

MRes Thesis

School of Life Sciences

University of Nottingham

2 Research Projects Looking at Different Ways of Combatting Antibiotic Resistance.

By Samuel Jacob Fenn

Supervisors: Professor Miguel Camara (University of Nottingham) and Dr. Alan McNally (University of Birmingham)

Completed as a requirement for the Wellcome AAMR
Programme

Supported by
wellcometrust

Acknowledgements

I would like to thank all students and staff encountered at both the University of Nottingham and University of Birmingham for making me feel so welcome and creating a great environment for me to complete my work.

Special thanks must go to Miss Ioanna Chatzigianni and Dr. Jean Dubern for all their technical help during my time at the University of Nottingham. Extra special thanks go to Dr. Emily Richardson and the entire MicrobesNG team based at the University of Birmingham, for putting up with my constant questioning whilst bioinformatics was blowing my mind.

Last but not least, I would like to thank the Wellcome Trust for their generous funding and thanks to Professor Paul Williams and Professor Ian Henderson for accepting me onto this very exciting and prestigious Wellcome AAMR training programme.

Summary of contents

1.0 Background and context	pg.5-16
1.1 Spread of AMR	pg.6
1.1.1 Selection Pressure	pg.6
1.1.1.1 Therapeutic pressure	pg.7
1.1.1.2 Animal husbandry pressure	pg.7
1.1.2 Mobile Genetic Elements	pg.9
1.1.2.1 Plasmids	pg.9
1.1.2.2 Transposons, Insertion sequences and Integrins	pg.10
1.2 Current approaches to diagnostics and treatment	pg.12
1.2.1 Future diagnostic	pg.13
1.3 Combatting AMR	pg.14
1.3.1 Alternative therapeutics	pg.14
1.3.1.1 Vaccine development	pg.15
1.3.1.2 Novel therapeutics	pg.16
1.4 Overview of Project Rotations	pg.16
 2.0 Rotation 1 – Global Dissemination of ST167 and ST617 as multi-drug resistant clones, a familiar story	 pg.17-30
2.1 Abstract	pg.17
2.2 Introduction	pg.17
2.3 Materials and methods	pg.20
2.3.1 Sequence analysis	pg.20
2.3.2 Phylogenetic analysis	pg.20
2.3.3 Core and Pan-genome analysis	pg.20
2.3.4 In Silico serotyping	pg.21
2.4 Results	pg.21
2.4.1 ST167 and ST617 are distinct monophyletic clusters that arose through clonal expansion	pg.21
2.4.2 Antibiotic resistance gene carriage	pg.22
2.4.3 Parallel, independent evolution towards novel O antigen type	pg.23
2.4.4 Parallel, independent evolution of both allelic genes and IGR ‘switches’	pg.25
2.5 Discussion	pg.26
2.6 Future perspectives	pg.29
 3.0 Rotation 2 – Protein expression and purification of the hydrogen cyanide self-protection protein PA4130, a potential anti-virulence target for <i>Pseudomonas aeruginosa</i>	 pg.33-50
3.1 Abstract	pg.33
3.2 Introduction	pg.33
3.3 Methods	pg.35
3.3.1 - Bacterial strains, plasmids and growth conditions	pg.35
3.3.2 – Cloning	pg.37
3.3.3 - Protein Overexpression and SDS-PAGE	pg.38
3.4 Results	pg.38
3.4.1 - Successful Production of Overexpression plasmids	pg.39
3.4.1.1- pSK67 Overexpression Vectors	pg.40
3.4.1.2 - pColdI Overexpression Vectors	pg.41

3.4.1.3 – <i>pCold-GST Overexpression Vectors</i>	pg.43
3.4.2 - <i>Presumptive successful PA4130 induction</i>	pg.45
3.5 Discussion	pg.46
3.5.1 <i>Success of PA4130 induction</i>	pg.46
3.5.2 <i>PA4130 function</i>	pg.47
3.6 Future perspectives	pg.48
3.6.1 <i>Determination of PA4129-PA4134 function</i>	pg.48
3.6.2 <i>Interplay between cioAB and PA4129-PA4134</i>	pg.49
 4.0 – Summary	 pg.50-51
 5.0 – Appendix 1 – ST10, ST167, ST617 strain list and metadata	 pg.52-64
 6.0 – Appendix 2 - R scripts for Gene accumulation and rarefaction curve analysis	 pg.65-67
6.1 – Gene accumulation plot script	pg.65
6.2 - Rarefaction curve plot script	pg.66
 7.0 – Appendix 3 – Pan genome visualisation of ST10, ST167 and ST617	 pg.68
 8.0 – Appendix 4 – Scoary Pan-Genome analysis top hits	 pg.69-78
 9.0 – Appendix 5 – SEER Pan-Genome analysis CDS and IGR breakdown	 pg.79-86
 10.0 – Bibliography	 pg.86-96

1.0 - Background and context

In 2014 the Review on Antimicrobial Resistance (AMR) was commissioned to gain insight into the impact antimicrobial resistance will have globally. This led to the startling figure that an estimated 100 trillion dollars of global productivity will be lost between 2016-2050 due to antimicrobial resistant infections (O'Neill 2014). By employing an economist (Jim O'Neill) to look at the problem of AMR, a new perspective can now be seen highlighting the wide-reaching implications AMR may have if left unchecked. The problem of AMR in a purely health context has been known and explored for a long time and current estimates put deaths caused by AMR infections at 700,000 per year (O'Neill 2014). This highlights the need for implementation of tactics to safeguard both the present and future antimicrobial usage.

Antimicrobial resistance is not a new phenomenon. All current antibiotics contain a biological core as their central structure (Davies and Davies 2010). Even before the therapeutic introduction of penicillin, penicillinases capable of hydrolysing the beta-lactam ring were already present in the bacterial population (Abraham and Chain 1940). This presence can now be fully appreciated in light that AMR genes are normal components in the microbial genetic reservoir (Davies and Davies 2010, D'costa *et al* 2006). Since these toxic compounds are produced naturally, intrinsic host mechanisms of defence evolved alongside antibiotic production to prevent self-intoxication (D'costa *et al* 2006). Therefore, it stands to reason that antimicrobial resistance has been present since the dawn of natural antimicrobial production, so why is AMR becoming such a threat now?

The exploitation of antibiotic for human clinical use was the greatest medical discovery of the 20th surgery. It enabled complete recovery from numerous simple bacterial infections; implementation of various surgeries which were previously untenable due to the likelihood of infection post-surgery and drastically reduced mortality in infants (Davies and Davies 2010). However, it appears that antibiotics are fast becoming a victim of their own success. Both overuse and misuse of antibiotics in clinical and animal husbandry settings has been followed by the emergence of widespread resistance, negating the efficacy of the antibiotic (Davies and Davies 2010). Previously this was combatted by developing a new antibiotic or variant of a current antibiotic (Aminov 2010, Piddock 2012). For example, as early as the 1950's widespread penicillin resistance was proving to be a threat (introduced in 1943), however this problem drove the discovery of methicillin (Spellman and Gilbert, 2014). This

trend continued through the 'golden age' of antibiotic discovery, however in recent years the antibiotic discovery and development pipeline has begun to wane meaning we can no longer keep ahead of the developing resistance (Piddock 2012). If allowed to continue a very likely outcome is a regression to a pre-antibiotic world.

1.1 - Spread of AMR

Emergence and dispersion of antibiotic resistant bacteria is facilitated by the plasticity of the bacterial genome. The simplicity of prokaryotes coupled with extremely fast duplication enables them to adapt to many environmental niches and pressures due to simple mutation (Piddock 2012). This mechanism holds true for some antibiotic resistances, for example the emergence of fluoroquinolone resistance which occurs as a result of mutation in DNA gyrase (Willmott and Maxwell 1993)

It is widely agreed that the most troubling pattern of resistance is occurring in Gram-negative bacteria (O'Neill 2016, Taconelli and Magrini 2017). Organisms such as *Escherichia coli*, *Pseudomonas aeruginosa* and *Klebsiella pneumoniae* are rapidly becoming multiply drug resistant (MDR) (Taconelli and Magrini 2017). MDR resistant infections are extremely problematic as the therapies that can be used are extremely limited. Often the only therapies that can be implemented are the 'last-line' antibiotics known as the Carbapenems and Colistin. However widespread Carbapenem resistance is now being reported at an alarming frequency (Corbella *et al* 2000, Leavitt *et al* 2007 and McGann *et al* 2016), and the recent emergence of mobile Colistin resistance is heralding an age of truly pan-resistant infections (McGann *et al* 2016 and Liu *et al* 2016). Numerous factors are driving the emergence of these MDR bacteria, however the 2 main mechanistic factors are an increasing selection pressure and the ability of bacteria to assimilate mobile genetic elements.

1.1.1- Selection pressure

Selection pressures drive the genomic context of all life. The theory of natural selection still holds true for bacterial life, however the speed at which they divide, grow and mutate mean that evolution happens over a much shorter time period in comparison to eukaryotic life (Andersson and Hughes 2010). It is as simple as '*The greater the volume of antibiotics used, the greater the chances that antibiotic-resistant populations of bacteria will prevail in the*

contest for survival of the fittest at the bacterial level – Helen Gelbard et al 2015'. When we consider the increasing selection pressure we tend to only consider the therapeutic/human context, however the animal husbandry sector is a massive contributor to the increasing selection pressure.

1.1.1.1 Therapeutic pressure

The implications of utilising antibiotics to treat every minor infection was not appreciated in the early years of use. There are various factors responsible for proliferation and spread of AMR in a clinical context such as unregulated access to antibiotics in developing countries and globalization of travel (O'Neill 2016). However, the largest driver of resistance in incorrect usage, which can be attributed to wrong drug usage or simply not following the recommended course duration. Even now incorrect prescription of antibiotics for an infection occurs in 30-50% of cases (Luyt *et al* 2015). Given the tonnage of antibiotics used in America alone (3,379,226 kg in 2012) this corresponds to an enormous, unneeded selection pressure (O'Neill 2015).

Incorrect usage is a huge factor driving the emergence of antibiotic resistance, however the prime cause of this is not necessarily the patient but the lack of a point of care diagnosis (McAdams 2017). Clinicians are unable to identify the organism and its associated resistances in a medically viable timeframe and thus will treat the patient with an antibiotic that has the best chance of working. For example, a *Staphylococcus aureus* infection is difficult to treat with Methicillin, or other penicillin derivatives, due to the high likelihood that it is beta-lactam resistant (McAdams 2017, Liu *et al* 2011). However, a large number of Methicillin Sensitive *Staphylococcus aureus* infections could be treated with beta-lactams successfully. With a point of care diagnostic, we could begin guiding the use of 1st and 2nd generation beta-lactams for use against *Staphylococci* infections safeguarding other last resort antibiotics such as Vancomycin and Daptomycin. This makes development of a point of care diagnostic even more important to enable safe-guarding of what should now be considered finite resources.

1.1.1.2 Animal Husbandry Pressure

AMR spread is not purely down to therapeutic use. Most of the world's antibiotics are actually used in animal husbandry as evidenced by the fact that 70% of antibiotics used in America (8,893,103 kg) are used in animals (O'Neill 2015). Whilst some of these are essential

for therapeutic use the majority are used for the practice of growth promotion (Allen *et al* 2013, Van Boeckal *et al* 2015). Due to the increased demand for meat in Western civilisation the use of antibiotics at sub-therapeutic levels was utilised to encourage growth promotion and boost the yield of the animal, thus driving down meat prices (Van Boeckal *et al* 2015). The mechanism by which this works is not known, although it is hypothesized to manipulate the animal's gut flora preventing establishment of economically harmful diseases such as diarrhoea.

Most concerning is the use of clinically relevant antibiotics for animal prophylaxis and disease treatment (McEwen *et al* 2002). Sub-therapeutic exposure to antibiotics has been shown to encourage the development or acquisition of antibiotic resistance both *in vitro* and *in vivo* (Chait *et al* 2016). Much of the microbial community harboured within the animal gastrointestinal tract is zoonotic in nature, meaning it can successfully be transferred to and colonize humans (Djordjevic *et al* 2013). This means that use of clinically relevant human antibiotics in animals can drive development of resistance which may impact upon the human antibiotic resistome and become clinically relevant. For example, the use of avoparcin (vancomycin derivative) was used extensively for prophylaxis in pigs and chicken in Europe. This was linked with an increase in vancomycin-resistant *Enterococci* due to cross-resistance between avoparcin and vancomycin via the *van* genes in up to 80% of chickens sampled by Aarestrup (1995). (van den Boggard and Stobberingh 2000, Klare *et al* 2003). However, the presence of vancomycin resistant *Enterococci* was not limited to the animal population and an increase was also seen in the human GI tract following the use of avoparcin in animals. It was later confirmed that transfer of the *vanA* gene from *Enterococcus faecium* of animal origin to *Enterococcus faecium* of human origin can occur in the human gut (Lester *et al* 2006). This made treating opportunistic *Enterococci* infections hard due to widespread beta-lactam and aminoglycoside resistance combined with the newly emerged vancomycin resistance (van den Boggard and Stobberingh 2000, Klare *et al* 2003). In regions where Avoparcin was not used (such as the USA), no rise was seen in the level of vancomycin resistant *Enterococci* in both animals and the community (Bonten *et al* 2001). Upon withdrawal of avoparcin the levels of vancomycin resistant enterococci dropped, however a low level of resistance remained highlighting the fact that simple withdrawal is not enough to combat AMR due to mechanisms such as co-selection and bacterial persister cells (Lauderdale *et al* 2007).

The use of antibiotics for prophylaxis has been used since the 1950's and has resulted in an extremely high level of background resistance, with numerous studies demonstrating high incidence of various resistances to multiple antibiotics (Van den Bogaard *et al* 2001, Weber *et al* 2017 and Osterberg *et al* 2016). In 2006 the use of antibiotics for prophylaxis and growth promotion was banned in Europe order to limit unnecessary antibiotic exposure (Anadon 2006). The precedence was taken from Scandinavian countries where the use of antibiotics has been banned since 1990's (Anadon 2006). Despite an initial slump in profits and an increase level of morbidity and mortality on both pig and chicken farms, they maintained competitiveness in the market by improving animal welfare and conditions returning to almost the same level as when antibiotics were in use (Bengtsson and Wierup 2006). Whilst this ban has resulted in a drop in the incidence of antibiotic resistance locally, background levels are still significant (Cogliani *et al* 2011, Marshall and Levy 2007). This is due in part to mechanisms such as co-selection and the ability of organisms to offset the cost of harbouring and expressing various resistances (Baker-Austin 2006 and McNally *et al* 2016). However, the real undermining factor is the lack of a uniform worldwide approach. The majority of the world still utilise antibiotics for growth promotion and prophylaxis, and given the globalisation of import and export the dissemination of antibiotic resistance from foreign food production is still a problem (Hoffman *et al* 2015).

1.1.2 Mobile genetic elements

Bacterial diversity is a key factor in the spread of AMR, especially in the case of Gram negative bacteria. These organisms can recombine and assimilate new pieces of DNA quickly enabling fast adaptation to a new environmental pressure. Various genetic elements facilitate the spread of AMR but in the context of MDR central to the story are Plasmids, Transposons and Integrins.

1.1.2.1 Plasmids

Plasmids are self-replicating circular pieces of DNA which can be transferred both intra and interspecies via conjugation. A highly diverse set of genes can be encoded, however in recent times multiple antibiotic resistance genes are being found more frequently on these plasmids (Carattoli 2009). This is extremely problematic in the sense that the organism can become resistant to multiple antimicrobials in a very short space of time. More troubling is the fact that selection against one of the encoded resistance genes will result in maintenance of the entire plasmid (Baker-Austin 2006 and Fang *et al* 2016). This is one of

the problems with the simple 'withdrawal' method of combatting antibiotic resistance. Continuing the previous example of Avoparcin. Following its withdrawal prophylactically, vancomycin resistance levels dropped but not as anticipated. It was shown that a small subset of these *vanA* genes were linked on the same plasmid as *ermB* (Macrolide) resistance (Aerestrup 2000). This meant that vancomycin levels only dropped to current levels following withdrawal of both Avoparcin and Macrolides. This example illustrates the mechanism that was observed on European farms following the 2006 ban for many antibiotic groups. Various factors played a part in this such as the practice of herd antibiotic treatment, however the key factor seemed to be the rise in the use of copper and zinc in animal feed for growth promotion (Bednorz *et al* 2013, Seiler and Berendonk 2012). An *in vivo* study by Bednorz and colleagues (2013) demonstrated that the proportion of MDR *E. coli* in the porcine gut increased drastically (18% vs 0%) in piglets treated with elevated levels of zinc versus normal background levels. Zinc resistance was shown to be genetically linked to antibiotic resistance, in numerous cases, on plasmids showing that co-selection does indeed play a part in antibiotic resistance maintenance post antibiotic use (Baker-Austin 2006, Bednorz *et al* 2013 and Seiler and Berendonk 2013). This means that the use of one antimicrobial alone will maintain resistance to multiple antibiotics, complicating the process of eliminating the antibiotic resistance gene reservoir.

1.1.2.2 – Transposon, Insertion Sequences and Integrations

Transposable elements are able to significantly alter the antimicrobial susceptibility profile of an organism. These elements can insert themselves either chromosomally or on conjugative elements facilitating the quick spread of various foreign gene cassettes enabling quick adaptation to environmental pressures (Bennett 2008).

Transposons are short DNA fragments which can 'jump' and insert themselves into specific DNA regions. These elements either encode a transposase which facilitates movement or a reverse transcriptase which generates RNA which is subsequently copied to DNA and inserted elsewhere (Bennett 2005 and Bennett 2008). Insertion sequences effectively function as simple transposases and allow the formation of structures called composite transposons. Transposons and insertion sequences are often associated with antibiotic resistance cassettes and have been implicated with the emergence of various resistances due to their ability to mobilise chromosomal elements (Wellington 2013, Karim *et al* 2001 and Kumarasamy *et al* 2010). This mobilisation enables the dispersion of native resistance

mechanisms utilised by antibiotic producing bacteria, or bacteria occupying the same ecological niche. This has shown to be the case for the emergence of a new type of extended spectrum beta-lactamase termed CTX. In the 1970's and 80's the predominant beta-lactamases found were TEM and SHV or variants of this genetic type (Canton *et al* 2012). However, in the 90's CTX type emerged as a novel beta-lactamase and has now become globally disseminated with CTX-15 dominant in the USA and Europe, whilst CTX-M-14 is most frequently associated with China (Canton *et al* 2012). It was shown that the CTX genes share a large degree of homology with the host beta-lactamase of *Kluyvera spp* (Rodrigues *et al* 2004, Canton *et al* 2012). Further work shows these CTX enzymes to be associated with insertion sequences, for example CTX-M-15 with ISEsp1 (Tzouveleakis *et al* 2000). This enabled the CTX genes to be mobilised onto plasmids which, when combined with global travel, facilitates the global dissemination of such resistance types in a very short period (Canton *et al* 2012).

Integrans are regions of DNA which enable site-specific recombination via an integrase gene, alongside a proximal recombination site which enables incorporation of various gene cassettes (Stokes and Hall 1989). Integrans are found throughout life and the vast majority are chromosomal elements. In the context of AMR Integrans have become major contributors to the spread of MDR due to the ability of Integrans to 'capture' multiple antibiotic resistance cassettes and control expression of these cassettes with an efficient promoter (Gillings *et al* 2008). These cassettes are often upregulated during the SOS response caused by environmental stress (i.e. antibiotics) ensuring stringent control and thus minimising the fitness cost of harbouring and expressing the resistance cassettes. Integrans themselves are not mobile, however they can become associated with Transposons and Insertion sequences enabling mobilisation onto conjugative plasmids (Stokes and Hall 1989, Gillings *et al* 2008). Only Integrans that are mobile have an impact on AMR with Class 1,2 and 3 Integrans the main culprits (Gillings *et al* 2008).

The emergence of Class 1 integrans typifies how selection pressure enables mobile genetic elements to facilitate the spread of AMR. Whilst Class 1 integrans can encode a wide range of antibiotic resistance cassettes they are all associated with a Tn402-like element indicating a recent common ancestor is responsible for the mobilisation and dispersion of these elements (Radstrom *et al* 1994, Gillings *et al* 2008). The ancestral state of Class 1 integrans was identified on the chromosome of a soil organism, sharing the characteristics of a

chromosomally encoded integron (Stokes *et al* 2006). Upon association with Tn402, which encodes *sul1*, the spread of class 1 integrons was likely facilitated by the wide use of sulphonamides. This resulted in the emergence and dispersion of a mobilised Class 1 integron which can assimilate new resistance gene cassettes allowing adaptation to the current environment i.e. what antibiotics were in use (Gillings *et al* 2008). These elements have become particularly problematic in clinical Gram negative *Enterobacteriaceae* where up to 70% of pathogens harbour such elements, complicating antibiotic treatment due to their MDR nature (Rao *et al* 2006).

Various selection pressures result in the transfer of multiple transposons encoding different resistances along with transfer of Integron elements which can encode numerous cassettes. This can lead to the development of large MDR plasmids via sequential addition of transposable elements. This pattern has already been demonstrated in *Salmonella* by using a 30-year retrospective analysis of IncF1 type plasmids by Caratolli *et al* (2001), where sequential addition of new integrons altered the antibiotic resistance profile. These large MDR plasmids are now becoming common place in both pathogenic and commensal Gram-negative organisms underpinning the wider antibiotic resistance gene reservoir. This means that adaption to multiple antibiotic challenges can now occur almost instantaneously via a conjugal transfer event. It was previously thought that harbouring these large MDR plasmids incurs a substantial fitness cost affecting the ability of both pathogens and commensal strains to establish themselves. However, recent work performed by McNally and colleagues (2016) has demonstrated that *E. coli* ST131 is able to offset the cost of harbouring and expressing such large extra-chromosomal elements via promoter 'switching', altering the expression profile of ST131.

1.2 Current Approach to Diagnostics and Treatment

As mentioned previously we are severely hindered by a lack of point of care diagnostics able to inform a clinician of relative resistances within a medically viable time frame. The majority of methods are still based upon rudimentary culture and antibiotic exposure methodologies i.e. disk diffusion assay (O'Neill 2015). These methods can take days and in the case of *Mycobacterium tuberculosis* as long as ~3 months. This presents huge problems as clinicians often need to begin treatment immediately. The ultimate goal has been set out by the longitudinal prize and would achieve 4 main objectives.

- 1 Is the infection causing the illness bacterial or viral?
- 2 If bacterial, what type of bacteria is causing the infection
- 3 Are the bacteria that are causing the infection resistant to available antibiotics?
- 4 Are the bacteria that are causing the infection susceptible to existing drugs?

Current methods are not readily available to answer all of these points yet, however awards such of the longitudinal prize have sparked a massive amount of research into development of such methodologies capable of answering these questions.

1.2.1 – Future Diagnostics

Step by step process is more likely than a ‘magic bullet’ answering all the points previously set out. Experimental point of care (POC) diagnostics have begun entering the market which have been shown to reduce the use for antibiotics. One such simplistic example allows the distinction between a viral or bacterial respiratory infection. The CRP (C-reactive protein) test alongside measuring white blood cell count, enables prediction of whether the infection is bacterial or viral and thus whether antibiotics are needed. In a study from the Netherlands antibiotic prescription for respiratory illness dropped from 53 to 31%. However, uptake of this methodology is poor with only the Netherlands and Scandinavian countries utilising this methodology extensively. It is no surprise that these countries have some of the lowest prescription rates worldwide, although a lack of a worldwide uniform approach will undermine any local success.

One methodology which has been showing promise as a point of care diagnosis is whole genome sequencing. Genome sequencing has enhanced our understanding of life at a genetic level. This has provided us with knowledge on the extent of the problem AMR poses and has enabled us to gain local insight into the spread of resistance and thus what regimes will be effective in a particular region. The rate at which sequencing has developed may allow us to sequence bacterial genomes and establish treatment regimens based to antibiotic resistance carriage in a matter of minutes. Previously sequencing has only been used as a purely retrospective epidemiological tool, however the use of the Min-Ion, which utilises oxford nanopore technology, has already begun to be implemented in viral outbreak regions such as Ebola and Zika (Quick *et al* 2016, Quick *et al* 2017 and Faria *et al* 2017. This was used in a purely epidemiological context to try and establish the origin of the outbreaks and to track its evolution in real time. However, more recently this methodology began

being implemented in bacterial outbreaks and informed clinicians of appropriate therapies without resorting to antibiotics of last resort on the island of San Tome (Quick *In Press*). The use of platforms such as the min-ion allow the generation of real-time data by linking the data generated to various genomic databases at a low cost in comparison to similar sequencing platforms. This means that you will have the ID of a causative organism within minutes along with any associated resistance genes following completion. The dropping cost of min-ion sequencing is making it an attractive possibility for clinical use in both economically developed and developing regions. However, significant streamlining of the process is needed to call it a POC diagnostic. Although detection of a pathogen in a clinical sample can be achieved within 6 hours a low bacterial or viral load makes it much harder. Significant bioinformatics training is also needed to interpret the data correctly, a skill which is lacking among clinicians and microbiologists alike.

1.3 – Combatting AMR

The development of antibiotics has slowed drastically in recent years with many ‘big’ pharma companies unwilling to invest in research and development due to low and unguaranteed returns (O’Neill 2016). This has resulted in the implementation of numerous financial incentives to readdress this problem. However, the development of new antibiotics is a risk since the ‘low hanging fruit’ have been discovered and any further development will be both technically and financially challenging. This is very worrying when we consider that only 3 new treatments for carbapenem resistant Gram-negative bacteria were under development in 2016 (O’Neill 2016).

The current model proposed by Jim O’Neill recommends a fluid approach where financial incentives are linked to the most urgently required treatments. For example, at present the largest financial incentives are currently aimed at new treatments for Tuberculosis and antibiotics effective against MDR Gram negative pathogens. On the back of these reports a Global Innovation fund was set up providing financial incentives for research and development into new antibiotics with a focus on early stage research which consumes vast financial resources with no returns.

1.3.1 – Alternative therapies

Along with pump-funding the antimicrobial pipeline alternative avenues for lowering the workload on antibiotics are now being explored. These have the added benefit that they reduce the 'workload' on antibiotics. The development of alternative therapies does not necessarily have to displace the use antibiotics, they can assist or simply lower the demand. 2 methods currently under investigation are vaccine development and novel therapeutics such as virulence factor inhibition.

1.3.1.1 – Vaccine development

The use of vaccines to treat infections has been around since 1768 when Jenner noted the protective effect of cowpox against smallpox. Various bacterial vaccines have been used to almost completely eradicate diseases and as such remove the need for antibiotic treatment. Successful examples include against *Clostridium tetanis*, the cause of tetanus (89% protection after 10 years) and *Corynebacterium diphtheria* (96% protection after 3 doses). In the context of antibiotic resistance, vaccines against *Neisseria meningitidis* and *Streptococcus pneumonia* have almost eliminated the AMR problem for these organisms lowering usage (Lipstich and Siber 2016). In the advent of continuous new antibiotic production, vaccine development began to wane due to their perceived uselessness with a narrow spectrum of treatment and often ineffective results (O'Neill 2016). However, in the advent of AMR these valuable tools have gained much attention.

One of the major problems with utilising vaccines as an alternative therapy is global dissemination of the vaccine. Many are required intravenously, are needed multiple times and thus, require the appropriate infrastructure and expertise to administer (Lipstich and Siber 2016 and O'Neill 2016). In order to have full effectiveness herd vaccination is needed to prevent spread of the AMR pathogen. The aim is to prevent dissemination of the bacteria lowering the chances of it transferring resistance elements, thus preventing infection and disrupting the spread of resistance simultaneously. Significant funding is needed to disseminate the vaccine and thus development of oral or intranasal vaccines should be encouraged to maximise effectiveness. Vaccine development is similar to antibiotic development in that it is both technically and financially challenging with numerous failed candidates (O'Neill 2016). Therefore, more 'arms' need to be explored to ensure a multi-pronged attack and maximise development of therapies to prevent the spread of AMR and any subsequent infections.

1.3.1.2 - Novel Therapeutics

Recent advancements in sequencing and analytical technologies have opened exciting new avenues in therapeutic development. One such methodology described by Dubern and colleagues (2015) utilised a combination of Transposon mutagenesis, bioinformatics, in vitro assays and multiple in vivo models to identify virulence factors key for host colonisation and virulence. These targets have now begun forming the basis of anti-virulence targets for therapeutic development utilising structural biology, a topic discussed in further depth in rotation 2. Utilising methodologies such as this and TRADIS has enabled identification of numerous novel targets which can be exploited for therapeutic development. These studies are still in their infancy and as such data is lacking on successful exploitation, however 'blue sky' research like this is needed to continuously expand avenues to treat AMR and disease in general.

1.4– Introduction to projects

The problems illustrated above demonstrate how complex the problem of AMR is. This means that a multi-faceted approach is needed to ensure success in combatting this global threat. The projects selected during this MRes year illustrate the need for multi-disciplinary training to enable microbiologists to understand the science behind the spread of AMR and how other scientific disciplines can enhance our understanding. Project 1 focuses on the use of Whole genome sequencing to provide mechanistic insight into the establishment of antibiotic resistance and its maintenance within an environment. Whilst project 2 focuses on the early stages of alternative therapy development against *Pseudomonas aeruginosa* by creating novel virulence inhibitors. Although these projects are not directly linked, they are parts to a much larger picture. Multiple small steps towards the ultimate target of reducing the incidence of antibiotic resistance are needed for both short and long-term success.

2.0 - Global dissemination of ST167 and ST617 as multi-drug resistant clones, a familiar story.

2.1- Abstract

The increasing level of multi-drug resistant infections caused by *Escherichia coli* is a global health crisis. The use of whole genome sequencing has revolutionised the understanding on how clones are able to stably maintain large MDR plasmids. Most studies focus on pathogenic lineages of *E. coli* however the recent emergence of ST167 and ST617, members of the ST10 clonal complex, as globally disseminated ESBL and carbapenemase associated sequence types highlights the need to explore carriage in the commensal population. Here we used a combined GWAS approach to compare ST167 and ST617 to central allelic profile of the ST10 clonal complex, ST10. This revealed parallel independent evolution towards the same novel O antigen type and patterns of metabolic adaptation previously reported in ST131. Our study demonstrates that clonal expansion of successful MDR clones is not just limited to pathogenic *E. coli* and that the commensal population are not just transient carriers of MDR resistance plasmids. It shows a common evolutionary mechanism is employed by ST167, ST617 and ST131 to become dominant MDR sequence types and highlights the need to more rigorously research commensal carriage as part of combatting the wider resistance gene reservoir.

2.2 – Introduction

The use of Antibiotics in the last ~70 years has revolutionized modern medicine. Paralleled with this increased usage of antibiotics is a rise in the incidence of resistance, and more worryingly, multi-drug resistant (MDR) bacteria (Chen *et al* 2013, Ridge *et al* 2011). Recently WHO listed carbapenemase and extended spectrum beta-lactamase (ESBL) producing *Enterobacteriaceae* as 'Priority one' for developing new antibiotics (Tacconelli and Magrini 2017). Amongst *Enterobacteriaceae*, *Escherichia coli* are a significant healthcare burden, causing a vast array of infections from diarrhoeal to extra-intestinal disease such as UTI's (Kaper *et al* 2004, de Kraken *et al* 2013). Extra-intestinal Pathogenic *E. coli* (ExPEC) are often linked to MDR infections with 50% of strains isolated from UTI and bacteraemia demonstrating resistance to 3 or more different antibiotic classes (Mathers *et al* 2015). The emergence of MDR *E. coli* is linked to acquisition of large MDR plasmids encoding multiple

resistance genes (Mathers *et al* 2015). Of particular concern are strains carrying these ESBL encoding MDR plasmids, since only carbapenems and colistin are treatment options (Mathers *et al* 2015, Livermore and Hawkey 2005). However recent reports of carbapenemase producing enzymes and mobile colistin resistance (*mcr-1*) are heralding the emergence of 'pan-resistant' bacteria (Liu *et al* 2016, McGann *et al* 2016 and Mediavilla *et al* 2016).

Despite *E. coli* being an extremely diverse species genetically, only a small number of lineages are responsible for the global dissemination of ESBL. ST73, ST95 and ST648 have all been linked with ESBL carriage (Alhashash *et al* 2013, Croxall *et al* 2011 and Ewers *et al* 2014); however, the most dominant lineage is ST131. ST131 is the most common cause of UTI in the developed world (Mathers *et al* 2015). Three distinct clades have been identified within ST131, of which clade H30Rx is associated with the ESBL CTX-M-15 (Price *et al* 2013). This clade of ESBL producing ST131 are extremely similar genetically indicating a clonal expansion is responsible for global dissemination (Price *et al* 2013). Similar patterns are observed in other ExPEC lineages such as ST73 and ST648 (Alhashash *et al* 2013 and Ewers *et al* 2014); and this holds true for other *Enterobacteriaceae* such as *Salmonella* and *Serratia* species (Butaye *et al* 2006 and Moradigaravand *et al* 2015). In the case of ST131 acquisition of specific virulence factors, such as *fimH30*, and increased usage of cephalosporin's and fluoroquinolones as therapy ~25 years ago was followed by acquisition of the IncFII type plasmid encoding CTX-M-15 and chromosomal mutation in *gyrA* and *parC*, conferring extensive resistance to both cephalosporin's and fluoroquinolones (Stoesser *et al* 2016 and Zakour *et al* 2016). Acquisition of such chromosomal mutations and large MDR plasmids is often associated with a fitness cost, compromising the ability of a pathogen or commensal to establish itself in a niche. This accounts for why only a limited number of *E. coli* sequence types can be both pathogenic and MDR. It seems the stable acquisition of both the plasmid and chromosomal mutations allowed ST131 to become the dominant ExPEC lineage due to the reduced spectrum of treatment (Harrison *et al* 2015). It was inferred that stable acquisition was achieved by allelic switching in gene regulatory regions, due to the apparent similarity between promoter regions in ST131 clades A and B regions whilst clade C (H30Rx containing clade) promoter regions exhibited far greater disparity (McNally *et al* 2016). This allows ST131 to maintain these resistances as they have managed to offset the 'cost' of MDR plasmids via compensatory mutations in promoter regions, despite the absence of immediate selection pressures. Recent work done by McNally and colleagues (McNally *In*

Press) has demonstrated that only a small subset of genes has come under selection between clade A and B ST131 and clade C (H30Rx). Many the genes that have undergone selection are involved with anaerobic respiration and metabolism and have been implicated in allowing the persistence of ST131 in the human gut for ~1 year. This may have facilitated the emergence of ST131 as the worldwide drug resistant clinical burden we see today. By ‘hanging around’ in the human GI tract ST131 was more likely to be exposed to extended spectrum beta-lactams and fluoroquinolones and thus select for resistance to these pressures. This coupled with the ‘switching’ in promoter regions allowed for selection of a *E. coli* capable of resisting treatment without compromising virulence.

A recent pre-print published by Zhou and colleagues identified ST167 and ST617 as emerging ESBL/CPE producing *E. coli* following a dedicated screen in the ICU of West China Hospital, Chengdu. ST617 was shown to be associated with CTX-M-15 and appears to have undergone a similar clonal expansion event as ST131. ST167 follows an unusual antibiotic resistance profile in that acquisition of resistance genes appears to be highly dynamic. Within the ST167 it was shown to encode 4 different carbapenemases (*bla*NDM-1, *bla*NDM-5, *bla*NDM-7, *bla*OXA-181 and *bla*KPC-3) and rapid resistance gene switching was identified in almost identical isolates from America and China (Zhou *et al* 2016). Despite the variability exhibited by ST167 it was shown to have undergone a similar clonal expansion event as ST131 and ST617 based on the SNP matrix distances (Zhou *et al* 2016). Both ST167 and ST617 belong to the ST10 clonal complex, where only transient carriage of ESBL and CPE had previously been reported (Alhashash *et al* 2013). The ST10 clonal complex are primarily host commensals lacking the known virulence factors required to cause both diarrhoeal and extra-intestinal infection (Kohler and Dobrint 2011). However, the study by Zhou *et al* indicates that members ST10 clonal complex may serve as an important genetic ‘reservoir’ of antibiotic resistance, potentiating the global dissemination of multi-drug resistance.

Currently very little data exists on carriage of ESBL and carbapenem resistance with the commensal *E. coli* population. In this study, we used a combined approach to genome wide association studies (GWAS) to determine any common genetic features which could account for the ability of ST167 and ST617 to offset the cost of carrying MDR. During this process, we identified an emerging sub-clade of ST10’s which demonstrate an MDR phenotype, and carry an ESBL gene. This ST10 sub-clade along with ST167 and ST617 appear to express the same novel O antigen which is synthesized by the *wzm/wzt* ATPase pathway.

In addition, a large proportion of the anaerobic respiration/metabolic genes identified in the ST131 study performed by McNally and colleagues (2016) appear to have come under selection in ST167 and ST617, providing a potentially exciting mechanistic insight in how MDR *E. coli* are able to establish themselves as globally disseminated sequence types.

2.3 – Methods

2.3.1 - Sequence analysis

Enterobase (www.enterobase.warwick.ac.uk) was interrogated on the 20/02/2017 and 163 ST10 *E. coli* genomes were downloaded from numerous independent studies with a diverse set of geographical locations. A further 96 ST10 genomes were provided by Sebastian Gunther containing commensal ST10 isolates of animal and human origin (Appendix 1). The ST167 and ST617 genomes from both the West China hospital screen and the original enterobase interrogation were provided by Alan McNally (Appendix 1), as used in the study from Zhou *et al.* (2016) All genomes were annotated using Prokka with the antibiotic resistance carriage of each strain determined by ABRicate (<https://github.com/tseemann/abricate>) and visualized using Phandango (Hadfield *et al* 2017).

2.3.2 - Phylogenetic analysis

Core genome phylogeny was inferred using Parsnp from the Harvest suite of phylogenetic tools (. This creates a phylogenetic tree based on co-linearized genomic blocks with the ST10 strain 216 acting as a reference genome. The tree produced was used to overlay metadata using ITOL and Phandango (Hadfield *et al* 2017).

2.3.3 - Core and Pan-Genome analysis

Separate pan genomes for each sequence type, and then all combined, were constructed using Roary with the `-e -mafft` setting to create an alignment of the core CDS of 2.2Mbp and a pan genome consisting of 51000 genes. Gene accumulation plots were created using adapted R scripts from Meric *et al* (2014) with the gene presence/absence matrix output generated by Roary (Appendix 2).

Two approaches to GWAS were used to ensure congruence between methods. Scoary uses the gene presence/absence output produced by Roary alongside a file containing a pre-

determined set of traits to correlate gene presence/absence to a sequence type (Brynildsrud *et al* 2016). SEER is independent of CDS constraints utilising distributed string mining to produce kmers of multiple lengths allowing identification of both gene presence/absence and any mutational event within these genes or IGR's, whilst correcting for population structure. This provides a higher level of sensitivity allowing identification of SNP's, however it is more computationally and memory intensive (Lees *et al* 2016). Kmers outputted as significant by the SEER analysis were Blasted against ST167 reference genome Y5 which was also utilised in all other analysis. Unfortunately, no complete annotated ST617 genome was available at the time of this study.

2.3.4 - In Silico Serotyping

Serotyping was performed using two independent methods with SRST2 and SerotypeFinder (Ingle *et al* 2016 and Joenson *et al* 2016). Both methods utilise WGS data to allow identification of O and H antigens expressed by an organism. SRST2 utilises an entirely nucleotide sequence orientated methodology to assign antigenic type from an internal database. This approach allows identification of novel antigenic types and updates the database accordingly. SerotypeFinder uses the WGS data to determine the genetic make-up of the antigen biosynthetic cluster. The locus identified is then correlated with classical serotyping results of strains containing the closest matching genetic make-up.

2.4 – Results

2.4.1 – ST167 and ST617 are distinct monophyletic clusters that arose through clonal expansion

Phylogeny for ST167, ST617 and ST10 was inferred using Parsnp from co-linearized blocks. It was previously reported that ST167 and ST617 were monophyletic clusters within the ST10 clonal complex as a result of a clonal expansion event, based on the number of SNP's between strains (Zhou *et al* 2017). This was confirmed with ST167 and ST617 forming independent clusters in relation to ST10 (Figure 1). Analysis of the pan-genome further supported this observation with both core genome (rarefaction) and gene accumulation curves for ST167 and ST617 demonstrating a closed pan genome in comparison to ST10 (Figure 2). This is indicative of a clonal expansion event since ST167 and ST617 have diverged less from each other and thus share a higher amount of core genes and lower level of accessory genes. ST10 is far more diverse forming multiple clusters within the tree and as

such exhibits greater diversity in both its core and pan-genome (Figure 2). Visualisation of the pan-genome yielded genomic patterns specific to both ST167 and ST617 independently, however no obvious shared regions between ST167 and ST617 were observed (Appendix 3).

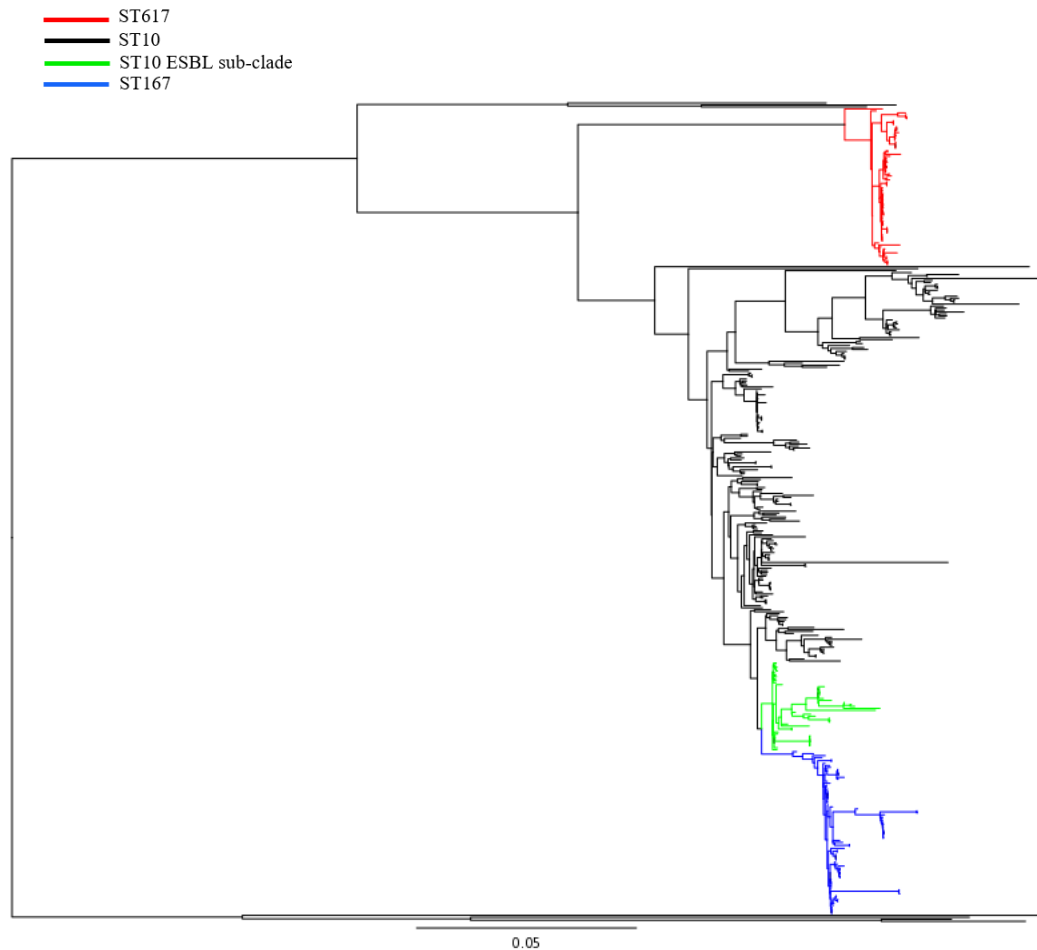


Figure 1) Phylogeny of ST167, ST617 in relation to ST10. Phylogenetic tree generated by Parsnp using co-linearized blocks of the core genome. Demonstrates that ST167 and ST617 emerged from ST10 independently and did not emerge from each other. Displays the clonal nature of ST167 and ST617 in comparison to the diversity exhibited by members of ST10 based on tree branch length.

2.4.2 - Antibiotic resistance gene carriage

Initially the antibiotic resistance gene profile was generated to confirm both the initial observations for ST167 and ST617 in Zhou *et al* and to determine the resistance profile for the ST10 isolates. As previously reported ST167 displays a highly dynamic resistance pattern encoding various ESBL's and carbapenemases, whilst ST617 displays a more stable resistance profile associated with CTX-M-15 and *blaOXA-1* (Figure 3). The ST10 genomes displayed transient resistance gene carriage, however a small sub-set of ST10 strains appear to display

a common resistance genotype (CTX-M-1 and *bla*TEM1-b). Visualisation of the resistance gene profile alongside the maximum likelihood phylogenetic tree revealed that those demonstrating this genotype seemed to be localised to a small sub-clade of the ST10 population (Figure 3). This small sub-clade of ST10 appears to branch off at the same point as ST167, presenting an interesting opportunity to study the evolution from a non-MDR to an MDR plasmid carrying strain

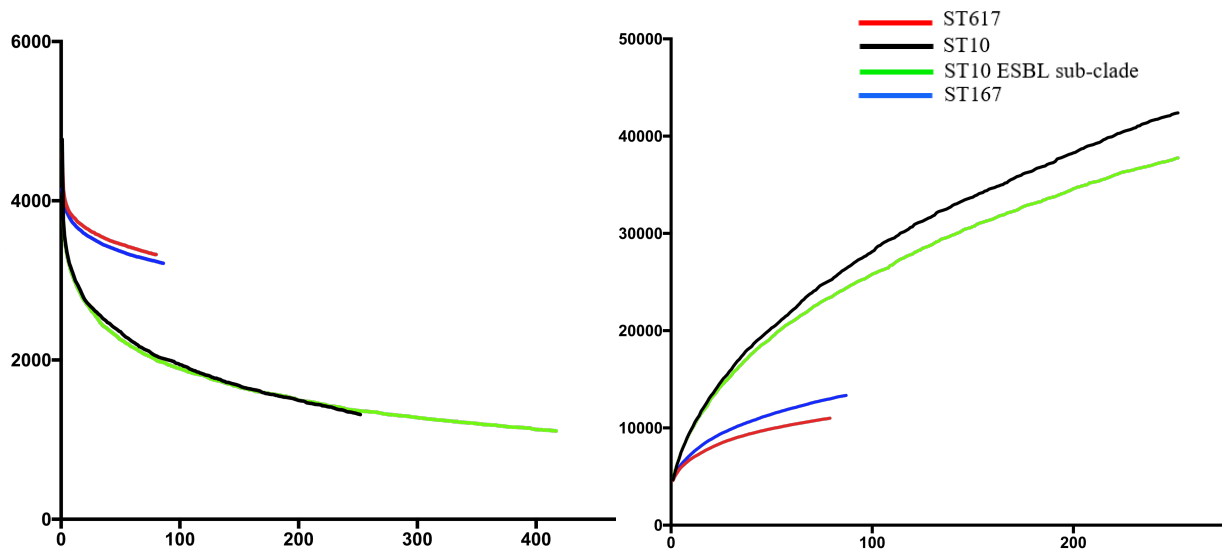


Figure 2) Gene accumulation and rarefaction (core genome) plots of ST167/ST617 in comparison to ST10. Graphs showing the number of core genes present (A) and total genes present (B) following 100 bootstraps with randomised permutations. Demonstrates the closed nature of the ST167 and ST617 pan-genome due to the high number of conserved genes (~3200) and the quick plateau exhibited in the gene accumulation graph (B).

2.4.3 - Parallel independent evolution to novel O antigen

A combined GWAS approach using both Scoary and SEER identified a wide range of genes associated with both ST167 and ST617 but not the ST10 population (Appendix 4). Most striking was the absence of the *wzzB* gene and *wca* biosynthetic cluster in both ST167 and ST617 whilst the majority of the ST10 population contained both. These genes are involved in LPS biosynthesis indicating rearrangement of the LPS biosynthetic cluster has occurred in both ST167 and ST617. Further analysis revealed that those ST10 strains which did not encode *wzzB* and the *wca* biosynthetic cluster belong almost exclusively to the previously mentioned emerging CTX-M-1 encoding ST10 sub-cluster.

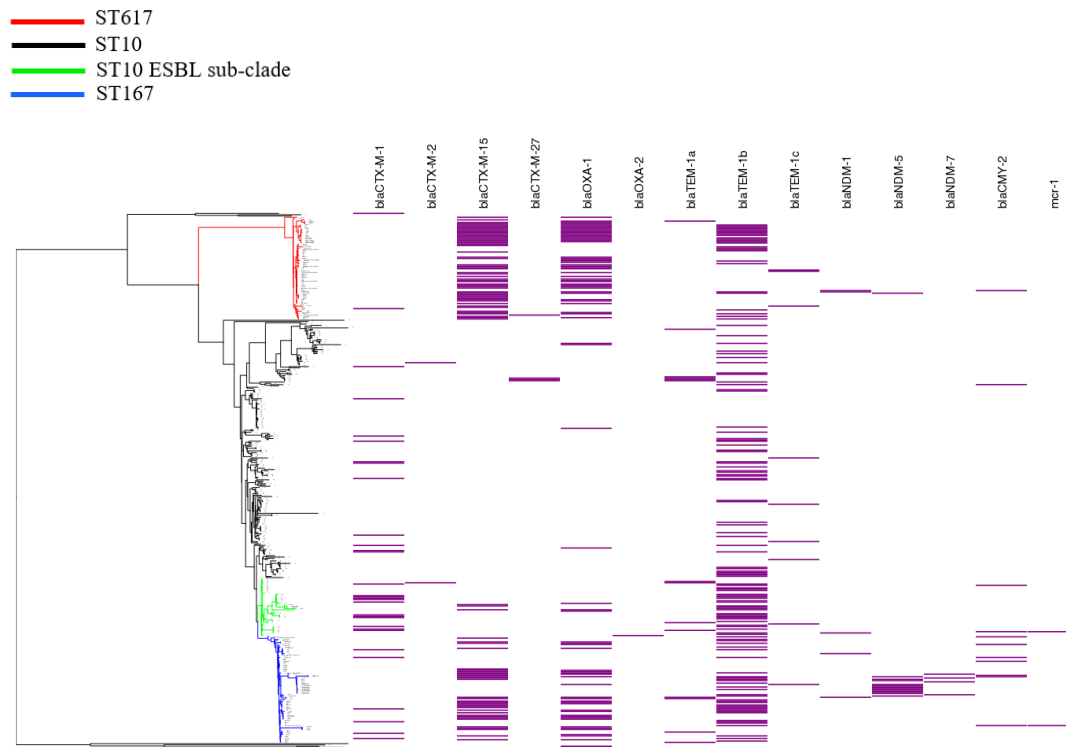


Figure 3) Beta-lactamase resistance profile. Core genome tree made using Parsnp using the entire collection of ST10, ST167 and ST617 strains. The phylogeny is inferred from co-linearized blocks in the genome of each strain. The tree is overlaid with the major beta-lactamase resistance genes and *mcr-1* as determined by ABRicate. Displays the strong association of ESBL resistance (CTX-M-15) associated with both ST617 and the highly dynamic profile previously reported in ST167. Unexpectedly an emerging ST10 ESBL subclade is enriched for CTX-M-1.

Since rearrangement of the LPS biosynthetic cluster has occurred in ST167 and ST617 we established the antigenic type using two *in silico* methods, SRST2 and SerotypeFinder. SRST2 established that ST167, ST617 and the ST10 sub-clade demonstrate the exact same O antigenic type (O32novel, based on levels of nucleotide identity) with similarity also seen in H antigen type (H9 or H10) (Figure 4). SerotypeFinder identified the strains as O89, since it uses WGS data cross-referenced with classical serotype data.

Sequence alignment of O89 and O32novel revealed homology of around 95% for both ATP binding protein *wzt* and the ATP-dependant permease *wzm*. In the case of *wzt* there is significant modification towards the C terminus, with this region being important in binding specificity of the O- antigen polysaccharide (Cuthbertson *et al* 2007). The ATP-binding cassette of *wzt* also contains a large amount of SNP's, although the majority do not occur in the 6 domains necessary for ATP binding except for the H-Loop (5 SNP's). *wzm* also displays significant modification towards the C –terminal end of the nucleotide sequence, however

the absence of any reliably annotated conserved domain features it is difficult to assign any biological significance to *wzm* transport permease.

2.4.4 - Parallel independent evolution of both gene and IGR 'allelic' switches

Both Scoary and SEER identified allelic differences in genes which could be implicated in host adaptation (Appendix 4, 5 and figure 5). However, the annotation by Scoary is inconsistent and it fails to identify intergenic regions (IGR's) which may have undergone allelic switching. Seer identified 46 IGR's, which were missed due to Scoary's CDS constraints, and identified another potential ~240 CDS which display allelic differences in the ST167/ST617 population (appendix 2 and 3). A high proportion of these CDS and IGR's appear to be linked to metabolism, with a large degree of crossover exhibited between these results and the results observed in the study performed by McNally *et al* 2016 on ST131, especially with relation to anaerobic metabolic/respiratory processes and colonisation (appendix 4, 5 and figure 5). Of note is the identification of shared *eut* and *pdu* allelic differences between ST167 and ST617. These genes enable the utilisation of ethanolamine and fucose as a sole carbon source in organelles known as carboxysomes (Garsin 2010 and Dogan *et al* 2014). Upregulated use of such substrates has been frequently linked to pathogenesis of enteric pathogens (Bertin *et al* 2011 and Thiennirmitr *et al* 2011) and has been implicated in the success of ST131 as an MDR clone (McNally *et al*, in press). Ethanolamine and fucose are poorly utilised by the enteric microbiota providing *E. coli* with niche carbon sources to facilitate colonisation potential modulation of survival and attachment of ST167 and ST617 in the GI tract (Figure 5 and appendix 5)(Thiennirmitr *et al* 2011). SEER also identified numerous loci involved with acid and bile salt tolerance alongside numerous fimbriae-like proteins indicating potential modulation of enteric attachment and survival, although extensive mechanistic work is required to confirm this hypothesis (Appendix 5).

Recent work undertaken by McNally *et al* (2016) highlighted the role that that allelic 'switching' in IGR regions of ST131 clade C MDR plasmid carriage cost to be offset. Far less allelic switching was detected in comparison to the ST131 dataset with only 40 IGR switches detected by SEER. This is likely due to the fact we are grouping together two distinct

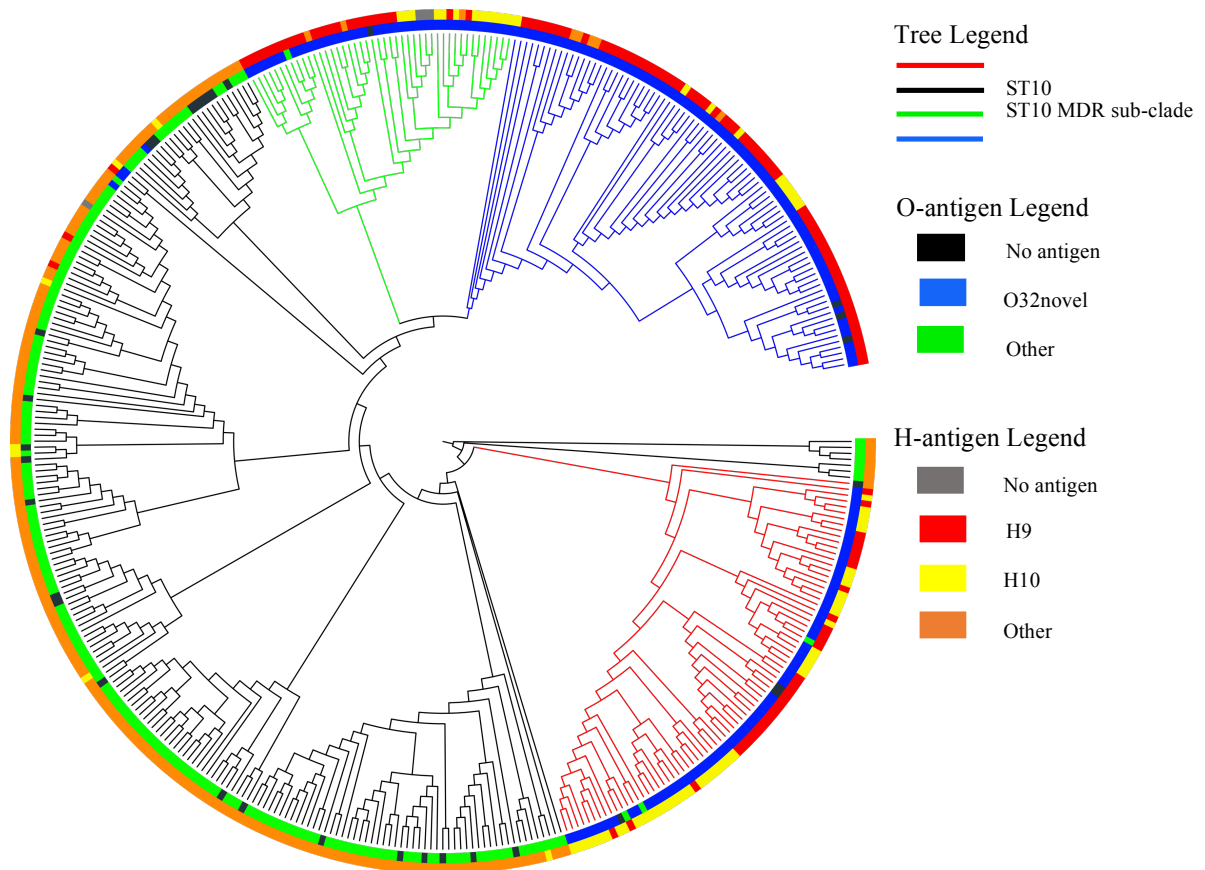


Figure 4) ST10, ST167 and ST617 serotype data. O-antigen and H-antigen information overlaid with Core genome tree generated by Parsnp. Demonstrates the antigenic switching to O32novel (inner ring) and H9/H10 (outer ring) linked to carriage of ESBL and MDR carriage in the case of ST167, ST617 and the ST10 ESBL cluster in comparison to the remaining ST10 strains. ST10 contained a highly diverse set of O and H antigens with no obvious pattern, except in the case of the ESBL ST10 subclade.

sequence types against ST10 as opposed to comparing clades within a single sequence type as done in McNally *et al* (2016). However, this makes the IGR switches detected more significant since they evolved by parallel evolution in ST167/ST617 as opposed to the clade or lineage associated artefacts detected as a result of a single clonal expansion event. These results indicate a parallel mechanism for offsetting MDR plasmid carriage for ST167, ST617 and ST131 enabling clonal expansion of these sequence types.

2.5 - Discussion

Currently most studies have focussed on establishment of MDR bacteria in pathogenic lineages. However, due to the continuous decrease in the cost of high throughput whole genome sequencing we can obtain more data on commensal *E. coli* strains. The emergence

of ST167 and ST617 as commensal ESBL and carbapenemase associated strains displays a similar clonal expansion event as seen with pathogenic *E. coli* such as ST131 and ST648 (Mathers *et al* 2015, Ewers *et al* 2014 and McNally *et al* 2016). These clonal expansion events may occur more frequently in the commensal populations than previously thought as demonstrated by the unexpected detection of a ESBL producing ST10 sub-clade. This is further compounded by shared IGR switched between ST167 and ST617 being implicated in offsetting the cost of MDR plasmid carriage as seen in ST131 (McNally *et al* 2016). Functional studies are needed to confirm this hypothesis; however, these are hindered by the lack of a reliable GI tract persistence model and the fact that ST167 and ST617 are not screened for routinely. If these sequence types can colonise both humans and animals for longer periods of time then their exposure to antibiotics is far higher than as an environmental commensal, placing a much larger selection pressure on them to acquire antibiotic resistance resulting in expansion of the 'fittest' clones. This pattern has been reported in ST131 where they first adapted to colonise and persist in a host before acquiring the resistance mechanism and offsetting the cost of CTX-M-15 and fluoroquinolone resistance by mutations in IGR's (Zakour *et al* 2016, Stoesser *et al* 2016 and McNally *et al* 2016).

Further work based on the discovery of the convergent evolution of the O and H antigen 'switch' allowed us to identify another *E. coli* sequence type (ST44) which is associated with ESBL CTX-M15. However, the available sequences are sparse (44 genomes on enterobase) with poorly annotated meta-data meaning only ~20 genomes could be used reliably. This meant we could not include ST44 in this study, demonstrating the need for increased surveillance on the commensal population to allow us to understand the extent of the antibiotic resistance gene reservoir. The resistance gene reservoir underpins the emergence of MDR pathogenic sequence types by ensuring prevalence of a wide range of resistances. If we can reduce the 'reservoir' then there is less chance of pathogenic sequence types both acquiring and adapting to harbour these MDR elements, potentially extending the lifespan of many of our antibiotics.

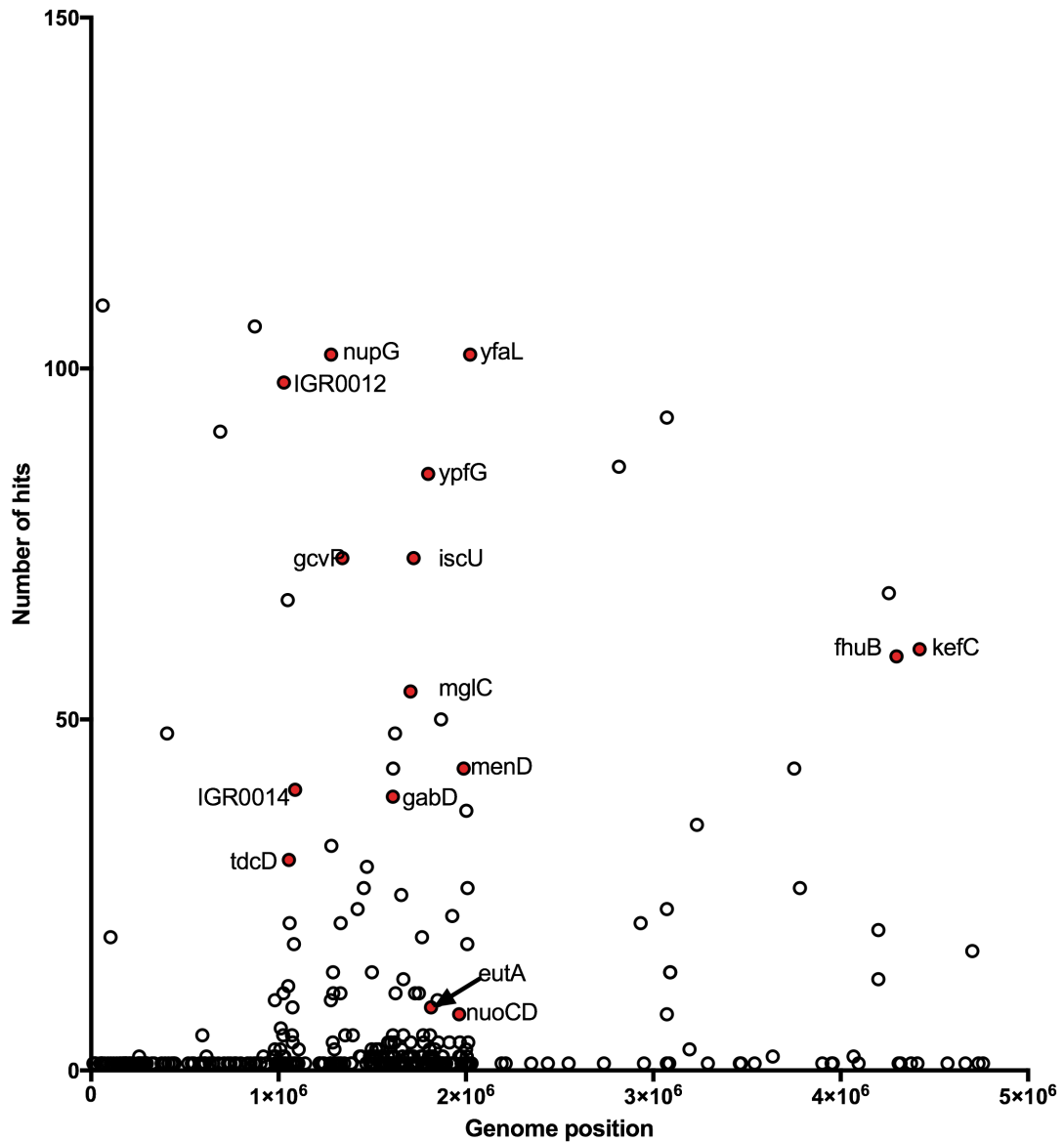


Figure 5. Manhattan skyline plot of SEER kmer hits. Significant Kmers shared between ST167 and ST617 overlaid with ST167 reference genome Y5 for positional reference and CDS annotation. Highlighted genes/IGR's have been directly implicated in anaerobic metabolism/respiration or in host survival during colonisation. Tabular breakdown available in appendix 9

A surprising find is the parallel evolution towards the exact same O and H antigenic type seen in both ST167 and ST617 alongside the MDR ST10 clonal cluster. The majority of *E. coli* produce their LPS utilising the O-unit translocation pathway encoded for by *wzx* and *wzy* (Iguchi *et al* 2016). This method utilises glycosyltransferases to assemble the O antigen in units at the cytoplasmic membrane. These units are then translocated by Wzx and polymerized by Wzy until the O antigen chain length is reached (Iguchi *et al* 2016). This mechanism is utilised by the majority of the ST10 isolates, however ST167, ST617 and the ST10 utilise the *wzm/wzt* ATP transporter pathway. This biosynthetic pathway assembles the

entire O-antigen on the cytoplasmic face before Wzt transports the O-chain across (Iguchi *et al* 2016).

O-antigen chain length plays a major role in gram negative organisms pathogenicity. It has a strong immune-stimulatory effect leading to the release of pro-inflammatory cytokines and nitric oxide from various components of the innate immune system (May and Groisman 2012). Conversely O-antigen chain length has also been shown to be essential for successful colonisation in a wide range of pathogens (*Salmonella enterica*, *Shigella flexneri* and *Pseudomonas aeruginosa*) due to its ability to inhibit complement and its role in bile salt resistance (May and Groisman 2012 and Tran *et al* 2014). The modal O antigen chain length is adapted to the specific organism with no 'one length fits all'. Expression of a long O antigen (80units) in *Shigella flexneri* increases resistance to bile salts whilst in *Salmonella enterica* it inhibits resistance to bile salts by interfering with antigen mediated bile resistance pathways (May and Groisman 2012, Tran *et al* 2014). Further work on this has demonstrated that loss of long O-antigen chains in *Salmonella* optimizes immune evasion and allows successful colonisation (Crawford *et al* 2013). The parallel evolution of ST167, ST617 and the ESBL producing ST10 clade with the same O-antigenic type point towards a form of enteric adaptation suited to these sequence types. Since the O-antigen detected is novel, no work has currently been done on chain length and LPS structure and in the absence of any pre-defined pattern between organisms and extensive work in *E. coli* it is hard to predict what effect this novel O-antigen is having.

2.6 - Future Perspectives.

This study demonstrates that commensal strains of the ST10 clonal complex are not just transient carriers of antibiotic resistance determinants as previously thought. Although ST167 and ST617 are not pathogenic the importance of combatting the expansion of MDR commensal strains cannot be underestimated. The current focus on combatting MDR pathogens, whilst essential in the short term, is only the tip of the iceberg. Facilitated by the easy transfer of resistance elements on both an inter and intraspecies level, successful clones will continually emerge harbouring these MDR resistance elements as shown by the independent emergences of ST131, ST73, ST95 and ST648. This means that more studies need to focus on MDR commensals to gain insight into the antibiotic gene 'reservoir' and the scale of the problem we face.

2.6.1 – The need for functional analysis following genomics

Global dissemination of ST167 and ST617 display parallels with ST131 indicating a common mechanism potentiating the spread of MDR *E. coli*, however the utilisation of genomics and bioinformatics to elucidate this trend is only half the story. Functional analysis of ST131, ST167 and ST617 is now needed to validate the theory of respiratory and metabolic adaptation (McNally in press). This could be achieved through further bioinformatics analysis via the application of RNA sequencing enabling us to determine the transcriptomic profile of ST131, ST167 and ST617. However, careful consideration must be taken to ensure that the correct environmental conditions are used to mimic the human gut so that we can generate a representative view following the switch to an enteric lifestyle. This will also allow us to determine the effect the promoter 'switches' are having on the expression profile by contrasting ST10 to ST167 and ST617 whilst simultaneously comparing ST131 Clade A and B with clade H30rx. Although none of the promoter region 'switches' share a common genetic basis between ST167/617 and ST131 it would be interesting to see if the offset of MDR plasmids follows a similar mechanistic process via convergent evolution or whether the offset is specific to each sequence type. The identification of common switches between ST167 and ST617 despite being phylogenetically distinct clades (Figure 1) indicates a similar mechanistic process via convergent evolution.

2.6.2 – Host adaptation followed by antibiotic resistance development limited to *E. coli*?

Similar work is now needed on other Gram-negative organisms to determine whether the theory of host adaptation followed by MDR plasmid adaptation holds true. For example, particular *Salmonella enterica* sv. Typhimurium sequence types have been linked with the ability to cause invasive disease whilst resisting multiple antibiotics (Kingsley *et al* 2009 and Ramachandran *et al* 2015). *S. Typhimurium* infection usually manifests as self-limiting gastroenteritis, however in the case of ST313 it has managed to transition to an epidemic MDR invasive non-typhoidal *Salmonella* (Kingsley *et al* 2009 and Ramachandran *et al* 2015). It has been demonstrated to be more 'fit' when causing infection as it has a defective flagellin which results in a much lower inflammatory cytokine response (Ramachandran *et al* 2015). ST313 is also able to survive within the host macrophage much more efficiently indicating further adaptation to its invasive lifestyle to which the genetics have yet to be explored (Ramachandran *et al* 2015). It would be interesting to observe whether the same pattern seen in ST131, ST167 and ST617 is also observed across other MDR pathogens and

commensals. First an adaptation to the host 'niche' i.e. the GI tract, inside a macrophage or in the lung enables enhanced survival within the host which subsequently increases exposure of that host-adapted strain to antibiotics. Following this acquisition of a plasmid or other antibiotic resistance elements proves to be advantageous in this environment and the sequence type evolves along with the plasmid to allow for minimal fitness cost ensuring stable acquisition. Antibiotic usage eliminates competitors leaving a host-adapted MDR pathogen/commensal free to spread and colonise. This pattern is often seen in Cystic Fibrosis patients infected with *P. aeruginosa* host adapts by becoming 'muroid' and adapting to a biofilm associated lifestyle (Winstanley *et al* 2016). Following detection of *P. aeruginosa* in the CF lung aggressive antibiotic therapy is used to combat the infection. Multiple mutational events result in resistance to numerous antibiotics and the fitness cost of these chromosomal mutation events is offset following prolonged exposure (Winstanley *et al* 2016 and Melnyk *et al* 2015). This generates a CF lung adapted MDR *P. aeruginosa* strain which eventually leads to severe respiratory decline and often death. If this theory proves to be true in some species, then a better tactic to eliminate these MDR sequence types may be to target colonisation factors to dislodge and MDR pathogens and commensal strains enabling 'normal' commensal strains to take hold. We can also then track sequence types undergoing a similar process as seen in ST131, ST167 and ST617, alerting clinicians and microbiologists about their potential, for example the emerging ESBL ST10 sub-cluster identified in this study may be in the early stages of such a process.

2.6.3 – Increase surveillance on commensal strains

Due to the initial high cost of sequencing the majority of sequencing data has been targeted at pathogenic strains of *E. coli*, therefore it is easy to understand why in the past only pathogenic sequence types of *E. coli* have been identified as MDR lineages. This work, following from Zhou *et al* (2016) are the first demonstrating that commensal lineages display a similar pattern to the pathogenic UTI strains in that it appears only specific sequence types can offset the cost of MDR plasmid carriage. Despite the availability of sequence data on sites such as Enterobase and the increasing number of data deposits made, the vast majority of commensal strains are underrepresented. This proved to be a problem when including the newly identified ESBL MDR lineages ST44 where the lack of reliable sequence and meta-data resulted in exclusion from the entire study. Far more dedicated sequencing to commensal strains is needed and a greater awareness raised about where to deposit data so that it can be used in studies such as this. The more we know about the development of

MDR, the better placed we will be placed to combat the threat and thus introduce appropriate guidelines to help combat such resistance.

3.0 – Protein expression and purification of the hydrogen cyanide self-protection protein PA4130, a potential anti-virulence target for *Pseudomonas aeruginosa*

3.1 – Abstract

Pseudomonas aeruginosa is amongst the top causes of healthcare associated infections. However due to its high level of intrinsic and acquired resistance, especially when in a biofilm, the use of antibiotics as a first line treatment has become problematic. A genome-wide transposon mutagenesis screening strategy was designed to identify core virulence factors with a view to developing novel inhibitors against them. Mutant selection was performed through *in vitro* and *in vivo* assays to identify mutants attenuated in multiple virulence factor production and infection models. PA4130, encoding a predicted nitrite/sulphite reductase was selected for further characterisation. Multiple PA4130 hexahistidyl tagged constructs were made, utilising the pCold, pCold-GST and pSK67 over-expression vectors in 2 variant protein over expression strains BL21 (DE3) and LOBSTR. Initial over-expression trials in pCGNT4130 revealed successful induction of PA4130, however solubility remains an issue. The ultimate aim is to overproduce soluble PA4130 for protein crystallisation studies. We can then use the molecular structure obtained by X-ray diffraction to implement a guided PA4130 inhibitor screen by targeting the active site of the enzyme.

3.2 – Introduction

Pseudomonas aeruginosa is one of the most common opportunistic pathogens causing 10-15% of all health care associated infections, making it a severe economic burden on the health care system (Lyksak *et al* 2000). At risk are cystic fibrosis patients, for which current treatment of *P. aeruginosa* in the CF lung is based on aggressive antibiotic therapy. An extremely high mutation rate and the ability to assimilate various transposable and conjugative elements, combined with antibiotic therapy results in a high level of intrinsic and acquired antibiotic resistance. The pliable nature of *P. aeruginosa* genome further complicates treatment as *P. aeruginosa* evolves to a pattern of chronicity. For example, *P. aeruginosa* mutates to overproduce the exopolysaccharide alginate demonstrating a 'mucoid' phenotype which promotes biofilm formation. This ability of *P. aeruginosa* to form biofilms *in vivo* further exacerbates the problem with AMR as biofilm cells are much more resistant to antibiotics their planktonic counterparts (Hoyle and Consterton 1993). However,

treatment is further complicated by the heterogeneous nature of *P. aeruginosa* in the CF lung. It has been shown that numerous phenotypic variants can co-exist in the CF lung, for example alginate overproducers are able to survive antibiotic treatment and re-establish an infection, whilst planktonically growing cells are able to move freely and 'seed' new areas of the lung. This results in numerous 'pockets' of *P. aeruginosa* colonisation, each of which has undergone its own unique CF lung adaptation. This heterogeneous nature of infection ensures that *P. aeruginosa* can survive any challenge and maintain a chronic infection. *P. aeruginosa* presents a unique problem within clinical microbiology and as such, novel methods are needed to combat this pathogen.

Anti-virulence factor targets present an interesting solution to the problem of antibiotic resistance. Targeting specific virulence factors may yield novel therapies with a very narrow spectrum of activity against *Pseudomonas* (Fenebro 2011). This is an extremely attractive prospect since new therapies designed against Gram negative organisms often fail to target *P. aeruginosa*. The aim would be to identify a virulence gene essential for human infection. This would limit the potential for mutation to counteract the therapy. Whole genome transposon mutagenesis has proved a valuable tool in identifying these anti-virulence targets. Numerous studies have utilised this methodology in *P. aeruginosa* yielding many attractive targets (Potvin *et al* 2003 and Kim *et al* 2008). However, many of these studies are limited by the number of *in vivo* models used and as such the targets may not represent virulence factors which are essential for human infection.

The work in this study follows on from a genome-wide transposon mutagenesis based screening was designed by Dubern *et al* (2015) to search for mutants attenuated in multiple virulence factor production. By filtering these transposon mutants through numerous *in vitro* assays and *in vivo* disease models only the most attenuated mutants were selected for further target characterisation. By utilising multiple infection models, we are able to focus in on the core factors required for virulence in all models, as opposed to host-specific virulence factors. This ensures that the virulence factors chosen for further characterisation are more likely to be essential in establishing and maintaining a human infection.

One of the mutants identified is the predicted nitrite/sulphite reductase protein PA4130. Both the Tn mutant and the clean PA4130 mutant displayed significant virulence attenuation in both *in vitro* assays and *in vivo* disease models (data not shown). Previous work done by

Frangipani et al (2014) showed that the *PA4129-PA4134* locus was strongly upregulated in the presence of HCN and aided self-protection alongside the well characterised *cioAB* and less well understood Rhodanases. *P. aeruginosa* produces HCN using a membrane bound HCN synthase complex consisting of HcnABC (Lavill *et al* 1998). Although the biological role of HCN is unclear, it has been shown to be an important virulence factor in the *Caenorhabditis elegans* and *Drosophila melanogaster* infection models (Gallagher and Manoil 2001, Broderick *et al* 2008). Further to this it is also detected at concentrations of around 130uM in the Cystic Fibrosis (CF) lung and implicated in reduced lung function when detected at such high concentrations (Ryall *et al* 2008).

In this study, we attempted to construct 6 separate protein overexpression constructs with a hexahistidyl Tag at both C and N terminus of *PA4130*. The aim is to over-produce *PA4130* with a view to obtaining the crystal structure. This structure will then form the basis of our *in-silico* inhibitor screen against *PA4130*. By targeting *PA4130* we aim to decrease the ability of *P. aeruginosa* to protect itself from HCN effectively promoting self-poisoning. The bacteriostatic nature of HCN would ensure that further progression of the infection does not occur and that further mutation cannot arise as the cells will not divide. This prevents further antibiotic resistance development facilitating clearance by some antibiotics and the host immune system. The production of HCN is well conserved amongst non-mucoid and mucoid strains making this an attractive prospect for therapy against both acute and chronic infections.

3 – Materials and methods

3.3.1 - Bacterial strains, plasmids and growth conditions

Summary of all strains, plasmids and primers used are shown in Table 1. *E. coli* strains DH5 and *P. aeruginosa* PA01 were routinely grown in LB (Sigma-Aldrich) at 37°C. Protein overexpression took place in Terrific Broth at a temperature specific to the plasmid. Ampicillin was added at 100ug/ml unless otherwise stated. A previously constructed *PA4130* overexpression strain was first utilised to produce *PA4130* utilising pSK67. However, upon sequencing it was revealed that several missense mutations had been introduced into the *PA4130* nucleotide sequence. This meant we had to reconstruct the overexpression system to obtain high yielding, soluble *PA4130* for crystallisation and structure determination.

Strains	Description	Origin or Source
<i>E. coli</i> strains		
DH5 alpha	F ⁻ ,φ80dlacZΔM15,Δ(lacZYA-argF)U169,deoR,recA1, endA1, hsdR17(rk ⁻ ,mk ⁺), phoA, supE44, λ ⁻ , thi1, gyrA96, relA1	Lab Stock
BL21 DE3	F ⁻ ompT gal dcm lon hsdSB(rB ⁻ mB ⁻) λ(DE3 [lacI lacUV5-T7p07 ind1 sam7 nin5]) [malB ⁺]K-12(λS)	Lab Stock
LOBSTR	Same as above, histidine residues mutated in <i>slyD</i> and <i>arnA</i> lowering affinity to nickel.	Andersen <i>et al</i> 2013
<i>P. aeruginosa</i> strains		
PA01	Laboratory wild type strain	Lab Stock
Plasmids		
Description		
pSK67	pTOPO type vector for protein overexpression. Under control of a T7 promoter and IPTG inducible.	Lab Stock
pCold1	Protein overexpression vector utilising an <i>E.coli cspA</i> to allow expression as low as 10 °C. Contains His Tag at N-terminus.	Lab Stock
pCold-GST	Identical to pCOLD1 except, labels vector with GST for lowering surface entropy in crystallisation studies	Lab Stock
pGEM T-easy	TA cloning plasmid	Promega
pGCT4130	Hexameric His tag PCR amplified at the C terminus of PA4130 and inserted into pGEM T-easy.	This Study
pGNT4130	Hexameric His tag PCR amplified at the N terminus of PA4130 and inserted into pGEM T-easy.	This Study
pG4130	PA4130 PCR amplified and inserted into pGEM T-easy.	This Study
pSKCT4130	CTHis PA4130 unit sub-cloned from pGEM::CTHisPA4130 into pSK67 using SacI and XhoI.	This study
pCGNT4130	PA4130 sub-cloned from pGEM::PA4130 and inserted upstream of the Hexameric His tag in pCold-GST using SacI and XhoI.	This study
pCGCT4130	CTHis PA4130 unit sub-cloned from Pgem::CTHisPA4130 using SacI and XhoI.	This study
Primers		
Sequence		
NT6PA4130F1	TTATTAGAGCTCATGTACCAGTACGATGAATACG	Sigma
NT6PA4130F2	TTATTAGAGCTCCATCACCATCACCATCACATGTACCAGTACGATGAATACG	Sigma
NT6PA4130R1	TTATTGCTCGAGTCATGCCGGTTCCTCCATCCT	Sigma
CT6PA4130F1	CTGCTGGAGCTCATGTACCAGTACGATGAATACG	Sigma
CT6PA4130R1	CTGCTGCTCGAGTCAATGGTGATGGTGATGGTGCCGGTTCCTCCATCCT	Sigma
pColdseqf1	ACGCCATATCGCCGAAAGG	Sigma
pColdseqr1	GGCAGGGATCTTAGATTCTG	Sigma

Table 1) All strains, plasmid vectors and primers used in this study for construction of overexpression vectors

3.3.2 – Cloning

We attempted to add a N and C terminal His-Tag onto *PA4130* using 3 different vectors. Cloning was achieved using traditional restriction endonuclease cloning. Restriction site *SacI* and *XhoI* were added to *PA4130* from *P. aeruginosa* PAO1 via PCR using primers relevant to each separate construct using Phusion polymerase (NewEngland Biolabs) as per the manufacturer's instructions with the following cycle conditions.

Step	Temp (°C)	Time (sec)
Initial DNA denaturation	98	600
DNA denaturation*	98	15
Primer annealing*	55	30
DNA extension*	72	90
Final extension	72	600

*Stages are looped and cycled 35 times

The subsequent PCR product was then run by gel electrophoresis on a 1% agarose gel and the *PA4130* fragments extracted by gel purification (Monarch). All plasmids and *PA4130* variants were digested with *SacI* and *XhoI* restriction endonucleases (NewEngland Biolabs) for 4 hours at 37°C followed by an inactivation step at 80°C for 10 minutes. Simultaneously all vectors were treated with alkaline phosphatase (Thermo Fisher) to remove the 5'-3' phosphate groups, inhibiting vector recircularization. Both plasmid and *PA4130* were gel purified following digestion and running on a 1% agarose gel. The relevant vector and *PA4130* fragment were then ligated together using DNA T4 ligase (NewEngland Biolabs) at 16°C for 16 hours and the ligation mixture dialysed against H₂O for 1 hour to remove salt. The ligation reaction was then electroporated into *E. coli* DH5, recovered in LB for 1 hour at 37°C and plated on LB with Ampicillin as a selection marker. Colonies were then screened for presence of *PA4130* using *Taq* polymerase (New England Biolabs) using the below cycle conditions, and presumptive colonies confirmed to contain plasmid via restriction digestion.

Step	Temp (°C)	Time (sec)
Initial DNA denaturation	95	600
DNA denaturation*	95	15
Primer annealing*	55	30
DNA extension*	72	120
Final extension	72	600

*Stages are looped and cycled 35 times.

Where standard restriction cloning failed an intermediate step was utilised with the TA cloning vector pGem T-easy (Promega), which contains a 5' Thymine overhang. *PA4130*, with *SacI* and *XhoI* restriction sites, had an A tail added at the 3' end using *taq* polymerase at 72°C in the presence of 1Mm Adenosine. T4 DNA ligase was then used to ligate the relevant *PA4130* variant into pGEM t-easy as described previously. The ligation reaction was then transformed into *E. coli* DH5 alpha and colonies screened for presence of *PA4130* using *taq* polymerase (as described above) and confirmed with restriction digestion (*SacI* and *XhoI*) following Vector mini-prepping. The confirmed vectors were then restriction digested and the *PA4130* fragment gel purified for insertion into the relevant over-expression vectors. Following confirmation of *PA4130* insertion into the overexpression vectors, sequencing of *PA4130* was performed to ensure correct cloning. Once confirmed, the *PA4130* overexpression vectors were electroporated into *E. coli* BL21 (DE3) and *E. coli* LOBSTR for protein overexpression trials.

A small-scale overproduction pilot was performed in 50 ml of Terrific broth. Strains were initially inoculated at an OD (570nm) of 0.05 and grown at 37°C until OD 0.6-0.8. Cultures were then transferred to 15°C and left to cool for 30 minutes. Following this each expression strain was induced at 1, 0.5 and 0.3 mM of IPTG. 1ml samples were taken at pre-induction the end of expression (48 hours). Both the 1ml samples and culture were harvested at 10000g for 10 minutes and stored at -20°C.

Whole cell lysates were screened for successful induction using SDS-PAGE analysis on a 10% Polyacrylamide gel. Samples were boiled in 1X SDS running buffer with 1% B-mercaptoethanol for 30 minutes and then run at 100V for 20 minutes, followed by 180V for 45 minutes. Samples were then visualised using Gel code Blue Protein Stain (Thermo Fisher). To assess the solubility of *PA4130* in samples, the harvested culture was re-suspended in 25mM Tris.HCl (pH 7.4) and sonicated for 6 minutes with a 15/15 second On/Off protocol. The cell lysate was pelleted and the supernatant containing the soluble fraction extracted. The pellet was re-suspended in 25mM Tris.HCl with 8M Urea for 1 hour at room temperature to extract the insoluble fraction. Both the soluble and insoluble fractions were re-suspended in a 1:1 ratio with 1x SDS running buffer with only the soluble samples boiled as described previously.

3.4 – Results

3.4.1 - Production of Overexpression plasmids

pSK67 had been used in previous work as a vector to overexpress *PA4130*, however the nucleotide sequence had developed missense mutations meaning we had to redesign the C terminally His tagged *PA4130* vector. Previous attempts to purify *PA4130* with both N and C terminally tagged *PA4130* (Figure 5D) resulted in low yield in the soluble fraction and impurities in the final product making crystallisation studies impossible. Therefore we decided to create multiple overexpression vectors with various properties to ensure we obtained pure *PA4130* at a high yield. We proposed to construct 6 overexpression vectors using 3 vector backbones (pSK67, pCold1 and pCold-GST) with both NT and CT permutations. In summary 5 of the 6 vectors were successfully produced during the project.

3.4.1.1 – pSK67 overexpression vectors

The proposed *PA4130* overexpression vectors pSKNT4130 and pSKCT4130, are outlined in Figure 1. Successful production of pSKCT4130 was achieved by direct restriction cloning, whilst the NT4130 construct failed to be sub-cloned from PGNT4130 (PGEM vector). Successful insertion of *PA4130* was confirmed by restriction digestion analysis and sequencing to ensure the native amino acid sequence will be expressed. pSKCT4130 was moved into *E. coli* BL21 for protein overexpression trials (data not shown) The pSK67 vectors were selected as they had already been demonstrated to produce soluble *PA4130*, albeit at low levels. With further work on expression and purification optimisation pSK67 can drive high yielding protein production due to the presence of a strong T7 promoter, however overexpression relies on a host containing T7 RNA polymerase.

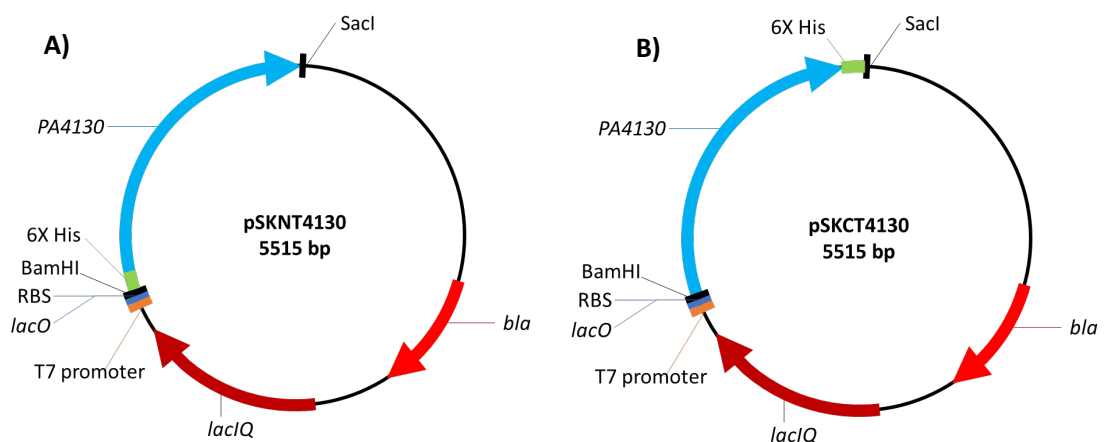


Figure 1) pSKNT4130 and pSKCT4130 vector map. Depicted are proposed structures of pSK67 derivative *PA4130* overexpression vectors. pSKNT4130 (A) and pSKCT4130 (B). His tags were added at the N and C terminus to facilitate nickel column purification of *PA4130* from the *E. coli* BL21 cell extract. High level expression is driven from a strong T7 promoter with induction achieved using IPTG. pSKNT4130 could not be produced following problems with sub-cloning from pGNT4130, however

3.4.1.2 – pColdI overexpression vectors

The pColdI overexpression constructs are displayed in figure 3. The pCold vector series enables protein expression to be achieved at very low temperatures (10°C) due to the use of *E. coli cspA* (cold shock protein) promoter in conjunction with *cspA* 3' and 5' untranslated regions (UTR). Use of *cspA* and *lacO* provides 2 levels of control, preventing any 'leaky' expression of the target protein, whilst the presence of the 3' and 5' UTR prevents translation at high temperatures due to the mRNA secondary structure inhibiting ribosome binding. Induction requires simultaneous cooling and addition of IPTG to drive target protein overexpression and the use of low temperature minimises secondary *E. coli* metabolism. This enables a much higher percentage of the overall protein produced under these conditions to be our target protein PA4130 and minimises the chances of contamination following His column purification. An additional advantage to using a low temperature is that protein folding kinetics slow down drastically. This has been shown to increase protein solubility in heterologous hosts potentially enabling a much higher yield of soluble PA4130 to be obtained. However, the use of such a low temperature means that an additional 24 hours is needed during overexpression in *E. coli*, due to the reduced growth rate, to obtain protein at a high enough concentration.

Production of both pCONT4130 and pCOCT4130 in *E. coli* DH5 was achieved, using pGCT4130 pG4130 to sub-clone PA4130 variants into pCold-GST. The use of pGEM was necessary due to problems with PA4130 amplification following addition of the his-tag and restriction sites by PCR (Figure 4). Use of pGEM does not require restriction digestion as it uses Thymine-Adenine overhang cloning (TA cloning) enabling us to transfer the entire PA4130 construct following PCR, preventing DNA loss following restriction digestion and DNA gel purification. This enabled us to sub-clone the PA4130 constructs from pGEM by restriction digest, allowing us to obtain a much higher yield of digested PA4130 (as compared to PCR amplification process) due to pGEM's high copy number. Whilst successful construction of both pColdI variants was achieved in *E. coli* DH5 (Figure 4E), movement into overexpression strains was not achieved in this project due to time constraints.

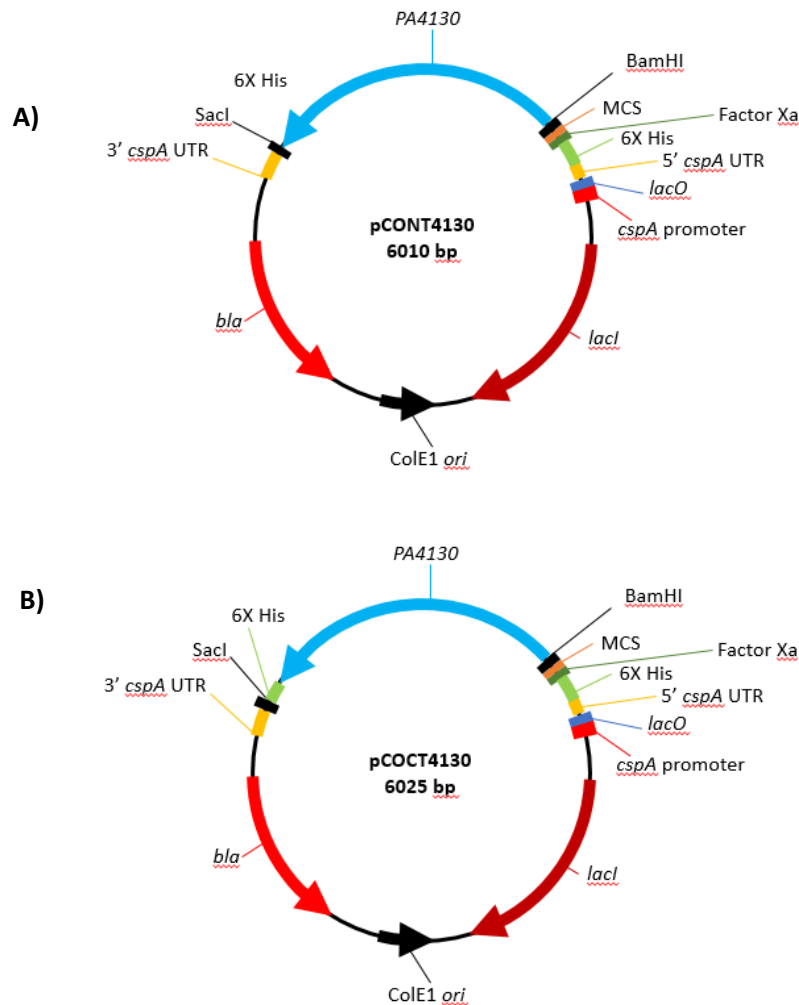


Figure 2) pCONT4130 and pCOCT4130 vector map. Shown are the proposed structures for pCONT4130 (A) and pCOCT4130 (B). His tag was added at the C terminus of PA4130 in pCOCT4130, however no N terminal His tag was needed in pCONT4130 due to the presence of 6x His tag in the pColdI vector backbone. Following expression using the pCOCT4130 vector, removal of the N terminal His tag can be achieved with factor Xa protease. These structures are presently in *E. coli* DH5 awaiting movement into protein overexpression strains *E. coli* BL21 and LOBSTR.

3.4.1.3 – pCold-GST overexpression vectors

The pCold-GST overexpression constructs pCGCT4130 and pCGNT4130 proposed structures are outlined in figure 3. This vector works in much the same way as pCold. It utilises the *cspA* promoter along with *lacO* to enable strict control of protein expression. Simultaneous cooling (10-15°C) and induction by IPTG enables expression of the target promoter. However, pCold-GST utilises a highly soluble protein fusion partner Glutathione S-Transferase (Moon *et al* 2008). Protein fusion is a common method used in protein overexpression for structural studies as often expression in a heterologous host results in a high proportion of the target protein in the insoluble fraction (Smith and Johnson 1998).

Presence of HRV-C3 protease site enables cleavage of the GST from the native target protein enabling them pursuit of structural studies unhindered by the presence of the large 25kDa GST tag (Moon *et al* 2008).

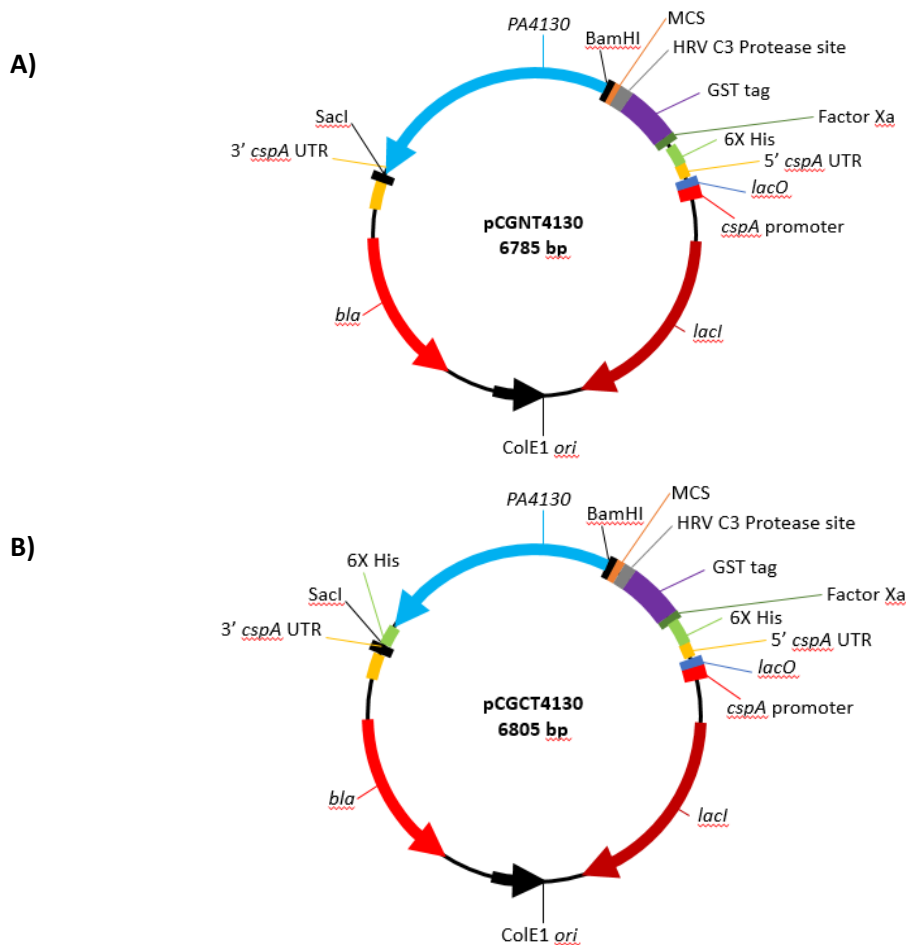


Figure 3) PCGNT4130 and PCGCT4130 vector maps. Displayed are the proposed structures for PCGNT4130 (A) and PCGCT4130 (B). Similar to pColdI, there is already a 6X His tag present in the pCold-GST vector backbone downstream of the GST tag. Since GST needs cleaving prior to structural studies we can remove both the GST tag and N terminal His tag simultaneously using the HRV C3 protease site for PCGCT4130. PCGNT4130 is present in E.coli BL21 LOBSTR whilst PCGCT4130 is in E.coli DH5 awaiting movement into the overexpression E. coli strains.

Both the N terminal and C terminal variants (Figure 3) were constructed successfully in E. coli DH5 as inferred from restriction digestion (Figure 4). pCGNT4130 was constructed successfully, however sequencing data demonstrated that only pCGCT4130 had been cloned incorrectly with a series of missense mutations present in *PA4130* (data not shown). At the time of writing two different pCGCT4130 versions had been sent for sequencing from distinct colonies to obtain the correct *PA4130* nucleotide sequence. The pCGNT4130 vector

was moved into *E. coli* BL21 and a BL21 derivative low background strain (LOBSTR) for protein overexpression trials.

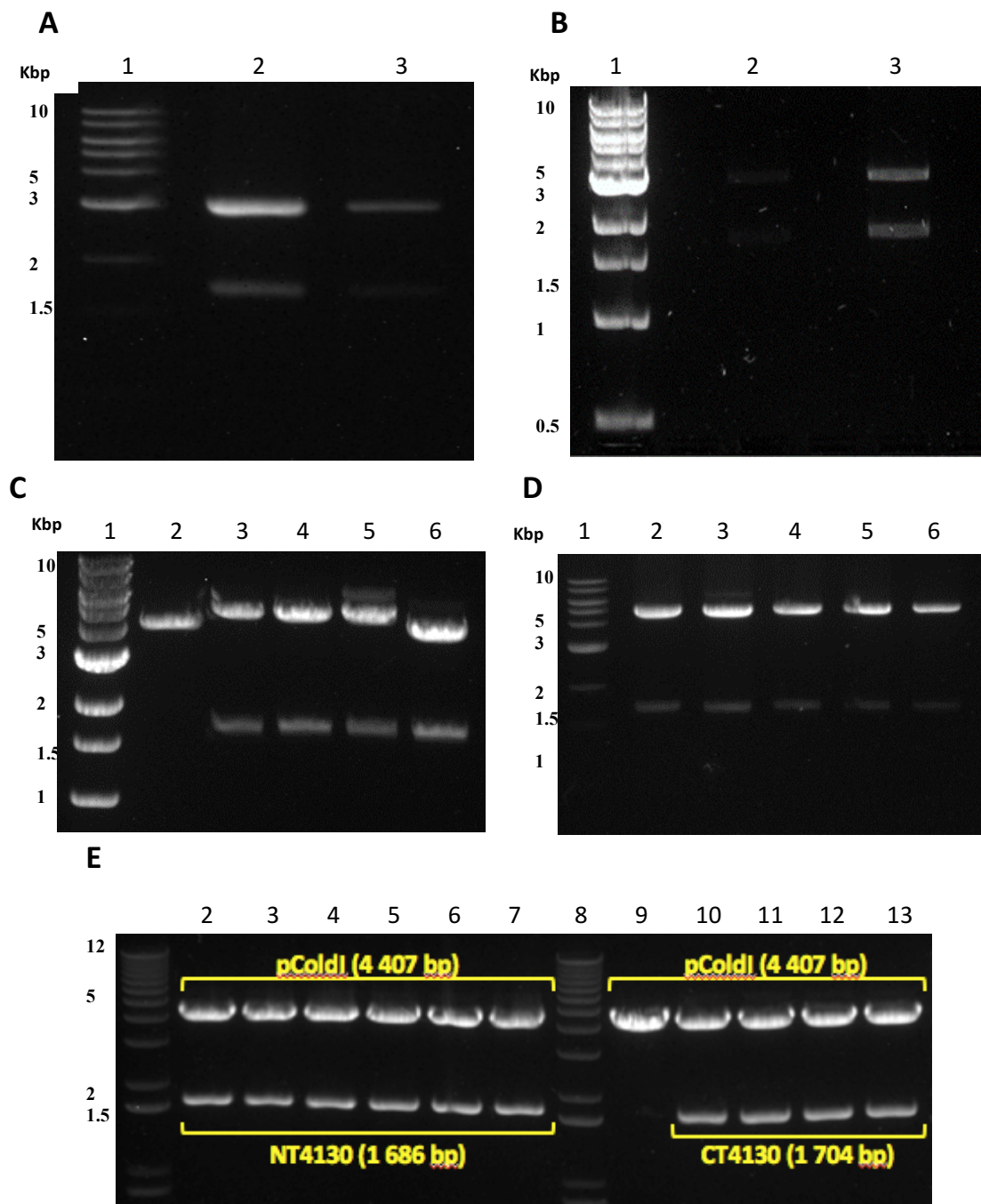


Figure 4) Restriction digestion confirming PA4130 insertion into vectors. Names and descriptions for each vector are in table 1 **(A)** Lane 1: 1kb ladder; Lane 2: pG4130 digestion; Lane 3: pGCT4130 digestion. These vectors were generated using a pGEM vector backbone to facilitate easy sub-cloning of PA4130. **(B)** Lane 1: 1 kb ladder; Lane 2: pGNSK4130(1); Lane 3: pGNSK4130 (2). **(C)** Lane 1: 1kb ladder; Lane 2: re-circularized pCold1 vector, Lane 3: PCGNT4130 digestion; Lane 4: PCGNT4130 digestion, Lane 5: PCGNT4130 digestion, Lane 6: pSKCT4130 digestion. **(D)** Lane 1: 1kb Ladder; Lane 2: pCGCT4130 digestion, Lane 3: pCGCT4130 digestion, Lane 4: pCGCT4130 digestion, Lane 5: pCGCT4130 digestion, Lane 6: pCGCT4130 digestion

3.4.2 - Presumptive successful PA4130 induction

The trial overexpression seemed to yield induction of PA4130-GST at around 85-87kDa for both BL21 (DE3) and LOBSTR (figure 2A and 2B). Negative cultures revealed no product in this region (figure 2C), providing reasonable evidence that this is *PA4130* in the absence of an Immunoblot which could not be achieved due to time restrictions. 1mM and 0.5mM IPTG induction seemed to result in a large amount present in the insoluble fraction, although 0.3 Mm induction for the BL21 (DE3) strain seemed to yield a higher percentage of presumptive PA4130 in the soluble fraction (Figure 5). This is not clear for the LOBSTR strain, although this may be due to the poor OD (600nm) achieved for these cultures in comparison to BL21.

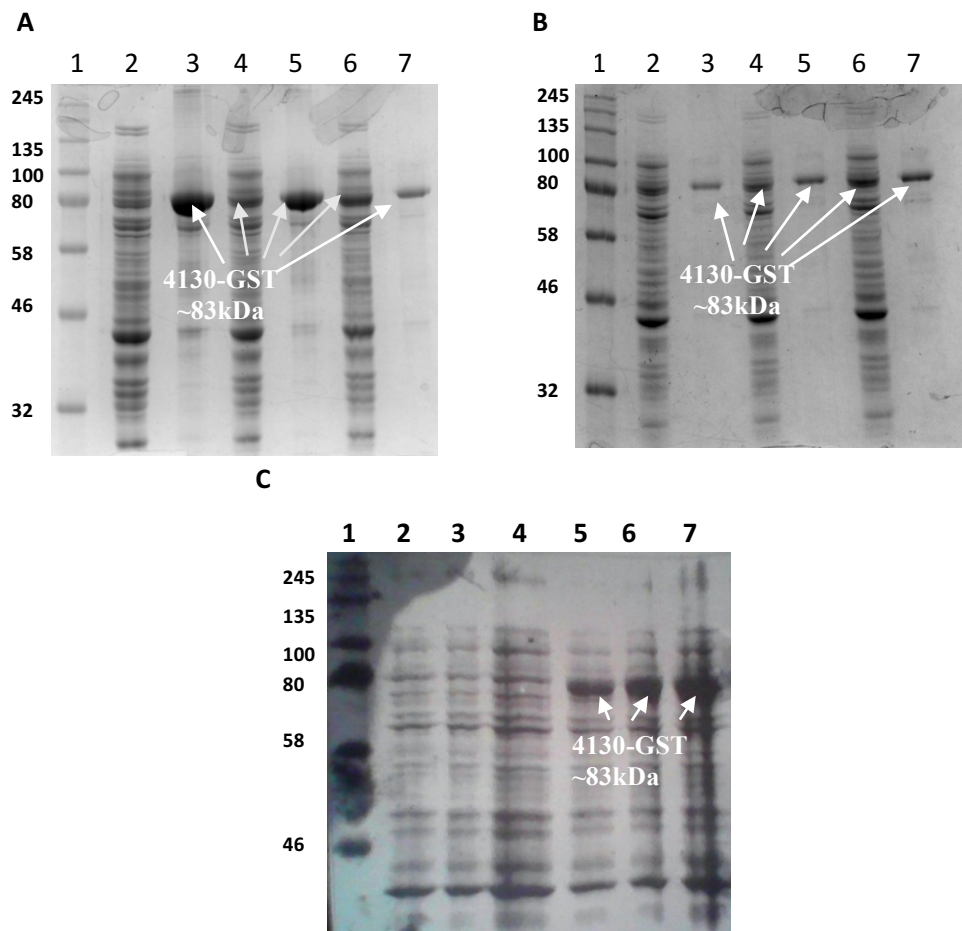


Figure 5) SDS-PAGE gels displaying trial overexpression of PCGNT4130 in BL21 and LOBSTR. A) BL21 overexpression PA4130-GST, Lane 1: 11-245 kDa protein ladder (New England Biolabs); Lane 2: Soluble fraction at 1mM IPTG induction; Lane 3: insoluble fraction at 1mM IPTG induction; Lane 4: Soluble fraction at 0.5mM IPTG induction; Lane 5: insoluble fraction at 0.5mM IPTG induction; Lane 6: Soluble fraction at 0.3mM IPTG induction; Lane 7: insoluble fraction at 0.3mM IPTG induction. **B)** LOBSTR overexpression PA4130-GST, Lane 1: 11-245 kDa protein ladder Lane 2: Soluble fraction at 0.3mM IPTG induction; Lane 3: insoluble fraction at 0.3mM IPTG induction; Lane 4: Soluble fraction at 0.5mM IPTG induction; Lane 5: insoluble fraction at 0.5mM IPTG induction; Lane 6: Soluble fraction at 1mM IPTG induction; Lane 7: insoluble fraction at 1mM IPTG induction. **C)** BL21 whole cell lysate analysis Pre-induction vs Harvest. Lane 1: 11-245 kDa protein ladder; Lane 2: Pre-induction sample for 0.3mM IPTG induction culture; Lane 3: Pre-induction sample for 0.5mM IPTG induction culture; Lane 4: Pre induction sample for 1mM IPTG induction culture; Lane 5: 48h post induction sample 0.3 mM IPTG; Lane 6: 48h post induction sample 0.5 mM IPTG; Lane 7: 48h post induction sample 1 mM IPTG.

3.5 – Discussion

3.5.1 – Success of PA4130 overexpression

During this project, we choose to use multiple vectors with varying properties to help both protein overexpression and downstream high resolution structural characterization. Displayed above is the presumptive overexpression of PA4130 fused to Glutathione S-Transferase. Initial work began by inducing *E. coli* BL21 harbouring pCGNT4130 with 1mM IPTG, however this resulted in PA4130 being almost exclusively in the insoluble fraction. We hypothesized that this was due to a strong induction resulting in the formation of PA4130 aggregates and inclusion bodies. In response to this we began to alter the induction strength ranging from 0.1 Mm IPTG to 1Mm IPTG, however this is a balancing act as the lower the level of IPTG the lower the expression of IPTG. We sought to strike a balance between high level expression and PA4130 solubility and observed that 0.3 Mm IPTG seemed to yield a higher level of PA4130 in the soluble fraction (Figure 2B). Whilst an immunoblot was not performed in this study due to time constraints, subsequent follow up studies have confirmed that this protein is PA4130 (Data not shown).

Protein overexpression and purification is hampered by numerous problems, thus optimisation of purification is needed to obtain PA4130 in a high enough and pure enough yield for structural characterization by crystallography. Utilization of the pCold based vectors as proved extremely successful for purification of previously insoluble proteins (Moon *et al* 2008 and Qing *et al* 2004). These vectors allow us to lower the overexpression temperature to as low as 10°C, vastly lowering background expression levels ensuring a higher proportion of total protein produced by *E. coli* is PA4130. Whilst it has been reported that lowering the temperature results in slower protein folding kinetics enabling high level production in the soluble fraction, in truth it is a balancing act (Qing *et al* 2004). Different proteins require different conditions for optimal expression in a heterologous host, therefore further study is needed altering...

1. Temperature: To strike a balance between levels of background expression and PA4130 folding kinetics.
2. Induction strength: Further optimisation of IPTG strength with the aim of obtaining the highest yield of monomeric PA4130 in the soluble fraction.

3. Induction timing: We need to consider when PA4130 is expressed in its native host. PA4130 is only expressed maximally upon entry into stationary phase and thus induction at an OD higher than 0.6-0.8 may be beneficial.
4. Codon Optimization: *P. aeruginosa* and *E. coli* have evolved a different codon bias (different tRNA's for the same amino acid) and thus alteration of nucleotide sequence to suit the codon usage of *E. coli* BL21 may be beneficial for expression and protein folding (less ribosome stalling).
5. Culture volume: Using pCold based vectors results in a lower OD of *E. coli* despite 48H incubation, therefore purification scale may need to be increased to obtain significant levels of PA4130.

Purification efforts are in their infancy and a great deal of experimental work needs to be done before achieving the level of PA4130 expression needed for crystallisation. The other successfully cloned vectors (pCONT4130, pCOCT4130 and pSKCT4130) need to be moved into *E. coli* BL21 and protein overexpression studies performed before deciding which vector to proceed with. Although the use of pCold-GST is proving successful at present it is possible that the addition of the large GST tag and the additional His tag is affecting protein folding resulting in excessive insoluble PA4130. Therefore, studies with the above vectors may prove more successful in obtaining soluble protein as the polypeptide chain is closer to the native configuration. Successful purification of PA4130 is central to this project not only for structural determination but also to elucidate PA4130's actual function.

3.5.2 – PA4130 function

Hydrogen cyanide production is an important virulence factor in *Pseudomonas aeruginosa* providing a selective advantage under low oxygen tension (Blumer and Haas 2000). However, self-protection from this potent toxin is even more fundamental to its success. By purifying PA4130 we hope, not only to obtain the crystal structure, but to determine its biological role. PA4130 is predicted to be a nitrite/sulphite reductase; however, work by *Frangipani et al* 2014 showed that it does not use nitrite or sulphite as a terminal electron acceptor when aerobic respiration is inactivated by HCN. The authors failed to consider that nitrite/sulphite reductases can be either assimilatory or non-assimilatory. Assimilatory nitrite reductases produce ammonia from nitrite and sulphite reductases generate sulphur containing compounds such as cysteine. Dissimilatory nitrite/sulphite reductases generate terminal electron chain carriers, enabling anaerobic respiration. The reductase genes can

play numerous roles and are extremely versatile, often enabling reduction of both nitrite and sulphite although generally a substrate is preferred.

Recent work done by Hirai and colleagues (2016) seemingly indicates that the PA4129-PA4134 region is not involved with anaerobic respiration, but respiration at low oxygen tension. They demonstrated the heterogenic nature of the *ccb3* type cytochrome c oxidases involved with aerobic respiration. *P. aeruginosa* encodes 2 main respiration chains encoded by *ccoN1O1P1Q1* and *ccoN2O2P2Q2*. The 4 components of the *cco* complex play different roles, *ccoN* is the core catalytic unit, whilst *ccoO* is a monoheme transmembrane subunit and *ccoP* a diheme transmembrane subunit. The role of *ccoQ* is not well understood however it stabilises formation of the CcoNOP complex. These 2 respiration chains work in synergy with *ccoN1O1P1Q1* continuously expressed and *ccoN2O2P2Q2* expressed under microaerophilic conditions due to its higher affinity to oxygen (Arai 2011). There are also 2 orphan cytochrome c oxidase subunits present PA1856-1855 and PA4133-PA4134 corresponding to *ccoN3Q3* and *ccoN4Q4* respectively (Arai 2011 and Hirai *et al* 2016). Hirai and colleagues (2016) demonstrated that these sub-units can recombine conferring different characteristics onto the chimeric respiration chains. The combination of PA4133-PA4134 (*ccoN4Q4*) with either CcoO1P1 or CcoO2P2 generates a low oxygen tension, cyanide resistant *ccb3* type cytochrome c oxidase chain (Hirai *et al* 2016). Therefore it seems that whilst PA4130 is not directly involved with the respiratory chain, simultaneous upregulation in response to cyanide from 3 separate promoters in the PA4129-PA4134 region indicates it is central to this story (Frangipani *et al* 2014).

Therefore, determining PA4130 exact biological function and how it is involved with self-protection from HCN is paramount. PA4130 is conserved alongside PA4129, PA4133 and PA4134 throughout *Pseudomonas* (as determined by blastn analysis), indicating that these genes are central to the story of HCN self-protection and further validating PA4130 as an inhibitor target. Obtaining PA4130 in high enough concentrations for both crystal generation and enzymatic assay development is crucial. By determining the product outputted by PA4130 after utilisation of nitrite/sulphite we can determine its function as an assimilatory or dissimilatory nitrite/sulphite reductase. This will inform us of avenues to explore although the work by Hirai and colleagues (2016) and Frangipani and colleagues (2014) indicate that PA4130 is likely to be of an assimilatory nature.

3.6 – Future Perspectives

So far the project is still in its infancy. The expression and purification of PA4130 for use in crystallisation studies and subsequent inhibitor development is only a small part of the entire project. Presumptive expression in the soluble fraction has been achieved in only one out of the six vector constructs highlighting many of the difficulties experienced with expression and purification. However, many more questions need to be answered about this poorly understood, yet seemingly vital for infection, region. Paramount to this study and the understanding of the function and the substrates utilised in the PA4129-PA4134 region is understanding the function of each gene.

3.6.1 – Determination of PA4129-PA4134 function

Thus far this project has focussed on PA4130 however characterisation of the whole PA4129-PA4134 region is needed if we are to fully understand the role this plays in HCN resistance. The work by Hirai and colleagues (2016) indicates that upregulation of the PA4133-PA4134 operon in response to HCN results in formation of a low oxygen tension, cyanide resistant *cbb3* type cytochrome oxidase respiration chain. The fact that other members of this region demonstrate the exact same upregulation by HCN from separate promoters indicates a role for the whole PA4129-PA4134 region. Conserved sequence and domain homology reveals PA4129 to be a hypothetical protein; PA4130 a nitrite/sulphite reductase; PA4131 a 4Fe-4S binding protein implicated as a *cbb3* cytochrome oxidase accessory protein implicated with oxygen use under microaerophilic conditions (*ccoG*); PA4132 a helix-turn-helix (HTH) type regulator; PA4133 is a *ccoN* type membrane bound component of a *cbb3* type cytochrome oxidase whilst PA4134 is a *ccoQ* type component of a *cbb3* type cytochrome oxidase (Frangipani *et al* 2014, Hirai *et al* 2016 and Winsor *et al* 2016). The fact that so many *cco* members are located in this region confirm its role in respiration with PA4132, PA4133 and PA4134 corresponding to *ccoG*, *ccoN* and *ccoQ*. The presence of *ccoG* lends weight to the theory put forward by Hirai and colleagues (2016) as this has been implicated in aerobic respiration under microaerophilic conditions (Colburn -Clifford and Allen 2010). Since PA4130 was identified as required for virulence in multiple *in vitro* and *in vivo* models and results in a loss of viability under anaerobic conditions indicates it plays a key role in microaerophilic, cyanide resistant respiration chain formation.

Key to understanding the role of this entire region will be the generation of in-frame deletion for each gene and each transcriptional unit. We can then follow a similar pattern as demonstrated in Dubern and colleagues (2015) and screen virulence inhibition to determine which genes are actually required for virulence in this region. Generation of mutants would also enable global transcriptional studies (RNA-sequencing) under aerobic and microaerophilic conditions, whilst screening of transcriptomics from *in vivo* models will provide insight into the importance of this region during an infection. By determining the transcriptome involved with *PA4130* expression and other members of the region we can build a better picture of how these members interact with each other and the impact seen globally.

3.6.2 – Interplay between *cioAB* and *PA4129-PA4134*

The best characterised cyanide resistant aerobic respiration chain in *P. aeruginosa* is *cioAB*, a cytochrome *bd* quinol oxidase that utilises quinol to catalyse the uptake of oxygen (Cunningham 1997). It has long been known that deletion of *cioAB* affects growth under cyanogenic conditions (Cunningham 1997 and Frangipani *et al* 2014). However, the fact that growth is not completely inhibited indicated that another cyanide resistant respiratory chain was active, later identified as *PA4133-34* (Cunningham 1997 and Hirai *et al* 2016). This presents an interesting opportunity to study the dynamic expression of *cioAB* and *PA4133-34* in biofilm formation and *in vivo*. Are these cyanide resistant aerobic respiration chains working concurrently? Or is does one predominate under a given set of conditions?

Usually cytochrome *bd* quinol oxidases are renowned for their ability to function under microaerophilic conditions, however *cioAB* in *P. aeruginosa* is unique in the fact that it has a relatively poor oxygen affinity in comparison to other cytochrome *bd* oxidases and *cbb3* type cytochrome oxidases (Arai *et al* 2014). This makes sense when we consider the context of HCN expression in *P. aeruginosa*. HCN is under control of both quorum sensing (*rhl* and *las* systems) and ANR (Anaerobic regulator) with maximal expression achieved under high cell density and low oxygen tension (Laville *et al* 1998). However, under conditions where cell density is high but molecular oxygen is present (inhibits ANR function) a significant amount of HCN is still produced allowing *CioAB* to function. Under periods of low oxygen tension ANR is active and enhances production of HCN (Laville *et al* 1998). The low availability of oxygen and increased HCN production *CioAB* would not function optimally and the HCN inducible *PA4129-PA4134* would be upregulated enabling enhanced formation of the

cyanide resistant, microaerophilic *cbb3* type cytochrome c oxidase. This theory is backed by the presence of an ANR box at the +13 positions of *cioA* potentially blocking transcription of the *cioAB* operon under microaerophilic conditions (Cooper *et al* 2013).

If we regard this in the context of an infection, for example the CF lung, the thick mucosal layer creates regions of both low and high oxygen tension. This is further exacerbated by the ability of *P. aeruginosa* to form biofilms *in vivo* resulting in an oxygen gradient throughout the biofilm (Borriello *et al* 2004). It would be interesting to speculate that at the surface of a biofilm or when the cells are growing planktonically in the lung that CioAB is the predominant cyanide resistant aerobic respiration chain. However, during biofilm growth, the cells at the lower depths under microaerophilic conditions may respire primarily via the PA4133-PA4144 chimeric cyanide resistant chain. Understanding this dynamic will be key in combatting any infection utilising inhibitors designed against any member of the PA4129-PA4134 region. If both CioAB and PA4129-PA4134 work concurrently in a biofilm then inhibitors against both will be needed to fully prevent proliferation of an infection. However, if PA4129-PA4134 is predominant during biofilm formation then any inhibitor designed will prevent extensive biofilm formation, enabling the usage of antibiotics since the biofilms intrinsic resistance to antibiotics is removed.

4.0 – Summary

Both projects illustrate varying aspects that need to be explored to fully understand the problem of AMR and subsequently treat it. If we can determine the mechanism by which MDR emerges then we can begin implementing safeguards and start to identify high risk organisms or sequence type on the precipice of becoming a much larger problem. This enables to adopt a much more pre-emptive approach to combatting the problem of AMR as opposed to the reactive system we employ at present. This is essential considering the fact it take up to 15 years before new drugs combatting a new threat emerge and can be utilized effectively. If we can get 'the jump' on development of AMR and MDR we can cut down this timeframe. Microbial genomics will be essential in allowing us to almost track evolution in real time, determine new threats allowing mitigation of risk factors leading to development of drug resistance. Alternative therapies are needed to antibiotics to both reduce the chance of antibiotic resistance developing by lowering usage and to allow treatment of pan-resistant infections. These pan-resistant infections often occur in chronic *P. aeruginosa*

infections in the CF lung leading to mortality. By targeting virulence factors of *P. aeruginosa* we hope to prove the concept of virulence inhibition and open research and the market to employing similar tactics with other pathogens.

This is a small view of the range of skills needed to combat AMR. Every scientific discipline from microbiologists to biophysicists will be needed to combat this problem. The projects displayed employ computer science, bioinformatics, microbiology and eventually structural biology to answer questions about AMR and subsequently treat it. As our understanding of the problem grows, so does our chance of limiting the damage by AMR and enable up to continue to use these valuable resources.

5.0 Appendix 1 – ST10, ST167 and ST617 strain list (*Escherichia coli*)

<i>E. coli</i> Sequence Type 10 metadata				
Name	Source	Country	Date	Accession number
0000-0054-2927	Human	2015	United States	SRP056445
0000-0055-3694	Human	2015	United States	SRP056445
0000-0057-1723	Human	2016	United States	SRP056445
2776	Human	2015	Germany	ERP009121
489	Human	2013	Germany	ERP009121
Jul-60	Soil/Dust	2007	United States	SRP046387
1.4-R4	Human	2008	Australia	SRP082992
1061_13	Human	2014	United Kingdom	SRP041317
1303	Cow	2012	United Kingdom	SRR3492219
131/07	Cow	2012		United Kingdom
132031	Human	2015	United Kingdom	SRP071789
143510	Human	2015	United Kingdom	SRP071789
147812	ND	2015	United Kingdom	SRP071789
1505721		2015		Luxembourg
155719	Human	2015	United Kingdom	SRP071789
157050	Human	2015	United Kingdom	SRP071789
1600368	Bovine	2016		Luxembourg
168491	Human	2015	United Kingdom	SRP071789
187554	Human	2015	United Kingdom	SRP071789
188152	Human	2015	United Kingdom	SRP071789
188156	Human	2015	United Kingdom	SRP071789
188346	Human	2015	United Kingdom	SRP071789
196378	Human	2015	United Kingdom	SRP071789
2-222-05_S3_C2	Human	stool; Homo sapiens	Africa	PRJNA233810
2-316-03_S4_C2	Human	2009	Tanzania	SRP039692
2011-70-42-2	Swine	2011	Denmark	ERP009669
209926	Human	2016	United Kingdom	SRP071789
21.1-R1	Human	2010	Australia	SRP082992
302900	Human	2016	United Kingdom	SRP071789
304567	Human	2016	United Kingdom	SRP071789
6409	Human			SRR17405557
8350/98	Human	1998	Germany	ERP010393
93260	Human	2014	United Kingdom	SRP071789
AA022	Human	2013	Denmark	ERP011986
AA034	Avian	2003	Denmark	ERP011986
AA039	Human	1991	Denmark	ERP011986
AA044	Human	2008	Denmark	ERP011986

AA071	Camelid	2009	Mali	ERP011986
AA075	Camelid	2009	Mali	ERP011986
AA083		2010	Germany	ERP011986
AA088	Human	1986	United Kingdom	ERP011986
AA090	Human	1978	Denmark	ERP011986
AZ_TG76954	Avian	2014	United States	PRJNA230968
AZ_TG77054	Avian	2014	United States	PRJNA230968
AZ-TG56070	Avian	2013	United States	PRJNA230968
AZ-TG71211	Bovine	2009	United States	PRJNA230968
AZ-TG71479	Swine	2009	United States	PRJNA230968
AZ-TG71511	Avian	2009	United States	PRJNA230968
AZ-TG73443	Avian	2013	United States	PRJNA230968
AZ-TG73471	Avian	2013	United States	SRP038995
AZ-TG73747	Avian	2013	United States	PRJNA230968
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia	Avian		Asia	PRJNA243101
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.	Pig		Asia	PRJNA243101
BIDMC112	Human			SRR2135193
blood-90187	Human	2009	United States	SRP042632
BW25993		2012		SRP072433
BW2952	Air	2013		SRP074322
C321.deltaA	Human	2015	United States	SRP097637
CDPHFDLB-F12M00493-2	Dairy	2012	United States	SRP071258
CDPHFDLB-F1602032-004B	Bovine	2016	United States	SRP071258
CFSAN051538	Air	2014	United States	SRP078859
CVM N33720PS	Fish	2011	United States	SRP061934
CVM N34562PS	Fish	2012	United States	SRP061934
DRS026941	Air	2015	Japan	DRP003069
DTU-1	Laboratory	2014	Denmark	SRP077690
DTU2016_462_PRJ1050_Escherichia_coli_EST2014_131		2014	Estonia	
DTU2016_488_PRJ1050_Escherichia_coli_14051236QM4DJ1		2014	Italy	
DTU2016_493_PRJ1050_Escherichia_coli_EC112	2014	Lithuania		
DTU2016_533_PRJ1050_Escherichia_coli_ZTA14_00134EC	2014	Spain		
E. coli_R10	Swine	2015	Denmark	ERP017320
E. coli_R7	Swine	2015	Denmark	ERP017320
ESC0167	Canine		United States	SRP022914
ESC0175	Human		Sweden	SRP022914
ESC0178	Human		Sweden	SRP022914
Escherichia coli				ERP009684
Escherichia_coli-S_140	Human	2014	Denmark	ERP015638
FSIS 210115784	Bovine	2014	United States	SRP066916
GM4792			Asia	SRR3910922
GM4792			Asia	SRR3865307
HICF105	Human	2009	United Kingdom	SRP067738
HICF8	Human	2010	United Kingdom	SRP067738

HVH 164 (4-5953081)	Human			SRR770131
IAI34				ERP022711
IEH-NGS-ECO-00205	Plant	2014	United States	SRP058582
IHD813_16	Human	2012	Cambodia	SRP053197
JEONG-9567	Cow		North America	PRJNA298331
K-12	Laboratory	2015	Israel	SRP102730
K-12	Laboratory	2015	Israel	SRP102730
K-12	Laboratory	2015	Israel	SRP102730
K-12	Laboratory	2015	Israel	SRP102730
K-12 BW2952	Laboratory			
KCJK4139	Wild Animal	Feces	North America	PRJNA298331
KCJK4162	Reptile	2015	United States	SRP064902
KCJK4201	Soil/Dust	2015	United States	SRP064902
KH012	Air	2011	Japan	DRP002628
KLY metG-T ampC-R	Laboratory	2013	Israel	SRP096781
KTE61		Human		SRR641855
MC4100				ERP003998
MC4100				ERP003998
MG1657	Air	2013		SRP033025
MGK3S	Laboratory	2015	China	SRP053595
MGY	Air	2014	Israel	SRP096781
MOD1-EC1663	Human	1995	Belgium	SRP058582
MOD1-EC3488	Bovine		North America	PRJNA230969
MOD1-EC3551	Air		North America	PRJNA230969
MOD1-EC3561	Air		North America	PRJNA230969
MOD1-EC3605	Livestock			
MOD1-EC3647	Air		North America	PRJNA230969
MOD1-EC3681	Air		North America	PRJNA230969
MOD1-EC3758	Air		North America	PRJNA230969
MOD1-EC3795	Bovine	2013	United States	SRP058582
MOD1-EC5052	Environment			
MOD1-EC5053		2003		SRR3957119
MOD1-EC5087	Air			SRR3407790
MOD1-EC5091	Human			SRR3410237
MOD1-EC5166	Bovine		North America	PRJNA230969
MOD1-EC5260	Swine	2000	China	SRP058582
MOD1-EC5272	Swine	2000	United States	SRP058582
MOD1-EC5383	Camelid	2003	United States	SRP058582
MOD1-EC5469	Water	1993	United States	SRP058582
MOD1-EC5500	Human	1992	India	SRP058582
MOD1-EC5504	Swine	1976	United States	SRP058582
MOD1-EC5535	Swine	1978	United States	SRP058582
MOD1-EC5606	Avian	1979	United States	SRP058582

MOD1-EC5676	Bovine	1980	United States	SRP058582
MOD1-EC5718	Bovine	1981	United States	SRP058582
MOD1-EC5789	Swine	1983	United States	SRP058582
MOD1-EC5880	Swine	1983	United States	SRP058582
MOD1-EC5965	Swine	1983	United States	SRP058582
MOD1-EC6206	Swine	1983	United States	SRP058582
MOD1-EC6219	Swine	1983	United States	SRP058582
MOD1-EC6396	Ovine		Oceania	PRJNA230969
MOD1-EC6399	Bovine		North America	PRJNA230969
MOD1-EC6560	Bovine		Asia	PRJNA230969
MOD1-EC6576	Bovine		North America	PRJNA230969
MOD1-EC6577	Bovine		North America	PRJNA230969
MOD1-EC6595	Air		North America	PRJNA230969
MOD1-EC6598	Air		North America	PRJNA230969
MOD1-EC6646	Human	1991	Saudi Arabia	SRP058582
MOD1-EC6707	Air		North America	PRJNA230969
MOD1-EC6748	Human	1994	Zambia	SRP058582
MOD1-EC6756	Human	1994	Zambia	SRP058582
MOD1-EC6791	Air		North America	PRJNA230969
MOD1-EC6812	Canine		North America	PRJNA230969
MOD1-EC6872	Bovine		North America	PRJNA230969
MOD1-EC6966	Wild Animal			
MOD1-EC6978	Canine			
MOD1-EC7010	Water	1991	United States	SRP058582
MOD1-EC720	Human	1985	United States	SRP058582
MOD1-ECOR25	Canine		North America	PRJNA230969
MOD1-ECOR5	Human		North America	PRJNA230969
MPEC4839	Bovine			
NCM3722	Human	2014		United States
NCTC86	Human	1885	Germany	SRP090911
OH-16-21268-B	Bovine	2016	United States	SRP083292
PH39	Human	2012	Pakistan	SRP048823
PNUSAE004170	Dairy		United States	SRP046387
SCK67-52	Human		Netherlands	ERP016935
SQ2203		2008		United States
SQ37		2008		United States
Test-H17_P12b	Human	1944		ERP011986
UMEA 3180-1	Human			SRR958619
UMEA 3318-1	Human			SRR786317
UMNF18	Air	2007		United States
upec-282	Human	2012	United States	SRP042632
WCHEC1609	Water	2015	China	PRJNA275512

710	Livestock		Europe	Gift from Sebastian Gunther
711	Livestock		Europe	Gift from Sebastian Gunther
712	Livestock		Europe	Gift from Sebastian Gunther
713	Livestock		Europe	Gift from Sebastian Gunther
714	Livestock		Europe	Gift from Sebastian Gunther
715	Livestock		Europe	Gift from Sebastian Gunther
716	Livestock		Europe	Gift from Sebastian Gunther
717	Livestock		Europe	Gift from Sebastian Gunther
718	Livestock		Europe	Gift from Sebastian Gunther
719	Livestock		Europe	Gift from Sebastian Gunther
720	Livestock		Europe	Gift from Sebastian Gunther
721	Livestock		Europe	Gift from Sebastian Gunther
722	Livestock		Europe	Gift from Sebastian Gunther
723	Livestock		Europe	Gift from Sebastian Gunther
724	Livestock		Europe	Gift from Sebastian Gunther
725	Livestock		Europe	Gift from Sebastian Gunther
726	Livestock		Europe	Gift from Sebastian Gunther
727	Livestock		Europe	Gift from Sebastian Gunther
728	Livestock		Europe	Gift from Sebastian Gunther
729	Livestock		Europe	Gift from Sebastian Gunther
730	Livestock		Europe	Gift from Sebastian Gunther
731	Livestock		Europe	Gift from Sebastian Gunther
732	Livestock		Europe	Gift from Sebastian Gunther

733	Livestock		Europe	Gift from Sebastian Gunther
734	Livestock		Europe	Gift from Sebastian Gunther
735	Livestock		Europe	Gift from Sebastian Gunther
736	Livestock		Europe	Gift from Sebastian Gunther
737	Livestock		Europe	Gift from Sebastian Gunther
738	Livestock		Europe	Gift from Sebastian Gunther
739	Livestock		Europe	Gift from Sebastian Gunther
740	Livestock		Europe	Gift from Sebastian Gunther
741	Livestock		Europe	Gift from Sebastian Gunther
742	Livestock		Europe	Gift from Sebastian Gunther
743	Livestock		Europe	Gift from Sebastian Gunther
744	Livestock		Europe	Gift from Sebastian Gunther
745	Livestock		Europe	Gift from Sebastian Gunther
746	Livestock		Europe	Gift from Sebastian Gunther
747	Livestock		Europe	Gift from Sebastian Gunther
748	Livestock		Europe	Gift from Sebastian Gunther
749	Livestock		Europe	Gift from Sebastian Gunther
750	Livestock		Europe	Gift from Sebastian Gunther
751	Livestock		Europe	Gift from Sebastian Gunther
753	Livestock		Europe	Gift from Sebastian Gunther
754	Livestock		Europe	Gift from Sebastian Gunther
755	Livestock		Europe	Gift from Sebastian Gunther
756	Livestock		Europe	Gift from Sebastian Gunther

757	Livestock		Europe	Gift from Sebastian Gunther
758	Livestock		Europe	Gift from Sebastian Gunther
759	Livestock		Europe	Gift from Sebastian Gunther
760	Livestock		Europe	Gift from Sebastian Gunther
761	Livestock		Europe	Gift from Sebastian Gunther
762	Livestock		Europe	Gift from Sebastian Gunther
763	Livestock		Europe	Gift from Sebastian Gunther
764	Livestock		Europe	Gift from Sebastian Gunther
765	Livestock		Europe	Gift from Sebastian Gunther
766	Livestock		Europe	Gift from Sebastian Gunther
767	Livestock		Europe	Gift from Sebastian Gunther
768	Livestock		Europe	Gift from Sebastian Gunther
769	Livestock		Europe	Gift from Sebastian Gunther
770	Livestock		Europe	Gift from Sebastian Gunther
771	Livestock		Europe	Gift from Sebastian Gunther
772	Livestock		Europe	Gift from Sebastian Gunther
773	Livestock		Europe	Gift from Sebastian Gunther
774	Livestock		Europe	Gift from Sebastian Gunther
775	Livestock		Europe	Gift from Sebastian Gunther
776	Livestock		Europe	Gift from Sebastian Gunther
777	Livestock		Europe	Gift from Sebastian Gunther
778	Livestock		Europe	Gift from Sebastian Gunther
779	Livestock		Europe	Gift from Sebastian Gunther

780	Livestock		Europe	Gift from Sebastian Gunther
781	Livestock		Europe	Gift from Sebastian Gunther
782	Livestock		Europe	Gift from Sebastian Gunther
783	Livestock		Europe	Gift from Sebastian Gunther
784	Livestock		Europe	Gift from Sebastian Gunther
785	Livestock		Europe	Gift from Sebastian Gunther
786	Livestock		Europe	Gift from Sebastian Gunther
787	Livestock		Europe	Gift from Sebastian Gunther
788	Livestock		Europe	Gift from Sebastian Gunther
789	Livestock		Europe	Gift from Sebastian Gunther
790	Livestock		Europe	Gift from Sebastian Gunther
791	Livestock		Europe	Gift from Sebastian Gunther
792	Livestock		Europe	Gift from Sebastian Gunther
793	Livestock		Europe	Gift from Sebastian Gunther
794	Livestock		Europe	Gift from Sebastian Gunther
795	Livestock		Europe	Gift from Sebastian Gunther
796	Livestock		Europe	Gift from Sebastian Gunther
71	Livestock		Europe	Gift from Sebastian Gunther
72	Livestock		Europe	Gift from Sebastian Gunther
73	Livestock		Europe	Gift from Sebastian Gunther
74	Livestock		Europe	Gift from Sebastian Gunther
75	Livestock		Europe	Gift from Sebastian Gunther
76	Livestock		Europe	Gift from Sebastian Gunther

77	Livestock		Europe	Gift from Sebastian Gunther
78	Livestock		Europe	Gift from Sebastian Gunther
79	Livestock		Europe	Gift from Sebastian Gunther
<i>E. coli</i> Sequence Type 167 metadata				
Name	Source	Country	Date	Accession number
BIDMC98	Human	USA	2014	SRR2131466
910-sc-1872589				ERR587005
upec-273	Human - clinical	USA	2012	SRR1314503
blood-09-0949	Human - clinical	USA	2009	SRR1314258
eo721				ERR486694
eo168				ERR486693
eo722				ERR474037
eo433				ERR474034
eo3128				ERR439728
eo578				ERR435071
eo1703				ERR435237
eo1317				ERR435383
eo652				ERR434298
E 1101				ERR119473
MS 175-1	Human - carriage	USA		SRR088913
MS 116-1	Human - carriage	USA		SRR088906
EuSCAPE-FR076				ERR1205075
EuSCAPE-SK035				ERR1204985
EuSCAPE-DE065				ERR1217011
C-0863N0030				ERR1148553
EC-5-M4	Human - clinical	UK	2015	ERR1263950
EC-5-M3	Human - clinical	UK	2015	ERR1263949
EC-5-M2	Human - clinical	UK	2015	ERR1263948
EC-5-M1	Human - clinical	UK	2015	ERR1263947
EC-5	Human - clinical	UK	2012	ERR1263946
MOD1-EC5069	Dog	USA	2004	SRR4101520
208738	Human - clinical	UK	2016	SRR3574326

0000-0058-7850	Human - clinical	USA	2016	SRR3615372
RiKo 2340/09	Cow	USA	2009	SRR3492241
RiKo 2308/09	Cow	USA	2009	SRR3492237
DHQPCR312	Human - clinical	USA	2015	SRR3222296
AR-0014	Environmental	USA		SRR3112265
AR-0011	Environmental	USA		SRR3112262
GN02169	Human - clinical	USA	2003	SRR3098780
0000-0055-8719	Human - clinical	USA	2015	SRR3095388
HICF661		UK		SRR3051075
HICF611		UK		SRR3051056
HICF588		UK		SRR3051052
C53-sc-2235935	UK			ERR987393
C33b-sc-2212426	UK			ERR926425
C30-sc-2212416	UK			ERR926417
2473-sc-2212378	UK			ERR926381
BCW-2999	Ireland			SRR1815446
H21	Human - clinical	Nepal	2012	SRR1508895
H17	Human - clinical	Nepal	2012	SRR1508893
C00009162	UK			ERR296631
494	Poultry	China	2015	Data directly uploaded to Enterobase
MLST-167	Human - clinical	USA	2016	SRR3929024
AR-0162		USA		SRR4026005
AR-0151		USA		SRR4025999
AR-0150		USA		SRR4025998
AR-0149		USA		SRR4025997
b03d3fa0-dfca-11e5-88b1-3c4a9275d6c6				ERR1585801
bab46300-dfca-11e5-88b1-3c4a9275d6c6				ERR1585712
2408		Germany	2014	ERR1195513
1352	Human	Germany	2014	ERR712769
1029	Human	Germany	2014	ERR712691
948	Human	Germany	2014	ERR712662
839	Human	Germany	2014	ERR712617
428	Human	Germany	2013	ERR712515
IIIBS4 MIAE 11630	Sewage	France	2014	Data directly uploaded to Enterobase
855	Human - carriage	Oman	2015	Data directly uploaded to Enterobase
1071	Human - clinical	Oman	2015	Data directly uploaded to Enterobase
1626	Human - clinical	Oman	2015	Data directly uploaded to Enterobase
ERS1032655	Germany			ERR1211731

ERS1105873	Germany			ERR1640767
33b644a0-fa7c-11e5-b612-3c4a9275d6c8				ERR1609215
0000-0060-5811	Human - clinical	USA	2015	SRR4294836
SCK52-19	Human	Netherlands		ERR1619303
SCP20-59	Human	Netherlands		ERR1619298
SCP13-20-2	Human	Netherlands		ERR1619102
SCP08-10	Human	Netherlands		ERR1619006
SCP03-55	Human	Netherlands		ERR1618874
SCP29-17	Human	Netherlands		ERR1618796
SCP25-17	Human	Netherlands		ERR1618453
SCP02-29	Human	Netherlands		ERR1618013
SCK12-17	Human	Netherlands		ERR1617892
SCK12-16	Human	Netherlands		ERR1617891
SCK12-44	Human	Netherlands		ERR1617754
243	Human - clinical	China	2014	SRR5153285
442	Human - clinical	China	2014	SRR5153286
936	Human - clinical	China	2014	SRR5153284
1237	Human - clinical	China	2014	SRR5153283
1222	Human - clinical	China	2014	SRR5153290
57	Human - clinical	China	2014	SRR5153287
215	Human - clinical	China	2013	SRR5153281
25	Human - clinical	China	2014	SRR5153282
<i>E. coli</i> Sequence Type 617				
Name	Country	Source	Year	Accession number
1744-sc-1872569				ERR586987
upec-99	USA	Human - clinical	2011	SRR1314598
blood-09-0626	USA	Human - clinical	2009	SRR1314250
eo1482				ERR486698
eo2350				ERR474145
eo1302				ERR435571
eo3387				ERR435015
Escherichia coli 13462				ERR271435
PB264				ERR1218541
PI085				ERR1202021
PI029				ERR1202017
DO27				ERR1202010
CLO29				ERR1202006
CLO047				ERR1202004
CLO040				ERR1202003
CAE13				ERR1202002

EuSCAPE_BU009				ERR1228232
VRES0049				ERR1036096
VRES0046				ERR1036094
VRES0256				ERR998824
VRES0360				ERR985173
MOD1-EC5117	USA	Human	2006	SRR3410780
RA194				ERR1366625
RA105				ERR1366618
RA051				ERR1366613
50739822	Norway	Human - clinical	2013	Data uploaded directly to Enterobase
0000-0057-7635	USA	Human - clinical	2015	SRR3345827
HICF685				SRR3051084
HICF660				SRR3051074
MSB1_4A-sc-2280390				ERR1023709
MSB1_1A-sc-2280383				ERR1023706
MSB1_3I-sc-2280410				ERR1015375
MSB1_8G-sc-2280388				ERR1015365
A3b-sc-2212439				ERR971994
B3-sc-2212437				ERR926435
PH94	Pakistan	Human	2012	SRR1610063
PH143	Pakistan	Human	2012	SRR1610050
H1485	Nepal	Human	2012	SRR1508884
81	UK			ERR277001
444	China	Poultry	2014	Data uploaded directly to Enterobase
455	China	Poultry	2014	Data uploaded directly to Enterobase
EuSCAPE_TR128				ERR1541386
MOD1-EC6042	USA	Pig	1983	SRR3988038
1601115	Luxembourg	Cow	2016	Data uploaded directly to Enterobase
AR_0114	USA			SRR4025978
b04778d0-dfca-11e5-88b1-3c4a9275d6c6				ERR1585802
a70aaa30-dfca-11e5-88b1-3c4a9275d6c6				ERR1585770
bc74d7b0-dfca-11e5-88b1-3c4a9275d6c6				ERR1585751
baeeace0-dfca-11e5-88b1-3c4a9275d6c6				ERR1585717
DHQP1401958	USA	Human	2014	SRR4035120
2461	Germany	Human	2014	ERR1195555
2455	Germany	Human	2014	ERR1195552
319	Germany	Human	2014	ERR712843
1034	Germany	Human	2014	ERR712686

900	Germany	Human	2014	ERR712619
837	Germany	Human	2014	ERR712612
318	Germany	Human	2013	ERR712490
228	Tanzania	Human		Data uploaded directly to Enterobase
693	Oman	Human - clinical	2015	Data uploaded directly to Enterobase
979	Oman	Human - carriage	2015	Data uploaded directly to Enterobase
5639	Oman	Human - clinical	2015	Data uploaded directly to Enterobase
ERS1105867				ERR1640764
ERS1105786				ERR1640702
1a246670-fa7c-11e5-b358-3c4a9275d6c8				ERR1681629
1c35ba40-fa7c-11e5-b5da-3c4a9275d6c8				ERR1622279
2c8cecb0-fa7c-11e5-9ab9-3c4a9275d6c8				Data uploaded directly to Enterobase
SCP29-76	Netherlands	Human		ERR1619556
SCP29-75	Netherlands	Human		ERR1619555
SCP28-30	Netherlands	Human		ERR1619537
SCK67-66	Netherlands	Human		ERR1619493
SCP24-17	Netherlands	Human		ERR1619439
SCK52-24	Netherlands	Human		ERR1619308
SCK46-32	Netherlands	Human		ERR1619140
SCK33-40	Netherlands	Human		ERR1619075
SCK32-69	Netherlands	Human		ERR1619065
SCP06-77	Netherlands	Human		ERR1618983
SCP06-58	Netherlands	Human		ERR1618965
SCP29-79	Netherlands	Human		ERR1618827
SCP29-28	Netherlands	Human		ERR1618805
SCP03-50	Netherlands	Human		ERR1618050
SCK63-04	Netherlands	Human		ERR1617959
SCK12-28	Netherlands	Human		ERR1617895
SCK62-15	Netherlands	Human		ERR1617792
SCK08-54-2	Netherlands	Human		ERR1617732
784	China	Human - clinical	2014	SRR5153289
1037	China	Human - clinical	2014	SRR5153288

6.0 Appendix 2 – R scripts for Gene accumulation and rarefaction curve analysis

6.1 – Gene accumulation plot script

= Directions for use and annotation

```
#Define working directory
setwd("/home/ubuntu");

#read the input file.
#The input file must have no header row and column.
#rows correspond to genes, columns correspond to isolates
data <- read.table("gene_presence_absence.tab", sep="\t");
data <- t(data);#we transpose the table to run a for loop row-wise
nr_rows <- nrow(data);
nr_cols <- ncol(data);
nr_iterations <- 100;#the number of iterations
#the matrix which will hold the results
core_results <- matrix(data=NA,nrow=nr_rows,ncol=nr_iterations);
#a vector the holds the updated results of each iteration
presence_line <- matrix(data=0,nrow=1,ncol=nr_cols);
# the times that we will run our calculation
for(times in 1: nr_iterations){
  #for all the rows of the table
  for (i in 1: nr_rows){
    sum <- 0;
    if(i==1){ #for the first row
      for(j in 1: nr_cols){#calculate the number of genes
        presence_line[1,j]<-data[i,j]
        sum <- sum + data[i,j];
      }
    }
    #and store the first result
    core_results[i,times] <- sum
    next;
  }
  #for every consecutive row, for every gene
  for(j in 1: nr_cols){
    #if we find e new gene count it
    if(presence_line[1,j] || data[i,j]){
      presence_line[1,j] <- 1;
    }
    else{presence_line[1,j] <- 0;}
  }
}
#count the total number of present genes
```

```

for(j in 1: nr_cols){
  sum <- sum + presence_line[1,j];
}
#store the result
core_results[i,times] <- sum;
}
#suffle the matrix and repeat the procedure
data <- data[sample(nrow(data),nrow(data)), ]
}
#print the results in a tab delemeter file
write.table(core_results, file = "pan_genome.txt", sep="\t");

```

6.2 – Rarefaction curve plot script

```

#Define working directory
setwd("/home/ubuntu");
#read the input file.
#The input file must have no header row and column.
#rows corespond to genes, columns corespond to isolates
data <- read.table("gene_presence_absence.tab", sep="\t");
nr_rows <- nrow(data);
nr_cols <- ncol(data);
nr_iterations <- 100; #the number of iterations
#the matrix which will hold the results
core_results <- matrix(data=NA,nrow=nr_rows,ncol=nr_iterations);
#a vector the holds the updated results of each iteration
presence_line <- matrix(data=0,nrow=1,ncol=nr_cols);
# the times that we will run our calculation
for(times in 1:nr_iterations){
  #for all the rows of the table
  for (i in 1: nr_rows){
    sum <- 0;
    if(i==1){ #for the first row
      for(j in 1: nr_cols){#calculate the number of genes
        presence_line[1,j]<-data[i,j]
        sum <- sum + data[i,j];
      }
      #and store the first result
      core_results[i,times] <- sum
      next;
    }
    #for every consecutive row, for every gene
    for(j in 1: nr_cols){
      #if the gene was and is present count it
      if(presence_line[1,j] && data[i,j]){
        presence_line[1,j] <- 1;
      }# else dont
    }
  }
}

```

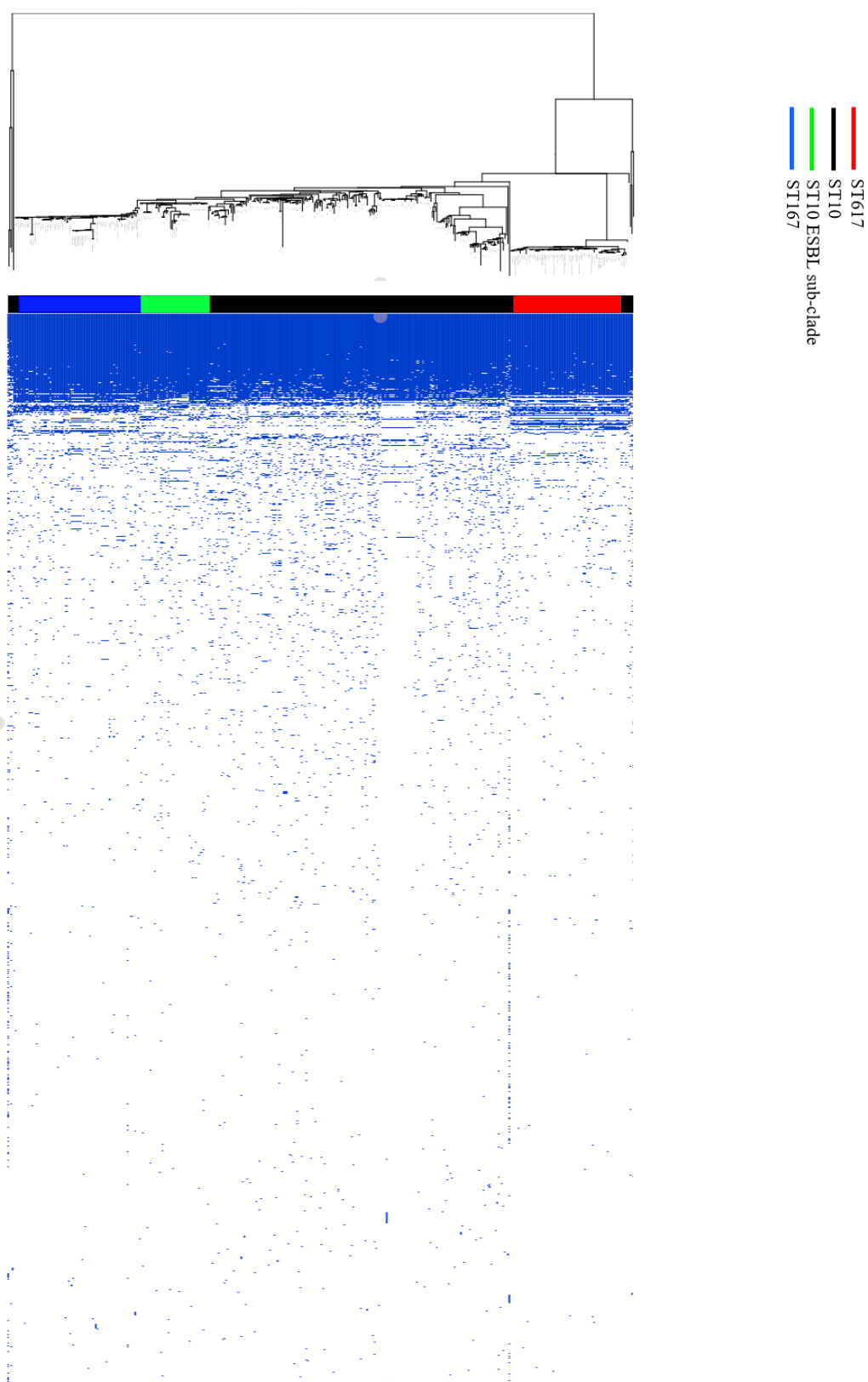
```

else{presence_line[1,j] <- 0;}
}
#count the total number of present genes
for(j in 1: nr_cols){
  sum <- sum + presence_line[1,j];
}
#store the result
core_results[i,times] <- sum;

}
#suffle the matrix and repeat the procedure
data <- data[sample(nrow(data),nrow(data)), ]
}
#print the results in a tab delemeter file
write.table(core_results, file = "core_genome.txt", sep="\t");

```

7.0 Appendix 3 – Pan-Genome visualisation of ST10, ST167 and ST617.



Appendix 3) Pan genome plot produced by Roary overlaid with phylogenetic tree inferred by Parsnp.

8.0 – Appendix 4 – Scoary analysis data breakdown

Top hits generated by Scoary for genes and alleles unique to ST10, Not present in ST167 and ST617.			
Gene_name	Annotation	Sensitivity (% association)	Bonferri_p_value
ytfA	HTH-type transcriptional repressor nicS	95.98394	8.23E-100
fimI	putative major type 1 fimbrial subunit LpfA	93.17269	1.08E-57
dsdC	HTH-type transcriptional regulator PerR	85.14056	1.08E-72
no homologue	D-tagatose 3-epimerase	83.93574	1.50E-70
wcaF	Galactoside O-acetyltransferase	83.13253	4.38E-71
wcaA	UDP-Glc:alpha-D-GlcNAc-diphosphoundecaprenol beta-1,3-glucosyltransferase WfgD	82.73092	1.10E-70
crt	Short-chain-enoyl-CoA hydratase	82.73092	3.48E-68
wzb	Low molecular weight protein-tyrosine-phosphatase Wzb	81.92771	2.50E-69
wzc	Tyrosine-protein kinase wzc	81.92771	2.50E-69
wcaI	putative glycosyltransferase	81.92771	2.50E-69
wcaJ	UDP-glucose:undecaprenyl-phosphate glucose-1-phosphate transferase	81.5261	8.57E-21
wcaB	Serine acetyltransferase	80.72289	3.61E-67
wcaL	GDP-mannose-dependent alpha-(1-6)-phosphatidylinositol monomannoside mannosyltransferase	80.32129	1.77E-64
wzxC	Lipopolysaccharide biosynthesis protein WzxC	79.51807	1.74E-65
gmd	GDP-mannose 4,6-dehydratase	79.51807	1.74E-65
wcaH	GDP-mannose mannosyl hydrolase	77.91165	6.03E-63
wcaG	GDP-L-fucose synthase	77.91165	6.03E-63
cspG	Cold shock-like protein CspJ	76.70683	4.07E-59
relB	Antitoxin RelB	76.70683	3.12E-57
cspB	Cold shock-like protein CspB	72.69076	1.03E-51
cspG	Cold shock-like protein CspG	73.89558	3.52E-55
relE	mRNA interferase RelE	71.48594	5.06E-52

Top hits generated by Scoary for genes and alleles associated with ST167 and ST617, not present in ST10			
Gene_annotation	Annotation	Sensitivity	Bonferroni_p
iclR_1	Acetate operon repressor	95.58233	1.32E-98
yagE	putative 2-keto-3-deoxy-galactonate aldolase YagE	95.58233	2.95E-96
sacA	Sucrose-6-phosphate hydrolase	93.17269	1.11E-87
scrK	Fructokinase	93.17269	1.11E-87
degA	HTH-type transcriptional regulator DegA	92.36948	1.05E-85
lacY_2	Lactose permease	91.16466	6.89E-83
group_16250	DsdX permease	93.5743	7.56E-87
group_6511	Glucosamine-6-phosphate deaminase	89.15663	1.03E-76
uxaA_1	Altronate dehydratase	95.58233	1.42E-90
axe1-6A	Carbohydrate acetyl esterase/feruloyl esterase	87.14859	4.79E-71
iraM	Anti-adapter protein IraM	81.1245	1.36E-60
group_11042	D-serine dehydratase	87.95181	1.59E-69
group_24461	Inorganic pyrophosphatase	86.34538	1.72E-66
chpS	Antitoxin ChpS	86.34538	1.72E-66
chpB	Endoribonuclease ChpB	86.34538	1.72E-66
crfC_1	Clamp-binding protein CrfC	85.94378	9.30E-66
group_17727	hypothetical protein	85.14056	2.52E-64
group_31789	hypothetical protein	82.73092	5.54E-60
kanE	Alpha-D-kanosaminyltransferase	81.5261	2.68E-58
group_28126	Thiosulfate sulfurtransferase PspE	98.79518	1.19E-96
group_19329	Psp operon transcriptional activator	98.79518	1.19E-96
group_13345	Multiple sugar-binding protein	93.17269	1.99E-79
group_28246	hypothetical protein	90.36145	5.45E-73
rusA	Crossover junction endodeoxyribonuclease RusA	86.74699	8.24E-66
group_22221	Ribonuclease E	98.79518	3.71E-95
group_31788	hypothetical protein	97.59036	1.19E-90
group_23454	Inner membrane protein YidI	97.59036	1.19E-90
group_27106	Flagellar hook-associated protein 3	95.18072	3.15E-83
group_29887	LexA repressor	91.56627	3.19E-74
mcrB	hypothetical protein	91.56627	3.19E-74
group_26781	hypothetical protein	91.16466	2.51E-73
hsdR_12	hypothetical protein	90.36145	1.38E-71
dnaJ_1	Chaperone protein DnaJ	88.75502	6.01E-67

group_30000	Magnesium and cobalt efflux protein CorC	86.34538	2.12E-62
group_15973	hypothetical protein	93.5743	2.94E-76
group_17174	Flagellar hook-associated protein 1	92.77108	2.63E-74
cma	hypothetical protein	89.95984	4.38E-68
group_25752	Toxin-antitoxin biofilm protein TabA	76.70683	5.72E-46
group_20763	sn-glycerol-3-phosphate transport system permease protein UgpA	99.19679	1.19E-92
group_8860	Flagellar basal-body rod protein FlgG	92.77108	5.18E-73
argF'	hypothetical protein	92.77108	5.18E-73
bshA_2	N-acetyl-alpha-D-glucosaminyl L-malate synthase	82.73092	7.98E-54
tagH	Teichoic acids export ATP-binding protein TagH	80.32129	4.34E-50
group_28483	Sensor protein PhoQ	98.39357	4.69E-88
rrrD_4	hypothetical protein	82.32932	4.98E-52
galE_2	UDP-glucose 4-epimerase	81.92771	2.86E-51
group_24605	Peptidoglycan hydrolase FlgJ	97.18876	2.55E-81
group_6302	Sensor protein EvgS	93.17269	2.79E-70
group_6498	Gamma-glutamyltranspeptidase	99.19679	1.58E-87
group_17125	Starvation-sensing protein RspA	85.94378	1.55E-54
pglA	N,N'-diacetylbacillosaminyl-diphospho-undecaprenol alpha-1,3-N-acetylgalactosaminyltransferase	82.32932	9.80E-49
galF_1	UTP--glucose-1-phosphate uridylyltransferase	82.32932	9.11E-47
group_1053	Actin cross-linking toxin VgrG1	77.10843	1.76E-39
recE	Exodeoxyribonuclease 8	96.78715	2.33E-75
nanS_1	hypothetical protein	95.98394	7.00E-72
narZ_1	Respiratory nitrate reductase 2 alpha chain	95.98394	8.48E-71
yehB_2	Outer membrane usher protein YehB	83.93574	3.60E-46
group_19483	Arsenical pump membrane protein	97.18876	2.90E-72
cmoB_2	putative S-adenosylmethionine-dependent methyltransferase/MSMEI_2290	85.14056	3.22E-45
rrrD	Lysozyme RrrD	96.78715	5.59E-67
group_20318	Polyol:NADP oxidoreductase	96.78715	5.25E-66
mec	CysO-cysteine peptidase	86.74699	1.17E-43
kdpE	KDP operon transcriptional regulatory protein KdpE	86.74699	1.17E-43
group_19743	Valine--tRNA ligase	98.39357	2.38E-68
intA_1	Prophage CP4-57 integrase	99.19679	7.66E-67
traC	hypothetical protein	95.98394	8.61E-55

tuaC	Putative teichuronic acid biosynthesis glycosyltransferase TuaC	87.14859	1.27E-35
ybdK	Putative glutamate--cysteine ligase 2	83.13253	4.91E-30
Top hits generated by Scoary for genes and alleles found in the ST167 population, but not present in the ST10 population. Genes and alleles shared by both ST167 and ST617 but not ST10 are not reported in this section.			
Gene	Annotation	Sensitivity	Bonferroni_p
araB_2	Ribulokinase	100	6.39E-71
rbsC_2	Ribose import permease protein RbsC	100	6.39E-71
rbsA_3	Ribose import ATP-binding protein RbsA	100	6.39E-71
rbsB_1	Ribose import binding protein RbsB	100	6.39E-71
cra_2	Catabolite repressor/activator	100	6.39E-71
yfcR	putative fimbrial-like protein YfcR	100	6.39E-71
yfcP	putative fimbrial-like protein YfcP	100	6.39E-71
yfcQ	putative fimbrial-like protein YfcQ	100	6.39E-71
	Invasin	100	6.90E-68
higB	mRNA interferase HigB	100	6.90E-68
xseA	Exodeoxyribonuclease 7 large subunit	100	6.90E-68
rne	Ribonuclease E	100	6.90E-68
licT	Transcription antiterminator LicT	100	1.41E-66
mdtG	Multidrug resistance protein MdtG	100	2.35E-65
higA	Antitoxin HigA	100	2.35E-65
mopA	Molybdenum-pterin-binding protein MopA	100	3.29E-64
ecdB	putative UbiX-like flavin prenyltransferase	100	4.29E-62
edcC	Phenolic acid decarboxylase subunit C	100	4.29E-62
ptsI_2	Phosphoenolpyruvate-protein phosphotransferase	100	4.29E-62
bsdD	Phenolic acid decarboxylase subunit D	100	4.29E-62
hosA	Transcriptional regulator HosA	100	4.15E-61
cysT	Sulfate transport system permease protein CysT	100	2.95E-59
rlmB_1	23S rRNA (guanosine-2'-O-)-methyltransferase RlmB	100	2.95E-59
iclR_1	Acetate operon repressor	100	2.95E-59
ygbN	Inner membrane permease YgbN	100	2.95E-59
manZ_2	Mannose permease IID component	100	1.11E-53

fucK	L-fuculokinase	100	5.65E-53
lexA_1	LexA repressor	100	1.30E-51
nagB_2	Glucosamine-6-phosphate deaminase	100	7.09E-48
ygcB	CRISPR-associated endonuclease/helicase Cas3	100	1.25E-45
glpT	Glycerol-3-phosphate transporter	100	1.25E-45
qorB	Quinone oxidoreductase 2	100	1.45E-44
rusA	Crossover junction endodeoxyribonuclease RusA	100	1.45E-44
crfC_1	Clamp-binding protein CrfC	100	1.54E-43
casE	CRISPR system Cascade subunit CasE	100	1.52E-42
cycA	D-serine/D-alanine/glycine transporter	100	4.63E-42
galF_2	UTP--glucose-1-phosphate uridylyltransferase	100	1.19E-40
kanE	Alpha-D-kanosaminyltransferase	100	1.95E-38
casA	CRISPR system Cascade subunit CasA	100	5.20E-38
ugpA	sn-glycerol-3-phosphate transport system permease protein UgpA	98.71795	6.18E-67
amiC	N-acetylmuramoyl-L-alanine amidase AmiC	98.71795	1.64E-65
malX_2	PTS system maltose- and glucose-specific EIICB component	98.71795	1.64E-65
kdsD_2	Arabinose 5-phosphate isomerase KdsD	98.71795	1.64E-65
patB_2	Cystathionine beta-lyase PatB	98.71795	1.64E-65
yfcV	putative fimbrial-like protein YfcV	98.71795	3.31E-64
yagU_2	Inner membrane protein YagU	98.71795	5.41E-63
ghoT_1	Toxin GhoT	98.71795	5.41E-63
yidI	Inner membrane protein YidI	98.71795	7.45E-62
rhmt_2	Inner membrane transport protein RhmT	98.71795	6.19E-57
yagE	putative 2-keto-3-deoxy-galactonate aldolase YagE	98.71795	6.19E-57
cysM	Cysteine synthase B	98.71795	6.19E-57
uxaA_1	Altronate dehydratase	98.71795	6.19E-57
sacA	Sucrose-6-phosphate hydrolase	98.71795	3.97E-52
scrK	Fructokinase	98.71795	3.97E-52
casC	CRISPR system Cascade subunit CasC	98.71795	3.97E-52
degA	HTH-type transcriptional regulator DegA	98.71795	1.05E-50
mcrB	hypothetical protein	98.71795	2.36E-49
lacY_2	Lactose permease	98.71795	1.06E-48
hsdR_12	hypothetical protein	98.71795	1.91E-47

axe1-6A	Carbohydrate acetyl esterase/feruloyl esterase	98.71795	6.67E-43
iraM	Anti-adaptor protein IraM	98.71795	6.60E-36
phoQ	Sensor protein PhoQ	97.4359	3.89E-62
bglF_2	PTS system beta-glucoside-specific EIIBCA component	97.4359	8.47E-60
ulaD	3-keto-L-gulonate-6-phosphate decarboxylase UlaD	97.4359	2.05E-52
dsdX	DsdX permease	97.4359	7.02E-51
rlmG	Ribosomal RNA large subunit methyltransferase G	97.4359	9.77E-44
dnaJ_1	Chaperone protein DnaJ	97.4359	3.60E-43
chpS	Antitoxin ChpS	97.4359	5.43E-40
ppa	Inorganic pyrophosphatase	97.4359	5.43E-40
chpB	Endoribonuclease ChpB	97.4359	5.43E-40
corC_3	Magnesium and cobalt efflux protein CorC	97.4359	5.43E-40
pspF	Psp operon transcriptional activator	96.15385	1.56E-61
pspE	Thiosulfate sulfurtransferase PspE	96.15385	1.56E-61
spaN	Surface presentation of antigens protein SpaN	96.15385	4.83E-59
spaL	putative ATP synthase SpaL	96.15385	4.83E-59
flgL	Flagellar hook-associated protein 3	96.15385	3.27E-52
fecD	hypothetical protein	96.15385	4.57E-49
msmE	Multiple sugar-binding protein	96.15385	2.44E-48
rfaP_1	Lipopolysaccharide core heptose(I) kinase RfaP	96.15385	1.45E-42
acrE_1	Multidrug export protein AcrE	96.15385	2.58E-37
rrrD_4	hypothetical protein	96.15385	1.02E-33
etfA_2	Electron transfer flavoprotein subunit alpha	94.87179	2.81E-57
yqcE	Inner membrane protein YqcE	94.87179	1.70E-50
flgK	Flagellar hook-associated protein 1	94.87179	5.91E-46
flgG	Flagellar basal-body rod protein FlgG	94.87179	5.91E-46
argF'	hypothetical protein	94.87179	5.91E-46
elaD_1	Protease ElaD	94.87179	1.03E-36
fliP_1	Flagellar biosynthetic protein FliP	93.58974	4.97E-61
dsdA	D-serine dehydratase	93.58974	3.06E-37
ggt	Gamma-glutamyltranspeptidase	92.30769	7.50E-58
ygcS_2	Inner membrane metabolite transport protein YgcS	92.30769	5.11E-54
invA	Invasion protein InvA	92.30769	5.11E-54
xylE_1	D-xylose-proton symporter	92.30769	7.14E-52
intA_1	Prophage CP4-57 integrase	92.30769	9.42E-46

bin3	Putative transposon Tn552 DNA-invertase bin3	92.30769	5.33E-45
bshA_2	N-acetyl-alpha-D-glucosaminyl L-malate synthase	92.30769	7.63E-30
tagH	Teichoic acids export ATP-binding protein TagH	92.30769	1.77E-27
orgB	Oxygen-regulated invasion protein OrgB	91.02564	1.71E-52
arsB	Arsenical pump membrane protein	91.02564	2.32E-50
flgJ	Peptidoglycan hydrolase FlgJ	91.02564	2.32E-50
hila	Transcriptional regulator HilA	91.02564	1.71E-35
galF_1	UTP--glucose-1-phosphate uridylyltransferase	91.02564	3.69E-28
galE_2	UDP-glucose 4-epimerase	91.02564	9.20E-28
mxlD	Outer membrane protein MxlD	89.74359	8.21E-55
guaB	Inosine-5'-monophosphate dehydrogenase	89.74359	8.21E-55
invF	Invasion protein InvF	89.74359	8.21E-55
elaD_2	Protease ElaD	89.74359	8.21E-55
ybdK	Putative glutamate--cysteine ligase 2	89.74359	9.63E-28
recE	Exodeoxyribonuclease 8	88.46154	1.55E-46
nanS_1	hypothetical protein	88.46154	9.89E-45
mec	CysO-cysteine peptidase	88.46154	1.85E-30
pglA	N,N'-diacetylbaicillosaminyldiphospho-undecaprenol alpha-1,3-N-acetylgalactosaminyltransferase	88.46154	8.92E-26
rhsC_5	hypothetical protein	88.46154	6.61E-24
yghQ	Inner membrane protein YghQ	88.46154	6.61E-24
bla_1	Beta-lactamase CTX-M-1	78.20513	2.87E-23
kdpE	KDP operon transcriptional regulatory protein KdpE	78.20513	5.66E-22
ccdB	hypothetical protein	78.20513	3.74E-17
Top hits generated by Scoary for genes and alleles associated with ST617, but not present in the ST10 population. Genes and alleles shared by both ST167 and ST617 but not ST10 are not reported in this section.			
Gene	Annotation	Sensitivity	Bonferroni_p
yagU_2	Inner membrane protein YagU	100	5.64E-70
pspF	Psp operon transcriptional activator	95.34884	1.91E-64
pspE	Thiosulfate sulfurtransferase PspE	95.34884	1.91E-64
rlmB_1	23S rRNA (guanosine-2'-O-)-methyltransferase RlmB	100	1.21E-63
yagE	putative 2-keto-3-deoxy-galactonate aldolase YagE	100	1.21E-63

iclR_1	Acetate operon repressor	100	1.21E-63
rhmT_2	Inner membrane transport protein RhmT	98.83721	2.56E-61
ulaD	3-keto-L-gulonate-6-phosphate decarboxylase UlaD	100	5.36E-61
valS	Valine--tRNA ligase	91.86047	2.63E-58
yjgN	Inner membrane protein YjgN	90.69767	3.99E-58
rne	Ribonuclease E	90.69767	3.99E-58
ggt	Gamma-glutamyltranspeptidase	89.53488	4.22E-58
uxaA_1	Altronate dehydratase	96.51163	1.93E-57
dsdX	DsdX permease	98.83721	4.54E-57
ugpA	sn-glycerol-3-phosphate transport system permease protein UgpA	88.37209	1.06E-56
scrK	Fructokinase	98.83721	2.72E-56
sacA	Sucrose-6-phosphate hydrolase	98.83721	2.72E-56
yidI	Inner membrane protein YidI	91.86047	5.92E-56
por_2	Polyol:NADP oxidoreductase	93.02326	2.32E-55
flgJ	Peptidoglycan hydrolase FlgJ	91.86047	7.07E-55
degA	HTH-type transcriptional regulator DegA	98.83721	8.43E-55
phoQ	Sensor protein PhoQ	88.37209	5.15E-54
flgL	Flagellar hook-associated protein 3	94.18605	3.36E-53
lacY_2	Lactose permease	98.83721	1.07E-52
evgS	Sensor protein EvgS	96.51163	1.73E-52
recE	Exodeoxyribonuclease 8	89.53488	5.33E-51
msmE	Multiple sugar-binding protein	95.34884	8.47E-51
nanS_1	hypothetical protein	88.37209	9.08E-48
narZ_1	Respiratory nitrate reductase 2 alpha chain	88.37209	9.08E-48
dsdA	D-serine dehydratase	98.83721	1.05E-47
flgK	Flagellar hook-associated protein 1	93.02326	5.80E-47
qorB	Quinone oxidoreductase 2	98.83721	5.05E-46
nagB_2	Glucosamine-6-phosphate deaminase	96.51163	8.44E-46
cycA	D-serine/D-alanine/glycine transporter	100	1.44E-45
flgG	Flagellar basal-body rod protein FlgG	91.86047	1.61E-45
argF'	hypothetical protein	91.86047	1.61E-45
rrrD	Lysozyme RrrD	83.72093	1.39E-44
arsB	Arsenical pump membrane protein	82.55814	2.29E-44
galF_2	UTP--glucose-1-phosphate uridylyltransferase	100	4.51E-44
mcrB	hypothetical protein	91.86047	1.79E-43
axe1-6A	Carbohydrate acetyl esterase/feruloyl esterase	96.51163	6.02E-43
lexA_1	LexA repressor	90.69767	4.18E-42

Int-Tn	Transposase from transposon Tn916	79.06977	6.13E-42
intS_2	Putative prophage CPS-53 integrase	83.72093	6.54E-42
hsdR_12	hypothetical protein	91.86047	1.43E-41
epsE_2	Type II secretion system protein E	93.02326	3.52E-41
cma	hypothetical protein	91.86047	5.78E-41
ppa	Inorganic pyrophosphatase	95.34884	2.76E-40
chpS	Antitoxin ChpS	95.34884	2.76E-40
chpB	Endoribonuclease ChpB	95.34884	2.76E-40
ptsI_4	Phosphoenolpyruvate-protein phosphotransferase	100	6.03E-40
dnaJ_1	Chaperone protein DnaJ	91.86047	3.20E-39
rspA	Starvation-sensing protein RspA	94.18605	2.78E-38
yfkM_1	General stress protein 18	69.76744	6.96E-37
eutN_2	Ethanolamine utilization protein EutN	69.76744	6.96E-37
ypdA_1	Sensor histidine kinase YpdA	69.76744	6.96E-37
pduL	Phosphate propanoyltransferase	69.76744	6.96E-37
pduA	Propanediol utilization protein PduA	69.76744	6.96E-37
pduB	Propanediol utilization protein PduB	69.76744	6.96E-37
ccmK	Carbon dioxide-concentrating mechanism protein CcmK	69.76744	6.96E-37
csdB	4-hydroxyphenylacetate decarboxylase large subunit	69.76744	6.96E-37
tdcD_1	Propionate kinase	69.76744	6.96E-37
dnaK_3	Chaperone protein DnaK	69.76744	6.96E-37
btr_1	HTH-type transcriptional activator Btr	69.76744	6.96E-37
metK_2	S-adenosylmethionine synthase	69.76744	6.96E-37
crfC_1	Clamp-binding protein CrfC	93.02326	7.14E-37
rusA	Crossover junction endodeoxyribonuclease RusA	91.86047	1.55E-36
lpxM2	Lipid A biosynthesis myristoyltransferase 2	89.53488	4.48E-36
corC_3	Magnesium and cobalt efflux protein CorC	91.86047	4.98E-36
epsF_2	Type II secretion system protein F	93.02326	6.89E-36
iraM	Anti-adaptor protein IraM	96.51163	1.10E-35
bshA_2	N-acetyl-alpha-D-glucosaminyl L-malate synthase	94.18605	1.76E-34
galE_2	UDP-glucose 4-epimerase	94.18605	1.33E-33
pglA	N,N'-diacetylbaicillosaminyl-diphospho-undecaprenol alpha-1,3-N-acetylglactosaminyltransferase	93.02326	1.11E-32

ghxP	Guanine/hypoxanthine permease GhxP	79.06977	2.13E-32
tagH	Teichoic acids export ATP-binding protein TagH	94.18605	6.49E-32
kanE	Alpha-D-kanosaminyltransferase	93.02326	8.00E-32
kdpE	KDP operon transcriptional regulatory protein KdpE	87.2093	1.23E-31
cmoB_2	putative S-adenosylmethionine- dependent methyltransferase/MSMEI_2290	88.37209	6.96E-31
yehB_2	Outer membrane usher protein YehB	89.53488	1.08E-30
rrrD_4	hypothetical protein	89.53488	5.71E-29
galF_1	UTP--glucose-1-phosphate uridylyltransferase	88.37209	7.83E-28
rhmr_1	putative HTH-type transcriptional regulator RhmR	89.53488	9.61E-28
mec	CysO-cysteine peptidase	77.90698	1.63E-23
icaB_2	Poly-beta-1,6-N-acetyl-D- glucosamine N-deacetylase	75.5814	1.73E-20
ypjF	putative toxin YpjF	70.93023	3.83E-16
yfjZ	Putative antitoxin YfjZ	70.93023	3.83E-16

9.0 – SEER analysis CDS and IGR breakdown and annotation

Breakdown of CDS and IGR sequences which had significant Kmers shared by both the ST167 and ST617 population, that are absent in the ST10 population. Highlighted in green are genes implicated in anaerobic respiration/metabolism whilst in yellow are genes involved with GI tract survival.			
start	end	CDS/IGR name	Description
11786	11796	dinF	DNA damage SOS response. Part of MATE-like family of proteins linked with production from osmotic and bile salt stress in e.COLI
19575	19585	yjbl	SopA like protein from TYPell secretion system acting as ubiquitin ligase enzymes
49138	49128	arpA	ankrin repeat protein A
57202	57193	yjaA	MreB family protein involved with cell division and elongation (envB in E.coli)
57661	57650	23s rRNA	
62119	62194	16S rRNA	
71980	71990	hemE	uropophoryin decarboxylase
103642	103698	16S rRNA	
104703	104693	MurI	Glutamate racemase used to create peptidoglycan precursor N-acetylmuramic acid
120674	120685	CptA	phosphoethanolamine transferase, toxin/antitoxin system, inhibits cytoskeletal proteins from polymerizing i.e ftsZ and mreB
159867	159877	cdh	CDP-diacylglycerol pyrophosphatase ;This is a family of CDP-diacylglycerol pyrophosphatases, EC:3.6.1.26. This enzyme catalyzes the reaction CDP-diacylglycerol + H2O <=> CMP + phosphatidate.
171184	171198	phage tail protein (pseudo)	
173775	173786	baseplate assembly protein	Phage related baseplate protein
182451	182439	YcjR	transcriptional regulator
185368	185357	hypothetical protein	blastp=parB like domains seemingly involved with DNA replication
190149	190138	replication endonuclease	blastp = Bacteriophage replication gene A
201196	201186	RhaT	Rhamanose/proton symporter
214579	214568	FdhD	sulfurtransferase which enables formation of formate in anaerobic conditions in the absence of nitrogen as an energy source
223303	223312	acetyltransferase	implicated in glutamate biosynthesis
250564	250574	yihA	GTPase required for normal cell division control and chromosomal partitioning
255174	255185	PlsC	involved with phospholipid synthesis
260233	260243	rrf	5S rRNA
266794	266784	trkH	potassium uptake protein
274676	274666	UbiD	catalyzes the decarboxylation of 3-octaprenyl-4-hydroxy benzoate to 2-octaprenylphenol
281746	281736	UbiE	ubiquinone/menaquinone biosynthesis C-methyltransferase
289920	289910	YigL	dephosphorylation of pyridoxal-phosphate (PLP) and moderately the dephosphorylation of 2-deoxyglucose 6-phosphate (2bGLU6P) and beta-glucose 6-phosphate (bGlu6P). Also hydrolyzes both purines (GMP and IMP) and pyrimidines as secondary substrates
290552	290562	pldB	catalyzes the conversion of 1-lysophosphatidylcholine to glycerophosphocholine

296217	296227	hypothetical protein	blastP = paal like thioesterase involved with degradation of phenyllytic acid
296587	296577	rarD	inner membrane protein drug/metabolite transporter
324604	324614	yigl	WecE family involved with biosynthesis and degradation of surface polysaccharide and LPS
336624	336614	GppA	guanosine-5'-triphosphate,3'-diphosphate pyrophosphatase
376618	376628	MnmG	tRNA uridine-5-carboxymethylaminomethyl(34) synthesis enzyme
395448	395458	bglF	used beta-glucoside-specific PTS enzymes: IIA component/IIB component/IIC component
404843	404862	IS6 family transposase (pseudo gene)	
429737	429727	dgoA	2-dehydro-3-deoxy-6-phosphogalactonate aldolase, functions in galactonate metabolism, upregulated in low oxygen conditions where pyruvate becomes limited 2-dehydro-3-deoxy-6-phospho-D-galactonate = pyruvate + D-glyceraldehyde 3-phosphate
439725	439715	glvC	arbutin specific enzyme IIC component of PTS;enzyme; Transport of small molecules: Carbohydrates, organic acids, alcohols~PTS system, arbutin-like IIB component~PTS system, arbutin-like IIC component
445804	445793	yidJ	blastP= YidJ putative sulfatase implicated in utilisation of sulfur for metabolism
519215	519225	WaaL	O-antigen ligase
537640	537629	lIdP	L-lactate permease
548399	548389	rhs	Type IV secretion protein, Uncharacterized
581338	581328	xylH	putative xylose transport, membrane component; membrane component of ABC superfamily
592281	592270		glycine-- tRNA ligase subunit beta
593602	593544	InsAB	IS1 transposase
604454	604464	yhjX	oxaloate/formate MFS transporter
608600	608590	DppA	Transport of dipeptides across membrane
615364	615415	Hypothetical protein	Pseudogene
633269	633280	ABC transporter	C4-dicarboxylate transport
649418	649363	GadA	glutamate decarboxylase subunit alpha, involved in production of GABA which aid survival in highly acidic conditions
652380	652392	yhjX	multidrug efflux rnd subunit, pyruvate-inducible inner membrane protein, putative transporter
679476	679465	YhiM	acid resistance protein, facilitates survival through digestive tract, required for RpoS, glutamine and lysine-dependent acid resistance pathways
689524	689615		ISAs1 family transposase ISEc1
709043	709032	Hypothetical protein	DUF2500 domain-containing protein
729326	729337	ggt	gamma-glutamyltranspeptidase
743922	743931	glgX	glycogen debranching enzyme
773432	773421	feoA	ferrous iron transporter A
798356	798346	dam	DNA adenine methylase
834946	834935	yhvV	DUF2387 family putative metal-binding protein
850509	850519	exeK	Type II secretion system protein, Uncharacterised
869267	869277	rpoL	50s ribosomal protein L15
873626	873725	rplQ	50s ribosomal protein L17
880881	880871	smf	May help load RecA onto ssDNA
903684	903674	yhdJ	adenine specific DNA methyltransferase
919819	919831	YhdR	unknown function

931773	931764	mdh	malate dehydrogenase
935643	935653	yhcB	Inner membrane protein shown to interact with RodZ in maintaining shape, link between cell membrane biogenesis and peptidoglycan biogenesis
973260	973270	LptC	forms channel through which LPS is transported to periplasm
980535	980620	MurA	UDP-N-acetylglucosamine 1-carboxyvinyltransferase
981066	981165	sfsB	malPQ operon transcriptional activator
983786	983774	Hypothetical protein	EaeM family transporter
985223	985212	ObgE	essential GTPase; exhibits high exchange rate for GTP/GDP; associates with 50S ribosomal subunit; involved in regulation of chromosomal replication
994649	994660	yhbX	eptAB family phosphoethanolamine transferase inner membrane protein, modification of LPS for protection against cations
1004197	1004186	TruB	tRNApseudouridine synthase
1009531	1009544	DeaD	ATP-dependant RNA helicase
1012182	1012281	yhbW	luciferase-like monooxygenase
1013495	1013585	YhbV	U32 family peptidase, PREDICTED COLLAGENASE
1023741	1023753	YraK	fimbrial-like adhesin protein
1024485	1024522	YraJ	fimbrial-like adhesin protein
1025204	1025192	yraH	fimbrial-like adhesin protein
1026297	1026308	FimD	fimbrial subunit usher protein
1028297	1028286		fimbrial-like adhesin protein
1029221	1029210	agal	galactosamine-6-phosphate isomerase
1031230	1031131	agaW	PTS N-acetylgalactosamine transporter subunit IIC
1031931	1032022		IS1 family transposase
1039199	1039208	tagatose-1,6-bisphosphate aldolase	involved in breakdown of N-acetylgalactosamine
1049837	1049870	yhaC	Hypothetical protein, response to DNA damage stimulus
1051545	1051644	TdcR	threonine dehydratase operon control, anaerobic threonine metabolism.
1055875	1055974	TdcD	propionate inase, part of threonine metabolism operon, anaerobic threonine metabolism
1059725	1059824	yphA	Inner membrane protein
1063035	1063024	FadR	fatty acid metabolism regulator
1072593	1072605	ygjR	NAD(P) dependant dehydrogenase
1074313	1074355	ygjP	Putative metal dependant hydrolase, involved with uridine metabolism
1075785	1075837	HigA	mRNA interferase toxin of the HigB-HigA toxin-antitoxin system
1077080	1077091	FadH	fatty acid metabolism
1086160	1086150	EbgA	beta-galactosidase subunit alpha
1092885	1092896	aer	chemotaxis receptor guiding motility
1099202	1099191	DnaG	DNA primase
1107135	1107146	BacA	undecaprenyl diphosphate
1141734	1141724	ygiS	ABC transporter, imports deoxycholate as important protection against bile salts, important in establishment in duodenum
1218570	1218559		phage baseplate protein
1219803	1219794	transcriptional regulator	blastp= yafY
1234353	1234364	maturase	group II intron reversetranscriptase/maturase

1242857	1242869	PduB	microcompartment protein
1278261	1278250		hypothetical protein
1279523	1279622	Hypothetical protein	ornithine decarboxylase
1280686	1280785	NupG	nucleoside permease
1282584	1282683	mltC	Exolytic cleavage of the (1->4)-beta-glycosidic linkage between N-acetylmuramic acid (MurNAc) and N-acetylglucosamine (GlcNAc) residues in peptidoglycan
1291116	1291215	YggU	unknown function
1292470	1292459	YggS	pyrodoxal phosphate (proline biosynthesis)
1292758	1292771	PilT	twitching motility
1296305	1296375	RsmE	Ribosomal small methyltransferase
1297659	1297671	Zinc metalloprotease	predicted transcription elongation factor
1298383	1298324	hypothetical	galactose/proton symporter
1302254	1302264	SpeA	arginine/ornithine decarboxylase
1325982	1325992	scpA	methylmanoyl-CoA MUTASE
1331206	1331305	ygfA	formatetrahydrofolate cyclo-ligase, may function in folate storage
1333894	1333906	pepP	peptidase M24 family
1340681	1340671	gcvP	glycine dehydrogenasen enables utilisation of carbon from glycine under anaerobic conditions, also implicated in acid stress response
1350549	1350643	RecJ	single stranded DNA specific endonuclease
1355446	1355545	UacT	uric acid transporter
1386426	1386436	XdhA	xanthine dehydrogenase molybdenum-binding subunit, involved with purine salvage under anaerobic conditions
1390776	1390765	hypothetical protein	
1421802	1421901	araE	arabinose/proton symporter
1441871	1441882	RecC	
1455590	1455689	tRNA-Met	
1467323	1467312	fucI	L-fucose isomerase, allows use offucose as carbon source in GI tract under anaerobic/microaerophilic conditions
1472057	1472156	sdaA	L-serine dehydratase, generates pyruvate in limiting conditions, requires Fe4S4 cluster indicating role in anaerobic metabolism
1487802	1487812	BarA	sensor histidine kinase, part of novel gluconeogenesis 2 component regulator system involved with carbon utilisation under limiting conditions. Deletion shown to affect fitness in vitro
1488181	1488169	RlmD	SAM-dependent methylation of U1939 in the 23S ribosomal RNA; binds an iron-sulfur cluster [4Fe-4S]
1496709	1496698	ygG	hypothetical protein
1498934	1498947		CRISPR repeat region
1511328	1511340	cysJ	sulphite reductase
1516912	1516900	Cas3	helicase associated cascade protein
1521386	1521399	CasC	ype I-E CRISPR-associated protein_
1539844	1539855	YgbK	SUGAR BINDING PROTEIN
1543023	1543035	apaH	serine/threonine phosphatase
1545169	1545179	MutS	DNA mismatch repair
1562365	1562353	PtsG	PTS arbutin-, cellobiose-, and salicin EIIBC component
1564145	1564157	PurR	transcriptional regulator in response to sugar
1571727	1571717	NorR	transcriptional regulator in response to sugar

1577974	1577984	murein transglycosylase	membrane-bound lytic murein transglycosylase B; catalyzes the cleavage of the glycosidic bonds between N-acetylmuramic acid and N-acetylglucosamine in peptidoglycan
1584439	1584427	tRNA-Arg	
1585215	1585314	YcjU	eta-phosphoglucomutase_
1592971	1592959	AzlC	branched chain amino acid transport permease
1594350	1594339	AraJ	MFS transporter of arabinose
1597642	1597632	glycine betaine transpoter	involved with osmoprotection
1599666	1599654	NrdA	ribonucleotide-diphosphate reductase subunit
1603514	1603503	AlaE	L-alanine exporter
1606957	1606948	GntR	transcriptional regulator of gluconate utilisation operon
1609839	1609826	GabT	4-aminobutyrate aminotransferase
1610581	1610569	gabD	NAD-dependent succinate-semialdehyde dehydrogenase
1611752	1611653	LhgO	L-2-hydroxyglutarate oxidase
1618224	1618234	hypothetical protein	TVSS protein involved with cell wall biogenesis
1622144	1622132	periplasmic autotransporter	outer membrane autotransporter barrel domain-containing protein
1641187	1641197	tyrA	atalyzes the formation of prephenate from chorismate and the formation of 4-hydroxyphenylpyruvate from prephenate in tyrosine biosynthesis
1648556	1648567	ClpB	protease
1654786	1654848	rRNA	5s rRNA
1657451	1657462	pssA	catalyzes de novo synthesis of phosphatidylserine from CDP-diacylglycerol and L-serine which leads eventually to the production of phosphatidylethanolamine; bounds to the ribosome
1661381	1661370	YfiP	unkown function
1663646	1663657	hypothetical	uracil DNA glycosylaase
1665928	1665916	hypothetical	LysR transcriptional regulator
1666676	1666686	SrmB	DNA dependant helicase involved in early stage formation of the 50s ribosome
1674735	1674773	LepA AKA EF4	elongation factor 4
1677974	1677963	Era	GTPase Era, couples activity of ribosome with cytokinesis during cell division
1680248	1680236	AcpS	Catalyzes the formation of holo-ACP, which mediates the essential transfer of acyl fatty acid intermediates during the biosynthesis of fatty acids and lipids
1688001	1688012	PurL	phosphoribosylformylglycinamide synthase
1704001	1704100	MglA	sugar ABC transporter ATPase, binds galactose/methyl-galactose
1704333	1704432	MglC	ABC transporter permease
1705471	1705461	yphC	Zn-dependent NAD(P)-binding oxidoreductase
1706898	1706908	yphB	Putative aldose-1-epimerase, function unknown; mutarotase family
1720555	1720654	iscU	Iron sulfur cluster scaffold protein
1720929	1720940	iscA	iron sulfur cluster accessory assembly protein
1729019	1729118	yfhM	bacterial alpha2-macroglobulin colonization factor ECAM; anti-host protease defense factor; periplasmic inner membrane-anchored lipoprotein
1744762	1744751	Der	ribosome biogenesis GTPase
1748634	1748623	guaA	GMP synthase, hydrolyses glutamine for purine biosynthesis
1757840	1757851	purN	phosphoribosylglycinamide formyltransferase
1765404	1765503	yfgO	putative UPF0118 family inner membrane permease
1766088	1766128	focB	formate transporter

1768487	1768476	HYDROGENASE-4 TRANSCRIPTIONAL ACTIVATOR	hydrogenase transcriptional activator
1769160	1769150	hyfJ	hydrogenase-4 component J
1769277	1769266	hyfI	hydrogenase component I
1770160	1770172	hyfH	hydrogenase component H
1772288	1772299	hyfF	hydrogenase component F
1775005	1775014	hyfD	hydrogenase component D
1783677	1783689	purC	phosphoribosylaminoimidazolesuccinocarboxamide synthase involved with purine biosynthesis
1784266	1784253	ypfJ	zinc metalloprotease
1792352	1792363	AcrD	AcrD family multidrug efflux RND transporter permease subunit, aminoglycoside/multidrug efflux system
1793463	1793473	NarQ	nitrate/nitrite two-component system sensor histidine kinase
1802534	1802524	maeB	malic enzyme: putative oxidoreductase/phosphotransacetylase
1806287	1806299	EutQ	ethanolamine usage protein
1809330	1809321	EutE	ethanolamine usage protein, catalyzes the formation of acetyl-CoA from acetaldehyde
1810543	1810531	EutJ	ethanolamine usage protein, possible chaperonin; may be involved in the assembly of the carboxysome
1813981	1813992	EutA	ethanolamine usage protein, reactivating factor for ethanolamine ammonia lyase EutBC
1816165	1816154	EutB	heavy chain of EutBC ammonia lyase, catalyzes the formation of acetaldehyde and ammonia from ethanolamine
1825403	1825412	PBP4B	DD-carboxypeptidase, cross-links pentapeptide in peptidoglycan
1826076	1826087	murP	PTS N-acetylmuramic acid transporter subunits IIBC
1829025	1829014	murR	repressor for murPQ; N-acetylmuramic acid 6-phosphate inducible
1833085	1833074	cysW	sulfate transport system permease W protein; membrane component of ABC superfamily
1834019	1834029	CysA	sulfate/thiosulfate import ATP-binding protein
1834777	1834788	cysM	cysteine synthase B, O-acetylserine sulfhydrylase
1835968	1835956	yfeZ	WGR domain containing protein
1842328	1842337	CysZ	sulfate transporter
1849030	1849101	xapA	xanthosine phosphorase, purine nucleoside phosphorylase 2; nicotinamide 1-beta-D-riboside synthase
1852442	1852518	tRNA-lys	
1862464	1862454	ypeC	putative periplasmic protein, function unknown
1867795	1867894	ypdF	Xaa-Pro aminopeptidase; limited methionine aminopeptidase
1868575	1868565	ypdE	aminopeptidase
1874335	1874408	ypdA	SENSOR HISTIDINE KINASE, sensor kinase regulating yhjX; pyruvate-responsive YpdAB two-component system
1876160	1876149	AlaC	glutamate--pyruvate aminotransferase, alanine biosynthesis
1880715	1880725	frc	formyl-CoA--oxalate CoA-transferase, functions in low pH adaption
1893006	1892994	EmrY	multidrug MFS transport; with EmrA is involved in resistance to hydrophobic antibiotics such as nalidixic acid
1895101	1895090	LacI family repressor	sucrose operon repressor
1912552	1912564	SteB	fimbrial protein
1921664	1921676	yfcA	putative TauE family membrane protein
1927220	1927319	yfcJ	putative arabinose efflux transporter
1942887	1942896	HisP	histidine transport ATP-binding protein_
1963690	1963701	NuoCD	NDH-1 shuttles electrons from NADH, via FMN and iron-sulfur (Fe-S) centers, to quinones in the respiratory chain; subunits NuoCD, E, F, and G constitute the peripheral sector of the complex

1964758	1964746	NuoE	
1965391	1965380	NuoF	part of NADH-ubiquinone oxidoreductase complex I; shuttles electrons from NADH, via FMN and iron-sulfur (Fe-S) centers, to quinones in the respiratory chain
1968280	1968269	NuoG	Catalyzes the transfer of electrons from NADH to quinone
1976475	1976542	NuoN	ADH-quinone oxidoreductase
1977018	1977027	yfbP	unknown function
1988569	1988668	menD	2-succinyl-5-enolpyruvyl-6-hydroxy-3-cyclohexene-1-carboxylate synthase; SEPHCHC synthase, forms
1994301	1994290	arnT	4-amino-4-deoxy-L-arabinose lipid A transferase (Lipid A modification)
281746	281736	UbiE	Menaquinone and electron carrier during anaerobic respiration
2001461	2001472	ais	putative LPS core heptose(II)-phosphate phosphatase, induced by aluminium
2002224	2002323	fliC	flagellar filament structural protein
2005414	2005425	rhmD	L-rhamnoate hehydratase
2006720	2006731	rhmT	putative L-rhamnonate transporter
2007962	2008061	rhmA	2-keto-3-deoxy-L-rhamnonate aldolase
2009354	2009453	yfaD	transposase
2012145	2012157	glpB	anaerobic glycerol-3-phosphate dehydrogenase subunit B MEMBRANE ANCHOR UNIT
2013378	2013366	glpA	anaerobic glycerol-3-phosphate dehydrogenase subunit A
2016532	2016521	glpQ	periplasmic glycerophosphodiester phosphodiesterase
2021014	2021028	nrdA	ibonucleoside-diphosphate reductase 1 subunit alpha
2030899	2030888	Pseudogene (frameshift)	DUF2138 domain-containing protein
2190346	2190336	gatZ	tagatose-biphosphate aldolase
2212605	2212617	Pseudogene (frameshift)	hypothetical protein
2349033	2349022	ybtQ	permease and ATP-binding protein of yersiniabactin-iron ABC transporter YbtQ, MULTIDRUG RESISTANCE TRANSPORTER
2548942	2548953	DNA topoisomerase III	decatenates replicating daughter chromosomes
2736753	2736763	ynfE	putative selenate reductase, periplasmic, produces terminal reductase during anaerobic growth
2816894	2816993	ddpB	D,D-dipeptide ABC transporter permease
2950795	2950784	dpbA	ATP-dependant RNA helicase specific to 23s rRNA
3072745	3072843	tRNA-tyr (i)	
3072803	3072902	rtt sRNA	processed from tyrT transcript
3073094	3073010	RtT sRNA ii	processed from tyrT transcript
3073214	3073186	RtT sRNA iii	
3086010	3086000	ychO	Invasin like protein
3090028	3090116	rdIB	sRNA antisense regulator affects LdrB translation; proposed addiction module in LDR-B repeat, with toxic peptide LdrB
3090563	3090651	rdIA	sRNA antisense regulator affects LdrA translation; proposed addiction module in LDR-A repeat, with toxic peptide LdrA
3193480	3193505	antitermination protein	
3232604	3232703	ycfH	metallo-dependant hydrolase
3291106	3291095	YcdY	chaperone protein, increase activity of YcdX, involved with motility
3461514	3461525	cydC	glutathione/cysteine ABC transporter export permease/ATPase
3469028	3469019	MacB	macrolide ABC transporter permease/ATP-binding protein MacB
3541334	3541324	dps	Fe-binding and storage protein; stress-inducible DNA-binding protein, global regulator, starvation conditions
3637605	3637616	ybgD	fimbrial-like adhesin protein

3752193	3752292	ybdN	PAPS reductase-like domain protein, actig on sulfur groups
3782649	3782738		IS6 family transposase IS15DIV
3951635	3951624		Lon protease
3958855	3958844	AmpG	muropeptide MFS transporter
4095735	4095745	pseudo gene	adhesin eaeH
4201880	4201864		ISAs1 family transposase (PSEUDO)
4202015	4201916		ISAs1 family transposase (Pseudo)
4257919	4258018	TilS	RNA lysidine(34) synthetase_
4297788	4297887	FhuB	e3+-hydroxamate ABC transporter permease
4319350	4319361	yadK	fimbrial like adhesin protein
4376042	4376031	murE	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase, formation of pentapeptide on peptidoglycan precursor
4410765	4410777	RluA	bifunctional tRNA pseudouridine
4422183	4422282	kefC	glutathione-regulated potassium-efflux system protein, electrophile protection. Protection from methyglyoxal produced as a result of growth on poor carbon source
4570111	4570101		IS3 family transposase
4665575	4665586		30S ribosomal protein S6
4703032	4703131	tRNA-gly	
4735556	4735566		IGR between hypothetical protein and IS1 family transposase
4760043	4760033	adiC	arginine:agmatine antiporter, protection from acidic conditions

10.0 – Bibliography

- Aarestrup, F.M., 1995. Occurrence of glycopeptide resistance among *Enterococcus faecium* isolates from conventional and ecological poultry farms. *Microbial Drug Resistance*, 1(3), pp.255-257.
- Aarestrup, F.M., 2000. Characterization of glycopeptide-resistant *Enterococcus faecium* (GRE) from broilers and pigs in Denmark: genetic evidence that persistence of GRE in pig herds is associated with coselection by resistance to macrolides. *Journal of Clinical Microbiology*, 38(7), pp.2774-2777.
- Abraham, E.P. and Chain, E., 1940. An enzyme from bacteria able to destroy penicillin. *Nature*, 146, p.837.
- Alhashash F, Weston V, Diggle M, McNally A. Multidrug-Resistant *Escherichia coli* Bacteremia. *Emerg Infect Dis* 2013; 19: 1699–701.
- Allen, H.K., Levine, U.Y., Looft, T., Bandrick, M. and Casey, T.A., 2013. Treatment, promotion, commotion: antibiotic alternatives in food-producing animals. *Trends in microbiology*, 21(3), pp.114-119.
- Aminov, R.I., 2010. A brief history of the antibiotic era: lessons learned and challenges for the future. *Frontiers in microbiology*, 1.
- Anadón, A., 2006. WS14 The EU ban of antibiotics as feed additives 2006. alternatives and consumer safety. *Journal of Veterinary Pharmacology and Therapeutics*, 29(s1), pp.41-44.
- Andersen K.R., Leksa N.C. and Schwartz T.U., 2013. Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins*. 81:1857-61.
- Andersson, D.I. and Hughes, D., 2010. Antibiotic resistance and its cost: is it possible to reverse resistance?. *Nature reviews. Microbiology*, 8(4), p.260.
- Arai, H., 2011. Regulation and function of versatile aerobic and anaerobic respiratory metabolism in *Pseudomonas aeruginosa*. *Frontiers in microbiology*, 2.
- Arai, H., Kawakami, T., Osamura, T., Hirai, T., Sakai, Y. and Ishii, M., 2014. Enzymatic characterization and in vivo function of five terminal oxidases in *Pseudomonas aeruginosa*. *Journal of bacteriology*, 196(24), pp.4206-4215.
- Baker-Austin, C., Wright, M.S., Stepanauskas, R. and McArthur, J.V., 2006. Co-selection of antibiotic and metal resistance. *Trends in microbiology*, 14(4), pp.176-182.
- Bednorz, C., Oelgeschläger, K., Kinnemann, B., Hartmann, S., Neumann, K., Pieper, R., Bethe, A., Semmler, T., Tedin, K., Schierack, P. and Wieler, L.H., 2013. The broader context of antibiotic resistance: zinc feed supplementation of piglets increases the proportion of multi-resistant *Escherichia coli* in vivo. *International Journal of Medical Microbiology*, 303(6), pp.396-403.

Ben Zakour N.L., Alsheikh-Hussain A.S., Ashcroft M.M., Nhu N.T.K., Roberts L.W., Stanton-Cook M., 2016. Sequential acquisition of virulence and fluoroquinolone resistance has shaped the evolution of *Escherichia coli* ST131. *bioRxiv*.

Bengtsson, B. and Wierup, M., 2006. Antimicrobial resistance in Scandinavia after a ban of antimicrobial growth promoters. *Animal biotechnology*, 17(2), pp.147-156.

Bennett PM (2005). Genome plasticity. In: Woodford N, Johnson A (eds). *Methods in Molecular Biology, Vol 266, Genomics, Proteomics and Clinical Bacteriology*. Humana Press Inc.: Totowa, NJ, pp 71–113.

Bennett, P.M., 2008. Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *British journal of pharmacology*, 153(S1).

Bertin, Y., Girardeau, J.P., Chaucheyras-Durand, F., Lyan, B., Pujos-Guillot, E., Harel, J. and Martin, C., 2011. Enterohaemorrhagic *Escherichia coli* gains a competitive advantage by using ethanolamine as a nitrogen source in the bovine intestinal content. *Environmental microbiology*, 13(2), pp.365-377.

Blumer C, and Haas D. Iron regulation of the *hcnABC* genes encoding hydrogen cyanide synthase depends on the anaerobic regulator ANR rather than on the global activator GacA in *Pseudomonas fluorescens* CHA0. *Microbio*. 2000; **146**: 2417–2424.

Bonten, M.J., Willems, R. and Weinstein, R.A., 2001. Vancomycin-resistant enterococci: why are they here, and where do they come from?. *The Lancet infectious diseases*, 1(5), pp.314-325.

Borriello, G., Werner, E., Roe, F., Kim, A.M., Ehrlich, G.D. and Stewart, P.S., 2004. Oxygen limitation contributes to antibiotic tolerance of *Pseudomonas aeruginosa* in biofilms. *Antimicrobial agents and chemotherapy*, 48(7), pp.2659-2664.

Boucher, Y., Labbate, M., Koenig, J.E. and Stokes, H.W., 2007. Integrons: mobilizable platforms that promote genetic diversity in bacteria. *Trends in microbiology*, 15(7), pp.301-309.

Broderick, K.E., Chan, A., Balasubramanian, M., Feala, J., Reed, S.L., Panda, M., Sharma, V.S., Pilz, R.B., Bigby, T.D. and Boss, G.R., 2008. Cyanide produced by human isolates of *Pseudomonas aeruginosa* contributes to lethality in *Drosophila melanogaster*. *The Journal of infectious diseases*, 197(3), pp.457-464.

Brynildsrud, O., Bohlin, J., Scheffer, L. and Eldholm, V., 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome biology*, 17(1), p.238.

Butaye, P., Michael, G.B., Schwarz, S., Barrett, T.J., Brisabois, A. and White, D.G., 2006. The clonal spread of multidrug-resistant non-typhi *Salmonella* serotypes. *Microbes and Infection*, 8(7), pp.1891-1897.

c Stokes, H.T. and Hall, R.M., 1989. A novel family of potentially mobile DNA elements encoding site-specific gene-integration functions: integrons. *Molecular microbiology*, 3(12), pp.1669-1683.

- Cantón, R., González-Alba, J.M. and Galán, J.C., 2012. CTX-M enzymes: origin and diffusion. *Frontiers in microbiology*, 3.
- Carattoli, A., 2009. Resistance plasmid families in Enterobacteriaceae. *Antimicrobial agents and chemotherapy*, 53(6), pp.2227-2238.
- Chait, R., Palmer, A.C., Yelin, I. and Kishony, R., 2016. Pervasive selection for and against antibiotic resistance in inhomogeneous multistress environments. *Nature communications*, 7.
- Chen, I.L., Lee, C.H., Su, L.H., Tang, Y.F., Chang, S.J. and Liu, J.W., 2013. Antibiotic consumption and healthcare-associated infections caused by multidrug-resistant gram-negative bacilli at a large medical center in Taiwan from 2002 to 2009: implicating the importance of antibiotic stewardship. *PloS one*, 8(5), p.e65621.
- Cogliani, C., Goossens, H. and Greko, C., 2011. Restricting antimicrobial use in food animals: lessons from Europe. *Microbe*, 6(6), p.274.
- Colburn-Clifford, J. and Allen, C., 2010. A cbb3-type cytochrome C oxidase contributes to *Ralstonia solanacearum* R3bv2 growth in microaerobic environments and to bacterial wilt disease development in tomato. *Molecular plant-microbe interactions*, 23(8), pp.1042-1052.
- Cooper, M., Tavankar, G.R. and Williams, H.D., 2003. Regulation of expression of the cyanide-insensitive terminal oxidase in *Pseudomonas aeruginosa*. *Microbiology*, 149(5), pp.1275-1284.
- Corbella, X., Montero, A., Pujol, M., Domínguez, M.A., Ayats, J., Argerich, M.J., Garrigosa, F., Ariza, J. and Gudiol, F., 2000. Emergence and rapid spread of carbapenem resistance during a large and sustained hospital outbreak of multiresistant *Acinetobacter baumannii*. *Journal of clinical microbiology*, 38(11), pp.4086-4095.
- Crawford, R.W., Wangdi, T., Spees, A.M., Xavier, M.N., Tsois, R.M. and Bäumlér, A.J., 2013. Loss of very-long O-antigen chains optimizes capsule-mediated immune evasion by *Salmonella enterica* serovar Typhi. *MBio*, 4(4), pp.e00232-13.
- Croxall G, Hale J, Weston V, et al. Molecular epidemiology of extraintestinal pathogenic *Escherichia coli* isolates from a regional cohort of elderly patients highlights the prevalence of ST131 strains with increased antimicrobial resistance in both community and hospital care settings. *J Antimicrob Chemother* 2011; 66: 2501–8.
- Cunningham, L., Pitt, M. and Williams, H.D., 1997. The *cioAB* genes from *Pseudomonas aeruginosa* code for a novel cyanide-insensitive terminal oxidase related to the cytochrome bd quinol oxidases. *Molecular microbiology*, 24(3), pp.579-591.
- Cuthbertson, L., Kimber, M.S. and Whitfield, C., 2007. Substrate binding by a bacterial ABC transporter involved in polysaccharide export. *Proceedings of the National Academy of Sciences*, 104(49), pp.19529-19534.
- D'costa, V.M., McGrann, K.M., Hughes, D.W. and Wright, G.D., 2006. Sampling the antibiotic resistome. *Science*, 311(5759), pp.374-377.

- D'mello, R., Hill, S. and Poole, R.K., 1996. The cytochrome bd quinol oxidase in *Escherichia coli* has an extremely high oxygen affinity and two oxygen-binding haems: implications for regulation of activity in vivo by oxygen inhibition. *Microbiology*, 142(4), pp.755-763.
- Davies, J. and Davies, D., 2010. Origins and evolution of antibiotic resistance. *Microbiology and molecular biology reviews*, 74(3), pp.417-433.
- de Kraker MEA, Jarlier V, Monen JCM, Heuer OE, van de Sande N, Grundmann H. The changing epidemiology of bacteraemias in Europe: trends from the European Antimicrobial Resistance Surveillance System. *Clin Microbiol Infect* 2013; 19: 860–8 diseases, 12(6), p.1011.
- Djordjevic, S.P., Stokes, H.W. and Chowdhury, P.R., 2013. Mobile elements, zoonotic pathogens and commensal bacteria: conduits for the delivery of resistance genes into humans, production animals and soil microbiota. *Frontiers in microbiology*, 4.
- Dogan, B., Suzuki, H., Herlekar, D., Sartor, R.B., Campbell, B.J., Roberts, C.L., Stewart, K., Scherl, E.J., Araz, Y., Bitar, P.P. and Lefébure, T., 2014. Inflammation-associated adherent-invasive *Escherichia coli* are enriched in pathways for use of propanediol and iron and M-cell translocation. *Inflammatory bowel diseases*, 20(11), pp.1919-1932.
- Dubern, J.F., Cigana, C., De Simone, M., Lazenby, J., Juhas, M., Schwager, S., Bianconi, I., Döring, G., Eberl, L., Williams, P. and Bragonzi, A., 2015. Integrated whole-genome screening for *Pseudomonas aeruginosa* virulence genes using multiple disease models reveals that pathogenicity is host specific. *Environmental microbiology*, 17(11), pp.4379-4393.
- Ewers, C., Bethe, A., Stamm, I., Grobbel, M., Kopp, P.A., Guerra, B., Stubbe, M., Doi, Y., Zong, Z., Kola, A. and Schaufler, K., 2014. CTX-M-15-D-ST648 *Escherichia coli* from companion animals and horses: another pandemic clone combining multiresistance and extraintestinal virulence?. *Journal of Antimicrobial Chemotherapy*, p.dkt516
- Fang, L., Li, X., Li, L., Li, S., Liao, X., Sun, J. and Liu, Y., 2016. Co-spread of metal and antibiotic resistance within ST3-IncHI2 plasmids from *E. coli* isolates of food-producing animals. *Scientific reports*, 6.
- Faria, N.R., Quick, J., Claro, I.M., Thézé, J., de Jesus, J.G., Giovanetti, M., Kraemer, M.U.G., Hill, S.C., Black, A., da Costa, A.C. and Franco, L.C., 2017. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. *Nature*, 546(7658), pp.406-410.
- Fernebro, J. Fighting bacterial infections-future treatment options. *Drug Resist Update* 2011; 14: 125–139.
- Frangipani, E., Pérez-Martínez, I., Williams, H.D., Cherbuin, G. and Haas, D., 2014. A novel cyanide-inducible gene cluster helps protect *Pseudomonas aeruginosa* from cyanide. *Environmental microbiology reports*, 6(1), pp.28-34.
- Gallagher L.A., and Manoil C. *Pseudomonas aeruginosa* PAO1 kills *Caenorhabditis elegans* by cyanide poisoning. *J Bacteriol* 2001; 183: 6207–6214
- Garsin, D.A., 2010. Ethanolamine utilization in bacterial pathogens: roles and regulation. *Nature Reviews Microbiology*, 8(4), pp.290-295.

- Gelbard, H., Miller-Petrie, M., Pant, S., Gandra, S., Levison, J., Barter, D., White, A and Laxminirayan, R. Center for Disease Dynamics, Economics & Policy. 2015. State of the World's Antibiotics, 2015. CDDEP: Washington, D.C.
- Hadfield, J., Croucher, N.J., Goater, R.J., Abudahab, K., Aanensen, D.M. and Harris, S.R., 2017. Phandango: an interactive viewer for bacterial population genomics. *bioRxiv*, p.119545.
- Harrison E, Guymer D, Spiers AJ, Paterson S, Brockhurst MA. Parallel Compensatory Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum. *Curr Biol*. 2015; 25: 2034–9. pmid:26190075
- Hirai, T., Osamura, T., Ishii, M. and Arai, H., 2016. Expression of multiple cbb3 cytochrome c oxidase isoforms by combinations of multiple isosubunits in *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences*, p.201613308.
- Hoffman, S.J., Caleo, G.M., Daulaire, N., Elbe, S., Matsoso, P., Mossialos, E., Rizvi, Z. and Røttingen, J.A., 2015. Strategies for achieving global collective action on antimicrobial resistance. *Bulletin of the World Health Organization*, 93(12), pp.867-876.
- Hoyle BD, Costerton WJ. Bacterial resistance to antibiotics: the role of biofilms. *Prog. Drug Res*. 1993; **37**: 91–105
- Iguchi, A., Iyoda, S., Kikuchi, T., Ogura, Y., Katsura, K., Ohnishi, M., Hayashi, T. and Thomson, N.R., 2015. A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA research*, 22(1), pp.101-107.
- Ingle, D.J., Valcanis, M., Kuzevski, A., Tauschek, M., Inouye, M., Stinear, T., Levine, M.M., Robins-Browne, R.M. and Holt, K.E., 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. *Microbial Genomics*, 2(7).
- Joensen, K. G., A. M. Tetzschner, A. Iguchi, F. M. Aarestrup, and F. Scheutz. 2015. Rapid and easy in silico serotyping of *Escherichia coli* using whole genome sequencing (WGS) data. *J.Clin.Microbiol*. 53(8):2410-2426
- Kaper, J.B., Nataro, J.P. and Mobley, H.L., 2004. Pathogenic *Escherichia coli*. *Nature Reviews Microbiology*, 2(2), pp.123-140.
- Karim, A., Poirel, L., Nagarajan, S. and Nordmann, P., 2001. Plasmid-mediated extended-spectrum β -lactamase (CTX-M-3 like) from India and gene association with insertion sequence IS Ecp1. *FEMS microbiology letters*, 201(2), pp.237-241.
- Keith W Ridge, Kieran Hand, Mike Sharland, Ibrahim Abubakar, David M Livermore 2011. Annual Chief Medical Officers Report. Chapter 5 pp76-84
- Kim, S.H., Park, S.Y., Heo, Y.J. and Cho, Y.H., 2008. *Drosophila melanogaster*-based screening for multihost virulence factors of *Pseudomonas aeruginosa* PA14 and identification of a virulence-attenuating factor, HudA. *Infection and immunity*, 76(9), pp.4152-4162.
- Kingsley, R.A., Msefula, C.L., Thomson, N.R., Kariuki, S., Holt, K.E., Gordon, M.A., Harris, D., Clarke, L., Whitehead, S., Sangal, V. and Marsh, K., 2009. Epidemic multiple drug resistant

Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome research*, 19(12), pp.2279-2287.

Klare, I., Konstabel, C., Badstübner, D., Werner, G., and Witte, W. (2003) Occurrence and spread of antibiotic resistances in *Enterococcus faecium*. *Int J Food Microbiol* 88: 269–290.

Kohler C-D, Dobrindt U. What defines extraintestinal pathogenic *Escherichia coli*? *Int J Med Microbiol* 2011; 301: 642–7.

Kumarasamy, K.K., Toleman, M.A., Walsh, T.R., Bagaria, J., Butt, F., Balakrishnan, R., Chaudhary, U., Doumith, M., Giske, C.G., Irfan, S. and Krishnan, P., 2010. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *The Lancet infectious diseases*, 10(9), pp.597-602.

Laville, J., Blumer, C., Von Schroetter, C., Gaia, V., Défago, G., Keel, C. and Haas, D., 1998. Characterization of the *hcnABC* gene cluster encoding hydrogen cyanide synthase and anaerobic regulation by ANR in the strictly aerobic biocontrol agent *Pseudomonas fluorescens* CHA0. *Journal of bacteriology*, 180(12), pp.3187-3196.

Leavitt, A., Navon-Venezia, S., Chmelnitsky, I., Schwaber, M.J. and Carmeli, Y., 2007. Emergence of KPC-2 and KPC-3 in carbapenem-resistant *Klebsiella pneumoniae* strains in an Israeli hospital. *Antimicrobial agents and chemotherapy*, 51(8), pp.3026-3029.

Lees, J.A., Vehkala, M., Välimäki, N., Harris, S.R., Chewapreecha, C., Croucher, N.J., Marttinen, P., Davies, M.R., Steer, A.C., Tong, S.Y. and Honkela, A., 2016. Sequence element enrichment analysis to determine the genetic basis of bacterial phenotypes. *Nature Communications*, 7, p.12797.

Lester, C.H., Frimodt-Møller, N., Sørensen, T.L., Monnet, D.L. and Hammerum, A.M., 2006. In vivo transfer of the *vanA* resistance gene from an *Enterococcus faecium* isolate of animal origin to an *E. faecium* isolate of human origin in the intestines of human volunteers. *Antimicrobial agents and chemotherapy*, 50(2), pp.596-599.

Liu, C., Bayer, A., Cosgrove, S.E., Daum, R.S., Fridkin, S.K., Gorwitz, R.J., Kaplan, S.L., Karchmer, A.W., Levine, D.P., Murray, B.E. and Rybak, M.J., 2011. Clinical practice guidelines by the Infectious Diseases Society of America for the treatment of methicillin-resistant *Staphylococcus aureus* infections in adults and children. *Clinical infectious diseases*, 52(3), pp.e18-e55.

Liu, Y.Y., Wang, Y., Walsh, T.R., Yi, L.X., Zhang, R., Spencer, J., Doi, Y., Tian, G., Dong, B., Huang, X. and Yu, L.F., 2016. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: a microbiological and molecular biological study. *The Lancet infectious diseases*, 16(2), pp.161-168.

Livermore DM, Hawkey PM. CTX-M: changing the face of ESBLs in the UK. *J Antimicrob Chemother* 2005; 56: 451–4

Lyczak J., Cannon C., and Pier G. Establishment of *Pseudomonas aeruginosa* infection: lessons from a versatile opportunist. *Microbes Infect.* 2000; 2: 1051–1060.

Marshall, B.M. and Levy, S.B., 2011. Food animals and antimicrobials: impacts on human health. *Clinical microbiology reviews*, 24(4), pp.718-733.

Mathers AJ, Peirano G, Pitout JDD. The role of epidemic resistance plasmids and international high-risk clones in the spread of multidrug-resistant Enterobacteriaceae. *Clin Microbiol Rev* 2015; 28: 565–91.

Maurer, L.M., Yohannes, E., Bondurant, S.S., Radmacher, M. and Slonczewski, J.L., 2005. pH regulates genes for flagellar motility, catabolism, and oxidative stress in *Escherichia coli* K-12. *Journal of bacteriology*, 187(1), pp.304-319.

May, J.F. and Groisman, E.A., 2013. Conflicting roles for a cell surface modification in *Salmonella*. *Molecular microbiology*, 88(5), pp.970-983.

McAdams, D., 2017. Resistance diagnosis and the changing epidemiology of antibiotic resistance. *Annals of the New York Academy of Sciences*, 1388(1), pp.5-17.

McEwen, S.A. and Fedorka-Cray, P.J., 2002. Antimicrobial use and resistance in animals. *Clinical Infectious Diseases*, 34(Supplement_3), pp.S93-S106.

McGann, P., Snesrud, E., Maybank, R., Corey, B., Ong, A.C., Clifford, R., Hinkle, M., Whitman, T., Lesho, E. and Schaecher, K.E., 2016. *Escherichia coli* harboring *mcr-1* and *blaCTX-M* on a novel IncF plasmid: first report of *mcr-1* in the United States. *Antimicrobial agents and chemotherapy*, 60(7), pp.4420-4421.

McNally, A., Oren, Y., Kelly, D., Pascoe, B., Dunn, S., Sreecharan, T., Vehkala, M., Välimäki, N., Prentice, M.B., Ashour, A. and Avram, O., 2016. Combined analysis of variation in core, accessory and regulatory genome regions provides a super-resolution view into the evolution of bacterial populations. *PLoS genetics*, 12(9), p.e1006280.

Mediavilla, J.R., Patrawalla, A., Chen, L., Chavda, K.D., Mathema, B., Vinnard, C., Dever, L.L. and Kreiswirth, B.N., 2016. Colistin-and carbapenem-resistant *Escherichia coli* harboring *mcr-1* and *blaNDM-5*, causing a complicated urinary tract infection in a patient from the United States. *MBio*, 7(4), pp.e01191-16.

Melnyk, A.H., Wong, A. and Kassen, R., 2015. The fitness costs of antibiotic resistance mutations. *Evolutionary applications*, 8(3), pp.273-283.

Méric, G., Yahara, K., Mageiros, L., Pascoe, B., Maiden, M.C., Jolley, K.A. and Sheppard, S.K., 2014. A reference pan-genome approach to comparative bacterial genomics: identification of novel epidemiological markers in pathogenic *Campylobacter*. *PloS one*, 9(3), p.e92798.

Moradigaravand, D., Boinett, C.J., Martin, V., Peacock, S.J. and Parkhill, J., 2016. Recent independent emergence of multiple multidrug-resistant *Serratia marcescens* clones within the United Kingdom and Ireland. *Genome research*, 26(8), pp.1101-1109.

O'Neill, J. (2015). Antimicrobials in agriculture and the environment: Reducing unnecessary waste. Amr-review.org Available at <https://amr-review.org/sites/default/files/Antimicrobials%20in%20agriculture%20and%20the%20enviro>

[nment%20-%20Reducing%20unnecessary%20use%20and%20waste.pdf](#). Date accessed: 15/08/17

O'Neill, J. (2015). Tackling a global health crisis: Initial steps. Amr-review.org. Available at https://amr-review.org/sites/default/files/Health%20infrastructure%20and%20surveillance%20final%20version_LR_NO%20CROPS.pdf. Date accessed: 14/08/2017

O'Neill, J. (2016) Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations. Amr-review.org. Available at https://amr-review.org/sites/default/files/AMR%20Review%20Paper%20-%20Tackling%20a%20crisis%20for%20the%20health%20and%20wealth%20of%20nations_1.pdf Date accessed: 12/08/2017

O'Neill, J. (2016). Tackling drug resistant infections globally: final review and recommendations. Amr-review.org. Available at https://amr-review.org/sites/default/files/160525_Final%20paper_with%20cover.pdf. Date accessed: 14/08/2017

Österberg, J., Wingstrand, A., Jensen, A.N., Kerouanton, A., Cibir, V., Barco, L., Denis, M., Aabo, S. and Bengtsson, B., 2016. Antibiotic resistance in *Escherichia coli* from pigs in organic and conventional farming in four European countries. *PloS one*, 11(6), p.e0157049.

Overdeest, I., Haverkate, M., Veenemans, J., Hendriks, Y., Verhulst, C., Mulders, A., Couprie, W., Bootsma, M., Johnson, J. and Kluytmans, J., 2016. Prolonged colonisation with *Escherichia coli* O25: ST131 versus other extended-spectrum beta-lactamase-producing *E. coli* in a long-term care facility with high endemic level of rectal colonisation, the Netherlands, 2013 to 2014. *Eurosurveillance*, 21(42).

Piddock, L.J., 2012. The crisis of no new antibiotics—what is the way forward?. *The Lancet infectious diseases*, 12(3), pp.249-253.

Poole K., *Pseudomonas aeruginosa*: resistance to the max. *Front Microbiol.* 2011; 2:65

Potvin, E., Lehoux, D.E., Kukavica-Ibrulj, I., Richard, K.L., Sanschagrin, F., Lau, G.W. and Levesque, R.C., 2003. In vivo functional genomics of *Pseudomonas aeruginosa* for high-throughput screening of new virulence factors and antibacterial targets. *Environmental microbiology*, 5(12), pp.1294-1308.

Price, L.B., Johnson, J.R., Aziz, M., Clabots, C., Johnston, B., Tchesnokova, V., Nordstrom, L., Billig, M., Chattopadhyay, S., Stegger, M. and Andersen, P.S., 2013. The epidemic of extended-spectrum- β -lactamase-producing *Escherichia coli* ST131 is driven by a single highly pathogenic subclone, H30-Rx. *MBio*, 4(6), pp.e00377-13.

Qing, G., Li-Chung, M., Khorchid, A., Swapna, G.V.T., Mal, T.K., Takayama, M.M., Xia, B., Phadtare, S., Ke, H., Acton, T. and Montelione, G.T., 2004. Cold-shock induced high-yield protein production in *Escherichia coli*. *Nature biotechnology*, 22(7), p.877.

Quick, J., Grubaugh, N.D., Pullan, S.T., Claro, I.M., Smith, A.D., Gangavarapu, K., Oliveira, G., Robles-Sikisaka, R., Rogers, T.F., Beutler, N.A. and Burton, D.R., 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *bioRxiv*, p.098913.

Quick, J., Loman, N.J., Duraffour, S., Simpson, J.T., Severi, E., Cowley, L., Bore, J.A., Koundouno, R., Dudas, G., Mikhail, A. and Ouédraogo, N., 2016. Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589), p.228.

Rådström, P., Sköld, O., Swedberg, G., Flensburg, J., Roy, P.H. and Sundström, L., 1994. Transposon Tn5090 of plasmid R751, which carries an integron, is related to Tn7, Mu, and the retroelements. *Journal of bacteriology*, 176(11), pp.3257-3268.

Ramachandran, G., Perkins, D.J., Schmidlein, P.J., Tulapurkar, M.E. and Tennant, S.M., 2015. Invasive *Salmonella* Typhimurium ST313 with naturally attenuated flagellin elicits reduced inflammation and replicates within macrophages. *PLoS neglected tropical diseases*, 9(1), p.e3394.

Rao, A.N., Barlow, M., Clark, L.A., Boring III, J.R., Tenover, F.C. and McGowan Jr, J.E., 2006. Class 1 integrons in resistant *Escherichia coli* and *Klebsiella spp.*, US hospitals. *Emerging infectious*

Rodríguez, M.M., Power, P., Radice, M., Vay, C., Famiglietti, A., Galleni, M., Ayala, J.A. and Gutkind, G., 2004. Chromosome-encoded CTX-M-3 from *Kluyvera ascorbata*: a possible origin of plasmid-borne CTX-M-1-derived cefotaximases. *Antimicrobial agents and chemotherapy*, 48(12), pp.4895-4897.

Ryall, B., Davies, J.C., Wilson, R., Shoemark, A. and Williams, H.D., 2008. *Pseudomonas aeruginosa*, cyanide accumulation and lung function in CF and non-CF bronchiectasis patients. *European Respiratory Journal*, 32(3), pp.740-747.

Seiler, C. and Berendonk, T.U., 2012. Heavy metal driven co-selection of antibiotic resistance in soil and water bodies impacted by agriculture and aquaculture. *Frontiers in microbiology*, 3.

Smith D.B and Johnson K.S. Single- step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. 1998; *Gene* 67: 31–40.

Spellberg, B. and Gilbert, D.N., 2014. The future of antibiotics and resistance: a tribute to a career of leadership by John Bartlett. *Clinical infectious diseases*, 59(suppl_2), pp.S71-S75.

Stoesser, N., Sheppard, A.E., Pankhurst, L., De Maio, N., Moore, C.E., Sebra, R., Turner, P., Anson, L.W., Kasarskis, A., Batty, E.M. and Kos, V., 2016. Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131. *MBio*, 7(2), pp.e02162-15.

Stokes, H.W., Nesbø, C.L., Holley, M., Bahl, M.I., Gillings, M.R. and Boucher, Y., 2006. Class 1 integrons potentially predating the association with Tn402-like transposition genes are present in a sediment microbial community. *Journal of bacteriology*, 188(16), pp.5722-5730.

Tacconelli E, and Magrini E, 2017. Global priority list of antibiotic-resistant bacteria to guide research, discovery and development on new antibiotics. Who.int. Available at <http://www.who.int/medicines/publications/global-priority-list-antibiotic-resistant-bacteria/en/>. Date accessed: 14/08/2017

- Thiennimitr, P., Winter, S.E., Winter, M.G., Xavier, M.N., Tolstikov, V., Huseby, D.L., Sterzenbach, T., Tsolis, R.M., Roth, J.R. and Bäuml, A.J., 2011. Intestinal inflammation allows *Salmonella* to use ethanolamine to compete with the microbiota. *Proceedings of the National Academy of Sciences*, 108(42), pp.17480-17485.
- Tran, E.N.H., Papadopoulos, M. and Morona, R., 2014. Relationship between O-antigen chain length and resistance to colicin E2 in *Shigella flexneri*. *Microbiology*, 160(3), pp.589-601.
- Tzouvelekis, L.S., Tzelepi, E., Tassios, P.T. and Legakis, N.J. (2000) CTX-M-type L-lactamases: an emerging group of extended-spectrum enzymes. *Int. J. Antimicrob. Agents* 14, 137-142.
- Van Boeckel, T.P., Brower, C., Gilbert, M., Grenfell, B.T., Levin, S.A., Robinson, T.P., Teillant, A. and Laxminarayan, R., 2015. Global trends in antimicrobial use in food animals. *Proceedings of the National Academy of Sciences*, 112(18), pp.5649-5654.
- van den Bogaard, A.E. and Stobberingh, E.E., 2000. Epidemiology of resistance to antibiotics: links between animals and humans. *International journal of antimicrobial agents*, 14(4), pp.327-335.
- Van den Bogaard, A.E., London, N., Driessen, C. and Stobberingh, E.E., 2001. Antibiotic resistance of faecal *Escherichia coli* in poultry, poultry farmers and poultry slaughterers. *Journal of Antimicrobial Chemotherapy*, 47(6), pp.763-771.
- Weber, N.R., Pedersen, K.S. and Nielsen, J.P., 2017. Association of resistance to ampicillin and other antibiotic agents as evidence of co-selection in *E. coli* isolates from diarrheic nursery pigs. In *The 9th European Symposium of Porcine Health Management*.
- Wellington, E.M., Boxall, A.B., Cross, P., Feil, E.J., Gaze, W.H., Hawkey, P.M., Johnson-Rollings, A.S., Jones, D.L., Lee, N.M., Otten, W. and Thomas, C.M., 2013. The role of the natural environment in the emergence of antibiotic resistance in Gram-negative bacteria. *The Lancet infectious diseases*, 13(2), pp.155-165.
- Willmott, C.J. and Maxwell, A.N., 1993. A single point mutation in the DNA gyrase A protein greatly reduces binding of fluoroquinolones to the gyrase-DNA complex. *Antimicrobial Agents and Chemotherapy*, 37(1), pp.126-127.
- Winsor, G.L., Griffiths, E.J., Lo, R., Dhillon, B.K., Shay, J.A. and Brinkman, F.S., 2015. Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the *Pseudomonas* genome database. *Nucleic acids research*, 44(D1), pp.D646-D653.
- Winstanley, C., O'Brien, S. and Brockhurst, M.A., 2016. *Pseudomonas aeruginosa* evolutionary adaptation and diversification in cystic fibrosis chronic lung infections. *Trends in microbiology*, 24(5), pp.327-337.
- Zhiyong, Z., Feng, Y. and McNally A., 2016. New kids on the block: Intercontinental dissemination and transmission of newly emerging lineages of multi-drug resistant *Escherichia coli* with highly dynamic resistance gene acquisition. *bioRxiv*.