J.S., Gouriou, S., Duriez, P., Brahimi, N., Bingen, E. et al. (1999) The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect Immun* 67(2): 546-553.

Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., and Slade, M.B. (2005) Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. *Environ Microbiol* 7(5): 631-640.

Pratt, L.A., and Kolter, R. (1998) Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Mol Microbiol* 30(2): 285-293.

Prigent-Combaret, C., Prensier, G., Le Thi, T.T., Vidal, O., Lejeune, P., and Dorel, C. (2000) Developmental pathway for biofilm formation in curli-producing *Escherichia coli* strains: role of flagella, curli and colanic acid. *Environ Microbiol* 2(4): 450-464.

Pupo, G.M., Lan, R., and Reeves, P.R. (2000) Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. *Proc Natl Acad Sci USA* 97(19): 10567-10572.

Rendueles, O., Garcia-Garcerà, M., Néron, B., Touchon, M., and Rocha, E.P.C. (2017) Abundance and co-occurrence of extracellular capsules increase environmental breadth: Implications for the emergence of pathogens. *PLOS Pathog* 13(7): e1006525.

Sakellaris, H., Hannink, N.K., Rajakumar, K., Bulach, D., Hunt, M., Sasakawa, C., and Adler, B. (2000) Curli loci of *Shigella* spp. *Infect Immun* 68(6): 3780-3783.

Savageau, M.A. (1983) *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. *Amer Nat* 122(6): 732-744.

Simoons-Smit, A.M., Verweij-van Vught, A.M.J.J., and MacLaren, D.M. (1986) The role of K antigens as virulence factors in *Klebsiella*. *J Med Microbiol* 21(2): 133-137.

Smati, M., Clermont, O., Le Gal, F., Schichmanoff, O., Jauréguy, F., Eddi, A. et al. (2013) Real-time PCR for quantitative analysis of human commensal *Escherichia coli* populations reveals a high frequency of sub-dominant phylogroups. *Appl Environ Microbiol* 79(16): 5005-5012.

Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R., and Palmer, C.J. (2000) Sources of *Escherichia coli* in a coastal subtropical environment. *Appl Environ Microbiol* 66(1): 230-237.

Struve, C., and Krogfelt, K.A. (2003) Role of capsule in *Klebsiella pneumoniae* virulence: lack of correlation between in vitro and in vivo studies. *FEMS Microbiol Lett* 218(1): 149-154.

U.S. EPA - United States Environmental Protection Agency (2012) Recreational water quality criteria. Office of Water. 820-F-12-058. https://www.epa.gov/sites/production/files/2015-10/documents/rwqc2012.pdf

Vidal, O., Longin, R., Prigent-Combaret, C., Dorel, C., Hooreman, M., and Lejeune, P. (1998) Isolation of an *Escherichia coli* K-12 mutant strain able to form biofilms on inert surfaces: involvement of a new *ompR* allele that increases curli expression. *J Bacteriol* 180(9): 2442-2449.

Weiner, R., Langille, S., and Quintero, E. (1995) Structure, function and immunochemistry of bacterial exopolysaccharides. *J Ind Microbiol* 15(4): 339-346.

Whitman, R.L., and Nevers, M.B. (2003) Foreshore sand as a source of *Escherichia coli* in nearshore water of a Lake Michigan beach. *Appl Environ Microbiol* 69(9): 5555-5562.

Whitman, R.L., Shively, D.A., Pawlik, H., Nevers, M.B., and Byappanahalli, M.N. (2003) Occurrence of *Escherichia coli* and enterococci in *Cladophora* (Chlorophyta) in nearshore water and beach sand of Lake Michigan. *Appl Environ Microbiol* 69(8): 4714-4719.

WHO-World Health Organization. (2017) Guidelines for drinking-water quality, Microbial aspects, 4th edn. incorporating the 1st addendum. World Health Organization, Geneva.

Williams, P., Lambert, P.A., Brown, M.R., and Jones, R.J. (1983) The role of the O and K antigens in determining the resistance of *Klebsiella aerogenes* to serum killing and phagocytosis. *Microbiology* 129(7): 2181-2191.

Williamson, D.A., Ralley, W.E., Bourne, A., Armstrong, N., Fortin, R., and Hughes, C.E. (2004) Principal factors affecting *Escherichia coli* at Lake Winnipeg beaches. Manitoba Water Stewardship Report No. 2004-01. Water Quality Management Section, Manitoba Water Stewardship, Canada.

Wyres, K.L., Wick, R.R., Gorrie, C., Jenney, A., Follador, R., Thomson, N.R., and Holt, K.E. (2016) Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom* 2(12): e000102.

Appendix. Publication

Diversity and distribution of *Klebsiella* capsules in *Escherichia coli*

Journal: Environmental Microbiology Reports (2019)

Available at
https://onlinelibrary.wiley.com/doi/abs/10.1111/1758-2229.12710

Diversity and distribution of *Klebsiella* capsules in *E. coli*

Buddhie S. Nanayakkara[1,2], Claire L. O'Brien[3,4] and David M. Gordon[1]

[1]Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT 2601, Australia.

[2]Department of Botany, Faculty of Science, University of Peradeniya, Peradeniya 20400, Sri Lanka.

[3]Medical School, The Australian National University, Canberra, ACT 2601, Australia.

[4]Gastroenterology and Hepatology Unit, Canberra Hospital, Canberra, ACT, Australia.

Corresponding author: David M. Gordon, Division of Ecology and Evolution, Research School of Biology, ANU College of Science, RN Robertson Building, 46 Sullivans Creek Road, The Australian National University, Canberra, ACT 2601, Australia.

Telephone: 61 2 6125 3552

Email: David.Gordon@anu.edu.au

Running Title: Occurrence of *Klebsiella* capsules in *E. coli*

Originality-Significance Statement

The acquisition of the capsule genes from *Klebsiella* is a key element in determining whether an *E. coli* strain can cause significantly elevated counts (blooms) in a water body. The distribution of encapsulated strains in *E. coli* is non-random and the genomic background of strains belonging to phylogroups B2, D, E, and F appear to be incompatible with the maintenance of the *Klebsiella* capsule. The type of *Klebsiella* capsule encoded, and within phylogroup A, the variable gene content of the strain do not appear to play a role in enabling an encapsulated strain to cause bloom events.

Summary

E. coli strains responsible for elevated counts in freshwater reservoirs in Australia carry a capsule originating from *Klebsiella*. The occurrence of *Klebsiella* capsules in *E. coli* was about 7% overall and 23 different capsule types were detected. Capsules were observed in strains from phylogroups A, B1, and C, but were absent from phylogroup B2, D, E, and F strains. In general, few A, B1, or C lineages were capsule-positive, but when a lineage was encapsulated multiple different capsule types were present. All *Klebsiella* capsule-positive strains were of serogroups O8, O9, and O89. Regardless of the phylogroup, O9 strains were more likely to be capsule-positive than O8 strains. Given the sequence similarity, it appears that both the capsule region and the O-antigen gene region are transferred to *E. coli* from *Klebsiella* as a single block via horizontal gene transfer events. Pan genome analysis indicated that there were only modest differences between encapsulated and non-encapsulated strains belonging to phylogroup A. The possession of a *Klebsiella* capsule, but not the type of capsule, is likely a key determinant of the bloom status.

Introduction

Contamination of drinking and recreational waters by animal faeces poses a significant human health risk, as faeces may serve as a source of harmful pathogens, such as the hepatitis A virus (Szewzyk *et al*., 2000; Gordon, 2001; Alm *et al*., 2011; U. S. EPA, 2012). Faecal-associated bacteria have long been used as faecal indicator species. An ideal faecal indicator, among other attributes, should not be present in a water body in the absence of faecal contamination and be unable to multiply outside a host (Bonde, 1966; Power *et al*., 2005; Bitton, 2011). *Escherichia coli* has been assumed to exhibit many of the attributes of an ideal indicator, and for many years the species has been widely used as an indicator of recent faecal contamination of drinking and recreational waters (Edberg *et al*., 2000; U. S. EPA, 2012; Gordon, 2013). However, there is a growing body of evidence suggesting that *E. coli* can not only survive for extended periods, but also proliferate in environments such as water, soil, algae, plants, and manure (Kudva *et al*., 1998; Jiang *et al*., 2002; Ishii and Sadowsky, 2008; Vital *et al*., 2008; val Elsas *et al*., 2011), thus confounding its use as a water quality indicator (Bonde, 1966; Alm *et al*., 2011).

In Australia, significantly elevated *E. coli* counts have been reported from fresh water reservoirs and recreational waters (Power *et al*., 2005). These elevated count events have been termed *E. coli* 'bloom' events, as counts from 10,000 – 100,000 cells/100 ml of water have been reported; counts well above the 'safe' level of 235 cfu/100 ml in a single sample (Ishii and Sadowsky, 2008; U. S. EPA, 2012). Sanitary surveys indicate that these elevated count events cannot be attributed to faecal contamination, and indeed, achieving cell counts this high would require an unachievable level of faecal contamination (Power, *et al*., 2005). The strains responsible for the bloom events can be isolated from the water bodies at any

time, and have not been detected in the faeces of humans or other animals. These outcomes suggest that the presence of these strains in a water body is largely independent of faecal inputs and indeed, the strains responsible may represent, free-living *E. coli* (Power *et al.*, 2005; Alm *et al*., 2011).

Relatively few strains have been found to be responsible for *E. coli* bloom events, and all of the strains isolated from bloom events have a mucoid phenotype and encode a group 1 capsule originating from *Klebsiella* (Power *et al*., 2005; unpublished data). Restriction fragment length polymorphism (RFLP) analysis of the capsule region has revealed that the group 1 capsules possessed by the strains are not identical (Power *et al*., 2005).

The polysaccharide capsule is the outermost layer that envelopes the bacterial cell surface (Whitfield and Roberts, 1999). In *E. coli*, capsules are classified into four major groups (Whitfield and Roberts, 1999). The group 1 capsule exhibits high similarity to those of *Klebsiella* in structural, genetic, and expression terms (Amor and Whitfiled, 1997; Rahn *et al*., 1999; Whitfiled, 2006). The biosynthesis of *Klebsiella* capsules and *E. coli* group 1 capsules occurs via a *wzy*-dependent polymerization pathway (Whitfiled, 2006). The *Klebsiella* capsule spans a 10-30 kbp region in the genome and has been implicated in virulence (Williams *et al*., 1990; Lawlor *et al*., 2005; Schembri *et al*., 2005; Wyres *et al*., 2016). It protects the cell from phagocytosis, complement-mediated killing and lethal effects of serum (Kabha *et al*., 1995; Hsu *et al*., 2011). Similarly, in the external environment the capsule protects the bacterial cell from adverse conditions encountered including desiccation, UV radiation, predation and bacteriophage infection (Weiner *et al*., 1995; Rendueles *et al*., 2017).

Given the strong association of the Australian 'bloom' status with the possession of a *Klebsiella* capsule, the present study aimed to examine the diversity and distribution of capsules in

known *E. coli* bloom strains, and in a large collection of *E. coli* isolated from a variety of sources.

Results

Diversity and distribution of capsule types

Of the 1194 Australian *E. coli* strains screened, 82 strains (6.9%) were *Klebsiella* capsule-positive. The encapsulated strains comprised 53 out of 332 phylogroup A (16%), 21 out of 300 phylogroup B1 (7%), and eight out of 19 phylogroup C (42%) strains. Capsules were not observed among the 261 phylogroup B2, 142 D, 72 E, and 68 F strains. Encapsulated strains detected in this study represented 23 of the 134 distinct capsule synthesis loci reported by Wyres and colleagues (2016). Forty different capsule types were found among the capsule-positive strains from the NCBI database. Among the 99 EnteroBase strains with a serogroup of O8, O9, O89, 27 strains (27%) were capsule-positive, representing 17 different capsule types. The distribution of capsules among the phylogroup A, B1, and C strains is depicted in Figure 1.

Among the Australian phylogroup A and B1 strains, the frequency of encapsulated strains varied with respect to the phylogroup membership of the strain and the source of the isolate (nominal logistic regression: phylogroup, $p = 0.727$; source, $p = 0.028$; phylogroup*source, $p < 0.0001$) (Table 1). Phylogroup C strains were excluded as the number of isolates was too small for meaningful analysis (Table 1). Phylogroup A strains isolated from water (excluding bloom strains) were more likely to be encapsulated while strains from native birds and poultry meat were unlikely to be encapsulated (Table 1). By contrast, B1 strains from water were rarely encapsulated, while encapsulated strains were most likely to be detected among the poultry

meat B1 isolates (Table 1). In general, encapsulated strains can be isolated from a variety of vertebrate hosts or from water samples. However, the bloom strains have never been isolated from a vertebrate host (Power *et al*., 2005).

A number of phylogroup B2, D, and A strains were identified by Kaptive as encoding the *Klebsiella* capsule deletion variant KL156-D1. These strains were not considered to be true capsule-positive strains because: KL156-D1 co-occurred with the colanic acid gene cluster, which encapsulated strains do not harbour (Jayaratne *et al*., 1993; Whitfield and Paiment, 2003; Whitfield, 2006); we detected a flanking region having *wzzB*, which determines the length of O-antigen polysaccharide in *wzy*-dependent O-serogroups, and is associated with non-encapsulated strains (Whitfield and Roberts, 1999; Iguchi *et al*., 2014); the serogroups of the strains were typically O-types other than the capsule-associated O8, O9, and O89; and they lacked the characteristic colony morphology of encapsulated strains.

A total of 35 bloom strains isolated at different time periods from multiple localities in Australia were used in the study (Supplemental Table 1). Each of the strains known to have been responsible for 'bloom' events exhibited a different capsule type. The eastern Australian bloom strains belonged to phylogroups A and B1. The phylogroup A sequence type 10 (ST10) strains had either capsule type KL16 or KL113, while the ST609 strains had capsule type KL49. The phylogroup B1 bloom strain (ST1494) had a KL53 capsule type. Five strains: four phylogroup A strains and one B1 strain, were isolated from a single bloom event in Western Australia and each strain had a different capsule type, i.e., KL31 (ST58), KL53, KL60 (ST10), KL63 and KL101 (ST227).

The encapsulated strains were present in a limited subset of all lineages, regardless of their phylogroup membership (Figure 1). When a clade was capsule-positive, it contained multiple capsule types. In other words, closely related lineages carried diverse capsule types.

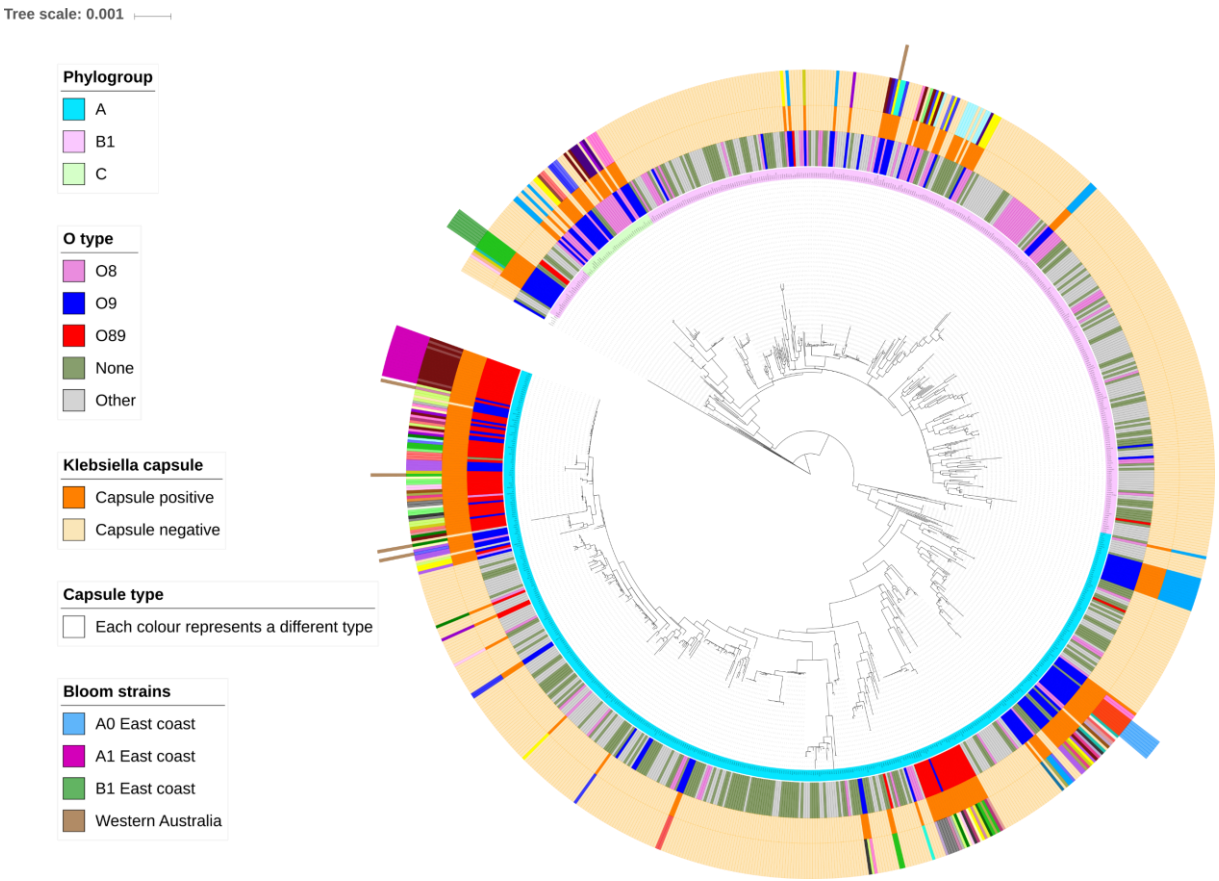


Figure 1. Distribution of *Klebsiella* capsules and O-types in *E. coli* phylogroups A, B1, and C. The phylogenetic tree was constructed using the whole genome sequences of 869 *E. coli* strains. These included environmental and host-associated strains from an Australian collection, and those downloaded from NCBI and EnteroBase. The innermost ring depicts the phylogroup membership of the strains; A, B1, or C. The second innermost ring denotes the O-type of the strains, while the third ring denotes the presence/absence of a *Klebsiella* capsule. The fourth ring indicates the capsule type, with each colour corresponding to a different capsule type. The outermost ring depicts the bloom strains. The tree has been rooted on the phylogroup B2 strain *E. coli* ED1a.

Table 1. The occurrence of *Klebsiella* capsules in terms of the phylogroup and source of isolation of the Australian *E. coli* strains.

Phylogroup	Source*	Number capsule-positive	Number capsule-negative	% Capsule-positive
A	Bird	0	56	0
	Human	12	77	13.5
	Non-human mammal	3	27	10.0
	Poultry meat	2	89	2.2
	Water	9	29	23.7
B1	Bird	1	48	2.0
	Human	3	38	7.3
	Non-human mammal	2	32	5.9
	Poultry meat	7	41	14.6
	Water	1	118	0.8
C	Bird	0	0	0
	Human	6	5	54.5
	Non-human mammal	0	1	0
	Poultry meat	1	4	20.0
	Water	1	0	100

* The bloom strains and 4 strains isolated from fish or reptiles were excluded from the analysis.

Capsule Region

The capsule region was flanked by *galF* (UDP-glucose pyrophosphorylase) and *ugd* (UDP-glucose 6-dehydrogenase). The region between *galF* and *ugd* among encapsulated strains varied from 13.4 kbp – 30.3 kbp and encoded an average of 15 genes.

Previous studies have shown that the group 1 capsule has been inserted between the *his* (histidine biosynthesis) operon and *galF* in *E. coli* (Whitfield, 2006) (Figure 2). In *E. coli* K-12 the O-antigen gene cluster is flanked by the *his* operon and *galF*, while the colonic acid gene cluster is located upstream of *galF*. Encapsulated strains notably lacked genes of the *wzy*-dependent O-antigen clusters, including *wzzB* (Batchelor *et al*., 1991; Iguchi *et al*., 2014). The *wzx/wzy* genes that typically determine the O-antigen in *E. coli* have, in encapsulated strains, been replaced by *wzt* and *wzm*. The best BLAST query hits (>95%identity) in NCBI for *wzt* from

encapsulated *E. coli* strains were for *Klebsiella* and encapsulated *E. coli*, indicating a likely *Klebsiella* origin for *wzt*. Transposases/insertion sequence (IS) elements were present on either side of the capsule cluster in most, but not all encapsulated strains. The position of the capsule region insertion was not identical in all encapsulated strains, but occurred within 1-9 genes upstream of the *his* operon. The arrangement of *his*, O-antigen region (*rfb*) and the capsule region (*cps*) of *Klebsiella,* was comparable to that of encapsulated *E. coli* strains (Figure 2). The gene region for colonic acid biosynthesis did not co-occur with the *Klebsiella* capsule.

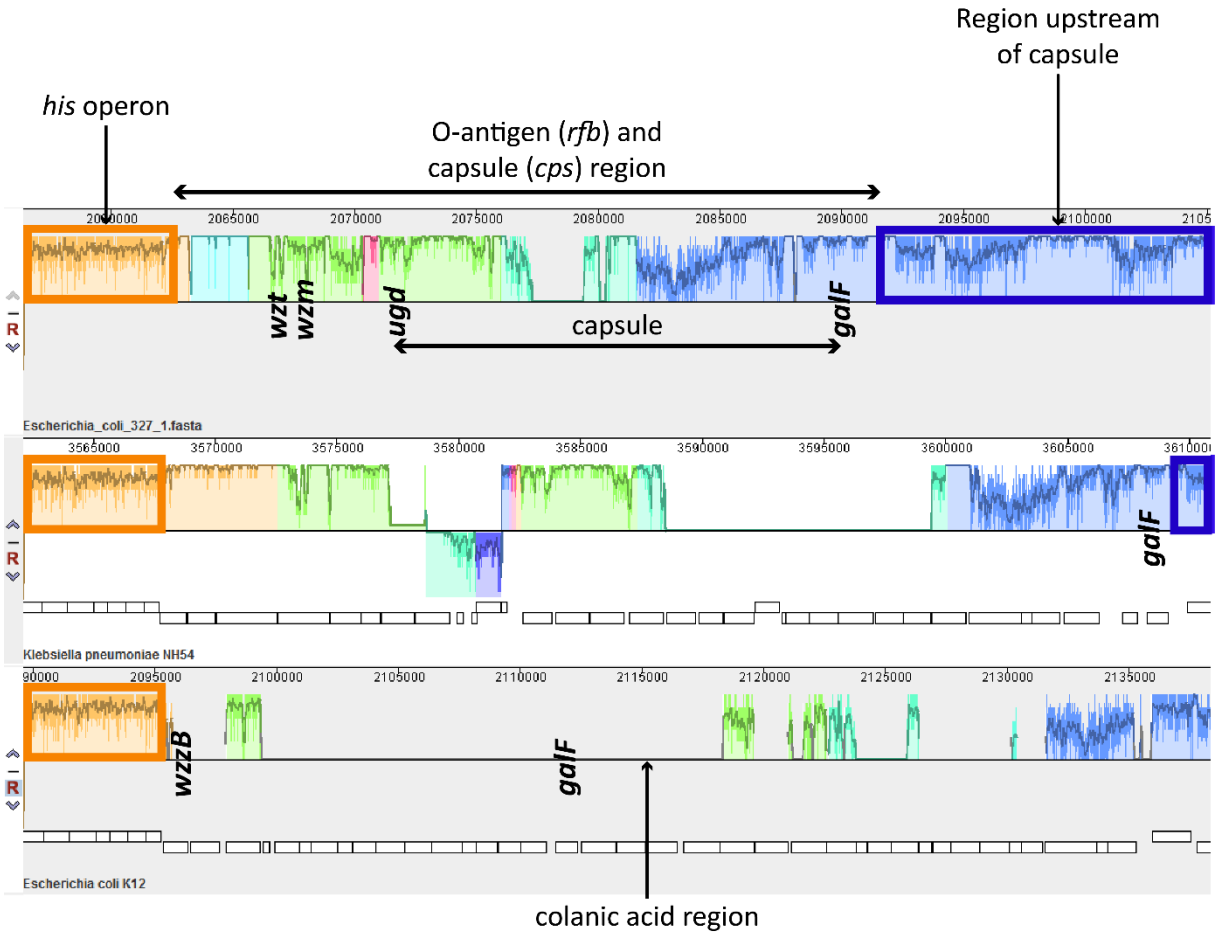


Figure 2. progressiveMauve alignment of the *his* operon (outlined in orange), O-antigen region, and the capsule locus of *E. coli* bloom strain 327_1, *Klebsiella pneumoniae* NH54 and capsule-negative strain *E. coli* K-12, in order from top to bottom. Coloured regions represent locally collinear blocks (LCBs). The upstream region of the capsule is outlined in blue and contains *yegH* and *asmA*. In *E. coli* K-12, the region with no homology to the other two strains is the colanic acid gene cluster.

Association of the capsule with the O-antigen and other capsule types

All capsule-positive strains were O8, O9, or O89 (220 out of 240 strains; 91.7%) or the serogroup could not be determined. However, not all of the O8, O9, and O89 strains were encapsulated (Table 2). Capsule-negative strains were of O serogroups which were both *wzy*-dependent and ABC transporter-dependent. Strains that were capsule-positive and O8/O9/O89 clustered together in multiple clades whereas those that were O8/O9/O89 and capsule-negative occurred throughout phylogroups A, B1, and C (Figure 1). Phylogroup C strains were most likely to have one of either an O8/O9/O89 serogroup (17 out of 19 strains; 89%), while the frequency of the strains with these serogroups was 20% (67 out of 332 strains) for phylogroup A strains and 17% (50 out of 301 strains) for B1 strains. O8/O9/O89 strains were uncommon among phylogroups B2, D, E, and F (Table 2). Phylogroup A strains were more likely to be O9/O89, while O8 strains were more common among phylogroup B1 strains. However, regardless of a strain's phylogroup membership, those with an O9 serogroup were more likely to be capsule-positive than the O8 strains (Figure 1) (Table 2). The H-types of the encapsulated strains were variable. Strains with the same capsule type had different H-types, while similar H-types occurred with different capsule types.

Strains encoding a *Klebsiella* capsule (228 strains) were next investigated for the presence of other capsule types reported in Bacteria and Archaea. When a strain was positive for any of the eight capsule types reported by Rendueles and colleagues (2017), the capsule types identified were Wzx/Wzy-dependent, Group IV_e, or Group IV_s (Table 3). Group IV_e was the most abundant (214 out of 228 strains; 93.9%) followed by Group IV_s (174 out of 228 strains; 76.3%), and Wzx/Wzy-dependent (167 out of 228 strains; 73.2%). The presence of these three capsule types was variable among the bloom strains.

Table 2. Presence/absence of *Klebsiella* capsules in O8, O9, or O89 *E. coli* strains of different phylogroups.

Phylogroup	Serogroup*	Capsule-positive n (%)
A (n=332)	O8 (n=16)	4 (25.0%)
	O9 (n=25)	22 (88.0%)
	O89 (n=26)	25 (96.2%)
B1 (n=301)	O8 (n=35)	2 (5.7%)
	O9 (n=14)	13 (92.9%)
	O89 (n=1)	0 (0%)
C (n=19)	O8 (n=10)	2 (20.0%)
	O9 (n=7)	6 (85.7%)
B2 (n=245)	O8 (n=13)	0 (0%)
D (n=142)	O8 (n=2)	0 (0%)
E (n=72)	O8 (n=1)	0 (0%)
	O9 (n=1)	0 (0%)
F (n=68)	O8 (n=2)	0 (0%)

* Strains with an O9 serogroup were not observed for phylogroups B2, D, and F, while strains with an O89 serogroup were absent from phylogroups C, B2, D, E, and F.

Table 3. Occurrence of Wzx/Wzy-dependent, GroupIV_s, and GroupIV_e capsule types in *Klebsiella* capsule-positive *E. coli*.

Wzx/Wzy-dependent	GroupIV_s	GroupIV_e	Number of strains
-	-	-	3
-	-	+	16
-	+	+	45
+	-	-	6
+	-	+	29
+	+	-	5
+	+	+	124

Variable gene content of bloom and other encapsulated E. coli

Pan genome analysis of the capsule region of encapsulated strains showed that there were no capsule-specific genes unique to all bloom strains compared to other encapsulated strains.

The majority of the bloom strains were members of phylogroup A, as were the bulk of the strains found to possess a *Klebsiella* capsule. Consequently, to determine if variable gene content varied between bloom strains and other encapsulated strains, or between encapsulated and non-encapsulated strains the pan genome analysis was restricted to phylogroup A strains.

No genes were found to be unique to bloom strains compared to other encapsulated strains, and apart from capsule genes, no genes were unique to encapsulated strains compared to non-encapsulated strains. However, all phylogroup A bloom strains encoded the ferric citrate uptake system (*fecIRABCDE*), while ≤60% of non-bloom encapsulated strains, and fewer than 39% of non-bloom phylogroup A strains encoded the *fec* genes. Genes of the *fim* operon and *cas* genes were also over-represented among the bloom strains compared to non-bloom encapsulated strains. Other genes were found to be over- or under-represented when bloom strains were compared to other encapsulated phylogroup A strains (Supplemental Table 2A,B) or when encapsulated strains were compared to non-encapsulated phylogroup A strains (Supplemental Table 3A,B). Overall, the variable gene content of bloom or encapsulated strains did not differ substantially from non-encapsulated strains (Supplemental Figure 1).

Discussion

Overall, about 7% of *E. coli* strains have acquired a group 1/*Klebsiella* capsule. However, the distribution of encapsulated strains is non-random and restricted to phylogroups A, B1, and C. *Klebsiella* capsules are very rare or absent among strains belonging to phylogroups B2, D, E, and F. The variable gene content of a strain varies with its phylogroup membership and, on average, each phylogroup has a distinct variable gene content (Touchon *et al*., 2009; Clermont

et al., 2013). This outcome suggests that the acquisition and maintenance of a group 1 capsule depends on the genomic background of the strains. Further, it suggests that there is some aspect of the genomic background of B2, D, E, and F strains that is incompatible with the possession of a group 1 capsule, although the nature of this incompatibility is unknown.

Another clearly non-random association between the possession of a group 1/*Klebsiella* capsule and the genomic background of a strain is the long recognized association between group 1 capsule and the serogroups O8, O9, and O89 (Kauffmann, 1947; Kido *et al*., 1995; Amor and Whitfield, 1997; Drummelsmith *et al*., 1997; Whitfield and Roberts, 1999). O-antigen biosynthesis in *E. coli* operates in two major pathways (Willis and Whitfield, 2013). One is *wzy*-dependent and carries Wzx (O-antigen flippase) and Wzy (O-antigen polymerase). The other pathway is ATP-binding cassette (ABC) transporter-dependent relying on Wzt (ABC transporter ATP-binding protein) and Wzm (ABC transporter permease). Among 182 O-antigen gene clusters, the vast majority (171) had *wzx*/*wzy* genes, the remaining 11 had *wzt*/*wzm* genes and O8, O9, and O89 are among these ABC transporter-dependent serogroups (Kido *et al*., 1995; Iguchi *et al*., 2014).

BLAST comparisons of *wzt* from O8, O9, and O89 *E. coli* strains, regardless of their capsule status, most often resulted in a very close match with the *wzt* gene from either a strain of *Klebsiella* or an encapsulated *E. coli* strain. This outcome suggests that the entire region comprising the O-antigen and capsule gene clusters is the consequence of a single recombination event. The fact that O8 and O9 of *E. coli* are identical to O5 and O3 of *K. pneumoniae*, respectively (Jansson *et al*., 1985; Saeki *et al*., 1993) and that O8, O9 co-express with the *E. coli* group 1/*Klebsiella* capsule (Whitfield and Roberts, 1999) also supports the notion that both the O-antigen and capsule region co-transfer. This is further corroborated

by *E. coli* O9a, which is suggested to have arisen due to a post-recombination mutation in an O-antigen gene upon transfer of *K. pneumoniae* O3 to *E. coli* (Sugiyama *et al*., 1998).

There are phylogroups B2, D, E, and F strains with an O8 or O9 serogroup, but these strains are capsule-negative. If it is assumed that the O-antigen and capsule region are transferred as a single block from *Klebsiella* to *E. coli*, then the absence of a capsule would suggest that these strains have subsequently lost the capsule region genes but maintained the O-antigen region from *Klebsiella*. In turn, this outcome also indicates that the genomic background of strains belonging to these phylogroups is incompatible with the maintenance of a *Klebsiella* capsule.

Acquisition of the *Klebsiella* capsule has occurred in a limited number of lineages, but when a lineage has acquired a *Klebsiella* capsule, there are often multiple *Klebsiella* capsule types in the lineage (Figure 1). A similar distribution can be observed in O-serogroups, where the occurrence of multiple O-serogroups within a single lineage is not unusual (Ingle *et al*., 2016). This outcome is unlikely to be the result of a single *Klebsiella* capsule acquisition event, followed by subsequent evolution of the capsule region in the lineage. Different capsule types are defined not only by their sequence similarity but also by the presence of a particular suite of capsule region genes (Wyres *et al*., 2016). Therefore, while capsule types within a lineage might diverge due to the loss of genes within the capsule region, the gain of genes can only be the consequence of horizontal gene transfer events. Further, the presence of transposase/IS elements on either side of the capsule cluster is also indicative of a horizontal gene transfer event, as reported previously (Rahn *et al*., 1999). Once a lineage has acquired the *Klebsiella* capsule region and likely the associated O-antigen region as well, then it may be that this lineage has an increased likelihood of experiencing subsequent lateral gene transfer events involving this region. The *Klebsiella* capsule region is typically 15 - 30 kbp and the

region would be larger if the O-antigen region is also considered. This region would represent a large target for homologous recombination with capsule regions from other encapsulated *E. coli* strains or from *Klebsiella*. Consequently, subsequent homologous recombination of the capsule region would be expected to occur at a higher rate than *de novo* acquisition of the capsule region through non-homologous recombination. The capsule (*cps*) region around the *rfb* locus has been identified as a recombination hotspot in multiple bacterial species including *Streptococcus pneumoniae*, *K. pneumoniae* and *E. coli* (Milkman *et al*., 2003; Didelot *et al*., 2012; Alqasim *et al*., 2014; Wright *et al*., 2014; Mostowy *et al*., 2017).

The encapsulated strains do not carry the colanic acid gene cluster, indicating that the acquisition of the capsule has been at the cost of colanic acid, as has been previously reported (Rahn *et al*., 1999). The great majority of non-encapsulated *E. coli* strains are positive for colonic acid. Colanic acid appears to be the default exopolysaccharide, particularly in phylogroups B2, D, E, and F. Given the different temperatures of expression and the genetic localization, the expression of the *Klebsiella* capsule/*E. coli* group 1 capsule and colanic acid are mutually exclusive events (Rahn *et al*., 1999; Whitfiled, 2006).

Three strains have been associated with bloom events in eastern Australia (Power *et al*., 2005) and five distinct strains were isolated during a recent bloom event in Western Australia. All bloom strains are encapsulated, and given that only about 7% of *E. coli* are encapsulated, then this indicates that the possession of a *Klebsiella* capsule is required if a strain is to be capable of causing bloom events. Among the eight bloom strains, there are seven distinct capsule types, again suggesting that it is the possession of the capsule *per se* that is important rather than the possession of a particular *Klebsiella* capsule variant. Of the two strains with the same capsule type, one is a phylogroup B1 strain while the other belongs to phylogroup A, further

suggesting that it is the possession of a capsule that is important rather than the genomic background of the strain. The pan genome analysis did not detect any genes either uniquely present or absent in bloom strains or other encapsulated *E. coli*. However, the pan genome analysis indicated that the iron uptake system encoded by the *fecIRABCDE* operon was present in all phylogroup A bloom strains. It is, however, absent in the B1 bloom strains. The *fec* operon is also present in many (39%) other non-bloom phylogroup A strains and its presence in the bloom strains may simply be due to chance. Genes of the *fim* operon that code for type 1 pilus synthesis, assembly and regulation (Orndorff and Falkow, 1984; Klemm, 1986) are more prevalent in bloom strains compared to non-bloom encapsulated strains. The type 1 pilus is required for the attachment of *E. coli* cells to abiotic surfaces (Pratt and Kolter, 1998; Cookson *et al*., 2002) and might enhance the free-living lifestyle of the bloom strains. CRISPR (clustered regularly interspaced short palindromic repeats)-associated *cas* genes are also over-represented in the bloom strains. The CRISPR/Cas system provides resistance against viral infection (Horvath and Barrangou, 2010) and would enhance the survival of the strains. Single nucleotide polymorphisms (SNPs) which can modulate gene expression, were not taken into account and this is a limitation of the present study.

Rendueles and colleagues (2017) revealed that capsules are more likely to occur in free-living species than in pathogens, reversing the long-held belief that capsules are associated more with virulence (Williams *et al*., 1990; Lawlor *et al*., 2005). Possession of a capsule is believed to enhance strain survival and persistence through overcoming predation and adverse environmental conditions including desiccation, osmotic stress, and UV radiation encountered in the external environment. The capsule may facilitate a strain's ecological transitions, thereby increasing its environmental range (Weiner *et al*., 1995; Rendueles *et al*., 2017). Thus, being encapsulated may enhance a strains' persistence in the external environment, relative

to non-encapsulated *E. coli*. Indeed, studies indicate that in eastern Australian reservoirs and recreational lakes, bloom strains can be isolated at any time, even when *E. coli* counts are low (non-bloom periods).

Although encapsulation appears to be essential for conferring bloom status on an *E. coli* strain, it seems unlikely that it can be the sole trait required. At any given point in time, multiple *E. coli* genotypes are present in a water body (Casarez *et al*., 2007; Higgins *et al*., 2007). Bloom events are associated with nutrient influx events such as dust storms or the autumn die-off of aquatic vegetation. Encapsulated strains might be more likely to persist in water bodies compared to other *E. coli* strains, due to their potentially enhanced survival, and can exploit these nutrient influx events and achieve high cell densities. However, this does not explain why other strains of *E. coli* present in the water body are also not capable of doing this. Further studies are required to identify traits that confer a growth rate advantage on the bloom strains relative to non-encapsulated *E. coli*.

Experimental Procedures

Diversity and distribution of capsule types

Strain selection

Whole genome sequence (WGS) data for 1194 Australian *E. coli* strains isolated from a wide range of sources were used to determine the frequency and diversity of *Klebsiella* capsules. The strains comprised 332 phylogroup A strains, 300 phylogroup B1 strains, 261 phylogroup B2 strains, 142 phylogroup D strains, 72 phylogroup E strains, 68 phylogroup F strains, and 19 phylogroup C strains. These strains were isolated from a variety of sources across Australia

(Gordon and FitzGibbon, 1999; Gordon and Cowling, 2003; Power *et al*., 2005; Blyton *et al*., 2014; Blyton *et al*., 2015; Vangchhia *et al*., 2016). To investigate where other encapsulated strains occur in the *E. coli* phylogeny, the study was extended to include encapsulated strains from two publicly available databases. First, the *Klebsiella* capsule core gene *galF* was used to query the NCBI (National Centre for Biotechnology Information) *E. coli* database using BLAST and 135 strains with >90% sequence similarity were selected at random and the assemblies were downloaded. Second, group 1 capsules are known to be associated with the ATP-binding cassette (ABC) transporter-dependent serogroups O8, O9, and O89 in *E. coli* (Kido *et al*., 1995; Amor and Whitfield, 1997; Drummelsmith *et al*., 1997). Using the *E. coli*/*Shigella* database in EnteroBase (https://enterobase.warwick.ac.uk/species/index/ecoli) a subset of 99 O8, O9, and O89 strains were selected representing one example of each O:H type combination present in the database. The phylogroups of the strains have been determined experimentally or *in silico* (Clermont *et al*., 2000; Clermont *et al*., 2013; Beghain *et al*., 2018).

Determination of capsule status and type

The WGS data of all 1194 Australian strains were screened for the *Klebsiella* capsule using Kaptive, a software tool used for the rapid identification of *Klebsiella* capsule loci in whole genome sequences by comparison to a reference database of known *Klebsiella* capsule types (Wyres *et al*., 2016). Strains predicted by Kaptive to possess a capsule were examined morphologically on MacConkey and Congo Red plates. The strains' morphology was found to be consistent with the Kaptive predictions and showed that strains having ≥90% coverage and identity with the best match capsule locus could be considered capsule-positive. These were the criteria used to score the WGS data downloaded from NCBI and Enterobase.

The Harvest suite of tools (Treangen *et al*., 2014) was used to infer phylogeny of 869 strains belonging to phylogroups A, B1, and C. The tree was rooted on the phylogroup B2 strain ED1a and was annotated using the web interface Interactive tree of life (iTOL) (Letunic and Bork, 2016).

Within-capsule screening

Kaptive extracts the nucleotide sequence of the capsule region from whole genome sequences. The capsule regions of the 237 capsule-positive strains were annotated using Prokka (Seemann, 2014) and a pan genome analysis was conducted using Roary (Page *et al*., 2015) and Scoary (Brynildsrud *et al*., 2016).

Capsule flanking region

Capsule flanking regions of strains representative of each capsule type were investigated by aligning the capsule-positive genomes in progressiveMauve (Darling *et al*., 2004) and determining the sequence similarity using the NCBI BLASTn. The draft genomes were reordered against the *E. coli* K-12 reference genome prior to alignment. *E. coli* K-12 has the colanic acid gene cluster and is negative for *Klebsiella* capsule. The arrangement of the *his* operon, O-antigen (*rfb*) gene region, and capsule region of *Klebsiella* strains identified using BLAST was compared against *Klebsiella* capsule-positive/negative *E. coli* using progressiveMauve genome alignments (Darling *et al*., 2004).

Association of the capsule with the O-antigen and other capsule types

Encapsulated strains are known to be associated with particular ABC transporter-dependent O-antigens (Kido *et al*., 1995; Amor and Whitfield, 1997; Drummelsmith *et al*., 1997). The serotype of all strains was determined *in silico* using SerotypeFinder 1.1 (Joensen *et al*., 2015) within the Centre for Genomic Epidemiology website (http://www.genomicepidemiology.org). The O-type data were incorporated in the phylogroups A, B1, and C *E. coli* phylogeny.

To infer the origin of the O-antigen genes, the *wzt*/*wzm* nucleotide sequences of encapsulated and capsule-negative *E. coli* strains representing O8, O9, and O89 were extracted from SerotypeFinder 1.1 (Joensen *et al*., 2015). These were queried against the Microbial Nucleotide BLAST database in NCBI.

Rendueles and colleagues (2017) reported the presence of multiple capsule types in Bacteria and Archaea. Their programme CapsuleFinder (https://research.pasteur.fr/en/tool/capsulefinder/) is based on the use of HMM profiles to detect essential proteins involved in capsule biogenesis and uses computational models to describe the capsule components and their organisation. CapsuleFinder can detect eight capsule types: Wzx/Wzy-dependent, Group II and III or ABC-dependent, PGA capsule (Poly-γ-d-glutamate proteic capsule), synthase-dependent-HAS (hyaluronic acid capsules), synthase-dependent CPS3 (capsules of type cps3), Group IV_f (based on *Francisella tularensis* GroupIV capsule), Group IV_e (based on *Escherichia coli* GroupIV capsule), and Group IV_s (based on *Salmonella* sp. GroupIV capsule). The WGS data for all strains found to possess a *Klebsiella* capsule were screened using CapsuleFinder.

Pan genome comparison

Phylogroup A strains represent the majority of bloom strains isolated to date. Hence, the pan genome of a collection of phylogroup A strains was compared, based on whether they were bloom-associated or not (52 strains), or harboured a capsule or not (341 strains). Genomes were annotated using Prokka (Seemann, 2014) and pan genome analyses were conducted using Roary (Page *et al*., 2015). Scoary (Brynildsrud *et al*., 2016) was used to assess the association between the capsule and other genomic components. For a gene to be considered over-represented in one group (e.g. encapsulated strains) compared to a second group (e.g. non-encapsulated strains), the gene had to be present in the first group at a frequency >75% and present in the second group at a frequency ≤60%; and the difference in the frequencies of the gene in the two groups had to be >30%. For a gene to be considered under-represented, then its frequency needed to be <50% in the first group and >75% in the second group; and the difference in the frequencies of the gene in the two groups had to be >30%.

Acknowledgements

This study was funded in part by an Australian Research Council Linkage Grant (LP120100327). Additional support was provided by Melbourne Water, Water Corporation of Western Australia, Queensland Bulk Water Authority, Hunter Water Corporation, South East Water, Yarra Valley Water, City West Water, Sydney Water Corporation, South Australian Water Corporation, Sydney Catchment Authority, Fitzroy River Water, and Victorian Department of Health. Portions of the research were also funded by Water Research Australia (2018) Management of Environmental *E. coli*, Project 1101.

Conflicts of interest

The authors declare that they have no competing interests.

References

Alm, E.W., Walk, S.T., and Gordon, D.M. (2011) The Niche of *Escherichia coli*. In *Population Genetics of Bacteria: A Tribute to Thomas S. Whittam*. Walk, S.T., and Feng, P.C.H. (eds). Washington, DC: ASM Press, pp. 69–89.

Alqasim, A., Scheutz, F., Zong, Z., and McNally, A. (2014) Comparative genome analysis identifies few traits unique to the *Escherichia coli* ST131 H30Rx clade and extensive mosaicism at the capsule locus. *BMC Genomics* **15**: 830.

Amor, P.A., and Whitfiled, C. (1997) Molecular and functional analysis of genes required for expression of group IB K antigens in *Escherichia coli*: characterization of the *his*-region containing gene clusters for multiple cell-surface polysaccharides. *Mol Microbiol* **26**: 145-161.

Batchelor, R.A., Haraguchi, G.E., Hull, R.A., and Hull, S.I. (1991) Regulation by a novel protein of the bimodal distribution of lipopolysaccharide in the outer membrane of *Escherichia coli*. *J Bacteriol* **173**: 5699-5704.

Beghain, J., Bridier-Nahmias, A., Le Nagard, H., Denamur, E., and Clermont, O. (2018) ClermonTyping: an easy-to-use and accurate *in silico* method for *Escherichia* genus strain phylotyping. *Microb Genom* **4**: e000192.

Bitton, G. (2011) *Wastewater Microbiology*, 4th edn. Hoboken, NJ: Wiley-Blackwell.

Blyton, M.D.J., Cornall, S.J., Kennedy, K., Colligon, P., and Gordon, D.M. (2014) Sex-dependent competitive dominance of phylogenetic group B2 *Escherichia coli* strains within human hosts. Environ *Microbiol Rep* **6**: 605-610.

Blyton, M.D.J., Pi, H., Vangchhia, B., Abraham, S., Trott, D.J., Johnson, J.R., and Gordon, D.M. (2015) Genetic Structure and Antimicrobial Resistance of *Escherichia coli* and Cryptic Clades in Birds with Diverse Human Associations. *Appl Environ Microbiol* **81**: 5123-5133.

Bonde, G.J. (1966) Bacteriological methods for estimation of water pollution. Health laboratory science **3**: 124-128.

Brynildsrud, O., Bohlin, J., Scheffer, L., and Eldholm, V. (2016) Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* **17**: 238.

Casarez, E.A., Pillai, S.D., and Di Giovanni, G.D. (2007) Genotype diversity of *Escherichia coli* isolates in natural waters determined by PFGE and ERIC-PCR. *Water Res* **41**: 3643-3648.

Clermont, O., Bonacorsi, S., and Bingen, E. (2000) Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl Environ Microbiol* **66**: 4555-4558.

Clermont, O., Christenson, J.K., Denamur, E., and Gordon, D.M. (2013) The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* **5**: 58-65.

Cookson, A.L., Cooley, W.A., and Woodward, M.J. (2002) The role of type 1 and curli fimbriae of Shiga toxin-producing *Escherichia coli* in adherence in abiotic surfaces. *Int J Med Microbiol* **292**: 195-205.

Darling, A.C.E., Mau. B., Blattner, F.R., and Perna. N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**: 1394-1403.

Didelot, X., Méric, G., Falush, D., and Darling, A.E. (2012) Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. *BMC Genomics* **13**: 256.

Drummelsmith, J., Amor, P.A., and Whitfield. C. (1997) Polymorphism, duplication, and IS1-mediated rearrangement in the chromosomal *his-rfb-gnd* region of *Escherichia coli* strains with group IA and capsular K antigens. *J Bacteriol* **179**: 3232-3238.

Edberg, S.C., Rice, E.W., Karlin, R.J., and Allen M.J. (2000) *Escherichia coli*: the best biological drinking water indicator for public health protection. *J Appl Microbiol* **88**: 106S-116S.

Gordon, D.M. (2001) Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. *Microbiology* **147**: 1079-1085.

Gordon, D.M. (2013) The Ecology of *Escherichia coli*. In *Escherichia coli: Pathotypes and Principles of Pathogenesis.* Donnenberg, M.S. (ed). San Diego, CA: Academic Press, pp. 3-20.

Gordon, D.M., and Cowling, A. (2003) The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. *Microbiology* **149**: 3575-3586.

Gordon, D.M., and FitzGibbon, F. (1999) The distribution of enteric bacteria from Australian mammals: host and geographical effects. *Microbiology* **145**: 2663-2671.

Higgins, J., Hohn, C., Hornor, S., Frana, M., Denver, M., and Joerger, R. (2007) Genotyping of *Escherichia coli* from environmental and animal samples. *J Microbiol Methods* **70**: 227-235.

Horvath, P., and Barrangou, R. (2010) CRISPR/Cas, the immune system of Bacteria and Archaea. *Science* **327**: 167-170.

Hsu, C.-R., Lin, T.-L., Chen, Y.-C., Chou, H.-C., and Wang, J.-T. (2011) The role of *Klebsiella pneumoniae rmpA* in capsular polysaccharide synthesis and virulence revisited. *Microbiology* **157**: 3446-3457.

Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M., *et al*. (2014) A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res* **22**: 101-107.

Ingle, D.J., Valcanis, M., Kuzevski, A., Tauschek, M., Inouye, M., Stinear, T., *et al*. (2016) *In silico* serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and between pathogenic lineages. *Microb Genom* **2**: 0.000064.

Ishii, S., and Sadowsky, M.J. (2008) *Escherichia coli* in the Environment: Implications for Water Quality and Human Health. ***Microbes Environ*** **23**: 101-108.

Jansson, P-E., Lönngren, J., Widmalm, G., Leontein, K., Slettengren, K., Svenson, S.B., *et al*. (1985) Structural studies of the O-antigen polysaccharides of *Klebsiella* O5 and *Escherichia coli* O8. *Carbohydr Res* **145**: 59-66.

Jayaratne, P., Keenleyside, W.J., MacLachlan, P.R., Dodgson, C., and Whitfield, C. (1993) Characterization of *rcsB* and *rcsC* from *Escherichia coli* O9: K30: H12 and examination of the role of the *rcs* regulatory system in expression of group I capsular polysaccharides. *J Bacteriol* **175**: 5384-5394.

Jiang, X., Morgan, J., and Doyle, M.P. (2002) Fate of *Escherichia coli* O157: H7 in manure-amended soil. *Appl Environ Microbiol* **68**: 2605-2609.

Joensen, K.G., Tetzschner, A.M., Iguchi, A., Aarestrup, F.M., and Scheutz, F. (2015) Rapid and easy *in silico* serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. *J Clin Microbiol* **53**: 2410-2426.

Kabha, K., Nissimov, L., Athamna, A., Keisari, Y., Parolis, H., Parolis, L.A., *et al*. (1995) Relationships among capsular structure, phagocytosis, and mouse virulence in *Klebsiella pneumoniae*. *Infect Immun* **63**: 847-852.

Kauffmann, F. (1947) The serology of the coli group. *J Immunol* **57**: 71-100.

Kido, N., Torgov, V.I., Sugiyama, T., Uchiya, K., Sugihara, H., Komatsu, T., *et al*. (1995) Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E. coli* O9 *rfb* gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. *J Bacteriol* **177**: 2178-2187.

Klemm, P. (1986) Two regulatory *fim* genes, *fimB* and *fimE*, control the phase variation of type 1 fimbriae in *Escherichia coli*. *EMBO J* **5**: 1389-1393.

Kudva, I.T., Blanch, K., and Hovde, C.J. (1998) Analysis of *Escherichia coli* O157: H7 survival in ovine or bovine manure and manure slurry. *Appl Environ Microbiol* **64**: 3166-3174.

Lawlor, M.S., Hsu, J., Rick, P.D., and Miller, V.L. (2005) Identification of *Klebsiella pneumoniae* virulence determinants using an intranasal infection model. *Mol Microbiol* **58**: 1054-1073.

Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**: W242-W245.

Milkman, R., Jaeger, E., and McBride, R.D. (2003) Molecular evolution of the *Escherichia coli* chromosome. VI. Two regions of high effective recombination. *Genetics* **163**: 475-483.

Mostowy, R.J., Croucher, N.J., De Maio, N., Chewapreecha, C., Salter, S.J., Turner, P., *et al*. (2017) Pneumococcal capsule synthesis locus *cps* as evolutionary hotspot with potential to generate novel serotypes by recombination. *Mol Biol Evol* **34**: 2537–2554.

Orndorff, P.E., and Falkow, S. (1984) Organization and expression of genes responsible for type 1 piliation in *Escherichia coli*. *J Bacteriol* **159**: 736-744.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., *et al*. (2015) Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**: 3691-3693.

Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., and Slade, M.B. (2005) Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. ***Environ Microbiol* 7**: 631-640.

Pratt, L.A., and Kolter, R. (1998) Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. *Mol Microbiol* **30**: 285-293.

Rahn, A., Drummelsmith, J., and Whitfield, C. (1999) Conserved organization in the *cps* gene clusters for expression of *Escherichia coli* group 1 K antigens: relationship to the colanic acid biosynthesis locus and the *cps* genes from *Klebsiella pneumoniae*. *J Bacteriol* **181**: 2307-2313.

Rendueles, O., Garcia-Garcerà, M., Néron, B., Touchon, M., and Rocha, E.P.C. (2017) Abundance and co-occurrence of extracellular capsules increase environmental breadth: Implications for the emergence of pathogens. *PLoS Pathog* **13**: 1-30.

Saeki, A., Kido, N., Sugiyama, T., Ohta, M., Iwashita, T., Uchiya, K., and Kato, N. (1993) Isolation of *rfb* gene clusters directing the synthesis of O polysaccharides consisting of mannose homopolymers and serological analysis of lipopolysaccharides. *Microbiol Immunol* **37**: 601-606.

Schembri, M.A., Blom, J., Krogfelt, K.A., and Klemm, P. (2005) Capsule and fimbria interaction in *Klebsiella pneumoniae*. *Infect Immun* **73**: 4626-4633.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068 – 2069.

Sugiyama, T., Kido, N., Kato, Y., Koide, N., Yoshida, T., and Yokochi, T. (1998) Generation of *Escherichia coli* O9a Serotype, a Subtype of *E. coli* O9, by Transfer of the *wb** Gene Cluster of *Klebsiella* O3 into *E. coli* via Recombination. *J Bacteriol* **180**: 2775-2778.

Szewzyk, U., Szewzyk, R., Manz, W., and Schleifer, K.-H. (2000) Microbiological Safety of Drinking Water. *Annu Rev Microbiol* **54**: 81-127.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., *et al*. (2009) Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* **5**: e1000344.

Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014) The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. ***Genome Biol* 15**: 524.

United States Environmental Protection Agency (U. S. EPA). (2012) Recreational water quality criteria. Office of Water 820-F-12-058. https://www.epa.gov/sites/production/files/2015-10/documents/rwqc2012.pdf.

van Elsas, J.D., Semenov, A.V., Costa, R., and Trevors, J.T. (2011) Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J* **5**: 173-183.

Vangchhia, B., Abraham, S., Bell, J.M., Collignon, P., Gibson, J.S., Ingram, P.R., *et al*. (2016) Phylogenetic diversity, antimicrobial susceptibility and virulence characteristics of phylogroup F *Escherichia coli* in Australia. *Microbiology* **162**: 1904-1912.

Vital, M., Hammes, F., and Egli, T. (2008) *Escherichia coli* O157 can grow in natural freshwater at low carbon concentrations. ***Environ Microbiol* 10**: 2387-2396.

Weiner, R., Langille, S., and Quintero, E. (1995) Structure, function and immunochemistry of bacterial exopolysaccharides. *J Ind Microbiol* **15**: 339-346.

Whitfield, C. (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu Rev Biochem* **75**: 39-68.

Whitfield, C., and Paiment, A. (2003) Biosynthesis and assembly of Group 1 capsular polysaccharides in *Escherichia coli* and related extracellular polysaccharides in other bacteria. *Carbohydr Res* **338**: 2491-2502.

Whitfield, C., and Roberts, I.S. (1999) Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol Microbiol* **31**: 1307-1319.

Williams, P., Ciurana, B., Camprubi, S., and Tomas, J.M. (1990) Influence of lipopolysaccharide chemotype on the interaction between *Klebsiella pneumoniae* and human polymorphonuclear leucocytes. *FEMS Microbiol Lett* **69**: 305-310.

Willis, L.M., and Whitfield, C. (2013) Capsule and lipopolysaccharide. In *Escherichia coli*: *Pathotypes and Principles of Pathogenesis*. Donnenberg, M.S. (ed). San Diego, CA: Academic Press, pp. 533-556.

Wright, M.S., Perez, F., Brinkac, L., Jacobs, M.R., Kaye, K., Cober, E., *et al*. (2014) Population structure of KPC-producing *Klebsiella pneumoniae* isolates from midwestern US hospitals. *Antimicrob Agents Chemother* **58**: 4961-4965.

Wyres, K.L., Wick, R.R., Gorrie, C., Jenney, A., Follador, R., Thomson, N.R., and Holt, K.E. (2016) Identification of *Klebsiella* capsule synthesis loci from whole genome data. *Microb Genom* **2**: 0.000102.

Supplemental Material

Supplemental Table 1. Isolation details of the bloom strains.

Bloom strain group	Number of isolates	Date of isolation	Geographic origin*
East coast A0	6	2002-2014	Hinze Dam, QLD; Hunter valley, NSW
East coast A1	17	2002-2014	Hinze Dam, QLD; Hunter valley, NSW; Lake Ginninderra, ACT; Shoalhaven, NSW; Sullivans Creek, ACT
East coast B1	7	2002-2015	Googong Dam, NSW; Hinze Dam, QLD; Shoalhaven, NSW; Wyaralong Dam, QLD
Western Australia A1	4	2015	WA
Western Australia B1	1	2015	WA

*ACT = Australian Capital Territory; NSW = New South Wales; QLD = Queensland; WA = Western Australia

Supplemental table 2A. Genes over-represented among phylogroup A bloom strains compared to phylogroup A non-bloom encapsulated *E. coli*.

Gene	Gene product/function	% occurrence in bloom strains (n = 27) (>75 %)	% occurrence in non-bloom encapsulated strains (n = 25) (≤60%)
yjhV	Conserved hypothetical protein	100	56.0
fecE	Ferric citrate transport	100	56.0
ydaM	Sensor domain-containing diguanylate cyclase	100	56.0
fecB	Ferric citrate transport system	100	56.0
fecD	Ferric citrate transport system	100	56.0
ybjL	Putative transmembrane potassium ion transporter	100	60.0
fecC	Ferric citrate transport system	100	60.0
fecI	Ferric citrate transport system	100	60.0
fecA	Ferric citrate transport system	100	60.0
fecR	Ferric citrate transport system	100	60.0
fimA	Major type 1 subunit fimbrin (pilin)	96.3	40.0
group_9550	Hcp1 family type VI secretion system	96.3	56.0
fimE	Controls phase variation of type 1 fimbriae	96.3	56.0
fimB	Type 1 fimbriae regulatory protein FimB	96.3	56.0
fimI	Type 1 pilus biosynthesis	96.3	56.0
fimC	Chaperone type 1 pilus biosynthesis	96.3	60.0
tamB	Putative outer membrane protein	96.3	60.0
group_6187	Hypothetical protein	92.6	28.0
group_4967	Hypothetical protein	92.6	32.0
casB	type I-E CRISPR-associated protein Cse2/CasB	92.6	48.0
group_1154	DUF4132 domain-containing protein	92.6	52.0
casE	Encodes CRISPR system Cascade subunit CasE	92.6	52.0
yehQ	Hypothetical protein	92.6	56.0
cstA	Carbon starvation protein A	92.6	56.0
yfaW	L-rhamnonate dehydratase	92.6	56.0
cas2	CRISPR-associated protein	92.6	56.0
group_13527	Conserved hypothetical protein	92.6	56.0
yaiW	Putative lipoprotein	92.6	56.0
group_10425	Conserved protein of unknown function	92.6	60.0
group_12337	Hypothetical protein	88.9	32.0
group_10439	Hypothetical protein	88.9	32.0
gspD_1	Type II secretion system protein GspD	88.9	36.0
ydeS	Fimbrial-like adhesin protein	88.9	36.0
group_12403	Hypothetical protein	88.9	40.0
yieH	6-phosphogluconate phosphatase	88.9	40.0
group_6978	Pyruvate:ferredoxin (flavodoxin) oxidoreductase	88.9	40.0
yagP	Hypothetical protein	88.9	48.0
gnsB	Protein GnsB	88.9	48.0
cas1	Type I-E CRISPR-associated endonuclease Cas1	88.9	48.0
casA	Type I-E CRISPR-associated protein Cse1/CasA	88.9	48.0

Gene	Gene product/function	% occurrence in bloom strains (n = 27) (>75 %)	% occurrence in non-bloom encapsulated strains (n = 25) (≤60%)
yegJ	DUF2314 domain-containing protein	88.9	48.0
gspE_1	Type II secretion system protein GspE	88.9	52.0
casC	Type I-E CRISPR-associated protein Cas7/Cse4/CasC	88.9	52.0
yghF	Type II secretion system protein GspC	88.9	56.0
yodB	Cytochrome b561	88.9	56.0
yghG	Hypothetical protein	88.9	56.0
pppA	Prepilin peptidase	88.9	56.0
gspF_1	Type II secretion system protein GspF	88.9	56.0
gspG_1	Type II secretion system protein GspG	88.9	56.0
group_1600	Lipopolysaccharide heptosyltransferase family protein	88.9	56.0
casD	Type I-E CRISPR-associated protein Cas5/CasD	85.2	48.0
bglF	PTS beta-glucoside transporter subunit EIIBCA	85.2	52.0
group_8165	DNA ligase B	81.5	16.0
dicB	Division inhibition protein DicB	81.5	40.0
nohB	Host specificity protein J	81.5	44.0
ydfD	DUF1482 family protein	81.5	44.0
yieL	Xylanase	81.5	48.0
group_3309	Sensor domain-containing phosphodiesterase	77.8	32.0

Supplemental table 2B. Genes under-represented among phylogroup A bloom strains compared to phylogroup A non-bloom encapsulated *E. coli.*

Gene	Gene product/function	% occurrence in bloom strains (n = 27) (<50%)	% occurrence in non-bloom encapsulated strains (n = 25) (>75%)
ligB	DNA ligase B	14.8	84.0
yebB	YebB family permuted papain-like enzyme	29.6	84.0
erfK	L,D-transpeptidase	29.6	84.0
group_7642	YjjI family glycine radical enzyme	29.6	76.0
ypdF	Aminopeptidase	29.6	76.0
potG	Putrescine transport ATP-binding protein PotG	33.3	96.0
betT	Choline transporter	33.3	92.0
betB	Betaine-aldehyde dehydrogenase	33.3	92.0
ykgF	Ferredoxin-like LutB family protein	33.3	92.0
group_11760	FUSC family protein	33.3	88.0
safA	Two-component-system connector protein SafA	33.3	80.0
group_5813	Capsule assembly Wzi family protein	37.0	100
hyfR_2	Hydrogenase-4 transcriptional activator	37.0	100
cheZ	Protein phosphatase CheZ	37.0	92.0
yeeE	YeeE/YedE family protein	37.0	88.0
yfdE	Acetyl-CoA--oxalate CoA-transferase	37.0	80.0
ydcM_1	Transposase	37.0	80.0
aspS	Aspartate--tRNA ligase	37.0	80.0
yfaL	AIDA-I family autotransporter YfaL	37.0	76.0
group_14522	DUF2057 family protein	40.7	100
yfgO	AI-2E family transporter	40.7	96.0
ykgH	Hypothetical protein	40.7	88.0
dadX	Alanine racemase catabolic	40.7	84.0
ykgC	Pyridine nucleotide-disulfide oxidoreductase	40.7	80.0
yahA	Cyclic di-GMP phosphodiesterase YahA	40.7	76.0
fliY	Cystine ABC transporter substrate-binding protein	48.1	96.0
elaD	Deubiquitinase	48.1	80.0

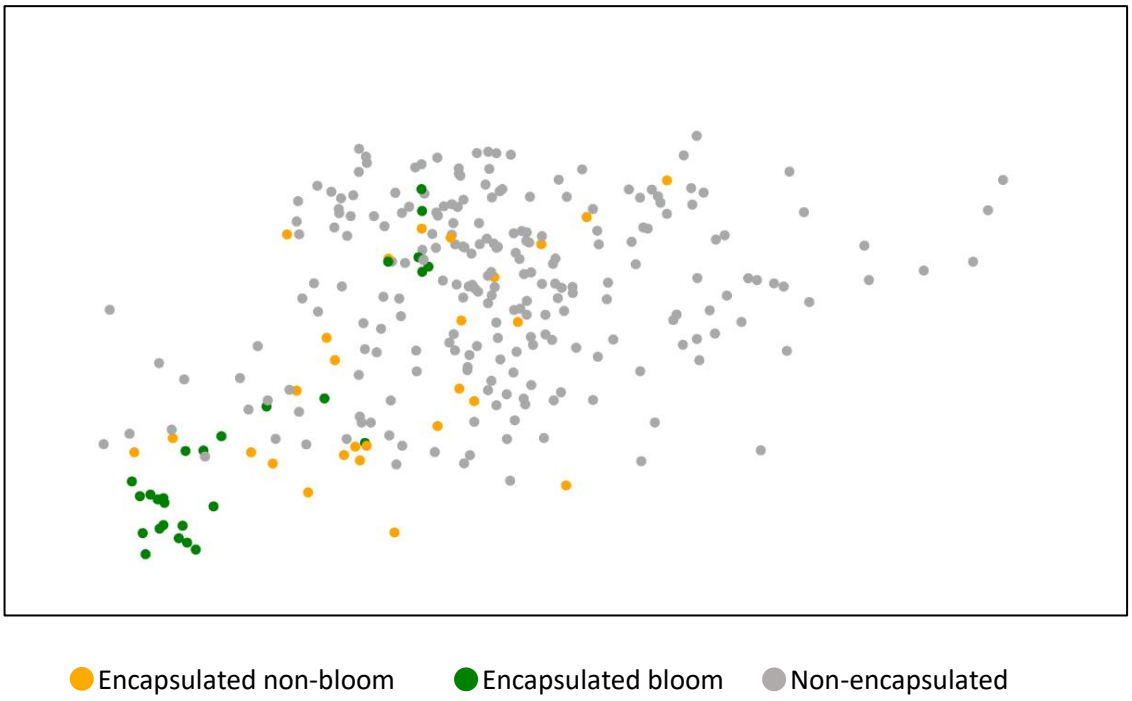
Supplemental table 3A. Genes over-represented among phylogroup A encapsulated strains compared to phylogroup A capsule-negative *E. coli.*

Gene	Gene product/function	% occurrence in encapsulated strains (n = 111) (>75%)	% occurrence in capsule-negative strains (n = 230) (≤60%)
galF	Putative uridylyltransferase subunit with GalU	99.1	0
kanE	Alpha-D-kanosaminyltransferase	95.5	8.7
ugd	UDP-glucose 6-dehydrogenase	93.7	0
hisG	ATP phosphoribosyltransferase	93.7	59.6
oppA	Peptide ABC transporter	92.8	55.7
rfaD	ADP-L-glycero-D-mannoheptose-6-epimerase	91.9	50.4
frmA	Formaldehyde dehydrogenase	91.0	55.7
yigG	Inner membrane protein	90.1	52.2
mfd	Transcription-repair coupling factor	90.1	57.8
arnB	UDP-L-Ara4O C-4 transaminase	90.1	59.6
arpA_1	Regulator of acetyl CoA synthetase	89.2	49.6
acrD	AcrAD-TolC multidrug efflux transport system	89.2	57.8
alsB	D-allose ABC transporter	88.3	57.4
yfaX	Putative DNA-binding transcriptional regulator	88.3	57.8
nepI	Purine ribonucleoside efflux transporter	88.3	57.8
iraD	Inhibitor of sS proteolysis	87.4	39.6
nirB	Nitrite reductase, large subunit	87.4	55.7
group_13623	Hypothetical protein	86.5	48.3
rhaD	Rhamnulose-1-phosphate aldolase monomer	86.5	53.5
tolQ	Colicin A import System	86.5	54.8
recQ	ATP-dependent DNA helicase	86.5	55.2
melB	Melibiose:H^+/Na^+/Li^+ symporter	86.5	55.7
group_11014	Hypothetical protein	86.5	56.5
insG	IS1 predicted transposase	85.6	49.1
ariR	Regulator of acid resistance	85.6	53.5
alsK	D-allose kinase	84.7	43.0
insG_2	Hypothetical protein	84.7	46.5
citF	Citrate lyase	84.7	51.7
bglH	Carbohydrate-specific outer membrane porin	84.7	52.6
yfgH	Putative lipoprotein	83.8	34.3
yjbL	Putative protein	83.8	43.9
leuA	2-isopropylmalate synthase	83.8	49.6
group_14921	Hypothetical protein	83.8	51.3
actP	Acetate/glycolate transporter	82.9	49.1
dtpB	Dipeptide/tripeptide:H^+ symporter DtpB	82.9	49.1
yneJ	DNA-binding transcriptional regulator	82.9	50.4
rdoA	Serine/threonine protein kinase	82.9	51.3
rsmC	16S rRNA methyltransferase	82.9	51.7
rarD	Putative chloramphenicol resistance permease	82.0	40.9
group_9715	Hypothetical protein	82.0	40.9

Gene	Gene product/function	% occurrence in encapsulated strains (n = 111) (>75%)	% occurrence in capsule-negative strains (n = 230) (≤60%)
ilvH	Acetolactate synthase	82.0	45.2
ynbD	Putative phosphatase	82.0	47.4
gapA_2	Glyceraldehyde 3-phosphate dehydrogenase A	82.0	47.4
yfgI	Putative membrane protein	81.1	38.3
ydeE	Putative transport protein YdeE	81.1	40.9
chbR	DNA-binding transcriptional dual regulator	81.1	44.8
rhsB	RhsB protein in *rhs* element	81.1	46.1
aldA	Aldehyde dehydrogenase A, NAD-linked	81.1	47.0
aphA	Acid phosphatase monomer	81.1	48.7
sdsP	SdsRQP multidrug efflux transport system	81.1	50.4
leuC	Isopropylmalate isomerase	80.2	45.2
leuO	LeuO DNA-binding transcriptional activator	80.2	46.1
trg	Methyl accepting chemotaxis protein	80.2	47.8
group_15111	Hypothetical protein	80.2	48.3
sgbH	3-keto-L-gulonate 6-phosphate decarboxylase	80.2	48.7
yiaK	2,3-diketo-L-gulonate reductase monomer	80.2	49.6
bcsQ	Putative cellulose biosynthesis protein	79.3	28.7
yhiS_1	Hypothetical protein	79.3	33.5
arpA_1	Hypothetical protein	79.3	43.9
group_25103	Hypothetical protein	78.4	33.9
ydjO	Putative protein	78.4	34.8
lexA	LexA DNA-binding transcriptional repressor	77.5	33.0
yjiA	P-loop guanosine triphosphatase	77.5	38.7
yafT	Putative aminopeptidase	77.5	40.0
group_14981	Hypothetical protein	77.5	40.4
ykfM	Hypothetical protein	77.5	43.9
sohB	Putative inner membrane peptidase	76.6	8.7
ygcU	Putative FAD-containing dehydrogenase	76.6	29.1
fecR	Ferric citrate transport system	76.6	40.9
group_2783	Hypothetical protein	76.6	42.2
group_8730	Hypothetical protein	75.7	34.3
yhaB	Putative protein	75.7	36.1
elfC	Putative outer membrane usher protein	75.7	39.6
fecI	Ferric citrate transport system	75.7	40.9
yjbS	Hypothetical protein	75.7	43.5

Supplemental table 3B. Genes under-represented among phylogroup A encapsulated strains compared to phylogroup A capsule-negative *E. coli*.

Gene	Gene product/function	% occurrence in encapsulated strains (n = 111) (<50%)	% occurrence in capsule-negative strains (n = 230) (>75%)
sohB	Putative inner membrane peptidase	22.5	91.7
caiB	CaiB monomer	35.1	81.7
ybeF	DNA-binding transcriptional regulator	35.1	78.3
ygcR	Putative flavoprotein	36.0	75.7
topA	DNA topoisomerase I	36.9	85.7
cysB	CysB-O-acetyl-L-serine transcriptional regulator	37.8	80.0
ligB	DNA ligase	40.5	81.7
ybdG	Mechanosensitive YbdG monomer	41.4	96.1
kch	K^+ channel Kch monomer	42.3	93.0
nfsB	Dihydropteridine reductase monomer	44.1	94.8
ybdK	Carboxylate-amine ligase	45.9	99.1
gltD	Glutamate synthase, small subunit	45.9	84.3
ybdJ	Putative inner membrane protein	48.6	99.6
group_16489	Hypothetical protein	48.6	99.1
yeeO	YeeO MATE transporter	49.5	80.0



Supplemental Figure 1. Principal coordinate (PCoA) plot of the variable gene content of phylogroup A encapsulated (n = 27; orange), bloom (n = 27; green), and non-encapsulated (n = 276; grey) *E. coli* strains. Analysis was done using PAST3 (Hammer *et al*., 2001). Axis 1 captures 9.7% of the variation and axis 2 captures 4.2% of the variation.