Campylobacter jejuni microevolution and phenotype:genotype relationships

A thesis submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy in Veterinary Science
at Massey University, Palmerston North,
New Zealand,
by

Anja Friedrich

2014

Massey University
mEpiLab | Infectious Disease Research Centre (IDReC)
Institute of Veterinary, Animal & Biomedical Science
Palmerston North, New Zealand

Abstract

Campylobacter spp. are a major cause of human gastroenteritis. Their wide host range, environmental distribution and high genetic diversity contribute to the complex molecular epidemiology of campylobacteriosis. The aim of this multidisciplinary thesis is to investigate the phenotype:genotype relationships of *C. jejuni* and how they influence the micro-evolution of these bacteria in New Zealand.

The first study used a time series of genotyped human campylobacteriosis cases from a region of New Zealand to investigate if the clonal complexes (CCs) identified in human cases showed a seasonal pattern. The analysis revealed a prevalent clonal complex (CC-45) which showed a consistent summer peak.

The second study applied phylogenetic and population genetic tools to describe the population structure and host associated genotypes within the *C. jejuni* population in wild and agricultural animals. The findings showed that the *C. jejuni* isolates from non-agricultural animals exhibit a higher number of mosaic alleles and fewer shared sequence types (STs) between the host groups, whereas the *C. jeuni* in agricultural animals show a higher number of shared STs and fewer occurrences of admixture.

The third study tested the ability of a variety of *C. jejuni* isolates to utilise 95 substrates as carbon sources and tested their tolerance to different osmotic conditions using phenotypic microarray (PM) technology. These phenotypic expressions were correlated with their genomes and a genome wide association study was used to identify genes associated with the observed phenotype.

The last study made use of data from a dual isolate chicken challenge. The study showed the out-competition of one challenge strain and genetic variations of 15 core single nucleotide polymorphisms (SNPs), 14 of which were non-synonymous point mutations. These SNPs were confined to nine genes all of which were associated with cell shape, chemotaxis or motility of the bacteria.

This thesis has furthered our understanding of the seasonality of human campylobacteriosis in New Zealand, the existing population structure of *C. jejuni*, its biochemical requirements and tolerance to osmolytes and novel insights into short-term evolutionary dynamics *in vivo*. Based on these findings and the recommendations for future directions, this could lead to a greater understanding of host-association and new intervention strategies.

Acknowledgements

Firstly, I would like to thank my supervisors, Professor Nigel P. French, Dr Patrick J. Biggs, Dr Anne C. Midwinter, Professor Mike Hendy, and Professor Paul B. Rainey for their help, encouragement and all their guidance throughout the PhD project.

I am also grateful to the staff and students of the mEpiLab team for their technical support and helpful discussions, especially to Dr Jonathan Marshall for his help with the statistical analysis.

I would like to thank the Allan Wilson Centre for Molecular Ecology and Evolution for funding this project. I would like to show my gratitude to the Institute of Veterinary, Animal and Biomedical Sciences for funding several conference travels and research projects.

I am also grateful to Steve Richie for his help with the software Primer, Jukka Corander for helpful comments concerning the software BAPS, Professor V. Ravindran and colleagues for the animal work. I would also like to thank Mauro Tuglio for his expertise in biostructural analysis.

I would like to thanks my parents for their love, support and encouragement.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 General background

Campylobacter spp. are the most common cause of acute gastroenteritis in the developed world (Adak et al., 2005; Coker et al., 2002; Friedman et al., 2000; Samuel et al., 2004). Three specific *Campyloabcter* spp. belonging to the group of thermophilic campylobacters, *Campylobacter jejuni* (*C. jejuni*), *C. lari* and *C. coli*, are able to thrive at 42°C and colonise the gastrointestinal tract GI of several agricultural and pet animals and wildlife (Moore et al., 2005). *Campylobacter* spp. have been isolated from a wide variety of domestic livestock, including poultry (Eberhart-Phillips et al., 1997; Müllner et al., 2009, 2010b), ruminants (Stanley and Jones, 2003) and pigs (Brown et al., 2004; Nesbakken et al., 2003). The majority of human infections are caused by *C. jejuni* (90%) and *C. coli* (10%) where most of the *C. jejuni* infections are believed to be caused through the consumption of contaminated food (Carter et al., 2009; Eberhart-Phillips et al., 1997; Gillespie et al., 2002; Lee and Newell, 2006; Nachamkin et al., 1992).

Campylobacter spp. display extensive genetic variation, which has arisen from intragenomic mechanisms as well as genetic exchange among strains and is well described within (Sheppard et al., 2014; Suerbaum et al., 2001) and between (Sheppard et al., 2008, 2011b) different species and lineages of *Campylobacter* spp. Apart from horizontal gene transfer HGT, intragenomic events such as point mutations, rearrangements, duplications, deletions and insertions have been proposed to contribute to the genetic diversity of the bacterial genome (de Boer et al., 2002). Previous studies (Cody et al., 2013; Dingle et al., 2005; Fearnhead et al., 2005; Jerome et al., 2011; McCarthy et al., 2007; Sheppard et al., 2008, 2013b; Thomas et al., 2014) have shown that the evolution of *C. jejuni* has been affected by both mutation and recombination. Recombination is the main driver for the diversity of *Campylobacter* (Wilson et al., 2009), whereas mutations are responsible for altering genes and as a result, potentially altering phenotypes (Jerome et al., 2011; Thomas et al., 2014). Previous studies using data from multilocus sequence typing (MLST)

estimated the range of the relative contributions of recombination and point mutation events to genetic diversity to be approximately 1:1 (Fearnhead et al., 2005; Feil et al., 1999, 2000). Yu et al. (2012), however, estimated recombination to contribute to the genetic diversity between 2.97 and 8.91 times more than mutation events (depending on the MLST locus examined). Studies examining members of the Campylobacterales family attempted to estimate the relative rate of recombination to mutation based on the whole genome level (Biggs et al., 2011; Falush et al., 2001) and identified a ratio of approximately 1:1 per event.

commentar 1, Mur- ray

Campylobacter spp. are able to colonise multiple hosts, and hence a number of different niches. Our understanding of the factors which influence the ability of *C. jejuni* to colonise specific hosts or to survive in the environment is still poor, therefore several studies in the past have attempt to determine the prevalence of specific clones among *C. jejuni* isolates from diverse sources (agriculture, wild birds and environment) by applying multilocus sequence typing (Colles et al., 2003; Dingle et al., 2001; French et al., 2005; Karenlampi et al., 2007; Manning et al., 2003; McCarthy et al., 2007; Sails et al., 2003; Sheppard et al., 2009; Taboada et al., 2008; Wilson et al., 2008). Large variation has been observed between the host distribution of MLST clonal complexes. A clonal complex (CC) is a cluster of closely related bacterial strains that group around a founder strain (Dingle et al., 2002).

Studies examining CCs related to *Campylobacter* infections in humans, have identified seasonal patterns for specific CCs in many temperate countries (Cody et al., 2012; McCarthy et al., 2007; Sopwith et al., 2008). Nylen et al. (2002) and Kovats et al. (2005) examined the incidence rates of human campylobacteriosis and identified a consistent peak in summer across all temperate countries where some countries also showed an early spring peak. Kovats et al. (2005) identified the timing of the peak as being least variable from year to year in England and Wales, Greece, Denmark and the Netherlands. These studies suggest differing seasonal phenotypes of clonal complexes, some being more associated with human *Campylobacter* infections.

commentar 2, Mur- ray

While colonising different niches *Campylobacter* spp. are subjected to unfavourable conditions. The environmental stress *Campylobacters* can be exposed to are extreme, with temperature, humidity and osmolarity, and the presence of sunlight and atmospheric oxygen being some of the stressors. The obvious success of *Epsilonproteobacteria* in colonising such a diversity of hosts and surviving in a huge range of environments is a sign of their ability to detect, adapt and evolve in response to such environments (Lee and Newell, 2006). To be 'successful in evolutionary terms', an organism needs to be able to continually adapt to changes in the environment. This process is called 'niche adaptation' which is accomplished by genetic mechanisms including HGT, gene duplication and gene shuffling.

Our understanding of the factors which influence the ability of *C. jejuni* to colonise specific

hosts or to survive in the environment is still poor, therefore several studies in the past have attempted to determine the prevalence of specific clones among *C. jejuni* isolates from diverse sources (agriculture, wild birds and environment) by applying multilocus sequence typing (Colles et al., 2003; Dingle et al., 2001; French et al., 2005; Karenlampi et al., 2007; Manning et al., 2003; McCarthy et al., 2007; Sails et al., 2003; Sheppard et al., 2009, 2011a; Taboada et al., 2008; Wilson et al., 2008). These studies have identified a large variation between the host distribution of MLST clonal complexes. CC-21 for example, is widespread, while others like CC-61 have a more restricted distribution (Colles et al., 2003; French et al., 2005). Sometimes, because of these different distributions, the bacteria are referred to as either 'generalists' or 'specialists' (Hepworth et al., 2011). McCarthy et al. (2007) studied the host associated genetic import in *C. jejuni* and confirmed the association of *C. jejuni* genotypes with host species and made the observation that host association appears to be stronger than temporal and geographical effects. Müllner et al. (2010b) suggested that *Campylobacter* spp. in New Zealand display a unique situation compared to *Campylobacter* spp. populations in other countries, exhibiting a relatively low genetic diversity, and revealing evidence of many niche-associated genotypes.

Evolutionary forces drive bacteria to grow in and adapt to as many niches as possible (Bochner, 2009). The adapting bacteria have to use the basic elemental nutrients (carbon, nitrogen, phosphorus, sulphur, oxygen, etc.) in that environment, and this differs between hosts (Bochner, 2009). Therefore, one way to determine why some genotypes show a higher degree of host-association than others, is to determine the specific phenotype of generalist and specialist genotypes based on nutritional requirements. Previous studies (Brás et al., 1999; Pajaniappan et al., 2008; Stintzi, 2003) have demonstrated that different temperatures, for example a chicken's body temperature at 42°C and a human's body temperature of 37°C, trigger expression of potential colonisation or virulence factors which could either lead to commensalism or pathogenesis. *C. jejuni* genotypes exhibit unique nutritional requirements and have been shown to utilise a limited range of carbon sources and amino acids for growth (Line et al., 2010; Tang et al., 2010).

Different '∼omics' techniques provide great insights into the biology of cellular organisms and, depending on the method, they address different steps in the information transfer, from coding DNA (genomics), via RNA (transcriptomics) to proteins (proteomics) to generate the cellular metabolites (metabolomics) (Buchanan et al., 2009; Vilchez-Vargas et al., 2010; Zhang et al., 2010). All of these addressed '∼omics' techniques study components which contribute to the phenotype (Papin et al., 2004). However, it is becoming increasingly clear that a single '∼omics' approach may not be sufficient to characterise the complexity of biological systems (Gygi et al., 1999). Taking advantage of fully sequenced bacterial genomes, a new technique, called microbial genome-wide association study GWAS is providing the means to investigate the correlation of genetic variants with phenotypic traits across several bacterial strains, enabling the potential discovery of virulence or host-associated genes. Moreover, once candidate genes have been

identified, one can perform knock-outs to see how these affect the phenotype.

Previous studies involving passage of a bacterial strain through a host attempted to identify *in vivo* genetic variation (de Boer et al., 2002; Espedido et al., 2013; Hänel et al., 2009; Koskiniemi et al., 2013; Thomas et al., 2014; Woodall et al., 2005), but only a few identified these variations at a nucleotide level (Espedido et al., 2013; Koskiniemi et al., 2013; Thomas et al., 2014) as opposed to pulse field gel electrophoresis PFGE analyses (de Boer et al., 2002; Hänel et al., 2009; Woodall et al., 2005). Single nucleotide polymorphisms (SNPs) in virulence genes can change the phenotype and determine the pathogen's virulence. SNPs are also used to identify differences between closely related strains (Chen et al., 2006b; Snyder et al., 2013).

The main theme of this thesis is the relationship between phenotype and genotype in *C. jejuni*, and how it influences the microevolution of these bacteria in New Zealand. Using different discriminatory levels of genotypes (e.g. MLST and whole genome sequences), this project is identifying specific phenotypes (seasonality, host-association, different nutritional requirements and short-term evolution) related to these genotypes.

1.2 The structure and aims of this thesis

All chapters of this thesis, except for the literature review and the discussion are written in the form of manuscripts for peer-reviewed publication. Therefore, each chapter has been written to stand alone. However, the bibliography has been combined for ease of reading. The aims of each chapter are stated below:

- ◇ This thesis begins with a literature review (Chapter 2) related to the genotypic and phenotypic characteristics of *Campylobacter* spp. and its evolution.

- ◇ Chapter 3 uses data from a sentinel site in the Manawatu area of New Zealand to determine whether the *Campylobacter* clonal complexes identified in human cases in this region show a seasonal pattern and if so, how do these compare with seasonality-associated CCs in other countries.

- ◇ Chapter 4 applies a combination of phylogentic and population genetic tools to MLST data, to examine how ecological factors such as niche association influence the population structure of *C. jejuni* in New Zealand.

- ◇ Chapter 5 examines phenotypic profiles based on carbon utilisation and sensitivity to osmolarity across two temperatures in 15 isolates associated with different hosts and multilocus

sequence types of *C. jejuni*. Additionally, using the whole genome sequences of the examined isolates and a genome-wide-association analysis, this study aimed to identify some of the genes associated with the observed phenotypes.

- ◇ Chapter 6 makes use of data from a previously performed chicken feed trial experiment by colleagues in IFNHH (Institute of Food Nutrition and Human Health, Massey University, Palmerston North, New Zealand) in 2010. The experiment used dual challenges with two *C. jejuni* isolates in chickens to examine the effects of caprylic acid on the colonisation of *C. jejuni*. This chapter uses *C. jejuni* isolated from faecal extracts to examine short-term evolutionary dynamics and distinct strain-to-strain variations.

This thesis concludes with a discussion (Chapter 7) of the findings of these studies and their implications for our knowledge of the ecology of *C. jejuni*. I present an outlook on future applications for the methodologies described and discuss the advantages and shortcomings of the key elements of this thesis.

CHAPTER 2

Literature review

2.1 Introduction

Campylobacteriosis, mainly caused by the bacterial species *Campylobacter jejuni* and *Campylobacter coli*, is a major cause of foodborne illness, causing human acute bacterial gastroenteritis worldwide (Adak et al., 2005; Coker et al., 2002; Samuel et al., 2004). *Campylobacter* spp. are microaerophilic pathogens belonging to the *Epsilonproteobacteria* which are a class of Gram-negative organisms found in a variety of habitats (e.g. water, wild birds, live stock, pets). These bacteria are genetically close to *Helicobacter* and *Arcobacter* (On, 2001; Vandamme et al., 1991) and primarily colonise the gastrointestinal tracts of a wide variety of host species. However, they have also been detected in oil fields and salt marshes (Lee and Newell, 2006), indicating an evolutionary adaptation to multiple ecological niches.

Most *Campylobacter* species are believed to live a commensal lifestyle in the gastrointestinal tracts of many animals including wild birds and poultry, however, previous research has shown that *C. jejuni* and *C. fetus* cause abortions in sheep (Sahin et al., 2012; Wu et al., 2014), *C. fetus* subs. *venerealis* causes abortions in cattle (Hum et al., 1991), *C. upsaliensis* and *C. helveticus* cause diarrhoea in dogs and cats (Acke et al., 2009; Chaban et al., 2010) and *C. jejuni* has recently been shown to cause infections in poultry (Humphrey et al., 2014). The main source of human campylobacteriosis is thought to be the consumption of chicken or chicken products (Adak et al., 2005; Müllner et al., 2009). *C. jejuni* and *C. coli* represent one of the most important emerging food pathogens in developed and developing countries (Cody et al., 2010; Gillespie et al., 2002). *Campylobacter* spp. are the most frequently isolated bacteria in cases of human gastroenteritis in industrialised countries, with the vast majority of reported cases attributed to *C. jejuni* (90% to 95%), *C. coli* (5% to 10%) and *C. lari* less than 1% (Carter et al., 2009; Gillespie et al., 2002; Lee and Newell, 2006; Nachamkin et al., 1992). However, emerging *Campylobacter* spp. (*C. concisus*, *C. curvus*, *C. fetus*, *C. hominis*, *C. hyointestinalis*, *C. lanienae*, *C.*

mucosalis, *C. rectus*, *C. showae*, *C. sputorum*, *C. ureolyticus* and *C. upsaliensis*) have been isolated from symptomatic and asymptomatic humans (Bullman et al., 2011; Lastovica, 2006; Miller et al., 2012). *C. jejuni*, *C. coli* and *C. lari* are among those termed as the 'thermophilic Campylobacters' (Lee and Newell, 2006).

Due to the similar disease epidemiology of *C. jejuni* and *C. coli* (Gillespie et al., 2002) and the predominance of *C. jejuni* among cases, most studies exploring the epidemiology of human *Campylobacter* infections have focused on *C. jejuni* or have treated *Campylobacter* infections as a single source of infection in the past. Through the development of new molecular diagnostic tools, other *Campylobacter* species including *C. upsaliensis* and *C. fetus* have also been identified to cause human infection (Gillespie et al., 2002; López et al., 2002; Sheppard et al., 2009). The interest in *Campyobacter* spp. has been growing constantly over the past two decades, and has been strengthened due to the emergence of antibiotic resistant strains. One of the interesting observations about *C. jejuni* is that the sequencing of these genomes revealed many regulation systems, fully understood in well studied bacteria, are absent in *Campylobacter* spp. (Garénaux et al., 2008b). The new insight into the biology of the genus *Campylobacter* provided by the genome sequences emphasised that *Campylobacters* offer an alternative and novel model for intestinal colonisation and pathogenicity.

2.2 History

By describing spiral organisms in the large intestines of children who had died of diarrhoeal disease (Escherich 1886), Theodor Escherich might have been the first one to describe the genus *Campylobacter* in 1880 (Moore et al., 2005; Samie et al., 2007). The first isolation of a *Vibrio*-like organism in foetal tissue of aborted sheep was made by McFadyen and Stockman in 1913 (cited in McFarlane et al. (1952)) and a confirmation test was carried out by Smith (1918, 1919) when they isolated similar organisms from aborted bovine foetuses. As the organisms were originally assigned to the *Vibrio* genus (due to their spiral appearance), Smith named the organism *Vibrio fetus* (Smith, 1919). The first time that a human infection was associated with the microaerophilic vibrios was in 1947 when a foetus died in a pregnancy related infection (King, 1957; Moore et al., 2005). In 1957, King (1957) was the first one to propose two different types of vibrios associated with enteric diseases, the first one *V. fetus* and the second one was of a thermophilic nature. When it was realised that the organism could not utilise sugars and had a different G+C content to that of *Vibrio* spp., Sebald and Veron (1963) (cited in Véron and Chatelain (1973)) proposed the genus *Campylobacter* (etymologically from the Greek kampulos = curved, bacter = rod). At this time the genus included only two species: *Campylobacter fetus* and '*Campylobacter bubulus*' (now *Campylobacter sputorum*) (On, 2001).

Since its first observation in 1880 by Escherich our interest in *Campylobacter* increased due to their high prevalence in human diarrhoea and our knowledge has expanded to include the curated complete genome sequences of *C. jejuni* NCTC 11168 (Gundogdu et al., 2007; Parkhill et al., 2000), *C. jejuni* 81116 (NCTC 11828) (Parsons et al., 2009), *C. jejuni* RM1221 (Parker et al., 2006) and *C. jejuni* 81-176 (Hofreuter et al., 2006). Fourteen other *Campylobacter* spp. genomes are now available on NCBI (`http://www.ncbi.nlm.nih.gov/genome/?term=campylobacter`, last accessed 06/09/2014), however only some of them are curated full genomes. Even though numerous *Campylobacter* spp. have been fully sequenced, many details regarding their pathogenicity, population diversity, host association and epidemiology still remain unclear.

2.3 Taxonomy

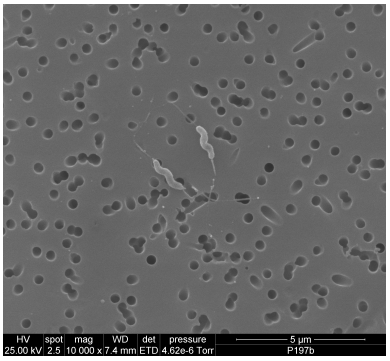
The genus *Campylobacter* has been classified as follows:

- Domain: Bacteria
- Phylum: Proteobacteria
- Class: *Epsilonproteobacteria*
- Order: *Campylobacterales*
- Family: *Campylobacteraceae*
- Genus: *Campylobacter*

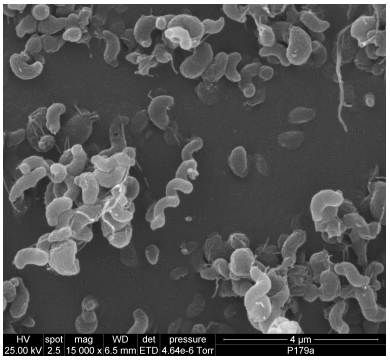
Within the class *Epsilonproteobacteria*, the genera *Campylobacter*, *Arcobacter*, *Helicobacter*, *Wolinella*, and '*Flexispira*' constitute as a separate phylogenetic branch identified as rRNA superfamily VI (Vandamme and De Ley, 1991; Vandamme and On, 2001). At present, the genus *Campylobacter* includes 25 recognised species of microaeorobic $\varepsilon-$ proteobacteria (excluding subspecies) (bacterio.net, 2014) (last accessed on the 5th of August 2014).

2.4 Morphological and biochemical characteristics

C. jejuni is a Gram-negative bacillus, fine and curved, 0.2 to 0.5 microns (μm) in diameter and 1 to 8 μm long. It can have the shape of a comma, an 'S', a helix or a spiral (Garénaux et al., 2008b). *Campylobacter* display usually one or several curves and have a single polar flagellum at



(a) *C. jejuni* isolate P197b.



(b) *C. jejuni* isolate P79a

Figure 2.2: Scanning electron microscopy (SEM) images showing the *C. jejuni* isolates P179a (ST-474) and P197b (ST-45).

one or both ends of the body (Garénaux et al., 2008b; Nachamkin et al., 1993; Song et al., 2004; Vandamme, 2000).This structure gives the bacterium a great mobility, moving in a characteristic rotating rapid corkscrew- like motion, using their unipolar or bipolar flagella (Parkhill et al., 2000).

commentar 3, Murray

C. jejuni is considered to be a microaerophilic bacterium (*i.e.* grows in the presence of an atmosphere poor in oxygen), but many authors find it to be more capnophilic (requires an atmosphere enriched in CO_2 to proliferate) than microaerophilic (Garénaux et al., 2008b). Comparative genomic analysis indicates that *Campylobacter* species lack many of the adaptive responses commonly seen in other foodborne pathogens. *C. jejuni*, for example, lacks a number of pathways for surviving acid attacks (Park, 2002). Despite the absence of some pathways, laboratory experiments have shown that campylobacters express a tolerance response regarding acid resistance as well as resistance to aerobic conditions (Murphy et al., 2003).

C. jejuni is the only *Campylobacter* species that hydrolyses hippurate (though *C. jejuni* subsp. *doylei* varies in its ability to hydrolyse hippurate)(Vandamme and Goossens, 1992). As a result, hippurate hydrolysis (Hwang and Ederer, 1975) has become the most widely used test to identify *C. jejuni* and also to differentiate it from *C. coli* that are phenotypically and genotypically similar (Walder et al., 1983). But even for some strains of *C. jejuni* variability in the hippurate reaction has been observed (Fermer and Engvall, 1999; Morris et al., 1985), therefore an additional number of biochemical tests are employed to differentiate *Campylobacter* spp.

2.5 Mechanisms of colonisation, virulence and pathogenesis

The bacterial colonisation of a host does not necessarily cause disease (commensal lifestyle), whereas virulence is defined as the ability of the bacteria to cause disease. This section describes the progress made in identifying mechanisms related to colonisation of a host. *Campylobacter* spp. do not ferment or oxidise carbohydrates, instead they acquire energy from amino acids, keto acids and tricarboxylic acid cycle intermediates (Vandamme, 2000). Each species of *Campylobacter* has a favoured reservoir, *C. jejuni* for example, the most common species associated with human campylobacteriosis, is notably associated with poultry (Muellner et al., 2013; Müllner et al., 2010a; van den Ent et al., 2001) and has evolved to preferentially colonise the chicken gut (Newell, 2001; van den Ent et al., 2001). Once the campylobacter are excreted into the environment, they commonly do not multiply (Jones et al., 1991; Rosef et al., 2001). One of the reasons is their relatively high minimal growth temperature ($> 30°C$ (Van de Giessen et al., 1996)). Instead, *Campylobacter* spp. may enter a 'viable but nonculturable' (VBNC) stage (Oliver, 2005, 2010; Rollins and Colwell, 1986). Campylobacter cells transform from a spiral form into a coccoid form when they enter the VBNC (Rollins and Colwell, 1986).

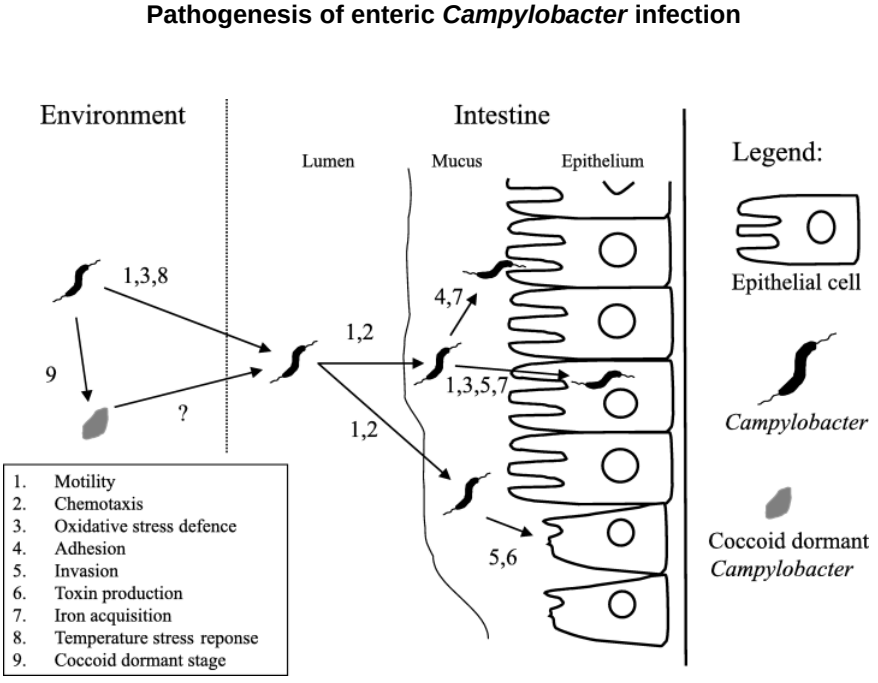


Figure 2.3: Overview of the different phases of *Campylobacter* colonisation of the intestine (van den Ent et al., 2001). 1: motility; 2: chemotaxis; 3: oxidative stress defence; 4: adhesion; 5: invasion; 6: toxin production; 7: iron acquisition; 8: temperature stress response; 9: coccoid dormant stage.

The exact mechanisms why *Campylobacter* are pathogenic in humans but not in animals are unknown but many virulence factors contribute to its pathogenicity in humans. Known virulence factors include flagellin which contributes to the motility of the organism and allows it to travel throughout the host (Nachamkin et al., 1993), chemotaxis which allows the bacteria to sense its environment and rotate the flagella accordingly (Grant et al., 1993), adhesion (allows the bacteria to stay on the host cell) and invasion which are important for colonising the host's intestinal cells and iron acquisition which is important for sustaining nutrients within the host (van den Ent et al., 2001).

comment 14, Murray

2.5.1 Motility

The polar flagellum is crucial for approaching attachment sites on intestinal epithelial cells (Nachamkin et al., 1993). The flagellum is composed of *O*-linked glycosylated flagellin and defects lead to loss of motility, decreasing the adherence and invasion of host cells. The flagellin locus contains two adjacent genes, *fla*A (encoding the major flagellin) and *fla*B (encoding a minor flagellin). A two-component system composed of the sensor FlgS and the response regulator FlgR are central for the regulation of the *Campylobacter* flagellum (Wösten et al., 2004). Previous experiments with mutants have shown that *fla*A, but not *fla*B is essential for colonisation of chickens (Jones et al., 2004; Wassenaar et al., 1993). Colonisation is also reduced for mutants of the motility accessory factor 5 (*maf*5) gene, which is important for the formation of flagella (Jones et al., 2004; Karlyshev et al., 2002).
Previous studies have shown that differentially expressed genes that are involved in the modification of the flagellum, are located in hypervariable regions of the *C. jejuni* genome (Fernando et al., 2007; Hiett et al., 2008; Karlyshev et al., 2005). This variability translates to the protein level, suggesting that it may contribute to the survival of *C. jejuni* in different environments and hosts. Karlyshev et al. (2005) showed that the flagellin *O*-linked glycosylation island, which is important for successful flagellin assembly and motility, is very diverse. Interestingly, three genes (*Cj1324-Cj1325/6*) located in this variable region, are highly prevalent among *C. jejuni* strains associated with poultry (Champion et al., 2005), and might therefore be important for the ability to colonise this host. Howard et al. (2009) supported this hypothesis and showed that particularly *Cj1324* is important for chick colonisation.

2.5.2 Chemotaxis

C. jejuni is a highly motile bacterium, and therefore chemotaxis may be an important factor for sensing favourable conditions and promoting migration towards them. This may be useful

for its survival in and colonisation of the intestinal mucosa. For successful chemotaxis, an intact gradient-sensing mechanism is essential. The chemotaxis genes *che*A, *che*W, *che*V, *che*Y, *che*R and *che*B have been identified in the *C. jejuni* genome of which some are also involved in commensal colonisation of chickens (Hendrixson and DiRita, 2004). The putative adaptation proteins CheB (a methylesterase) and CheR (a methyltransferase) are both involved in a methylation-dependent chemotaxis pathway (Stephens et al., 2006), and a ΔcheBR mutant showed reduced ability to colonise the chick cecum (Kanungpean et al., 2011). *Che*Y, which codes for a response regulator controlling flagellar rotation, is involved in the same signal transduction pathway as CheBR (Stephens et al., 2006) and is crucial for virulence in the ferret model (Yao et al., 1997). *C. jejuni* is attracted by the gycoprotein mucin, several amino acids (aspartate, cysteine, serin, glutamate) and salts of organic acids (citrate, fumarate, α-ketoglutarate, malate, pyruvate and succinate) (Hugdahl et al., 1988). Vegge et al. (2009) also identified L-asparagine, formate and D-lactate as chemoattractants for *C. jejuni*. All of these chemicals are detected by the transmembrane methyl-accepting chemotaxis protein (MCP) (Vegge et al., 2009).

2.5.3 Oxidative stress defence

C. jejuni, as a microaerophilic organism, requires low oxygen levels for its growth and unlike other enteric pathogens, it lacks many adaptive responses to the environment, such as RpoS and SoxRS (Kelly et al., 2001; Park, 2002). The global regulator RpoS, which is the basis for the survival of many Gram-negative bacteria during exposure to different types of environmental stress (Lombardo et al., 2004; Lushchak, 2011; Ramos et al., 2001), is absent in the genome of *C. jejuni* (Parkhill et al., 2000). However, they still need to resist oxidative stress they encounter in the environment (during transmission) and in the invaded host (stressors like superoxide anion, hydrogen peroxide and biotoxic hydroxyl radicals). These stressors can be induced by the host's immune defence, during normal metabolism, or result from incomplete oxygen reduction by *C. jejuni* (Atack and Kelly, 2009).

C. coli and *C. jejuni* share the same oxidative stress defence system, which is generally split into superoxide stress defence and peroxide stress defence (Storz and Imlayt, 1999). The superoxide dismutase (SOD) protein SodB (Pesci et al., 1994; Purdy and Park, 1994), which is the main part of the superoxide stress defence, is an iron-containing SOD that is involved in converting superoxides into hydrogen peroxides. The cytochrome *c* peroxidases (CcPs) are generally responsible for converting the hydrogen peroxide to water (Atack and Kelly, 2009), however, the two CcP loci present in *C. jejuni* do not contribute to hydrogen peroxide resistance (Hendrixson and DiRita, 2004). Instead, it seems that in *C. jejuni* hydrogen peroxide defence is mediated by the cytoplasmic catalase KatA, which breaks it down to water and oxygen (Atack and Kelly, 2009; Bingham-Ramos and Hendrixson, 2008). PerR is the peroxide-sensing regulator and a

mutation of *per*R significantly reduced *C. jejuni* motility and lowered colonisation capacity in chicken (Palyada et al., 2009). This study also discovered a functional network including the antioxidant proteins encoded by *sod*B, defending *C. jejuni* against the superoxide anion, the reductase *ahp*C and *kat*A, their regulators *fur* and *per*R and the regulatory pathways that connect them. This suggests a link between oxidative stress (PerR regulated) and iron metabolism (Fur regulated), and that the defence mechanisms and their proper regulation are essential for successful and efficient colonisation of chickens. Knockout mutations of the Δ*ahp*C reduced colonisation ability in chickens 50 000-fold, whereas Δ*per*RΔ*fur*, Δ*kat*A and Δ*sod*B mutants failed to colonise the chickens. This indicates that all members of the oxidative stress (PerR regulated) and iron metabolism (Fur regulated) need to be intact for normal colonisation of *C. jejuni* in chicks.

2.5.4 Adhesion

Previous studies have identified putative adhesion or binding factors of *C. jejuni*, including the fibronectin-binding outer membrane protein CadF (Konkel et al., 1997), the autotransporter CapA, the periplasmic binding protein PEB1 (Pei and Blaser, 1993) and the surface-exposed lipoprotein JlpA (Jin et al., 2001). Ashgar et al. (2007) created a mutant of the adhesion protein A (*cap*A) and showed that its mutant resulted in reduced capacity to adhere to human and chicken intestinal epithelial cells and abolished its colonisation in a chicken model. However, another study showed that mutation in *cap*A did not result in reduced colonisation (Flanagan et al., 2009) and as the gene is absent in many *C. jejuni* poultry isolates, the actual contribution of *cap*A to successful chick isolation is unclear (Flanagan et al., 2009; Friis et al., 2010). The identified fibronectin-like protein A (FlpA), as well as CadF and PEB1, (Flanagan et al., 2009) have been shown to be important for full binding capacity of *C. jejuni* to chicken epithelial cells. An earlier study suggested that different *C. jejuni* strains compete for colonisation in broilers (Konkel et al., 2007) and hypothesised that this is due to shared adhesions among the isolates and limited host epithelial cell binding places.

2.5.5 Invasion

Studies conducted on different animal models concluded that early mucosal damage is a result of invasive *C. jejuni* in the colonic epithelial cells (Babakhani et al., 1993; Field et al., 1986b; Humphrey et al., 1985; Welkos, 1984) and confirmed the invasive ability of *C. jejuni* as an important pathogenicity-associated factor. Earlier studies have shown that invasive *C. jejuni* strains enter the cells via a microtubule-dependent and actin filament independent invasion mechanism (Biswas et al., 2000, 2003; Monteville et al., 2003; Oelschlaeger et al., 1993). Studies

were conducted on genes which are thought to play a role during invasion on human cell lines, however, experiments on chicken primary epithelial caecal cells are still lacking. The invasion process of *C. jejuni* into cells is unique in such a way that the bacterium enters the cell with its tip followed by the flagellar end (Krause-Gruszczynska et al., 2007).

2.5.6 Toxin production

Even though a variety of toxic activities have been reported for *C. jejuni*, cytolethal distending toxin (CDT) is the only verified *Campylobacter* toxin. It is produced by several *Campylobacter* spp., including *C. jejuni*, *C. lari*, *C. coli*, *C. fetus* and *C. upsaliensis* (Johnson and Lior, 1988). CDT was found to cause cell distension in the mammalian cell lines HeLa, Chinese hamster ovary (CHO), and Caco-2. This distension is characterised by elongation, swelling, and eventual cell death (Whitehouse et al., 1998). Pickett and Whitehouse (1999) hypothesised that within the Cdt ABC complex, CdtA and CdtC are important for binding to the host cell, whereas CdtB is the active moiety. Additional research has confirmed this subunit structure and revealed that CdtA and CdtC interact with CdtB to form a holotoxin necessary for the delivery of the enzymatically active subunit, CdtB (Lara-Tejero and Galán, 2001). The CdtB protein shows similar activity to the enzyme DNaseI (Lara-Tejero and Galán, 2000) and causes cell cycle arrest in the transition phase. The wild-type of the reference *C. jejuni* strain NCTC 11168 produces more toxin and is highly invasive in SCID (severe combined immunodeficiency) mouse tissue compared to isogenic *cdt*B mutants. Interestingly CDT induces the production of neutralising antibodies only in humans, but not in chickens, which might indicate host specific recognition of *C. jejuni* antigens (Young et al., 2007). The specific role of CDT in *C. jejuni* pathogenesis has not been determined yet, but it is hypothesised that it plays a role in invasiveness and modulation of the immune response (Purdy et al., 2000).

2.5.7 Iron acquisition

The intracellular iron concentration is an important factor for *C. jejuni* to insure colonisation success. *C. jejuni* is not able to produce iron itself, therefore it has to acquire it from the host. The concentration of free iron in host tissue is too low to ensure bacterial growth, which might constitute a non-specific host defence (van den Ent et al., 2001). Several iron-uptake mechanisms are essential for the survival and the colonisation of *C. jejuni* in the chicken host. *C. jejuni* requires iron for electron transfer processes, to act as a cofactor for several enzymes and it is responsible for the generation of hydroxyl radicals. *C. jejuni* is able to utilise relatively low numbers of iron compounds (Field et al., 1986a). *C. jejuni*, for example, is able to use the siderophores ferrichrome and enterochelin produced by other organisms (Pickett et al., 1992)

and it is able to use haem compounds (Field et al., 1986a; Pickett et al., 1992).
C. jejuni expresses multiple ferric iron acquisition systems upon growth in iron-restricted conditions. To date, a haemin/haemoglobin uptake system (*chu*ABCD) (Ridley et al., 2006), an enterochelin transporter system lacking an outer membrane (OM) receptor (*ceu*BCDE) (Richardson and Park, 1995), a *cfr*A gene which encodes a homologue of an OM receptor, a ferrichrome uptake system (*fhu*ABD) and a system that has only been identified in *C. jejuni* NCTC 11168 have been identified.
The inactivation of *cfr*B in a *cfr*A- negative *C. jejuni* strain fully eliminated its ability to utilise ferric enterobactin (FeEnt) as an iron source for growth (Xu et al., 2010). Furthermore, the reduced colonisation phenotype could not be restored by the presence of a functional *cfr*A gene, suggesting that CfrB plays an important role during colonisation of chicks and can not be replaced by other iron uptake mechanisms without diminishing the colonisation potential.

2.6 Characterisation of bacterial strains

2.6.1 Antibiotic sensitivity test

Antibiotic susceptibility testing (AST) suggests which antibiotic therapy should be used and how effectively it may work against a bacterial infection *in vivo*. The most frequently used test is the disk diffusion procedure (Kirby-Bauer method) (Winn et al., 2006). Disk diffusion has been standardised and been widely used for rapidly growing pathogens such as Enterobacteriaseae, however, the failure to achieve inter-laboratory reproducibility in measuring inhibition zone sizes has delayed the development of standard disk-based methods for *Campylobacter* (Van Der Beek et al., 2010). Other methods to test antimicrobial susceptibility include the Stokes method (Stokes, 1971) and E-test (also based on antibiotic diffusion) (Luber et al., 2003). Even though the disk diffusion method is a convenient and cost-effective way for generating minimal inhibitory concentration (MIC) values, because of the proprietary nature this technology, the method has not yet been approved by the Clinical and Laboratory Standards Institute (CLSI) (Ge et al., 2013).

commentar 4, Mur- ray

2.6.2 Phenotyping

Phenotype is from the Greek ‘phainein’ (to show) and ‘tupos’ (type) and refers to the set of observable characteristics of an individual resulting from the interaction of its genotype with the environment. This means that phenotypes reflect the nature and the nurture of the organism

(Mayr, 1982; Wanscher, 1975). Phenotyping includes biotyping, serotyping and phage typing (Fitzgerald et al., 2001). Due to the lack of specific antisera when serotyping is employed, unavailability of standard reagents, cross- reactivity between strains and because of emergence of high numbers of non-typeable strains, strain characterisation using phenotyping techniques becomes difficult and often ambiguous (Garénaux et al., 2008b; Jackson et al., 1996). Additionally, the identification of bacterial strains through morphology of colonies on various culture media, biochemical tests, serology, pathogenicity and others are not variable enough to discriminate between closely related strains (Li et al., 2009b; On, 2013)

Biotyping

The first biotyping scheme was developed utilising 12 biochemical test for *Campylobacter* spp. (Bolton et al., 1984). Today, several biochemical properties specific to *Campylobacter* are known and some of them are of taxonomic and diagnostic interest. The main biochemical characteristics of *C. jejuni* are (Garénaux et al., 2008b):

- Oxidase +
- Catalase +
- Nitrate +
- Glucides are neither oxidised nor fermented in Hugh and Leifson's medium
- The Voges-Proskauer reaction is negative (VP-)
- The methyl red reaction is negative (MR-)
- Indoxyl acetate hydrolysis +
- Selenite reduction negative
- Urease negative
- Hydrogen sulfide (H_2S) production is negative in triple sugar iron medium
- Generally, hippurate hydrolysis +

The list is not exhaustive and many authors have combined several characteristics in different ways, proposing numerous biotyping schemes, but no scheme has emerged as a reference (Garénaux et al., 2008b). The variety of responses obtained for the strains of a single species

and the absence of correlation with other typing methods (serotyping, phage typing) resulted in less and less use of biotyping today (Garénaux et al., 2008b).

Serotyping

Two serotyping schemes for *Campylobacter* which have been developed in Canada in the 1980s have been widely used, either separately or together (Penner, 1988; Penner and Hennessy, 1980). The Penner scheme (Penner and Hennessy, 1980) is based on soluble heat-stable (HS) antigens, while the Lior scheme (Lior et al., 1982) detects variation in heat-labile (HL) antigens. The application of these two techniques to *C. jejuni* isolates from farms, humans and the environment has proven useful in the investigation of outbreaks and in studies of potential reservoirs for human infection (Patton et al., 1985; Woodward and Rodgers, 2002). However, its routine application has been inhibited by the requirement of a large panel of antisera that is labour intensive to produce and maintain and not commercially available (Fitzgerald et al., 2005). Another disadvantage of serotyping is the potential non-typability of bacterial isolates. However, despite these disadvantages, earlier studies have demonstrated that a combination of HS and HL provides a greater level of discrimination for outbreak investigation (Patton et al., 1993) and is a useful approach for investigating the epidemiology of *C. jejuni* (Jackson et al., 1998; Woodward and Rodgers, 2002).

Phage typing

Single strains of bacteria can be detected through phage typing, which can also be used for tracing the source of the outbreak of infections (Baggesen et al., 2010). Bacteria can be infected by viruses called bacteriophages (phages) and some of these can only infect a single strain of bacteria. These phages are used to identify different strains of bacteria within a single species. Phage typing schemes using bacteriophages to classify *C. jejuni* isolates have been described in the UK, USA and Canada (Grajewski et al., 1985; Khakhria and Lior, 1992; Salama et al., 1990). Phage typing is often used in combination with HS serotyping, however, many isolates are non-typeable or possess unique phage reaction patterns (Miller and Mandrell, 2005).

2.6.3 Sequence based typing techniques

Genotyping is the discrimination of bacterial strains based on their genetic content (sequence). Strain typing methods based on the sequence level can be divided into three categories: DNA banding pattern, DNA sequencing, and DNA hybridisation based methods. Due to its high

resolution, genotyping is more widely employed for the identification and subtyping of microorganism than phenotyping (Li et al., 2009b).
DNA banding pattern-based genotyping distinguishes the studied strains based on the size of the DNA band (fragment). The fragment can be generated by amplification of DNA, by cleavage of DNA using restriction enzymes (REs) or both.
DNA sequencing-based genotyping methods reveal the underlying sequence of nucleotides and discriminate directly from polymorphisms in the examined sequence.
DNA hybridization-based methods are also known as DNA macroarray and microarray studies. The bacterial strains are distinguished by analysing the hybridization of their DNA to probes of known sequences.

DNA banding-based methods

Digestion of DNA with REs and amplification of DNA can produce millions of copies of a fragment and has many advantages including sensitivity, speed and applicability to a wide range of environmental samples and human specimens (Li et al., 2009b; Lo and Chan, 2006). REs precisely recognise and cut DNA at a defined sequence and the large variety of REs makes it possible to apply them to a wide range of samples. The different DNA banding-based methods are discussed briefly in the following section.

Enzymatic restriction
Enzymatic restriction can be categorised into two main techniques: Pulse-field gel electrophoresis (PFGE) and restriction fragment length polymorphism (RFLP). PFGE is an electrophoretic technique used to separate large DNA molecules (10kB-10Mb). Even though DNA fragments >20kB show the same mobility in a conventional constant electric field, making it impossible to differentiate them, previous experiments (Herschleb et al., 2007; Schwartz and Cantor, 1984) have shown that by applying alternating electric fields at different angles, PFGE can separate large DNA molecules in a flat agarose gel. Due to the high resolution and macro-restriction analysis at the genome level, it is considered to be a 'gold standard' for subtyping many bacteria (Gerner-Smidt et al., 2006; Tenover et al., 1995). Nevertheless PFGE should be validated with other genotyping results (Li et al., 2009b) and although it is widely used, it has several limitations. It is time and labour consuming, lacks reproducibility and inter-laboratory comparability. The different restriction enzymes used to digest the chromosomal DNA and varying electrophoresis conditions in different studies, make it difficult to compare inter- laboratory obtained PFGE profiles (Wassenaar and Newell, 2000). However, using more than one RE increases the discriminatory power significantly (Gibson et al., 1994; Imai et al., 1993; Matsuda et al.,

1994). Adequate results have been obtained with *Sma*I, *Sal*I, *Kpn*I, *Apa*I and *Bss*HII (Wassenaar and Newell, 2000). *Xho*I appears to be useful for *C. upsalenensis* (Bourke et al., 1996). The largest PFGE database, PulseNet, currently tracks ten foodborne pathogens: *Clostridium botulinum*, *Campylobacter jejuni*, *Escherichia coli* O157:H7 and non-O157, *Listeria monocytogenes*, *Salmonella*, *Shigella sonnei*, *Shigella flexneri*, *Vibrio cholerae* and *Vibrio parahemolyticus* (http://www.pulsenetinternational.org/protocols/; last accessed on 09/08/2014).

RFLP detects variations in DNA and measures the size of restriction fragments separated by agarose gel electrophoresis. RFLP produces hundreds of short restriction fragments, making it difficult to separate them using agarose gel electrophoresis. However this can be simplified using Southern blotting with radioactively labeled probes (Southern, 1975). Ribotyping which is a variation of RFLP, uses rRNA probing. Most bacteria have more than two ribosomal gene sets (5S, 16S, 23S) located in the chromosome which show variability. However, most *Campylobacter* spp. contain only three ribosomal gene copies limiting the discriminatory power of this technique (Wassenaar and Newell, 2000). For example, Denes et al. (1997) showed that ribotyping cannot differentiate within or between subspecies of *C. fetus*, however, Kiehlbauch et al. (1991) showed that ribotyping can be useful to determine the species of *Campylobacter* isolates that are aerotolerant or phenotypically hard to analyse.
The digestion enzymes (*Pst*I, *Hae*III, *Hind*III, *Pvu*II) used in ribotyping studies can vary substantially between laboratories, which can hinder the direct comparison of the obtained ribotype profiles (Wassenaar and Newell, 2000). Unlike PFGE, RFLP does not require expensive equipment and is therefore more cost effective than PFGE is, apart for automated ribotyping (riboprinting) which involves high cost for equipment and consumables (Wassenaar and Newell, 2000).

DNA amplification

AP-PCR (arbitrarily primed PCR), also called 'Random Amplified Polymorphic DNA' is a type of PCR reaction, involving segments of DNA that are amplified at random. The method is to create several arbitrary, short primers (<10 bp), which combined with low temperatures, enables genomic DNA to be amplified at multiple loci. This method does not require prior knowledge of the DNA sequence for the target genome. AP-PCR is inexpensive and fast, however, there are still limitations, such as the reproducibility, which hinders the comparison of RAPD patterns within and between laboratories (Power, 1996; Tyler et al., 1997; Wassenaar and Newell, 2000). Additionally, RAPD results may be influenced by several factors such as annealing temperature and sequence of arbitrary primers; inconsistency in thermal cyclers; DNA template purity and concentration; and PCR equipment and reagents (Ellsworth et al., 1993; Li et al., 2009b; MacPherson et al., 1993; Penner et al., 1993; Tyler et al., 1997).

High-resolution melting (HRM) analysis aims to discriminate DNA alleles through melting temperature analysis. HRM is based on real time PCR amplification and melting curve analysis. It allows detection of sequence variants without sequencing or hybridisation procedures (Gundry et al., 2003). Single nucleotide polymorphisms (SNPs) can be typed without probes and HRM can be combined with various PCR strategies for bacterial strain typing (Li et al., 2009b). The limitation is that HRM requires real time PCR equipment and may not be equally discriminatory for all bacterial species.

Multi-locus variable number tandem repeat analysis (MLVA) is a genotyping method based on the polymorphic analysis of multiple variable tandem repeat (VNTR) loci (Lindstedt, 2005; van den Berg et al., 2007). VNTR are DNA sequences which are variable in copy number and are distributed across the bacterial and human genome (Lupski and Weinstock, 1992; Vergnaud and Denoeud, 2000). They can be found in coding and non-coding regions. PCR amplification can be used to determine the number of repeats for each locus, and the banding patterns can be used to reveal the genotype and infer phylogenetic relationships (Keim et al., 2000; Tenover et al., 2007). The identified variable loci in conjunction with tandem repeat-finding programs (TANDEM REPEAT FINDER (Benson, 1999), the Tandem Repeat database and the Microsatellite database) can be used to screen bacterial genomes for potential VNTR loci.
Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) have been used in various bacteria for genotyping. A CRISPR consists of near-perfect direct repeat sequences of approximately 24−48 bases long interspersed with similarly sized nonrepetitive spacer sequences. The CRISPR region is widely used to genotype *M. tuberculosis* (Fang et al., 1998; Filliol et al., 2003; Groenen et al., 1993) and to a lesser extend for *Streptococcus pyogenes* (Hoe et al., 1999), *Y. pestis* (Pourcel et al., 2005), *C. jejuni* (Kovanen et al., 2014; Price et al., 2007; Schouls et al., 2003). However, a study by Schouls et al. (2003) employing three molecular typing methods (AFLP, MLST, CRISPR typing) to 184 *Campylobacter* isolates, showed that MLST and AFLP analysis yielded more than 100 different profiles and patterns whereas the CRISPR analysis allowed distinction between two-thirds of the typeable strains. Therefore Schouls et al. (2003) concluded that the three typing methods are equally powerful in identifying strains from outbreaks of human campylobacteriosis. Price et al. (2007) combined a high-resolution melt (HRM)-based assay with the hypervariable CRISPR locus of *C. jejuni*. The study showed that a combination of binary gene typing methods (Price et al., 2006) and the CRISPR HRM assay equals or surpasses the power of the 'gold standard' PFGE (Price et al., 2007). MLVA has been shown to be a high-resolution method for discriminating bacteria and although it may not be useful to infer phylogenetic relationships because the loci evolve too quickly, it is useful for tracking outbreaks of bacterial infections (Lindstedt, 2005; Schouls et al., 2003).

Enzymatic digestion following DNA amplification

A method to overcome the shortcoming of restriction fragment length polymorphism (RFLP) is PCR-RFLP (Wichelhaus et al., 2001). This method engages RFLP at a specific locus amplified by PCR. The combination of PCR-RFLP with capillary electrophoresis (PRACE) shows higher resolution than traditional PCR-RFLP (Chang et al., 2007; Ho et al., 2004). It also produces a limited number of restriction fragments from RE digestion which can be separated and visualised by gel electrophoresis without having to employ hybridisation. The limitation is that the genetic information is often based on only one locus which limits the discriminatory power. To overcome this limitation, several polymorphic genes can be combined and analysed by a multiplex PCR. One multiplex PCR for *C. jejuni* for example, uses the polymorphic genes *gyr*A and *pfl*A and reaches a discriminatory power similar to PFGE (Ragimbeau et al., 1998). Other candidate genes would be *fla*A and *flg*E (Wassenaar and Newell, 2000).

Amplified fragment length polymorphism combines the accuracy of restriction enzyme analysis with the precision of PCR (Vos et al., 1995). The way the method works is that the digestion is carried out with two restriction enzymes, one with a four-base pair cutter (*Csp*6I, *Hha*I, *Mfe*I) and the other one with a six-base recognition pattern (*Bgl*II or *Hin*dIII) (Klena and Konkel, 2005). The restriction fragments are ligated with end-specific adapters followed by selective amplification. A variation of the conventional AFLP technique is a method that uses fluorescent labelled primers for PCR amplification (FAFLP) and automated DNA sequencers. FAFLP has a higher resolution compared to traditional AFLP as it is able to detect size differences as small as 1 bp (Lindstedt et al., 2000; Mortimer and Arnold, 2001; Zhao et al., 2000). AFLP analysis has widely been used to type *C. jejuni* as well as other pathogens (Duim et al., 1999; Kokotovic and On, 1999; Schouls et al., 2003; Siemer et al., 2004) and has been described to be as discriminatory as MLST, RFLP, RAPD and PFGE (Melles et al., 2007; Schouls et al., 2003; Zhao et al., 2000) . The AFLP assays are cost-effective and can be automated (Duim et al., 1999). Its limitations lie in the complexity of the technique (comparable to PFGE), in the combination of REs and the possibly necessary automated analysis equipment (an automated DNA sequencer and appropriate software) (Wassenaar and Newell, 2000).

commentar 7, Mur- ray

2.6.4 Whole genome sequencing technologies

A DNA sequence is defined by the precise order of the four nucleotides (adenine, guanine, thymine and cytosine) in a strand of DNA. Each species and its individuals are uniquely defined by their sequence which makes it essential for the research of the organism. DNA sequencing technologies could be used in a broad range of applications, including molecular cloning, breeding, finding pathogenic genes and comparative and evolutionary studies. The genomic DNA information can be used directly for phylogenetic analysis and identification of bacterial strains. It is highly reproducible and can therefore be shared between laboratories and

stored in databases. The largest online DNA sequence database available to date is GenBank `http://www.ncbi.nlm.nih.gov/genbank`. The possibility to identify SNPs, insertions, deletions and genes under selective pressure has become crucial for biological research, as well as in numerous applied fields such as diagnostic, biotechnology and forensic biology.
DNA sequencing technologies should ideally be fast, accurate cheap and easy to operate. During the course of the last 30 years, DNA sequencing technologies and applications have undergone an incredible development. In order to review next generation sequencing (NGS) systems, and to compare their advantages and disadvantages, it is necessary to look back at the history of sequencing technology development.

First generation DNA sequencers

In 1974, two methods were developed independently by an American team and an English team to find a technique to sequence DNA. The American team, lead by Maxam and Gilbert, used a 'chemical cleavage protocol', while the English team, lead by Sanger, aimed to develop a procedure which would simulate the natural process of DNA replication. Even though both teams shared the 1980 Nobel Prize, Sanger's method became the 'gold standard' because of its practicality, high-efficiency and lower radioactivity. It is also referred to as dideoxy sequencing or chain termination DNA sequencing (Li et al., 2009b).
The steps within the procedure are outlined below (Sanger et al., 1977):

1. Desired region of the DNA is amplified and denatured to produce a single strand of DNA.

2. A sequencing primer is annealed to the single stranded DNA.

3. The method takes advantage of the fact that a growing chain of nucleotides, extending in the 5' to 3' direction, will terminate if, instead of a conventional deoxynucleotide, a 2'3' dideoxynucleotide becomes incorporated. Repeating this four times, each time with four different dideoxynucleotides in addition to all four deoxynucleotides, four different sets of chain-terminated fragments will be produced.

4. The chain terminated fragments will remain attached to the single stranded DNA molecule which has been used as a template. By adding a denaturing agent and heating the partially double stranded molecules, the chain terminated fragments will be released from the template (single stranded DNA molecule) and separated using high resolution denaturing gel electrophoresis.

5. The original sequence of DNA is then deduced by examining the relative positions of the dideoxynucleotide chain termination products in the four lanes of the denaturing gel.

Second generation DNA sequencers

Roche 454 Pyrosequencer

Pyrosequencing is a nonelectrophoretic sequencing method based on the sequencing-by-synthesis" (SBS) principle. The method relies on the detection of pyrophosphate release during nucleotide incorporation as opposed to chain termination with dideoxynucleotides (Sanger).

DNA is amplified inside water droplets in an oil solution (emulsion PCR) where each water droplet contains one DNA template that is attached to a primer-coated bead. Each combination of DNA template and bead forms a clonal colony. The technique is sequencing one strand at a time by synthesising the complementary strand alongside. This way it is synthesising one base pair at a time and detecting which base was added at each step. The template DNA is immobile, and solutions of dATP, dCTP, dGTP, and dTTP nucleotides are sequentially added and unmatched bases are degraded by *apyrase*. The sequencing reaction is monitored by luciferase which generates light (by transforming luciferin into oxyluciferin) for the detection of the individual nucleotides upon primer extension.

In 2005, the initial read length of Roche 454 was 100-150 bp, 200,000 reads and an output of 20 Mb per run (Mardis, 2008). This increased with the release of the 454 GS FLX Titanium system in 2008. The read length could now reach 700 bp with an accuracy of 99.99% (with filter) and an output of 0.7 Gb of data per run within 24 hours. In 2009, by combining the GS Junior (a bench top system) with the 454 pyrosequencing, Roche managed to simplify the library preparation and increase the output to 14 Gb per run (Huse et al., 2007). The greatest advantage of Roche 454 sequencing is its speed (10 hours from sequencing start till completion), however the high cost of the reagents and the relatively high error rate in terms of homopolymer runs longer than 6 bp (Liu et al., 2012) are disadvantageous.

Pyrosequencing has been used for a variety of applications such as sequencing of one gene (Jonasson et al., 2002), identification of different taxa in stool specimens (Dethlefsen et al., 2008), detection of antibiotic resistance and drug resistance screening (Arnold et al., 2005; Jureen et al., 2006). A parallelised version of pyrosequencing was developed by 454 Life Science and has since been acquired by Roche Diagnostics.

Ion Torrent

Ion Torrent Systems Inc., which is now owned by Life Technologies, developed a method which is based on standard sequencing chemistry, but with a novel, semiconductor-based detection system. As opposed to the optical methods in other sequencing systems, this technique is based on the detection of hydrogen ions that are released during the polymerisation of DNA. A high-density array of micro-machined wells, where each well holds a different DNA template, is flooded with a single type of nucleotide. If the introduced nucleotide is complementary to the template nucleotide, it is incorporated into the growing strand. For example, if nucleotide A is added to a DNA template and it is incorporated into the growing strand, then a hydrogen ion will be released. Underneath the wells is an ion-sensitive layer and beneath that a proprietary

ion sensor. The charge from the release upon incorporation into the strand, will change the pH of the solution and can be detected by the ion sensor without cameras.
The error rate for substitutions is at 0.1%, which is similar to that of the Illumina systems (Merriman et al., 2012). However, the main criticism is the homo-polymer errors the system endures. Regardless of many improvements, the 5-mer homo-polymer error rate is still at 3.5% (Merriman et al., 2012).
The initial output of the first Ion-314 chip were 10 Mb, this has increased to 1 Gb with the newest Ion-318 chip. Ion torrent has announced a new sequencer, the Proton-II chip which will have four times the number of sensor wells and an expected output of 32 Gb per chip. This should generate a whole human genome at ∼10x coverage with a run time of 4 h. This output puts the Ion torrent technology on par with a paired-end run on a single HiSeq2000 lane.

SOLiD

The SOLiD (Sequencing by Oligonucleotide Ligation and Detection) system employs a sequencing-by-ligation scheme which has been commercially available since 2006 (Shendure et al., 2005). The sequencer adopts the technology of two-base sequencing based on ligation sequencing. The templated beads, which are based on a flow cell, are sequenced by eight base-probe ligation. These contain a ligation site (the first base), a cleavage site (the fifth base), and four different flourescent dyes (linked to the last base) (Mardis, 2008). The first step is the hybridisation of a primer complementary to the adapter which is then followed by the hybridisation of octamer probes (Buermans and den Dunnen, 2014). The first two bases (starting at the 3' end) are complementary to the nucleotides being sequenced and the fluorescent signal is recorded (by bases 6-8) during the annealing of adjacent probes (first two bases). The last three bases are removed and new octamer probes are allowed to hybridise and ligate.
The SOLiD system, similar to the Illumina system, is able to create paired-end reads. However, due to the limitation of the ligation scheme, the maximal read length is limited to 75 bases for read 1 and 35 bases for read 2. While the number of reads generated by a SOLiD run is comparable to a HiSeq run, the total output in Gb per run is only half of that of a HiSeq run, due to the shorter read length. The main advantage of the ligation scheme is the higher accuracy as each base is interrogated by two octomer ligations.

Illumina

The Illumina platform uses bridge amplification and a sequencing-by-synthesis approach. All enzymatic processes and imaging steps occur in a flow cell, which, depending on the specific Illumina platform may be partitioned into one (MiSeq), two (HiSeq2500) or eight (HiSeq2000, HiSeq2500) separate lanes (Buermans and den Dunnen, 2014). The library is denatured to single strands and attached to the entire inside surface of the flow cell lanes which is followed by bridge amplification of the DNA to form clusters which contain clonal DNA fragments. In order to determine the sequence, the four nucleotide types of reversible terminate bases (RT-bases) are

added. Each of the bases is fluorescently labelled with a different colour and has a blocking group attached. The added bases compete for the binding sites on the template, and non-incorporated molecules are washed away. In contrast to pyrosequencing, the technique employed by Illumina, extends the DNA chain one base at a time. The signal is captured by a CCD camera, but as it can be taken at a delayed moment, large arrays of DNA colonies can be captured by sequential images by a single camera. Therefore, the reachable sequencing throughput by Illumina is limited solely by the analog-to-digital conversion rate of the camera.

In 2006, when Solexa released the genome Analyzer (GA) (which was purchased by Illumina in 2007), the output was 1 GB per run. Through improvements in buffer, flow cell, polymerase and software, this output increased to 20 Gb per run in August, 30 Gb per run in October and 50 Gb per run in December of 2009. The latest GAIIx series can output 85 Gb per run. Early in 2010, Illumina released the HiSeq2000, which is based on the same sequencing strategy as the GA. The initial output of the HiSeq2000 was 200 Gb per run, but this increased to 600 Gb which can be finished in 8 days. The cycle time for the HiSeq2000 is approximately 1 h which is mainly due to the imaging of the flow cell. By reducing the imaging process, the whole process can be sped up considerably, this is implemented in the MiSeq (bench top system) and HiSeq2500 platforms. By providing the option to decrease the total surface area to be imaged, the cycle time can be reduced to 5 and 10 min for the MiSeq and HiSeq2500, respectively (Buermans and den Dunnen, 2014).

Early in 2014, Illumina announced the release of two new sequencers, the NextSeq500 and the HiSeq X Ten. The NextSeq500 was designed to be a highly flexible, smaller version of the HiSeq2500 with runtime of under 30 h. The HiSeq X Ten was designed to enable whole human genome sequencing and reaching the $1000 genome cost in run cost (Buermans and den Dunnen, 2014).

commentar 5, Mur- ray

Third generation DNA sequencers

The third generation of DNA sequencers is characterised by two main features. Firstly, no PCR is required before sequencing, which shortens the DNA preparation time; and secondly, the signal is captured in real time. This means that the signal is monitored during the enzymatic reaction of adding nucleotides to the complementary strand.

Pacific Bio

Single-molecule real-time (SMRT) is the third-generation sequencing method developed by Pacific Biosystems. The underlying principles are quite different from those of the above mentioned sequencers. SMRT cells consist of millions of zero-mode waveguides (ZMWs), which are embedded with one complex consisting of template molecule, sequencing primer and DNA polymerase attached to the bottom of the ZMW (Korlach et al., 2008). SMRT works with single molecule

detection, which means the optics used are sensitive enough to detect incorporation of one fluorescently labelled nucleotide. Therefore the template preparation does not require amplification, and the prepared library molecule is used as the sequencing template (Buermans and den Dunnen, 2014). The sequencing reaction of a DNA fragment is executed by a single DNA polymerase molecule that is attached to the bottom of the 150,000 ZMW wells (Levene et al., 2003). During the reaction, the enzyme in each ZMW well will incorporate the nucleotide into the complementary strand and cleave off the flourescent dye previously linked with the nucleotide (Buermans and den Dunnen, 2014). The camera inside the machine is capturing the signal in real time (Timp et al., 2010), which will also produce the signal difference over time. This may be useful for the prediction of structural variance in the sequence, especially in epigenetic studies such as DNA methylation (Branton et al., 2008).

Several advances at the technological, bioinformatics and chemistry level increased the average output per SMRT cell considerably since the release of the system in early 2011. The main achievement was the increase of longer average read length, which started at the time of the release, with 1 kb. The maximal achievable read length obtained by SMRT is directly related to the length of the sequencing time but not all the polymerase complexes reach identical read length as shown in Figure 2.4. The main reason for this is photo-damage of the phi29 polymerase which terminates the sequencing reaction (Buermans and den Dunnen, 2014). The current maximal movie length of 180 min produces a read length of 40,000 bases. This translates into an output per SMRT cell of 400 Mb for the PacBio RSII (Buermans and den Dunnen, 2014). The high single pass error rate of 10-15% with its majority being insertion/deletion errors is a much debated aspect. However, with the PacBio Quiver software (Chin et al., 2013) the consensus can reach an accuracy of 99.99%. Despite the relatively low output per SMRT cell compared to other NGS platforms, PacBio's long read length, absence of GC bias and insight into the kinetic state of the polymerase during sequencing, enable it to occupy a niche for specific applications that can not be covered by other currently available platforms (Buermans and den Dunnen, 2014).

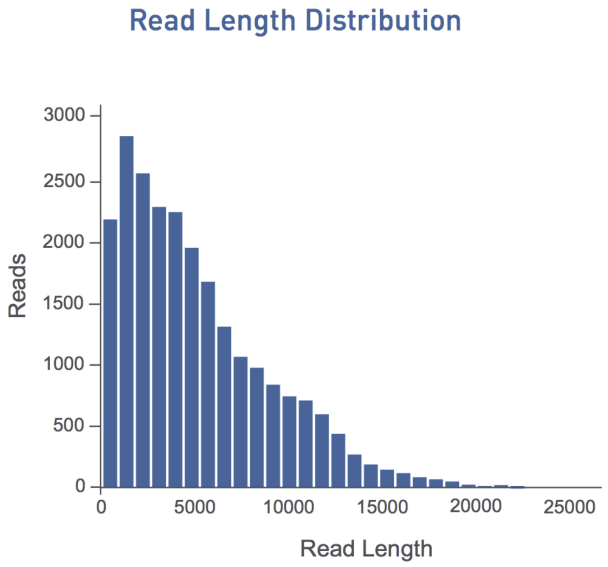


Figure 2.4: Typical PacBio C2Xl raw read length distribution.
From `http://pacificbiosciences.com/brochure` (February 2013)

Nanopore sequencing

Expectations for sequencing a DNA molecule with the nanopore DNA sequencers are for single molecule, amplification free, base detection without labels, long reads and low GC bias. The underlying principle of this technology is to tunnel (polymer) molecules through a pore that separates the two compartments, and to convert the electrical signal of nucleotides that pass through this pore. The physical presence of the molecules passing through the pore causes a characteristic change which allows identification of the specific molecule (Ashkenasy et al., 2005). The detection is achieved through a standard electro-physiological technique.

Two versions of nanopore DNA sequencers are being developed; one using the natural pore forming protein α-hemolysin (a 33 kD protein isolates from *Staphylococcus aureus* (Deamer and Akeson, 2000)) (Clarke et al., 2009), the other one is using manufactured solid state pores (Fologea et al., 2005; Li et al., 2003a).

One of the companies working on building nanopore sequencing devices, is Oxford Nanopore Technologies (ONT). In late November 2013, Oxford Nanopore announced the opening of registration for a MinION Access Programme so that a large number of customers will have access to MinION technology, however even though the MinION has been introduced at the American Society of Human Genetics annual meeting in Boston (Figure 2.5), it has not been commercially available yet and specifics on read length, accuracy or run times are difficult to obtain.

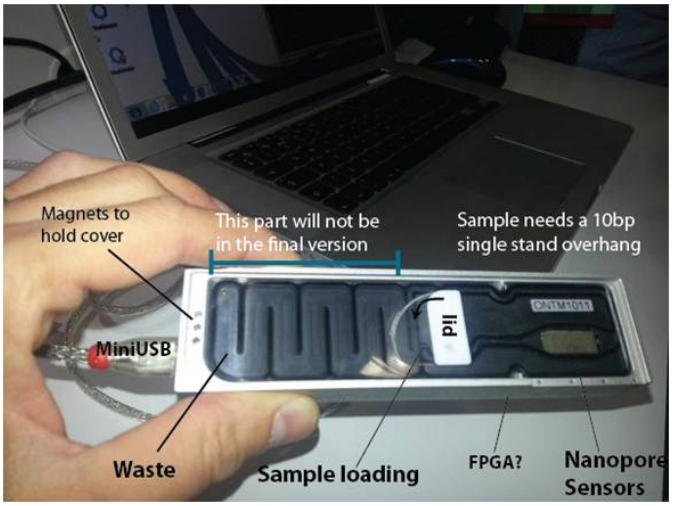


Figure 2.5: The MinION introduced at the American Society of Human Genetics (ASHG) annual meeting in Boston. From http://www.ngsleaders.org/Blog/At-ASHG,-Oxford-Nanopore-Execs-Talk-Non-Stop-but-Say-Little/

Table 2.1: Comparison of Next Generation Sequencing technologies

Method	Single-molecule real-time sequencing (Pacific Bio)	Pyrosequencing (454)	Ion semiconductor (Ion Torrent sequencing)	Sequencing by synthesis (Illumina)	Sequencing by ligation (SOLiD sequencing)	Chain termination (Sanger sequencing)
Read length (bp)	min. 8,500 bp; N50: 10,000 bp	400- 700	200-400	2 x 100 (HiSeq2500)	75 + 35 bp	500-600 bases$^{\flat}$
Accuracy	99.99% consensus accuracy; 85-90% single-read accuracy$^{\perp}$	99.5%	99.94%	>98.5%	99.9%	99.9%
Reads per run	0.8 million	0.1- 10 million	4 million bp	3 billion	2.7 billion	N/A
Run time	2 days	10-23 hours	4 hours	12 days	10 days	20min to 3 hours*
Advantages	long reads, absence of GC bias, high consensus accuracy	low error rate; medium read length ($\sim$ 400 bp to 1 kB)	low start-up costs; low error rate; fast runs	low error rate; lowest cost per base	low cost per base	lowest error rate, long read length ($\sim$ 750 bp)
Disadvantages	less data generated per run; high error rate (for single read); high start-up costs	relatively low throughput, error-prone at homopolymers; high cost per base	developing technology; higher cost than Illumina; read length only at $\sim$ 100-200 bp	short read length; long running times; high start-up costs	slower than other sequencers, problem sequencing palindromic sequences	high cost; long time to generate data; need for cloning

$\diamondsuit$(Pacific Biosciences of California, Inc., 2013); $\natural$(Corporation, 2013);
$\flat$(Wetterstrand, 2014); $^{\perp}$(Buermans and den Dunnen, 2014); $\star$(Liu et al., 2012)

2.6.5 Gene sequencing approaches

commentar 6, Mur- ray, set genes are not in- cluded any- more

Some genes are highly conserved among bacteria and can therefore be used for identification and phylogenetic classification of bacteria (Stackebrandt and Goebel, 1994; Woese, 1987). The gene of choice when only sequencing one gene is the 16S rRNA gene. It is highly conserved among bacteria because it is essential for the bacteria's survival (Woese, 1987). Other less conserved genes, particularly the ones under purifying selection pressure, have been used for species identification and strain typing. *Rpo*b, which encodes the β subunit of RNA polymerase, has been used for bacterial species identification and subtyping, as it often shows higher discriminatory power than 16S (Adékambi et al., 2006; Korczak et al., 2006; Mollet et al., 1997). Another gene, *gro*EL (also called *cpn*60 or *hsp*60), which encodes the universal 60-kDa chaperonin, has been used widely to distinguish between *Campylobacter* species and closely related bacteria (Chaban et al., 2009; Hill et al., 2006; Kärenlampi et al., 2004). The limitation of using genes encoding surface proteins or virulence factors is that they evolve too quickly, which can lead to misleading results. A combination of genes encoding surface antigens and housekeeping genes might be a more rational approach for strain typing (Cai et al., 2002). Multilocus sequence typing (MLST) (Maiden et al., 1998) has been successfully applied to the characterisation of numerous bacteria (Achtman et al., 1999; Dingle et al., 2001; Enright and Spratt, 1998; Maiden et al., 1998; Suerbaum et al., 1998). It employs the same philosophy as multilocus enzyme electrophoresis (Selander et al., 1986) in the way that neutral genetic variation from multiple chromosomal locations is indexed, but in contrast to MLEE, MLST uses nucleotide sequence variation to identify the variation. One observation in MLST studies of different bacteria was that stretches of 400 to 600 bp from seven loci provide roughly the same discrimination as obtained with 15 to 20 loci in MLEE analyses (Maiden et al., 1998). Another advantage of MLST is that it allows the comparison of results between different laboratories and the electronic storage and distribution of the results. To date, 29 MLST schemes for different bacteria have been developed (`http://www.mlst.net/databases/default.asp`). Each gene fragment (one of the housekeeping genes/alleles) is characterised by a unique number in the order of its discovery. For example *aspA*-1 is the first unique MLST allele identified for the *aspA* locus (Maiden et al., 1998). For each of the seven loci (for *C. jejuni/coli*), distinct allelic sequences are assigned an arbitrary number and each isolate is therefore characterised by seven numbers building an allelic profile. Isolates with identical sequences were assigned the same allele number. The multilocus sequence type (ST) is defined for each strain by the combination of alleles at each locus (allelic profile). The following seven loci (Figure 2.6) were chosen for the MLST scheme for *C. jejuni* and *C. coli*: *asp*A (aspartase A), *gln*A (glutamine synthetase), *glt*A (citrate synthetase), *gly*A (serine hydroxymethyltransferase), *pgm* (also called *glm*M; phosphoglucosamine mutase), *tkt* (transketolase), *unc*A (also called *atp*A; ATP synthase α subunit) (Dingle et al., 2001). The first definition of a clonal complex (CC) was a group of two or more independent isolates with

a ST that shared identical alleles at ≥ 4 loci (Dingle et al., 2001). A CC is named after the ST identified as the putative founder of the group, using BURST (Dingle et al., 2001). At the time of writing (August 2014), 7318 STs were identified and the numbers of alleles within ranged from 407 for the *asp*A locus to 719 for the *pgm* locus (Table 2.2).

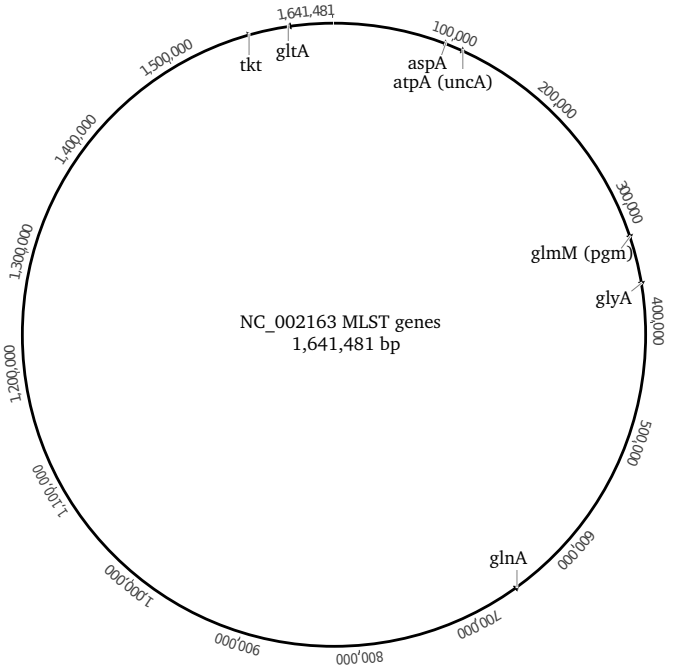


Figure 2.6: Distribution of the seven housekeeping loci in the *C. jejuni* genome NCTC 11168.

Locus	Fragment size (bp)	No. of alleles
*asp*A	477	407
*gln*A	477	551
*glt*A	402	472
*gly*A	507	632
pgm	498	719
tkt	459	568
*unc*A	489	484

Table 2.2: Genetic diversity of the seven housekeeping genes used in MLST scheme for *C. jejuni* and *C. coli* (last accessed 13/08/2014)

MLST analysis has shown that in many species recombination contributes more to clonal diversification of the bacteria, than point mutation does (Feil and Spratt, 2001; Wilson et al., 2009). So far it has been successfully applied to more than 29 bacterial species and is seen as a

reference genotyping method for many bacteria. Since the first description of an MLST scheme for *C. jejuni* (Dingle et al., 2001), additional schemes for other *Campylobacter* spp. have been developed, including *C. coli*, *C. lari*, *C. upsaliensis*, *C. helveticus* (Miller et al., 2005), *C. fetus* (Van Bergen et al., 2005), *C. sputorum*, *C. hyointestinalis*, *C. curvus* and *C. consisus* (Miller et al., 2005). The phylogenetic signal obtained from MLST analysis, combined with analytical methods, made attribution of sporadically reported infections to sources possible (On, 2013).
MLST has been successfully applied to a wide variety of investigations, including, source attribution studies in New Zealand (Müllner et al., 2009) and the Netherlands (Gras et al., 2012); to identify seasonal patterns associated with specific CCs or STs (Cody et al., 2012; McCarthy et al., 2012; Sopwith et al., 2008); exploring the origin of antibiotic-resistance in *C. coli* in antimicrobial-free swine production systems (Quintana-Hayashi and Thakur, 2012), for the characterisation of the genetic diversity in *C. upsaliensis* (Parsons et al., 2012) and MLST, PFGE and serotyping were used to explore the distribution of a highly pathogenic clone of *C. jejuni* associated with sheep abortion across both time and space (Sahin et al., 2012).
Multispacer typing (MST) is a genotyping method that was first applied to *Yersinia pestis* in 2004. The typing is based on the highly variable intergenic spacers (Drancourt et al., 2004). The method is based on the assumption that the noncoding regions are under less selection pressure than genes are, and that the spacers should therefore vary more between bacterial strains (Drancourt et al., 2004). The intergenic spacer are chosen to be the noncoding regions varying the most between aligned genomes of one species (Drancourt et al., 2004; Fournier et al., 2004; Li et al., 2006). After amplification and sequencing, similar to MLST, any variation in a spacer can provide a unique ST (Li et al., 2009b). The technique has been succesfully applied to several bacteria including *Y. pestis* (Drancourt et al., 2004), *Rickettsia conorii* (Fournier et al., 2006), *Coxiella burnetti* (Glazunova et al., 2005) and others. Studies comparing MST and MLST have shown that MST has a higher resolution and the potential to be used directly on non-cultured samples (Fournier et al., 2004; Li et al., 2006). However, MST has so far only been used on intracellular human pathogens, and the fact that other bacteria (e.g. *Campylobacter* spp.) experience a high amount of recombination and mutation, may make this method unsuitable for other bacteria.

Ribosomal MLST (rMLST) can be seen as an extension of MLST in the sense that this method is using 53 housekeeping genes instead of seven. It is based on the 53 genes encoding the bacterial ribosome protein subunits (*rps* genes) (Jolley et al., 2012). It invokes the same principle as MLST employing curated reference sequences to identify the gene variants. However, it is synthetic, meaning that the whole genome has to be sequenced in order to extract the (rMLST) genes. *Rps* loci are distributed around the genome and similar to housekeeping genes under selective pressure, which makes them highly conserved. This method has been shown to exhibit significantly more resolution than 16S rRNA (Jolley et al., 2012).

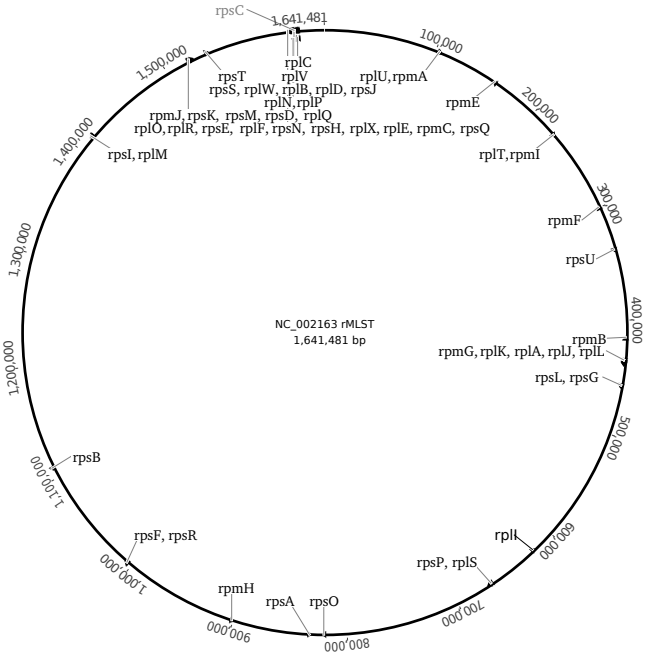


Figure 2.7: Distribution of the 52 ribosomal housekeeping loci in the *C. jejuni* genome NC_002163.

Depending on the questions, there are different 'ideal' genotyping approaches to answer them. 16S rRNA is a suitable typing method for bacterial differentiation on the levels of Phylum, Class, Order, Family and Genus, for further discrimination different methods are more suitable. MLST is able to differentiate bacteria on the levels of the Genus, Species and the Lineage or Clonal Complex, whereas rMLST covers the range from Species to Strain. The highest discriminatory power lies in the whole genome sequencing method which can go down to the levels of Meroclone (group of organisms that are descended from a single cell but have started to diversify by recombination) and Clone (Maiden et al., 2013). MLST could be seen as the new 'gold standard' for subtyping the majority of *Campylobacter* species. Its epidemiological value, portability, taxonomic range and international usage make it the ideal typing method (On, 2013). However, MLST is not the choice for all *Campylobacter* species. *C. fetus* for example, appears to be highly clonal (Van Bergen et al., 2005) and therefore MLST may not provide the discriminatory power needed to identify subspecies. Other methods such as PFGE or AFLP may be better suited (On and Harrington, 2000, 2001).

Nowadays there is a high demand for fast and low-cost sequencing which has driven the development of high-throughput sequencing (or next-generation sequencing) technologies that parallelise this process. The Sanger sequencing method produces long reads with a maximum of 0.44 Mb

per run whereas pyrosequencing produces short reads with an output of approximately 4Gb per run (Li et al., 2009b). Pyrosequencing is more difficult to use for *de novo* assemblies due to the shorter read length. Although the sequencing technologies have increased greatly (reduced in price, increased in read length and speed), genome sequencing remains too expensive and time consuming to be used in routine genotyping.
Third generation NGS platforms have a number of advantages over the second generation NGS platforms. Not only do they not rely on PCR amplification, they also produce higher throughput; increased read length (essential to enhance *de novo* assembly); require less starting material (theoretically only a single molecule); faster turnaround time and higher consensus accuracy. However, some of the third generation sequencers are still under development (Oxford Nanopore) or show high error rate for single reads (SMRT). The choice of the sequencer depends highly on the application it is intended for, the budget available and the availability of sequencers.

2.6.6 Phenotype MicroArray (PM) system

Different '∼omics' techniques provide great insights into the biology of cellular organisms and, depending on the method, they address different steps in the information transfer, from coding DNA (genomics), via RNA (transcriptomics) to proteins (proteomics) to generate the cellular metabolites (metabolomics) (Buchanan et al., 2009; Vilchez-Vargas et al., 2010; Zhang et al., 2010). All of these addressed '∼omics' techniques study components which contribute to the phenotype (Papin et al., 2004). However, the phenotype itself was until recently not accessible with high-throughput '∼omics' techniques.
Taking advantage of fully sequenced bacterial genomes, a new technique, called microbial genome-wide association study (GWAS) is providing the means to investigate the correlation of genetic variants with phenotypic traits across several bacterial strains enabling the potential discovery of virulence or host-associated genes. In order to correlate phenotypes to specific genotypes, for example via presence/absence of genes, one must be able to consistently measure the phenotypes. A simple way to do this would be to characterise an organism's growth behaviour under specific conditions (Neysens et al., 2003; Sikorski and Nevo, 2007). A few studies attempted to correlate phenotype and genotype via comparative genome hybridisation (CGH) (Malloff et al., 2001; Salama et al., 2000; Willenbrock et al., 2006) but the limitation is that the genes in the query genome can only be detected in relation to the reference sequence.
While the cost of genome sequencing dropped significantly over the years, only a recent development of a Phenotype MicroArray (PM) system enabled global phenotyping of bacterial strains (Bochner, 2003; Bochner et al., 2001). The OmniLog instrument developed by Biolog (Hayward, CA, (Bochner et al., 2001) is an integrated system of cellular assays, instrumentation, and bioinformatic software for high throughput screening of up to 2000 cellular phenotypes simultaneously

(Bochner, 2003, 2009). A physiological reaction (respiration) produces Nicotinamide Adenine Dinucleotide Hydrogen (NADH) causing a redox potential and flow of electrons to reduce a dye, for *Campylobacter*, tetrazolium violet (Bochner and Savageau, 1977). This results in a colour change over time which is recorded by a CCD camera every 15 min. The metabolic flow, i.e. cellular respiration, is linearly correlated to the formation of a purple colour (Bochner, 2009; Bochner et al., 2001). The single steps of how to culture, measure the turbidity, inoculate the bacteria and finally obtain the results, are shown in the workflow in Figure 2.8. Figure 2.9 shows an overview of the 20 PM plates and the associated tests that can be used with the Omnilog instrument.

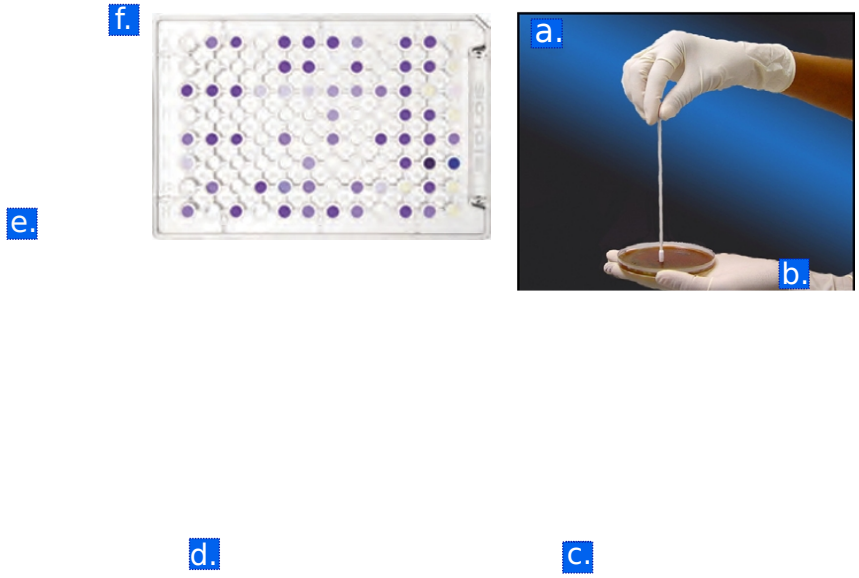


Figure 2.8: This figure shows an overview of steps how to obtain high-throughput results from the Omnilog. Panel a) to d) shows how the cells are being cultured (a.), the cell suspension being prepared (b.), the turbidity is being measured (c.) the turbidity is being measured (d.) the suspension is inoculated into the microarrays (PM plates). Panel (e.) shows the PM panel set being placed inside the OmniLog- where a CCD camera cycles every 15 min to record the colour formed from reduction of the tetrazolium dye in each well. Panel (f.) shows an example of a PM plate after the run.

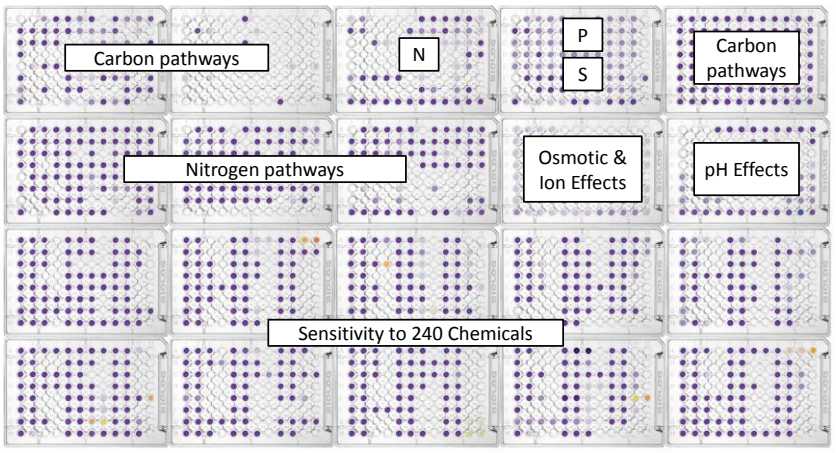


Figure 2.9: Overview of the 20 PM plates. Plot re-created from (Bochner, 2011)

C. jejuni is known to be sensitive to environmental stress (Park, 2002), therefore, questions remain as to how *C. jejuni* is able to adapt to and colonise different hosts. A few studies have used the Omnilog system to test for specific phenotypes of *C. jejuni* reference genomes. Line et al. (2010) tested the ability of the *C. jejuni* strain NCTC 11168 (GS) (a variant of the original NCTC 11168) to utilise 190 different substrates as sole carbon sources at 38°C and 42°C using phenotypic microarray (PM) technology. Their results indicate that several amino acids, as well as a number of organic acids were utilised, and that the respiration was generally greater at 42°C as compared to 38°C (Line et al., 2010). Tang et al. (2010) investigated the phenotypic profiles of *C. jejuni* strain ATCC 33560 based on carbon source utilisation and sensitivity to osmolytes at 42°C and 30°C using the same phenotypic microarray system. Utilisation was observed for carbon sources from amino acids and carboxylates but not from sugars. No phenotype loss was observed for incubation at 30°C (suboptimal temperature) within the carbon source plate, but less tolerance was observed in the sensitivity to specific osmolytes and growth in pH 5 (Tang et al., 2010). Previous experiments report that *C. jejuni* is sensitive to 2% and greater concentrations of NaCl (Cameron et al., 2012; Tang et al., 2010) but showed tolerance to a wide range of food preservatives (sodium lactate, sodium phosphate, sodium benzoate, ammonium sulphate and sodium nitrate) (Tang et al., 2010).

2.7 Epidemiology

New Zealand had an average annual rate of (notified) human campylobacteriosis cases of 353.8/100,000 between 2002-2006 (The Institute of Environmental Science and Research Ltd., 2013) with a peak of >380 per 100,000 population in 2006 (The Institute of Environmental Science and Research Ltd., 2007). The complex epidemiology of campylobacteriosis (Dingle et al., 2002; French et al., 2005; Müllner et al., 2010a; Wilson et al., 2008) has largely complicated the development of effective disease-control measures in New Zealand and elsewhere. However, a partial success was achieved by introducing, in collaboration with the poultry industry, an intervention involving a range of control measures to reduce the *Campylobacter* spp. contamination of poultry (New Zealand Food Safety Authority, 2008; Sears et al., 2011). *C. jejuni* and *C. coli* are the major cause of human campylobacteriosis cases (>90%) (Carter et al., 2009; Gillespie et al., 2002; Lee and Newell, 2006; Nachamkin et al., 1992). *Campylobacter* spp. are commensals in many wild, farmed and domestic animals. The risk factors for human infection range from consumption of undercooked chicken (Eberhart-Phillips et al., 1997; Gras et al., 2012; Harris et al., 1986; Kapperud et al., 1992; Müllner et al., 2009, 2010b; Neal and Slack, 1997), consumption of offal (mainly chicken and sheep liver) (Lake, 2006), raw dairy products (Blaser et al., 1979; Eberhart-Phillips et al., 1997; Evans et al., 1996; Heuvelink et al., 2009; Lake, 2006) and untreated water (Adak et al., 1995; Close et al., 2008; Savill et al., 2001), to contact with pets and farmed animals (Adak et al., 1995; Eberhart-Phillips et al., 1997; Gras et al., 2013; Kapperud et al., 1992; Lake, 2006; Neal and Slack, 1997). Identifying the most important sources of human foodborne disease is important for prioritising food-safety interventions and setting public health goals (Pires et al., 2009).

New Zealand has a unique ecological situation: the extensive agricultural usage of land and the geographically remote location of the islands, result in a relatively isolated human, animal and pathogen population. The high ratio of food-producing animals to humans and the frequent use of rural water may be contributing to the high enteric infectious disease rate in New Zealand (Crump et al., 2001).

MLST has been used to reveal differences in the epidemiology of individual bacterial genotypes (Dingle et al., 2001) and allowed for the identification of source-associated strains (Sopwith et al., 2008) by informing epidemiological and population-genetic source attribution models (Mullner et al., 2009; Müllner et al., 2010b; Sheppard et al., 2009).

2.7.1 Sources of *Campylobacter*

The *Campylobacter* spp., especially *C. jejuni* and *C. coli* are remarkably successful organisms. They are detectable in a broad range of host species where they usually colonise the intestinal

tract but have also been found in a range of other tissues, including spleen and liver (Cox et al., 2007). *Campylobacter* spp. can be isolated in the absence of disease (e.g. faecal sample from cattle or chickens) (Garénaux et al., 2008b). *Campylobacter* isolates can not grow outside the host but, they are ubiquitous in the environment and recoverable from a wide range of sources such as water and soils (beach sand) where they successfully survive for long periods (Lee and Newell, 2006; Newell, 2002) often entering a 'viable but nonculturable' state (Oliver, 2005, 2010; Rollins and Colwell, 1986). During this state bacteria reduce their metabolic activity to a minimum and can not divide but have the ability to become culturable ones resuscitated (Oliver, 2005).

commentar 8, Mur- ray

Various *Campylobacter* species can be found colonising the gut and are shed in the faeces of domestic livestock including sheep, cattle, pigs (Devane et al., 2005; Newell, 2002) and poultry (Müllner et al., 2009, 2010b). *C. jejuni*, *C. coli* and *C. lari* are the most common thermophilic species colonising domestic livestock, but *C. fetus* subs. *fetus* a non-thermophilic species is also common (Garénaux et al., 2008b). Gut infections in animals are usually asymptomatic but in pregnant cattle or sheep, *C. fetus* and specific pathogenic strains of *C. jejuni* (Sahin et al., 2012) can cause abortion or infertility (Garénaux et al., 2008b; Herve et al., 2004; Krause et al., 2002). *C. jejuni* has been identified as successful commensal in chicken, cattle and sheep (Devane et al., 2005; Newell, 2002) whereas *C. coli* is more associated with pigs and sheep (Brown et al., 2004; Nesbakken et al., 2003). *C. hyointestinalis* and *C. mucosalis* are found in pigs whereas *C. upsalinesis* and *C. helviticus* are more associated with cats and dogs (Chaban et al., 2010; Gow et al., 2009; Hald and Madsen, 1997; Moreno et al., 1993). *Campylobacter* species have also been identified in wild birds, rabbits, other wild mammals, molluscs, rivers and sewage (Humphrey et al., 2007; Jones, 2001). Recently Humphrey et al. (2014) has shown that *C. jejuni* can cause infections in poultry.

2.7.2 Source attribution

A recent systematic review (Domingues et al., 2012) aimed to identify the main risk factors of human campylobacteriosis through a meta-analysis of case-control studies of sporadic infections. This study included case-control studies from 14 different countries conducted between 1983 and 2004. Domingues et al. (2012) estimated international travel to be the highest risk factor for human campylobacteriosis, followed by consumption of undercooked chicken, environmental exposure to *Campylobacter* and direct contact with farm animals. Amongst the food transmission routes, consumption of undercooked chicken and unpasteurised dairy products were identified as the main sources for human campylobacteriosis (Domingues et al., 2012).

A sentinel surveillance site was founded in the Manawatu region of New Zealand's North Island in 2005. Human campylobacteriosis cases and potential sources were sampled and *C. jejuni*

isolates typed by multilocus sequence typing (MLST) simultaneously over consecutive years (McTavish et al., 2008; Müllner et al., 2009) thereby informing source attribution studies and providing strong evidence that a large proportion of human cases were linked to poultry consumption. Additionally, a dominant poultry-associated *C. jejuni* sequence type (ST-474) was identified, that has been reported rarely from other countries. Molecular epidemiological studies, largely based on the data from the sentinel site, have contributed to the understanding of the epidemiology of campylobacteriosis (McTavish et al., 2008; Muellner et al., 2013; Mullner et al., 2009; Müllner et al., 2010a,b) in New Zealand. Müllner et al. (2010b); Spencer et al. (2012) identified the association between ruminant-associated genotypes, and pre-school-age children (0-5 years) in rural areas, as being direct contact with faecal material, evidence of which has also been reported in other countries (Strachan et al., 2013).
Source attribution studies are important for the development and evaluation of appropriate, country-specific control strategies to lessen the human disease burden. After the poultry intervention, the number of poultry-associated cases have fallen and source attribution studies estimate that there has been a relative increase in the importance of ruminant-associated strains of *C. jejuni* (French et al., 2010; Muellner et al., 2011, 2013). Ongoing work is investigating the complex epidemiology of *Campylobacter* in ruminants (Muellner et al., 2011) and wildlife sources (French et al., 2009). Additionally to MLST, other typing approaches are being investigated to increase resolution of the molecular analysis (Biggs et al., 2011; Colles and Maiden, 2012; Jolley et al., 2012). Incorporating next generation sequence data into epidemiological studies and surveillance means, specific questions such as the pathogen's evolution, the timing of the emergence, the origin and spread of the pathogen and aspects of between-host transmission can be answered (Kühnert et al., 2011). Biggs et al. (2011) used next generation sequencing to obtain insights into the evolution of *C. jejuni* ST-474, which is one of the most important human enteric pathogens in New Zealand, and to identify potential markers for host-association.

2.7.3 Seasonality of *Campylobacter jejuni*

Despite the complex epidemiology of campylobacteriosis (Dingle et al., 2002; Müllner et al., 2010b) and the difficulties of identifying the reasons for the seasonality of human *Campylobacter* infections (Sopwith et al., 2008), a number of aspects are known about human campylobacteriosis. Studies examining the incidence rates identified a consistent peak in summer across most temperate countries (Kovats et al., 2005; Nylen et al., 2002) and some countries also showed an early spring peak (Kovats et al., 2005; Nylen et al., 2002). Countries with milder winters display peaks earlier in the year (Kovats et al., 2005), whereas countries with tropical climates present little variation throughout the year (Strachan et al., 2013). Those findings suggest that climate might be a contributing factor to the timing of the seasonal peak. Not only is the peak

loosely associated with the highest temperature of the year (Louis et al., 2005) but other climatic variables like sunshine and precipitation have also been associated with increased human *Campylobacter* cases (Febriani et al., 2010; Louis et al., 2005; Weisent et al., 2014). However, this seems not to be true for all regions (Arsenault et al., 2012; Lal et al., 2013).
The pattern of seasonality is quite different in Australia and New Zealand, as compared to England and Wales, showing less consistency over time and a prolonged summer peak (McCarthy et al., 2012; Nylen et al., 2002). Kovats et al. (2005) also identified differences between Australia and New Zealand, showing that the seasonality was less pronounced in Australian cities and that there are differences between the North and South Island of New Zealand. Hearnden et al. (2003) identified marked differences in seasonality patterns across New Zealand, in particular, three different groups of seasonality and incidence patterns between the North and South Island of New Zealand. The first group, the Far North and rural areas of the North Island, showed a relatively low summer incidence and small inter-seasonal variation. The second group, a group of urban areas (Auckland, Hamilton, Napier) in the North Island as well as some areas in the South Island, exhibited a higher summer incidence and more seasonality compared to the first group. The highest summer incidences and strongest inter-seasonal variations were found in Wellington and surroundings (North Island) and Christchurch, Dunedin and much of the rest of the South Island (Hearnden et al., 2003). These differences could be due to the proportion of poultry- to non-poultry associated cases in these areas, with urban areas showing a higher risk of poultry-associated infections (Muellner et al., 2011).
A recent study by McCarthy et al. (2012) examining a United Kingdom, Finnish, Australian and New Zealand dataset, sought to explain the seasonality and other aspects of *C. jejuni* epidemiology. The study identified a clade formed by CC-45 and CC-283 which shows a summer peak and is present in the UK and the Finnish dataset but not in the Australian or New Zealand dataset (McCarthy et al., 2012). However, the McCarthy et al. (2012) study used a small-scale dataset from New Zealand which was focused on a prolonged outbreak in the winter of one year (McTavish et al., 2008) instead of one of the larger published datasets (Müllner et al., 2009).

2.8 Bacterial Evolution

Evolutionary biology is the study of the diversity of life at every level of biological organisation, and the processes responsible for shaping it. Evolution is the change of inherited characteristics of populations over consecutive generations. The evolutionary processes shape diversity at every biological level, including species, individual organisms and molecules. Evolution is generally acting on two different levels and timescales. On the population level and over a considerably short timescale (days to month or years), studies investigate how traits become well adapted to

the environment. Over a longer timescale (decades to hundreds or thousands of years), evolutionary processes produce the diversity of life (divergence of eukaryotes, bacteria and archaea). The evolutionary processes that occur rapidly within population over a short time scale are called microevolution, whereas the processes that occur slowly and over a longer timescale are called macroevolution. Evolutionary biology aims to explain the diversity, adaptation and complexity of a given population.

2.8.1 Mechanisms of bacterial evolution

Genetic mechanisms

Genetic recombination is the process of exchanging genetic material between bacterial strains. Three mechanisms contributing to recombination are known in bacteria. Horizontal gene transfer (HGT) which is sometimes called lateral gene transfer (LGT), is the transfer of DNA between and within species and has been well documented (Medini et al., 2005; Sheppard et al., 2008; Suerbaum et al., 1998; Wilson et al., 2009). HGT in bacteria is, in contrast to meiosis-associated recombination, unidirectional and not necessarily associated with cell division. The importance of HGT in evolutionary processes has been debated, but it is now generally accepted and seen as an evolutionary 'fast route', which enables an organism to quickly adapt to a changing environment (Medini et al., 2005). The three main mechanisms used by bacteria to acquire new DNA are transformation, conjugation, and transduction (Figure 2.10). Transformation involves acquisition of DNA from the environment, conjugation involves acquisition of DNA directly from another bacterium, and transduction involves acquisition of bacterial DNA via a bacteriophage.

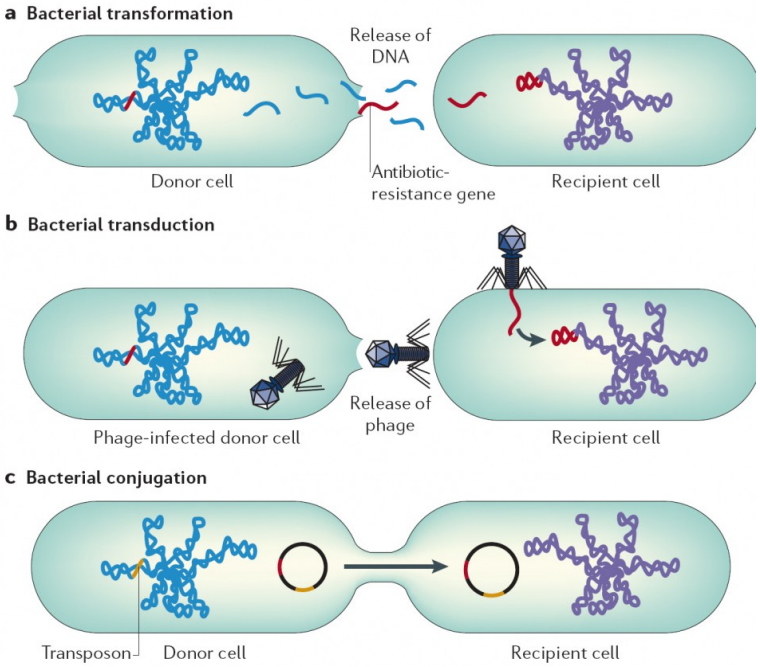


Figure 2.10: Overview of the three evolutionary mechanisms ((a) transduction, (b) transformation, (c) conjugation) for acquisition of foreign DNA used by bacteria. Image obtained from: `http://scienceofacne.com/wp-content/uploads/Horizontal-Gene-Transfer-in-Bacteria-Furuya-et-al-950x1247.jpg`

Another way the sequence of a bacterium can be altered is through mutation. Mutations can result from unrepaired damage to DNA or errors in the process of replication. Mutations can be divided into small-scale mutations (point mutations, indels (either insertions or deletions)) and large-scale mutations (duplications, inversions, translocations). Point mutations (also called single nucleotide polymorphisms (SNPs)) can be either synonymous (code for the same amino acid), nonsynonymous (code for a different amino acid) or nonsense mutations (translate into a stop codon) and may therefore not change the translated amino acid and have no functional influence on the phenotype of the organism (Freese, 1959a,b). SNP insertions or deletions in protein coding sequences will result in a frameshift that will affect the downstream region, so that the codons are being translated in a different reading frame (Hu and Ng, 2012). This can lead to significant alterations of the encoded protein and therefore to gain or loss of the protein's function. Point-mutations in non-protein-coding sequences can lead to functional consequences when they affect a regulatory element. Depending on the nature of the mutation (deleterious, advantageous or neutral mutations), the fitness of the organism may be affected. By studying the type, function, rate and distribution of SNPs in a genome, one can obtain insights into the evolutionary pressure affecting bacterial genomes.

Phase variation, contingency loci, bet hedging and bi-stability

The term 'contingency' locus refers to a region of hypervariable DNA that mediates high-frequency, stochastic genotypic switching. These DNA stretches consist of iterations of contiguous nucleotides (such as guanines (G's)) termed homopolymeric tracts. The hypervariable regions are prone to high rates of mutation through slipped strand mispairing during replication (Levinson and Gutman, 1987), which can change the open reading frame of a gene. 'Phase variation' is defined as the high-frequency gain or loss of a phenotype resulting from changes of expression of single or multiple genes. Therefore the mutations in the homoploymeric tracts are generally considered to be a mechanism for phase variation, controlling the ON/OFF switch for phenotypic expressions. Contingency loci are thought to be a bet-hedging strategy frequently involved in adaptations to new environments and hosts (Rainey et al., 2011), as cells derived from a single ancestral cell can possess several inherited phenotypes (Moxon et al., 2006), this was also been shown experimentally (Beaumont et al., 2009).
When the *C. jejuni* genome NCTC 11168 was sequenced in 2000, it was found to have a high level of genetic variation affecting translation of over 23 contingency loci (such as *Cj0031*, *Cj0032*, *Cj1420*, *Cj1426*) (Dorrell et al., 2001; Parkhill et al., 2000). These genes group into three regions of the genome and are mainly associated with lipo-oligosaccharide (LOS) biosynthesis, capsule biosynthesis and flagellar modification (Parkhill et al., 2000). Jerome et al. (2011) re-sequenced the *C. jejuni* genome NCTC 11168 after several passages through a C57BL/6 IL-$10^{-/-}$ mouse model which allowed them to compare the variation in the 23 contingency loci before and after *in vivo* adaptation. They noticed that mutations were largely restricted to the homoploymeric tracts of thirteen contingency loci. These mutations caused frameshifts in genes associated with surface structure biosynthesis loci and genes with unknown functions (Jerome et al., 2011).

Molecular clock

The first scientists to notice an evolutionary rate of proteins and therefore the existence of a molecular clock were Zuckerkandl and Pauling (Zuckerkandl and Pauling, 1965, 1962) cited in (Lemey and Posada, 2009). Zuckerkandl and Pauling made the observation that the genetic distance for two sequences coding for the same protein isolated from different species, appeared to increase linearly with the divergence time. This observation seemed to be true for different proteins and they hypothesised that the rate of evolution is constant over time. This also means that if the rate of evolution for a gene can be calculated, then this information can provide the divergence time between two species.
The idea is termed as the molecular clock which can formally be used to describe the substitution process which is important for our understanding of the evolutionary process (Lemey and

Posada, 2009), and is in agreement with the neutral theory of evolution published by Kimura (1968, 1984). The neutral theory of molecular evolution states that on the molecular level most changes are caused by random genetic drift of alleles that are neutral or nearly neutral (Kimura, 1984). This also implies that deviances from molecular clock-like behaviour may imply adaptive evolution (Lemey and Posada, 2009). However, this null-model may not be appropriate for bacteria such as *C. jejuni* as the calculated recombination/mutation ratios argue against the neutrality theory but rather for natural selection. The assumption of natural selection has also been used by (Wilson et al., 2008) who attempted to calculate a molecular clock using a population genetic model of the forces of drift, mutation, recombination and natural selection. The general approach to calculate a molecular clock is to take a measure of genetic distance between species and then use a calibration rate to convert the distance into time. The reliability of the molecular clock depends on the accuracy of the genetic distance and the appropriateness of the calibration rate (Bromham and Penny, 2003). There are different ways to estimate the genetic distance. The simplest way is to count the number of changes between the DNA (or protein) sequences. However, this approach does not consider repeated changes (Figure 2.11 b-f).

commentar 9, Mur- ray

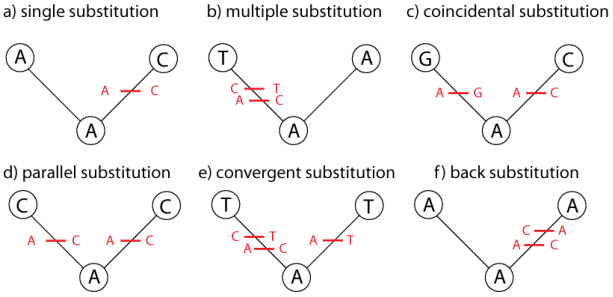


Figure 2.11: The different substitutions that can occur in a nucleotide sequence.

Therefore, a more complex substitution model is necessary to calculate the true number of changes that have occurred (Page and Holmes, 1998). A substitution model has not only to consider the relative frequency of different types of substitutions, but also the variation in substitution rates in sites. Therefore the selection of the appropriate substitution model is crucial for the accuracy of molecular dating (Bromham and Penny, 2003). However, all the substitution models share the same underlying assumptions (Lemey and Posada, 2009):

1. in a sequence, the rate of change from base i to base j is independent from the previous base (Markov property)

2. the rates do not change over time (homogeneity)

3. relative base frequencies ($\pi_A, \pi_G, \pi_T, \pi_C$) are at equilibrium (stationarity)

The simplest possible nucleotide substitution model is the Jukes-Cantor model (JC69, Jukes and Cantor, 1969) (Jukes and Cantor, 1969; Lemey and Posada, 2009).It not only assumes equal base frequency ($\pi_A = \pi_G = \pi_T = \pi_C = \frac{1}{4}$) but also equal mutation rates (nucleotides have the same probability to be replaced by any other nucleotide). In order to obtain the transition matrix p, $P_{ii(t)}$, the probability of a nucleotide remaining the same during the evolutionary time t, and $P_{ij}(t)$, the probability of replacement have to be calculated.

comment 11, Mur-ray

commentar 10, mur-ray

P=

1-3α	α	α	α
α	1-3α	α	α
α	α	1-3α	α
α	α	α	1-3α

(a) transition matrix for the Jukes-Cantor model

$$P_{ii} = 1 - \sum_{j \neq i} P_{ij} \tag{2.1}$$

(b) probability estimation for the Jukes-Cantor (69) model

Figure 2.13: transition matrix and probability estimation for the Jukes-Cantor (69) model

However, if all of the eight parameters (different rates for transition, transversion and different mutation rates) of a reversible nucleotide rate matrix P are specified, it results in the general time reversible model (GTR) (Figure 2.14)

P=

·	π_Ca	π_Gb	π_Tc
π_Aa	·	π_Gd	π_Te
π_Ab	π_Cd	·	π_Tf
π_Ac	π_Ce	π_Gf	·

Figure 2.14: transition matrix for general time-reversible models

Every model can be derived from the GTR model by imposing different constraints on it. For example, nucleotide exchanges can be grouped into two groups. Substitutions from A $\leftrightarrow$ G and C $\leftrightarrow$ T are called transitions and all other substitutions are called transversions (Figure 2.15). If $\gamma = 1$ and therefore the purine and pyrimidine transitions have the same rate, this model reduces to the HKY85 model. For equal base frequencies of $p_i = \frac{1}{4}$ the HKY85 model (Hasegawa et al., 1985) reduces further to the Kimura 2-Parameter model (Kimura, 1980).

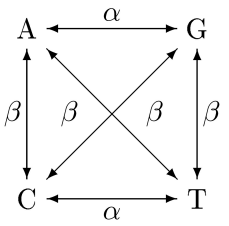


Figure 2.15: transitions (α): A $\leftrightarrow$ G and C $\leftrightarrow$ T; transversions (β): every other nucleotide substitution

For many analyses, especially for longer evolutionary distances, evolution is modelled on the amino acid level. However, because not all changes on the nucleotide level alter the amino acid, information is lost. Nevertheless, there are advantages to estimate molecular evolution on the amino acid level. The most important advantage is the 'alphabet size' of the amino acids compared to the nucleotides (22 (including selenocystein and pyrrolysine) compared to 4) as it is easier to identify back substitutions with 22 characters compared to four. Another advantage is that DNA is more inclined to show compositional bias (the occurrence of the four bases in unequal proportions) than amino acids (Foster and Hickey, 1999). However, the sequence differences of two proteins over time is not linear (as for nucleotides) but takes the course of a negative exponential (Lemey and Posada, 2009). Therefore the observed percentage of differences is not proportional to the actual evolutionary difference and underestimations will occur at larger evolutionary distances.

Amino acid models are, unlike DNA models, empirical (modelling based on empirical observations instead of the mathematical relationships of the system modelled). The original measure to investigate the true evolutionary distance between two proteins is based on the Point-Accepted-Mutation (PAM) value which was introduced by Dayhoff in 1978 (Lemey and Posada, 2009). The PAM value is the number of mutations per 100 amino acids, therefore PAM-1 reflects an average change of 1% of all amino acid positions. The values for the PAM-1 matrix were estimated using 1572 changes in 71 groups of protein sequences that were at least 85% identical. Other matrices, e.g. PAM-250 (20%), are obtained by multiplying PAM-1 250 times by itself. If two compared homologous proteins have a PAM distance of $250-300$ (250-300 mutations per residue), they can evolutionary not be distinguished from two randomly chosen proteins. PAM-1 is the basis to calculate other commonly used substitution matrices, e.g. PAM-70. An example of the PAM250 matrix is shown below (Figure 2.16).

C	Cys	12																	
S	Ser	0	2																
T	Thr	-2	1	3															
P	Pro	-3	1	0	6														
A	Ala	-2	1	1	1	2													
G	Gly	-3	1	0	-1	1	5												
N	Asn	-4	1	0	-1	0	0	2											
D	Asp	-5	0	0	-1	0	1	2	4										
E	Glu	-5	0	0	-1	0	0	1	3	4									
Q	Gln	-5	-1	-1	0	0	-1	1	2	2	4								
H	His	-3	-1	-1	0	-1	-2	2	1	1	3	6							
R	Arg	-4	0	-1	0	-2	-3	0	-1	-1	1	2	6						
K	Lys	-5	0	0	-1	-1	-2	1	0	0	1	0	3	5					
M	Met	-5	-2	-1	-2	-1	-3	-2	-3	-2	-1	-2	0	0	6				
I	Ile	-2	-1	0	-2	-1	-3	-2	-2	-2	-2	-2	-2	-2	2	5			
L	Leu	-6	-3	-2	-3	-2	-4	-3	-4	-3	-2	-2	-3	-3	4	2	6		
V	Val	-2	-1	0	-1	0	-1	-2	-2	-2	-2	-2	-2	-2	2	4	2	4	
F	Phe	-4	-3	-3	-5	-5	-5	-4	-6	-5	-5	-2	-4	-5	0	1	2	-1	9
Y	Tyr	0	-3	-3	-5	-3	-5	-2	-4	-4	-4	0	-4	-4	-2	-1	-1	-2	7	10
W	Trp	-8	-2	-5	-6	-6	-7	-4	-7	-7	-5	-3	2	-3	-4	-5	-2	-6	0	0	17
		C	S	T	P	A	G	N	D	E	Q	H	R	K	M	I	L	V	F	Y	W

Figure 2.16: PAM250 scoring matrix. The highest values for replacing a residue into itself are shown for Trp and Cys, indicating that these two residues are the most conservative in homologous sequences. Image obtained from: `http://www.cryst.bbk.ac.uk/pps97/assignments/projects/leluk/project.htm`

Another method developed more recently (Henikoff and Henikoff, 1992) is BLOSUM (BLOck SUbstitution Matrix). BLOSUM methodology is calculating the probabilities based on the observed amino acid substitutions in blocks of conserved sequences found in multiple alignments (large sets of 2000 conserved amino acids). Two sequences that are identical to each other for more than a specific threshold T are clustered. Sequences are added to the cluster if they are identical to any sequence within the cluster at the same level. All sequences are then averaged leading to a series of matrices analogous to the PAM series of matrices. The number associated with a given BLOSUM substitution matrix refers to the identity level of the amino acid alignment blocks. However, the identity levels between the PAM and BLOSUM are inverted, which may lead to confusion comparing them. A PAM250 matrix which refers to approximately 80% of changed amino acids is equivalent to a BLOSUM20 (20%) matrix. A comparison between the PAM and BLOSUM matrices is displayed in Figure 2.18 below.

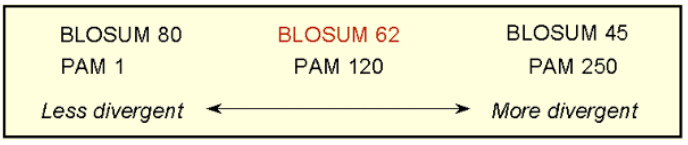


Figure 2.17: A comparison between the PAM and BLOSUM matrices. Image obtained from: www.ctu.edu.vn/ dvxe/Bioinformatic%20course/mod4/mod4_0.html

There are a few major differences between PAM and BLOSUM matrices which have to be taken into consideration when interpreting the results (Mount, 2008). For example, PAM matrices are more commonly used to score alignments between closely related protein sequences, while BLOSUM matrices are used to score alignments between evolutionary divergent protein sequences. PAM is based on global alignments, whereas BLOSUM is based on local alignment blocks. The higher numbers in the PAM matrix naming denotes greater evolutionary distance, while the higher numbers in the BLOSUM matrix refer to the sequence similarity and therefore smaller evolutionary distance. These differences have to be taken account when calculated molecular clocks.

2.9 Bacterial core- and pan genome

Due to next generation sequencing technologies and large-scale comparative genomics sequencing projects, the availability of whole genome sequences for strains of several bacterial species is steadily increasing. And due to the emerging field of singe cell genomics, even unculturable species can now be sequenced.
Comparative genomics analysis between multiple strains of a bacterial species have revealed extensive genomic intra-species diversity (Pallen and Wren, 2007). One major question is how many genomes need to be sequenced for any given bacterial species to accurately represent its entire gene repertoire. Tettelin et al. (2005) pioneered in the attempt to answer this by sequencing multiple genomes of *Streptococcus agalactiae*, followed by Hogg et al. (2007) who studied *Haemophilus influenza* genomes. One major finding of the studies was the realisation that a significant percentage of each genome was specific to each individual strain and therefore each newly sequenced genome added new genes. This was the first time the concept of a bacterial pan-genome or supragenome was mentioned. The pan-genome is the sum of the core genome and the dispensable (or accessory) genome that is composed of genes that are present in some, but not all of the sequenced strains, as well as strain-specific genes. The dispensable genome consists mostly of the genes that are not essential to the bacteria's basic lifestyle but contribute to the species' diversity and probably confer selective advantages including niche adaptation, antibiotic resistance, and the ability to colonise new hosts (Medini et al., 2005; Tettelin et al., 2008). The core genome includes all genes that are responsible for the major phenotypic characteristics of a species.

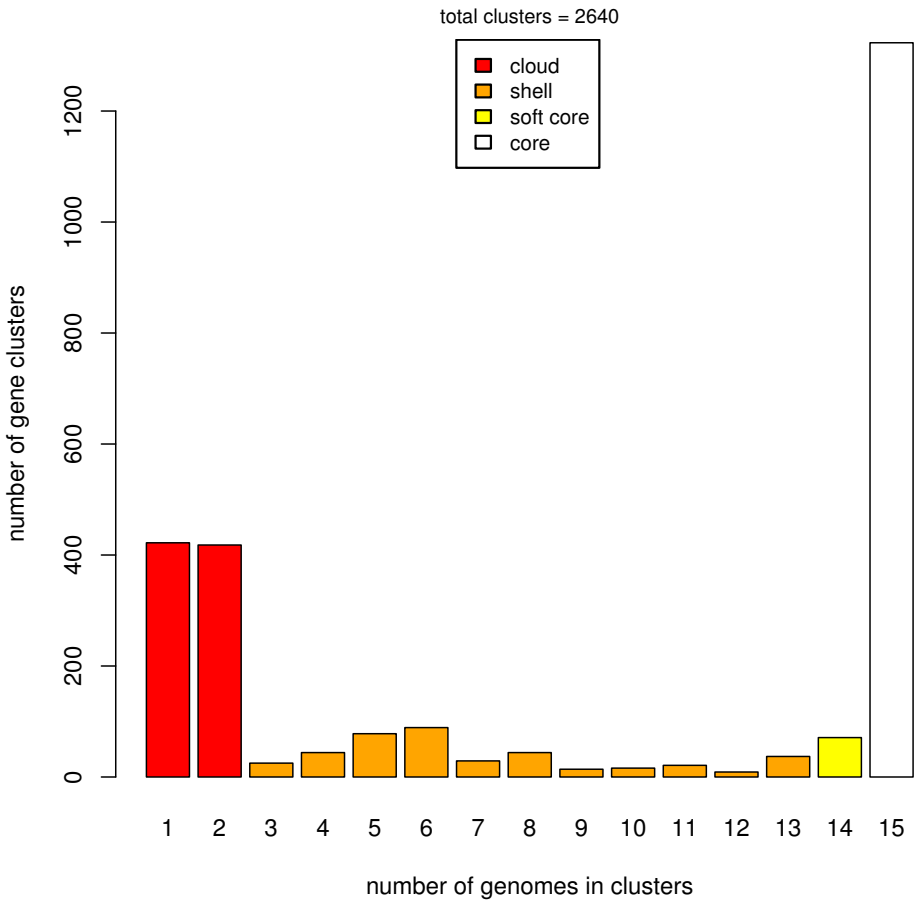


Figure 2.18: Distribution of shared genes across 15 *C. jejuni* isolates and four STs.

Multiple strains of some microbial species share only about 60% of their genes and the other 40% differ from strain to strain. The difference between the single pan-genomes of the microbial species (some have a large core genome but fewer genes in the dispensable genome, others the other way around), seem to be associated with the environment the species lives in. For example, species living in a highly variable environment will more likely have a much larger dispensable genome than species living in an isolated and static environment, which will have a much smaller dispensable genome (Reid and Buckley, 2011). Keeping a diverse pool of genetic resources has great evolutionary advantages. Due to the access to the variety of genes in the accessory genome, the species is able to respond to environmental changes without requiring each individual to carry the whole accessory genome. The dispensable genome can be used to characterise the gene sets in organisms living in a given environment, and identify the genes or functions related to niche adaptation. Previous studies have revealed that many bacterial genomes have a bias towards deletion events (Mira et al., 2001; Moran, 2002), which can result in pseudogenes (non-functional relatives of genes that have lost their protein-coding ability) and gene loss (Toft and Andersson, 2010). Patterns of genome reduction due to gene loss are generally observable in bacteria that have made a transition from facultative to obligate pathogens. This concept has only become

comment 12, Murray

known with whole genome sequencing and is therefore only been tested passively (McCutcheon and Moran, 2012; Moran, 2002).

comment 15, Murray

Pan-genome of *C. jejuni* and *C. coli*

There have been many discussions and speculations about the origin, composition and size of bacterial pan-genomes, and whether they are of finite size or not (Lapierre and Gogarten, 2009; Tettelin et al., 2008). Lefébure et al. (2010) studied the pan-genomes of *C. jejuni* and *C. coli* and demonstrated that their pan-genomes are finite and that they contain unique and cohesive features which define their genomic identity. They have noticed that both species have a similar pan-genome size, but *C. coli* has acquired a larger core-genome than *C. jejuni*, and both species have evolved a number of species-specific core genes which could reflect the different adaptive strategies (Lefébure and Stanhope, 2009). Each taxon of *C. jejuni* and *C. coli* (Lefébure et al., 2010) has a pan-genome size of around 2,600 genes, while the pan-genome of *C. jejuni* and *C. coli* combined reaches roughly 3,000 genes. Lefébure et al. (2010) showed that while sequencing 40 genomes, for each newly sequenced genome only about two new *C. coli* and three *C. jejuni* genes were discovered. Based on their analysis Lefébure et al. (2010) concluded that the pan-genome of *C. coli* and *C. jejuni* is finite and that the sequencing of any further genomes (80+) would only reconstitute a few putative pseudogenes. Comparing the pan-genomes of the two sister species, there is an indication that the greatest overlap involves the shared core genome, followed by the shared dispensable genes and more distantly by the specific components of each species. In conclusion, the two sister species have a gene repertoire which is almost completely overlapping.

Sheppard et al. (2008) reported that approximately 18.6% of the unique alleles of the seven housekeeping genes (loci) found in *C. coli* isolates may have been recently imported (through HGT) from the sister species *C. jejuni*. The results were based on 4,507 *Campylobacter* isolates, which were genotyped based on the MLST scheme (Dingle et al., 2001)). These isolates contained 2,917 unique STs. Sheppard et al. (2008) used the program STRUCTURE (Falush et al., 2003) to assign the unique STs to either *C. coli* or *C. jejuni*. When a ST could not be assigned to one species it was marked as a hybrid (the result of interspecies transfer of the gene) and the number of hybrids was reported. Based on this study Sheppard et al. (2008) suggested that *C. jejuni* and *C. coli* are converging or "despeciating".

The core genome hypothesis (Lan and Reeves, 2001) suggests that a bacterial species is defined by the core genome as a primary cohesive unit. To test this hypothesis, Lefébure et al. (2010) applied several statistical methods. The overall results of these applied methods support the idea that in the genus *Campylobacter*, recombination is frequent within the core genome of the species, rare between the two sister species and extremely rare with other species. This is in

contrast to the findings from Sheppard et al. (2008) and has been debated by several groups since (Caro-Quintero et al., 2009; Cohan and Koeppel, 2008; Doolittle, 2008; Lefébure et al., 2010). Caro-Quintero et al. (2009) used the same *Campylobacter* MLST data set as Sheppard et al. (2008) and showed that the majority of the STs do not contain imported alleles and therefore, there is no obvious support for the species convergence hypothesis. Additionally, and in agreement with these findings, analyses of the available *Campylobacter* spp. MLST sequences revealed that the interspecies genetic exchange is limited and strongly biased toward a few genes under purifying selection (mostly housekeeping genes).This housekeeping genes have been found to be exchanged between the two sister species only in rare "hitch-hiking" events (Caro-Quintero et al., 2009) which are associated with the horizontal gene transfer of adaptive genes. Therefore, according to Caro-Quintero et al. (2009), there is a clear species separation between *C. jejuni* and *C. coli* and it is unlikely that this boundary is being eroded. However, another study by Sheppard et al. (2013a) aimed to strengthen the hypothesis stated in (Sheppard et al., 2008). Sheppard et al. (2013a) used 30 whole genome sequences of *C. jejuni* and *C. coli* and showed that results from three different analyses (results of the STRUCTURE analysis, pairwise comparison of nucleotide differences between genomes and maximum likelihood trees based on ClonalFrame analysis) provided support for their earlier findings in (Sheppard et al., 2008). Sheppard et al. (2013a) explained that the disagreeing results with the study conducted by Lefébure et al. (2010) were due to differences in sampling and methodology rather than biological. Lefébure et al. (2010) rarely included *C. coli* isolates belonging to clade 1 which is where the introgression is happening (Sheppard et al., 2008). This matter will require further investigation.

comment 13, Murray

2.10 Niche adaptation and host association of *Campylobacter jejuni*

Bacteria have complicated systems for detection of nutrient levels in the environment and to adjust their metabolisms to adapt to changes in food sources, temperature and environmental stressors. Many bacteria produce metabolites and other repressive molecules, called bacteriocins (Lee and Newell, 2006), which help them to compete for a niche in the microbiota (community of microorganisms). The interaction with the microbiota is essential for effective colonisation, therefore bacteria must be able to competitively interact with it. The environmental stress *Campylobacters* can be exposed to are extreme, with temperature, humidity and osmolarity, and the presence of sunlight and atmospheric oxygen being some of the stressors. The obvious success of episilonproteobacters in colonising such a diversity of hosts and surviving in a huge range of environments is a sign of their ability to detect, adapt and evolve in response to such environments (Lee and Newell, 2006). To be evolutionary successful, an organism needs

to be able to continually adapt to changes in the environment. This process is called 'niche adaptation' which is accomplished by genetic mechanisms including HGT, gene duplication and gene shuffling. *Campylobacter* species can usually be found in the gastrointestinal tract of a wide variety of hosts, e.g. poultry, ruminants, wild birds, pets but also in environmental sources like water or beach sand (Devane et al., 2005; Humphrey et al., 2007; Jones, 2001; Müllner et al., 2009, 2010b; Newell, 2002). Why *Campylobacter* spp. are virulent in humans but not animals has not been discovered yet, but it is believed that both bacterial and host factors may contribute to it (Zilbauer et al., 2008). Our understanding of the factors which influence the ability of *C. jejuni* to colonise specific hosts or to survive in the environment is still poor, therefore several studies in the past have attempt to determine the prevalence of specific clones among *C. jejuni* isolates from diverse sources (agriculture, wild birds and environment) by applying multilocus sequence typing (Colles et al., 2003; Dingle et al., 2001; French et al., 2005; Karenlampi et al., 2007; Manning et al., 2003; McCarthy et al., 2007; Sails et al., 2003; Sheppard et al., 2009; Taboada et al., 2008; Wilson et al., 2008). Large variation has been observed between the host distribution of MLST clonal complexes. CC-21 for example, is widespread, while others like CC-61 have a more restricted distribution (Colles et al., 2003; French et al., 2005). Sometimes the different distributions are also referred to as 'generalists' or 'specialists' (Hepworth et al., 2011). McCarthy et al. (2007) studied the host associated genetic import in *C. jejuni* and confirmed the association of *C. jejuni* genotypes with host species and made the observation that host association appears to be stronger than temporal and geographical effects (McCarthy et al., 2007).

It is generally assumed that *Campylobacter* species are poor survivors in the environment without their animal host. Nevertheless a study of *C. jejuni* in a cattle farmland area in the UK (French et al., 2005) showed that isolates of the CC-45 appear to survive better in the environmental water as they were more frequently isolated from it than other clonal complexes. Studies focusing on the environment and wildlife have also discovered novel sequence types from environmental water and wildlife (rabbits/birds) which have not been identified among human isolates (French et al., 2005; Levesque et al., 2008). These water and wildlife (WW) isolates form a novel *C. jejuni* group and the transmission between water and wildlife of strains is consistent with the distribution of the group, which indicates that these strains may have adapted to survive in these niches (Hepworth et al., 2011). Williams et al. (2010) identified isolates specifically associated with bank voles (*Myodes glareolus*) characterised by a novel sequence type (ST-3704). The WW group and the clonal complex ST-3704 seem to represent niche specialists, in contrast to 'generalists' like CC-45, which can be isolated from both diverse environmental and animal hosts and humans.

Bacteria which undergo a specialisation or adaptation to a niche/host experience a significant reduction of their genome (gene loss). Multi-host bacteria on the other hand tend to keep more

genes in order to respond quickly to environmental changes. *C. jejuni*, in contrast to some other *Campylobacter* species (such as *C. fetus, C. curvus and C. hominis*) has a relatively small genome size, which indicates a genome-reduction reflecting a lifestyle closer to specialised than generalist bacteria (Moran, 2002). Researchers have shown interest in characterising genetic variation between isolates of *C. jejuni* concerning those genes relevant to niche association, host specification and severity of disease. High levels of genome variety but low levels of plasticity in *C. jejuni* have been identified in the genome of *C. jejuni* strain NCTC 11168 (Dorrell et al., 2001; Leonard II et al., 2003; On et al., 2006; Pearson et al., 2003). Despite its small genome, *C. jejuni* exhibits significant levels of variation which could be an indicator of evolution leading to niche specialisation (Hepworth et al., 2011). Hepworth et al. (2011) demonstrated for the first time that the WW (water and wildlife) and ST-3704 isolates contain significant diversity in *C. jejuni*, which enabled them to enlarge the pan-genome of *C. jejuni*. Based on the CGH data used in this study they discovered that the separation between WW and ST-3704 isolates was due to a number of genes or gene clusters where deletion or divergence was obvious. The WW isolates clustered closely with isolates of a non-livestock clade (Champion et al., 2005), in particular the environmental beach isolates 1791, 1792 and 1793 (Hepworth et al., 2011).

Previous studies exploring the genomes of *C. jejuni* strains (Champion et al., 2005), identified a region within the flagellin glycosylation locus (*Cj1321- Cj1325*) which is characteristic of a 'livestock' clade. The genomes of *C. jejuni* strain 1336 and 414 differed with respect to the same region, however it was completely deleted from the genome of strain 1336 and partly from the genome of strain 414 (Hepworth et al., 2011). This could indicate that this region plays a role in adapting to a new niche, especially because a recent study has shown that a *Cj1324* mutant is less able to colonise and persist in chickens (Howard et al., 2009). Hepworth et al. (2011) verified a number of deletions from the genomes of strain 1336 and 414 in regions which are known to be associated with host- bacterial interactions, including the ability to colonise chickens. Interestingly, strain 1336 was, in contrast to strain 414, able to colonise 3-week-old chickens, despite the fact that the regions associated with host colonisation are missing in its genome. This suggests that the regions missing are not essential for the colonisation phenotype, as other loci might have compensated for the loss, or the function of the deleted genes may have been replaced by alternative genes (Hepworth et al., 2011).

An earlier study (Colles et al., 2008) concentrating on the comparison of *Campylobacter* populations in wild geese with those in starlings and free-range poultry on the same farm observed that genotypes of *C. jejuni* isolates from geese are strongly host associated. The *C. jejuni* STs formed separate clusters on a ClonalFrame tree (Didelot and Falush, 2007), suggesting that *C. jejuni* isolates from geese are more closely related to *C. jejuni* isolates from chickens than from starling (Colles et al., 2008). The comparison of the results obtained in the study with the MLST *Campylobacter* database revealed that the ST-692, ST-702, ST-1034 and ST-1332 complexes may be associated with geese. The ClonalFrame analysis also suggested that the geese

genotypes are not monophyletic and that there was shared ancestry between chickens, starlings and geese in the past. Colles et al. (2008) found only little evidence for shared isolates of *C. jejuni* between starlings, geese and chickens on the same farm, which is in agreement with McCarthy et al. (2007) who suggested that *Campylobacter* genotypes may circulate more commonly among different farm animal species than among different wild bird species. Another study led by Sheppard et al. (2011b) applied multilocus sequence typing (MLST) to a large number of *C. jejuni* and *C. coli* isolates from diverse locations and sources, and revealed associations of certain STs with host species. Müllner et al. (2010b) suggested that *Campylobacter* spp. in New Zealand display a unique situation compared to *Campylobacter* spp. populations in other countries, exhibiting a relatively low genetic diversity, and revealing evidence of many niche-associated genotypes.

It is still unknown as to how some strains of *C. jejuni* are able to live the life of a generalist as the challenges of colonising organisms with such distinct environments (e.g. different gastrointestinal tracts, body temperature and immune systems) is substantial. However, possible factors which have been identified in other pathogens, and which could potentially reduce the cost of adaptation are the import of DNA by HGT from already adapted lineages present in each new host species (McCarthy et al., 2007); coordinated genetic regulation of host-specific factors (Gottesman, 1984; Killiny and Almeida, 2011) and genetic change via contingency loci (alleles that increase mutation rates in nearby DNA sequences, not across the whole genome) (Kim et al., 2012; Moxon et al., 2006). The mechanism of virulence in *Campylobacter* species are still understood poorly, but Fouts et al. (2005) suggested that loci responsible for adherence, cellular invasion, toxin production and flagellar motility are important virulence factors.

MLST studies of *C. jejuni* conducted by different groups and in different countries have confirmed the genetic diversity of this bacterium (Cohan and Koeppel, 2008; Coker et al., 2002; Moore et al., 2005; Ogden et al., 2009; Sheppard et al., 2011a). Despite the fact that there is extensive recombination both within and between *Campylobacter* populations (Sheppard et al., 2008, 2011b; Suerbaum et al., 2001), certain groups of related genotypes (clonal complexes), persist over time and geographical spread and are associated with specific host groups.

CHAPTER 3

Seasonality of *Campylobacter jejuni* isolates associated with human campylobacteriosis in the Manawatu, New Zealand

Abstract

A nine year time series of genotyped human campylobacteriosis cases from the Manawatu region of New Zealand was used to investigate strain-type seasonality. The data were collected from 2005 through to 2013 and the samples were multi-locus sequence typed (MLST). The four most prevalent clonal complexes (CCs), consisting of 1215 isolates, were CC-48, CC-21, CC-45 and CC-61. Of the four examined CCs, only CC-45 showed a marked seasonal (summer) peak. The association of CC-45 with summer peaks has been observed in other temperate countries, but has previously not been identified in New Zealand. This is the first in depth study over a long time period employing MLST data to examine strain-type associated seasonal patterns of *C. jejuni* infection in New Zealand.

3.1 Introduction

Campylobacteriosis, mainly caused by the zoonotic bacterial species *Campylobacter jejuni* and *Campylobacter coli*, is the most frequent form of acute bacterial gastroenteritis in humans worldwide and is therefore a major public health burden (Adak et al., 2005; Coker et al., 2002; Samuel et al., 2004).

New Zealand had an average annual rate of notified human campylobacteriosis cases of 353.8/100,000 between 2002-2006 (The Institute of Environmental Science and Research Ltd., 2013) with a peak of >380 per 100,000 population in 2006 (The Institute of Environmental Science and Research Ltd., 2007). This was followed by a sharp decline which has persisted to the present day. The reduction in the notified campylobacteriosis cases occurred after an intervention was introduced, involving a range of control measures to reduce the *Campylobacter* spp. contamination in the poultry supply (New Zealand Food Safety Authority, 2008; Sears et al., 2011). Major objectives were the Poultry Processing Code of Practice and mandatory *Campylobacter* performance targets for broiler chicken carcasses at the end of primary processing. The regulatory target was officially implemented on 1^{st} April 2008 (New Zealand Food Safety Authority, 2008). However, initiatives such as the development of voluntary Broiler Growing Biosecurity Manual by industry (building on existing manuals and codes of practice), improvements in procedures for catching and transporting birds and monitoring/reporting prevalence of *Campylobacter* spp. in caecal samples took place throughout 2007 (Sears et al., 2011).

Campylobacter spp. are commensals in many wild, farmed and domestic animals. The risk factors for human infection range from consumption of undercooked chicken (Eberhart-Phillips et al., 1997; Gras et al., 2012; Harris et al., 1986; Kapperud et al., 1992; Müllner et al., 2009, 2010b; Neal and Slack, 1997), consumption of offal (mainly chicken and sheep liver) (Lake, 2006), raw dairy products (Blaser et al., 1979; Eberhart-Phillips et al., 1997; Evans et al., 1996; Heuvelink et al., 2009; Lake, 2006) and untreated water (Adak et al., 1995; Close et al., 2008; Savill et al., 2001), to contact with pets and farmed animals (Adak et al., 1995; Eberhart-Phillips et al., 1997; Gras et al., 2013; Kapperud et al., 1992; Lake, 2006; Neal and Slack, 1997).

Seasonal patterns in *Campylobacter* infections in humans have been identified in many temperate countries. Studies examining the incidence rates identified a consistent peak in summer across all temperate countries (Kovats et al., 2005; Nylen et al., 2002), some countries also showed an early spring peak (Kovats et al., 2005; Nylen et al., 2002). Kovats et al. (2005) identified the timing of the peak as being least variable from year to year in England and Wales, Greece, Denmark and the Netherlands.

The pattern of seasonality was shown to be quite different in Australia and New Zealand, showing less consistency over time and a prolonged summer peak (McCarthy et al., 2012; Nylen et al., 2002). Hearnden et al. (2003) identified marked differences in seasonality patterns across New

Zealand, in particular, three different groups of seasonality and incidence patterns between the North and South Islands of New Zealand. The first group, the Far North and rural areas of the North Island, showed a relatively low summer incidence and small inter-seasonal variation. The second group, comprised of urban areas (Auckland, Hamilton, Napier) in the North Island as well as some areas in the South Island, exhibited a higher summer incidence and more seasonality when compared to the first group. The highest summer incidences and strongest inter-seasonal variations were found in Wellington and surroundings (North Island) and Christchurch, Dunedin and much of the rest of the South Island (Hearnden et al., 2003). These differences between the rural and urban areas could be due to the proportion of poultry- to non-poultry associated cases in these areas, with urban areas showing a higher risk of poultry-associated infections (Muellner et al., 2011). Spencer et al. (2012) examined spatial and temporal determinants of campylobacteriosis notifications over seven years (2001 to 2007) in New Zealand. Spencer et al. (2012) not only described differences in the spatial distribution between Canterbury (most seasonal variation with severe but short epidemics in the summer), Manawatu (lowest notifications rate but increased numbers of winter cases) and Auckland (notifications appear to come in short bursts during summer); they also found an association between social deprivation and a decreased risk of notification.

A recent study by McCarthy et al. (2012) examining a United Kingdom, Finnish, Australian and New Zealand dataset, sought to explain the seasonality and other aspects of *C. jejuni* epidemiology. The study discovered a clade formed by CC-45 and CC-283 which shows a summer peak and is present in the UK and the Finnish dataset but not in the Australian or New Zealand dataset (McCarthy et al., 2012). However, the McCarthy et al. (2012) study used a small-scale dataset from New Zealand which was focused on a prolonged outbreak in the winter of one year (McTavish et al., 2008) instead of one of the larger published datasets (Müllner et al., 2009).

In order to examine possible transmission routes of campylobacteriosis in New Zealand, a source attribution study was initiated in the Manawatu sentinel surveillance site of New Zealand in 2005 (French, 2008). The data collected as part of the Manawatu study between 2005-2013 is used in this study to explore strain-type-associated seasonality. The main questions addressed in this study are: do the *Campylobacter* clonal complexes identified in human cases in this region show a seasonal pattern; and how do these compare with seasonality-associated CCs in other countries?

3.2 Materials and methods

3.2.1 Data

clarification who provided the data

Multi locus sequence typing (MLST) data of laboratory-confirmed cases of human campylobacteriosis were obtained from the sentinel surveillance site in the Manawatu region of New Zealand's North Island between 2005 and 2013. The data collection and sequencing was carried out by students and staff of the mEpiLab team.

Human faecal swabs were cultured on modified Charcoal Cefoperazone Deoxycholate Agar (mCCDA) plates (Fort Richard Laboratories, Auckland, New Zealand) and in Boltons Campylobacter enrichment broth (Lab M, Bury, England) and incubated under microaerobic conditions (85% N_2, 5% O_2, 10% CO_2) generated by MACS-VA500-microaerphilic workstation (VA500, Don Whitley Scientific, Yorkshire, UK) at 42°C for 48 h. Single colonies resembling *Campylobacter* spp. were subcultured onto Columbia Horse Blood Agar (BA) (Fort Richard Laboratories, Auckland, New Zealand) and incubated microaerobically at 42°C for an additional 24 h. Boiled lysate DNA preparations were made and cultures were frozen in glycerol broth at -80°C. *Campylobacter* species isolates were identified by the method outlined in Linton et al. (1996) and further speciated using the *map*A gene which has shown to be unique in *C. jejuni* (Stucki et al., 1995). The forward (5'-CTTGGCTTGAAATTTGCTTG-3') and reverse (5'-GCTTGGTGCGGATTGTAAA-3') primers were designed to target this gene and the amplification protocols were based on the methods outlined in Linton et al. (1996) and Stucki et al. (1995) with slight modifications (Müllner et al., 2009).

The isolates were sequenced at seven housekeeping loci (*asp*A, *gln*A, *glt*A, *gly*A, *pgm*, *tkt* and *unc*A) by the MLST technique outlined in Dingle et al. (2001) and Maiden et al. (1998). A total of 3,309 nucleotides were sequenced per isolate. The 1215 isolates were grouped by clonal complex and the four most prevalent clonal complexes were used for further analysis. Due to the limited number of isolates only the four most prevalent clonal complexes were examined.

comment 16, Murray

3.2.2 Time series analysis using ARIMA methods

The time series of numbers of isolates per month for each CC was modelled using an ARIMA (auto-regressive, integrated, moving average) model. In an ARIMA(p,d,q) process, the observations (counts per month) are fitted through a three step process in which the components of the model are estimated. The auto-regressive component (AR) (symbol p) represents previous observations, the integrated component (I) (d) stabilises the data and makes it stationary (eliminating trends thus making forecasting easier), and a moving average (MA) (q) estimates the

past errors. In order to fit an ARIMA model to a time series, the order of each model component must be estimated.
Where seasonality was evident, a seasonal ARIMA (SARIMA (p,d,q)(P,D,Q)$_m$) process was fitted to the data. The seasonal components of the model (P,D,Q) consists of terms that are very similar to the non-seasonal part of the model, but they involve backshifts of the seasonal period (e.g. SARIMA (p,d,q)(P,D,Q)$_4$ for quarterly data ($m = 4$)). The function auto.arima in the R package *forecast* was used to model the time series (Hyndman and Khandakar, 2007; R Development Core Team, 2013). This function uses a variation of the Hyndman and Khandakar algorithm which combines unit root tests and minimisation of AIC (Akaike Information Criterion) and MLE (Maximum-Likelihood Estimation) to return the best ARIMA model (Hyndman and Khandakar, 2007). Additional constraints (stepwise=FALSE and seasonal=TRUE) were invoked to insure all possible models, seasonal and non-seasonal, were considered.
In order to assess the effect of the intervention (the range of control measures to reduce the *Campylobacter* spp. contamination of poultry), the data was divided into two periods: pre and post-intervention. The cut-off date was chosen to be January 2008, instead of the official date of the implementation, 1^{st} of April 2008, as some poultry companies had already started to trial different interventions, resulting in a lower number of contaminated carcasses (Muellner et al., 2011).

3.2.3 Linear Regression Models with Generalised Least Squares

In addition to the ARIMA model, a non-linear mixed effect model with an autoregressive error term was fitted to the case numbers. This was done, to provide further support for the results obtained by the ARIMA model, and to model the relationship between different explanatory variables (season and intervention) stratified by CC. To account for potential autocorrelation of residuals, a corAR1 process was fitted. The variable "season" was fitted as a factor (spring= September- November, summer= December- February, autumn= March-May, winter= June-August).
The gls model was fitted with counts per CC (i=1,...,4) as the outcome variable; season (k=1,...,4), interaction of season and CC and interaction of intervention (j=1,2) and CC as the dependent variables. Additionally, an autocorrelation-moving average correlation structure of order (form= ∼uniqueID|CC) was used as a correlation term and the model was fitted by maximizing the restricted log-likelihood (REML). The function gls in the R package nlme was used to fit the model (Pinheiro et al., 2013; R Development Core Team, 2013).

$$\mathsf{Counts}_{ijkt} = \mathsf{CC}_i + \mathsf{intervention}_j + (\mathsf{CC*intervention})_{ij} + \mathsf{season}_k + (\mathsf{CC*\ season})_{ik} + \varepsilon_{ijkt}$$

with $\varepsilon_{ijkt} = \rho\varepsilon_{ijk(t-1)} + \eta_{ijkt}$ and $\eta_{ijkt} \sim N(0, \sigma^2)$

3.3 Results

The 1215 isolates were grouped by clonal complex; the four most prevalent CCs, were CC-48 (n=354, 29%), CC-21 (n=317, 26%), CC-45 (n=170, 14%) and CC-61 (n=65, 5%). In New Zealand, CC-48 is a poultry-associated clonal complex (Müllner et al., 2010a), whereas CC-45 and CC-21 are associated with a wider range of hosts and environmental sources (Gripp et al., 2011; Levesque et al., 2008) and CC-61 is a ruminant-associated clonal complex (Manning et al., 2003).

Figure 3.1 displays the multiple time series plots visualising the number of cases for each examined CC for each month over nine years. The time of the poultry intervention is indicated by a dotted line at the beginning of 2008. Comparing the pattern of the four CCs for the pre and post intervention period, it is noticeable that the intervention had the greatest influence on CC-48 where the human campylobacteriosis incidences dropped notably (Figure 3.1, panel A). The intervention appears to not be associated with the number of CC-21, CC-45 or CC-61 cases, showing no apparent change in the pattern of disease notifications.

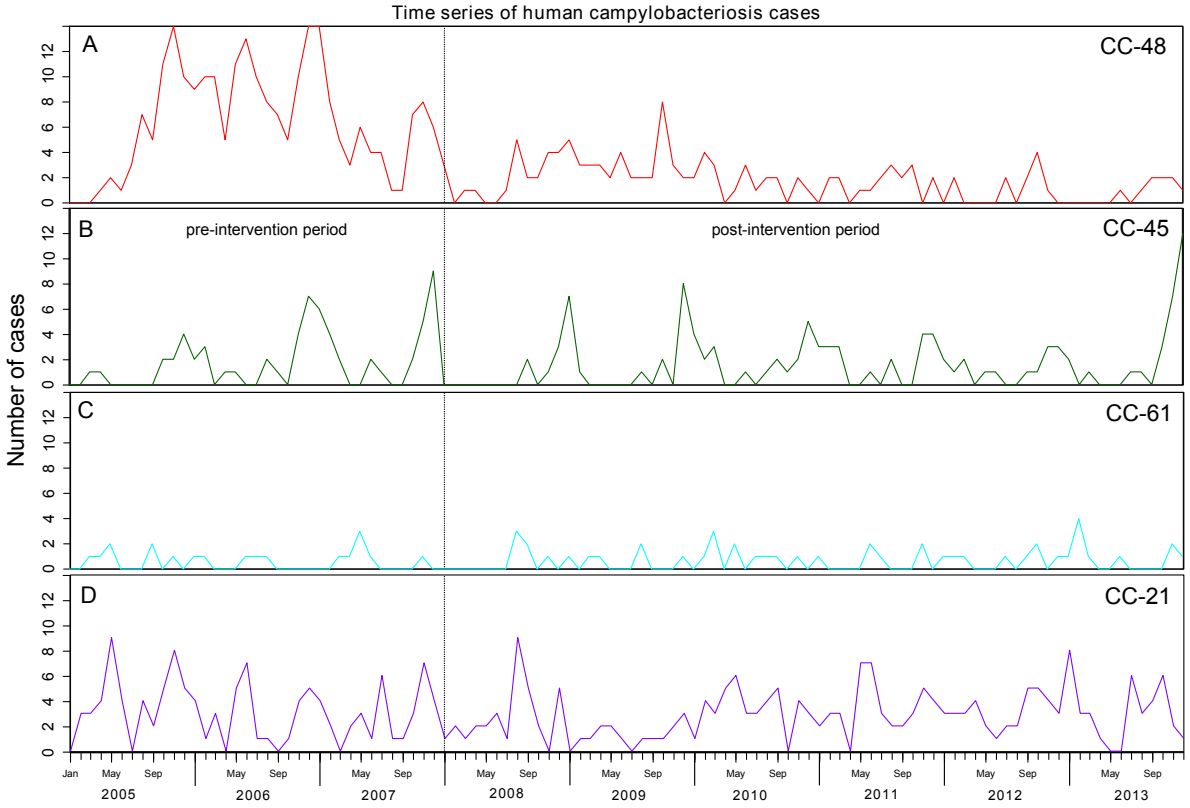


Figure 3.1: Each panel displays the pattern of human campylobacteriosis cases over nine years (2005-2013). Panel A shows CC-48, panel B displays CC-45, panel C displays CC-61 and panel D displays CC-21. The dotted line at the beginning of 2008 indicates the defined starting point of the intervention.

The numbers of isolates of the four CCs were summarised over nine years and plotted as a proportion compared to the number of all cases occurring in the same month over nine years (2005-2013) are displayed in Figure 3.2. The summarised number of cases per month for CC-45 shows a pattern of seasonality with a rising number of the proportion of cases in the late spring to early summer (November and December, Figure 3.2). For CC-21 the month with the highest number of cases on average was April, but this varied considerably between years (from zero to one). Both plots for CC-45 (Figure 3.2 and Figure 3.1) show signals of a seasonal peak in late spring to summer (Nov-Dec-Jan) each year, whereas the pattern for the other CCs is less obvious.

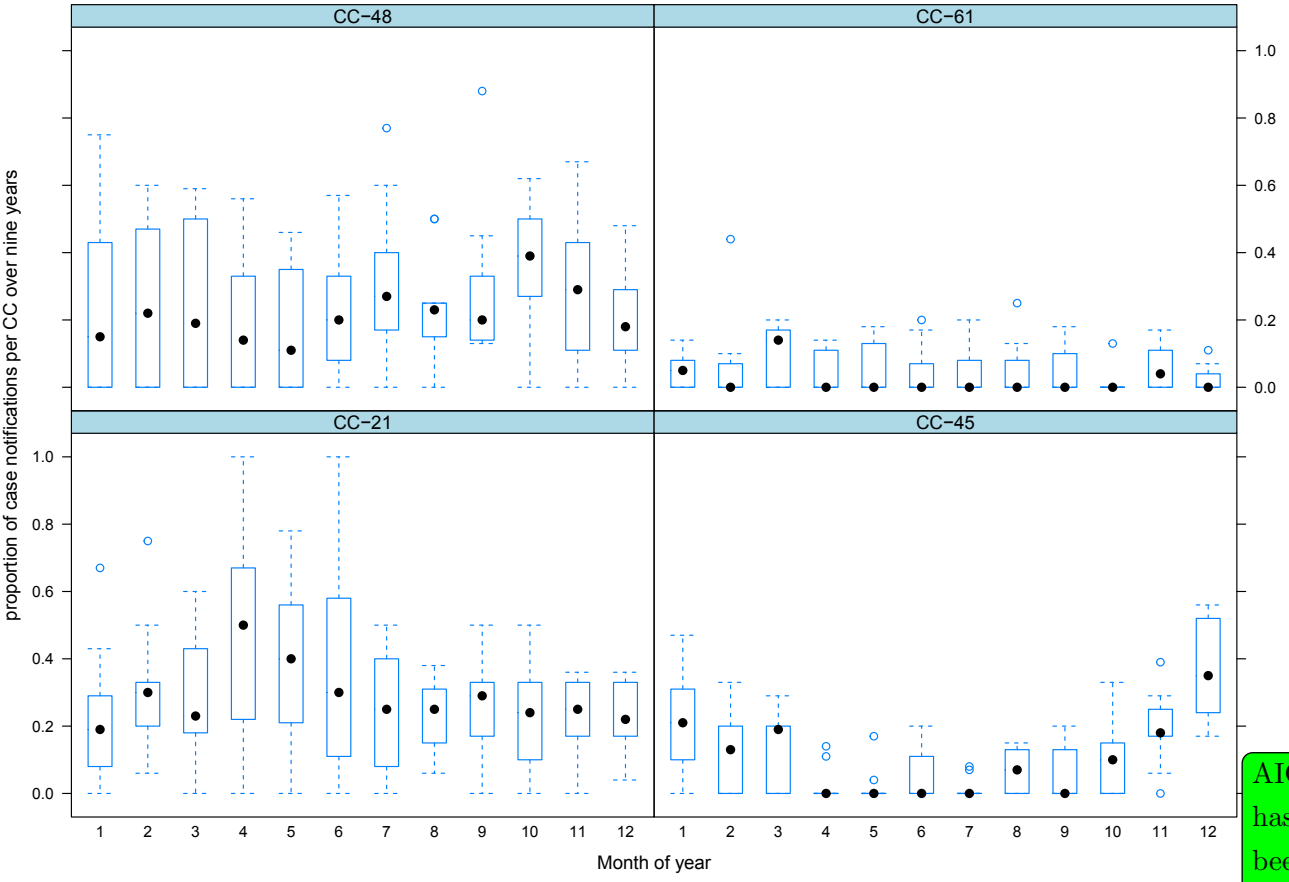


Figure 3.2: The box plots are displaying the pattern of notified human campylobacteriosis incidences divided by the four most common CCs (CC-48, CC-61, CC-21 and CC-45). Each box summarises the proportions of case notifications for a given CC and month between 2005 and 2013. Outliers are represented by blue circles.

AICc has been changed to AIC, comment 17, Murray

3.3.1 Time series modelled with ARIMA

Both seasonal (SARIMA) and non-seasonal (ARIMA) models were considered for each clonal complex. The automated model selection function auto.arima in the R package *forecast* was used, and confirmed by manually building models, using AIC to determine the optimal model. For each CC, based on the autocorrelation function (acf) and the result of the auto.arima function, several seasonal and non-seasonal models were assessed to find the best fitting model. The procedure is illustrated on the CC-45 data below. The aim was to find an appropriate ARIMA or SARIMA model based on the autocorrelation (acf) and partial autocorrelation (pacf) functions shown in Figure 3.3. Both functions show a sinusoidal pattern, implying a possible ARMA model with both autoregressive (AR) and moving average (MA) terms (no integrated term 'I'). However, the values for p (AR term) and q (MA term) were not intuitive, and therefore multiple models had to be evaluated (starting with ARIMA(1,0,1)). The model fit was evaluated

based on the AIC's and log likelihood value. Assuming non-seasonality of the data, the model ARIMA(1,0,0) was the best fit. However, we also tried fitting a seasonal model for each clonal complex, and for CC-45 the SARIMA$(1,0,0)(2,0,0)_{12}$ was the best fit, implying an underlying seasonal pattern. The auto.arima function for CC-45 suggested the same model, however, sometimes the automated function returned a more poorly fitted model and therefore each model had to be evaluated manually. This procedure was repeated for each CC. A marked seasonal pattern was identified only for CC-45. Figure 3.4 displays the fit of the model SARIMA$(1,0,0)(2,0,0)_{12}$. All residuals are within the 95% confidence interval and were therefore considered as random white noise. The estimated model fits are summarised in Table 3.2.

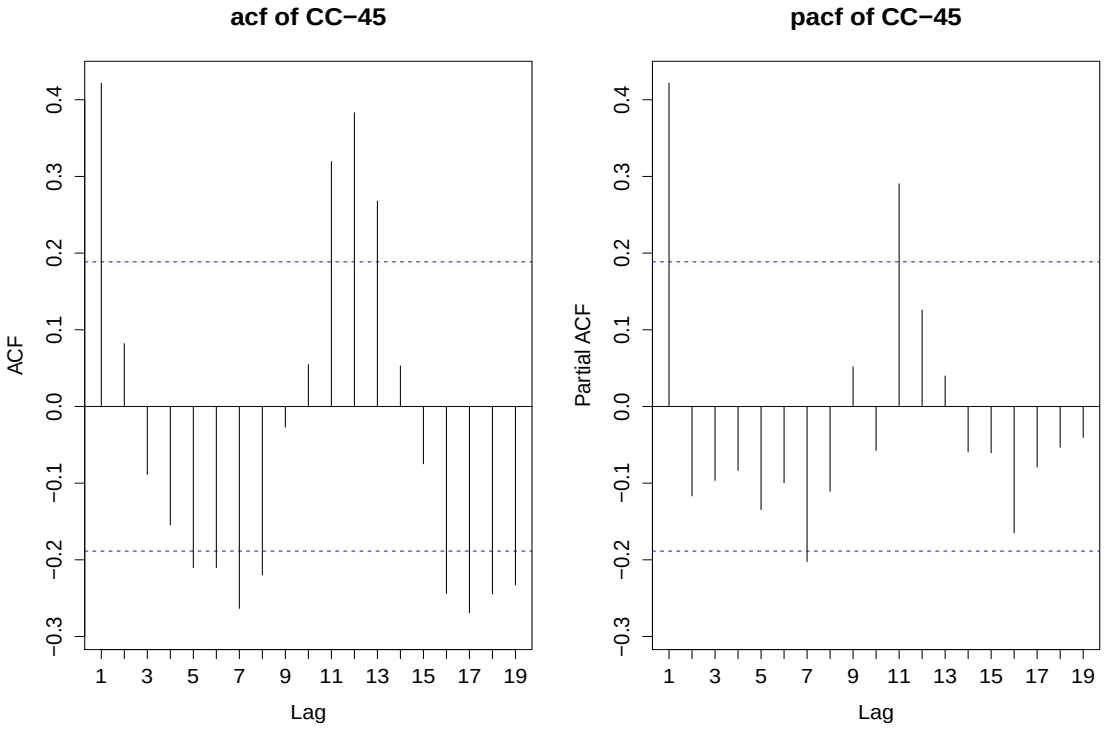


Figure 3.3: The acf and pacf functions plotted for CC-45, displaying several significant spikes at lag 1, 5- 8, 11- 13, 1, 16- 19 for acf and lags 7 and 11 for pacf. The blue dashed lines symbolises the 95% confidence intervals.

		MAE^a	$RMSE^b$	AIC^c
CC-45	SARIMA(1,0,0)(2,0,0)$_{12}$	1.22	1.73	438.83
	ARIMA(1,0,0)	1.36	1.92	454.00
CC-48	SARIMA(0,1,2)(1,0,0)$_{12}$	1.57	2.11	473.19
	ARIMA(0,1,2)	1.58	2.11	471.22
CC-61	SARIMA(0,0,3)(1,0,0)$_{12}$	0.63	0.80	270.42
	ARIMA(0,0,3)	0.64	0.80	269.00
CC-21	SARIMA(1,0,0)(2,0,2)$_{12}$	1.57	1.97	472.05
	ARIMA(2,0,2)	1.57	1.97	470.10

[a] Mean Absolute Error, [b] Root Mean Square Error,
[c] Akaike Information Criterion

Table 3.1: Results comparing the MAE, RMSE and AIC estimated for the ARIMA and SARIMA time series. Each model is significant ($|z| > 1.96$ which is equivalent to $p < 0.05$).

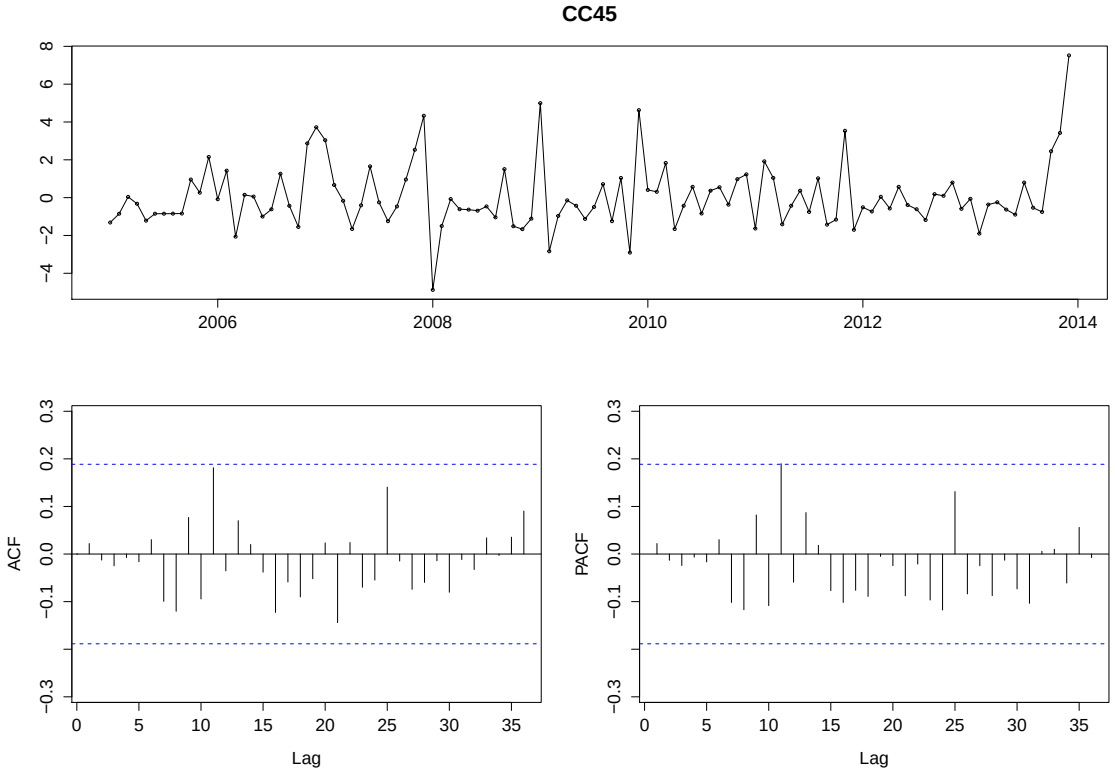


Figure 3.4: The first plot shows the time series of the raw residuals of CC-45 over the nine years. The second and third plot display the acf and pacf of the CC-45 residuals respectively. The acf and pacf functions plotted for CC-45, show that the residuals are within the 95% confidence interval and therefore a good fit of the SARIMA model $(1,0,0)(2,0,0)_{12}$.

The mean absolute error (MAE) is the absolute difference between the predicted forecast value and the actual value. It states how large of an error we can expect from the forecast on average, therefore the smaller the error (closer to zero), the better the model fits the data. MAE does not account for large but infrequent errors, however, by squaring the errors before calculating their mean and then taking the square root of the mean (RMSE), we get a measurement that gives more weight to the large but infrequent errors than the mean. RMSE and MAE can be compared to determine whether the forecast contains large but infrequent errors as the larger the difference the more inconsistent is the error size (Table 3.1).
The most intuitive way to assess the model fits for the different CCs is by comparing the AIC values. The lower the AIC, the better the data is described by the model. Comparing the AIC between the seasonal and non-seasonal ARIMAs, it is apparent that CC-45 is the only clonal complex that has a lower AIC value for a SARIMA model fit compared to the ARIMA model fit. The opposite is true for the other clonal complexes, suggesting that only the data for CC-45 shows an underlying seasonal pattern.

3.3.2 Generalised Least Squares

The results for the fitted generalised least squares model are summarised in Table 3.2. The main effects for CC-48, CC-61 and CC-45 are compared to the baseline CC-21. The linear regression model indicates that the intervention was associated with a sharp decline in CC-48 (p-value < 0.00001), mainly associated with sequence type 474 (a member of CC-48, which is estimated to account for approximately $25-30\%$ of human campylobacteriosis cases in New Zealand) (Muellner et al., 2013; Müllner et al., 2009). The coefficient (-4.10) for the interaction between intervention and CC-48 means that four times fewer cases occurred after 2008. This drop can also be seen in the TS plot (Fig. 3.1, panel A).
The only other significant interaction is CC-45 in summer (p-value < 0.01), providing further support for a seasonal peak. The coefficient for the interaction between CC-45 and summer (2.12) means that there were two times more cases in summer compared to the baseline (CC-45 in autumn). This also agrees with the results of the SARIMA model and the box plots.

	Coefficient $(SE)^a$	p-value
(baseline)	3.72 (1.25)	0.00**
intervention	−0.43 (0.69)	0.53
CC-45	−1.46 (1.77)	0.41
CC-48	6.38 (1.77)	0.00***
CC-61	−3.24 (1.77)	0.07
intervention:CC-45	−0.18 (0.97)	0.85
intervention:CC-48	−4.10 (0.97)	0.00***
intervention:CC-61	0.55 (0.97)	0.57
CC-45:(season)spring	−0.19 (0.92)	0.84
CC-48:(season)spring	1.14 (0.92)	0.22
CC-61:(season)spring	−0.02 (0.92)	0.98
CC-45:(season)summer	2.12 (0.84)	0.01*
CC-48:(season)summer	0.96 (0.84)	0.25
CC-61:(season)summer	−0.08 (0.84)	0.92
CC-45:(season)winter	0.03 (0.84)	0.98
CC-48:(season)winter	1.05 (0.84)	0.21
CC-61:(season)winter	0.22 (0.84)	0.79

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

[a] Standard error

Table 3.2: Results of the generalised least squares (gls) estimating the effect of intervention on CC and prediction of seasonality based on season and CC.

3.4 Discussion

This study presents a nine year time series of genotyped human campylobacteriosis cases from the Manawatu region of New Zealand. It provides the means to link identified seasonal patterns to specific genotypes.

Prior to 2008, New Zealand had a particularly high and increasing incidence of human campylobacteriosis, peaking in 2006 (McTavish et al., 2008). The high numbers of *Campylobacter* cases led to recommendations for new measures to reduce the risk (Baker et al., 2006) and thus the New Zealand Food Safety Authority announced the implementation of the *Campylobacter* Risk Management Strategy. After the intervention the cases linked to the poultry-associated ST-474 (CC-48) dropped to $< 5\%$ (Muellner et al., 2013). This drop can be seen in the time series of CC-48 (Figure 3.1, panel A). The significant drop in CC-48 associated cases was also detected by the regression model (p-value < 0.00001). The intervention seems to have had the

most profound effect on CC-48, mainly associated with ST-474, and little effect on the other clonal complexes. In New Zealand, CC-48 is a poultry associated clonal complex (Müllner et al., 2010a) whereas CC-45 and CC-21 are associated with a wider range of hosts and environmental sources (Gripp et al., 2011; Levesque et al., 2008; Sheppard et al., 2014) and CC-61 is a ruminant associated clonal complex (Manning et al., 2003). This might explain why the intervention was not associated with a decline in the other most prevalent CCs.

The fitted linear regression model with generalised least squares based on nine years of data displayed a seasonal peak for CC-45 in summer (Table 3.2). This is in concordance with the estimated ARIMA model for CC-45, the time series (Figure 3.1) and boxplots (Figure 3.2), showing a seasonal peak around December each year. Models for ARIMA and SARIMA time series were constructed, however, as shown in Table 3.1, the AIC's values were similar for the three non-seasonal clonal complexes. Therefore, from the model fit alone, it is not apparent whether the data is seasonal or not. However, neither the plots nor the regression model indicate seasonality in CC-61, CC-21 or CC-48 and therefore, choosing the most parsimonious model, the three non-seasonal models were a better fit. One explanation for the lack of differential ability of the ARIMA/SARIMA models for CC-61 may be the relatively low number of cases over the nine years. However, this explanation is less likely to hold true for CC-48 and CC-21 which were the first and second most frequently isolated clonal complexes (n=354 and n=317 respectively) over this time period.

CC-21 showed a relatively high degree of variability in proportions of notified human campylobacteriosis incidences over these nine years (Figure 3.2) and CC-48 was strongly influenced by the intervention which may have obscured or eliminated existing patterns. The plot for CC-21 (Fig. 3.2) shows an apparent peak in April, however the interval (dashed lines) is quite large, indicating a high amount of variation over the nine years. The highest amount of variation is over the months of April, May and June (late autumn to early winter). Precipitation has previously been shown to be associated with an increased number of *Campylobacter* incidences (Febriani et al., 2010) and environmental surface water has been shown to carry higher numbers of *Campylobacter* spp. during the winter months (Whiley et al., 2013). However, examination of the association between meteorological data and human campylobacteriosis cases was beyond the scope of this study and has also not been observed in previous studies (Lal et al., 2013).

comment 20, Murray

MLST data have been used to examine seasonal patterns of *C. jejuni* and *C. coli* in human disease (Cody et al., 2012; McCarthy et al., 2012; Sopwith et al., 2008) in several countries, identifying a consistent CC-45 peak in the summer across all examined temperate countries. These regular, re-occurring patterns could indicate environmental influence (Altizer et al., 2006; Febriani et al., 2010) or frequent host-pathogen interactions (Gilpin et al., 2008; Zappe Pasturel et al., 2013). Seasonality has previously been shown in New Zealand (Hearnden et al., 2003; Kovats et al., 2005; Nylen et al., 2002; Spencer et al., 2012) but only one study (McCarthy et al., 2012) has used MLST data to identify the specific genotypes involved. McCarthy et al.

(2012) examined the data from a 3-year longitudinal study in the UK and compared these to international datasets (Finland, Australia, New Zealand). They identified CC-45 and CC-283 being associated with a summer peak, but although they reported that the genotypes isolated in the UK are more similar to the New Zealand and the Australian samples than to the Finnish samples, (McCarthy et al., 2012) they did not identify a marked seasonal peak in New Zealand or Australia. The present study did not identify any isolates belonging to the CC-283, but the non-identification of the CC-45 summer peak by McCarthy et al. (2012) can be easily explained by the fact that they used a New Zealand dataset which was based on an outbreak in the Southern hemisphere winter of 2006, whereas our data span nine years. Nylen et al. (2002) who examined the seasonal distribution of campylobacter infection across nine European countries and New Zealand concluded that the seasonality in New Zealand was less consistent as the week in which the peak occurred was more variable from year to year and the summer increase more prolonged. As displayed in Figure 3.2 the proportions of cases of CC-45 over nine years shows a distinct peak in December with a comparatively small interval varying over the nine years, which indicates a relatively strong and consistent seasonal pattern. Possibly, the inconsistent seasonality detected by Nylen et al. (2002) is related to delays between the illness, sampling, processing and official submission. Our data is subject to similar constraints, but whereas Nylen et al. (2002) reported weekly incidences, our data is summarised monthly and therefore minimises this variation (e.g. it is not important if the incident occurred in the first or last week of the month).

Several studies examining the seasonality of *Campylobacter* isolates from human cases identified CC-45 as being strongly associated with an early summer peak (Cody et al., 2012; McCarthy et al., 2012; Sopwith et al., 2008) and to be frequently isolated from chickens, wild birds and environmental water (Colles et al., 2003; French et al., 2009; Sheppard et al., 2014; Sopwith et al., 2008). These findings suggest that CC-45 is adapted to survival outside the host and makes this adaptation a key driver of transmission between livestock, environment and humans (Sopwith et al., 2008). The present study showed that CC-45 was not affected by the poultry industry-specific intervention, suggesting that, even though it is frequently isolated from poultry (Müllner et al., 2010a), the isolates causing the human campylobacteriosis may have been of different origin or survived the intervention. Sopwith et al. (2008) sampled recreational surface water and urban and rural river systems and suggested that there may be an association between the CC-45 isolated from water and reported campylobacter incidences in humans, presenting a potential environmental transmission route for CC-45. Additionally, Sopwith et al. (2008) found that persons infected with CC-45 were more likely to live in rural areas, to be < 5 years old, to have been involved in outdoor activities (e.g. fishing), to have consumed home-delivered milk, and were less likely to have eaten chicken in the two weeks before the illness. CC-45 has also, compared to another generalist CC-21, more resilience to oxidative and freezing stress although a poorer survival response to heat or chilling (Habib et al., 2010). These finding support the hypothesis that CC-45 is more available to infect humans through transmission routes other

than food, e.g. through indirect exposure to pets or cattle or through exposure to untreated water during outdoor activities. This is further supported by the association of human CC-45 with more rural areas of residence (Sopwith et al., 2008).
After the intervention the number of poultry-associated cases has fallen, and source attribution studies have estimated that there has been a relative increase in the importance of ruminant-associated strains (French et al., 2010; Muellner et al., 2011). However, although the intervention reduced the number of poultry associated cases, the risk has not been eliminated. New strategies have to be developed to control campylobacteriosis cases acquired from both poultry- and non-poultry sources (New Zealand Food Safety Authority, 2008) and an examination of the seasonality of the CCs associated with these cases may help to focus on specific time frames to determine the sources and transmission routes. Once source-attribution studies identify specific sources, new intervention strategies can be developed to reduce the human disease burden.
In conclusion, this study has shown that genetic analytical approaches (MLST) in epidemiological analysis can successfully identify seasonality of campylobacteriosis. CC-45 was identified as a prevalent clonal complex showing a consistent summer peak over nine years, in contrast to previous studies that found limited seasonality in New Zealand. It has also shown that the intervention had a significant effect on the poultry associated CC-48.

Acknowledgments

We are grateful to all members of the Hopkirk Molecular Epidemiology Team for their support and helpful discussions. This work was funded by the Allan Wilson Centre for Molecular Ecology and Evolution.

CHAPTER 4

Evidence of host associated *Campylobacter jejuni* genotypes in wild and agricultural hosts in New Zealand

Abstract

Host association and niche adaptation play a key role in the evolution of bacterial populations such as *Campylobacter* spp. Despite their importance and extensive study, there is still uncertainty as to how mechanisms such as mutation and horizontal gene transfer (HGT) influence the high level of genetic structuring displayed by populations of *Campylobacter*. In this study, combining phylogenetic and population genetic tools with multilocus sequence data, we examined whether ecological factors such as niche association influence the population structure of *C. jejuni*. Specifically, we are interested whether high or low levels of HGT are related to niche specialism and the extent to which *C. jejuni* sequence types (STs) can be predictors for host origin. Thus the genetic structure and pattern of ancestry of the *C. jejuni* population in New Zealand was reconstructed using BAPS and ClonalFrame. Additionally, gene flow was investigated to infer underlying genetic admixture within the *C. jejuni* population. The findings of this study show host associated *C. jejuni* genotypes and lineages displaying either low or high levels of horizontal gene transfer. The variation between *C. jejuni* STs appears to be less pronounced within those isolated from domesticated animals as compared to those isolated from non-agricultural sources. This could suggest that the farm environment constitutes a specific niche for *Campylobacter* and that, keeping different agricultural species in close proximity to each other, may influence the population structure of *C. jejuni*.

4.1 Introduction

Campylobacteriosis, mainly caused by the bacterial species *Campylobacter jejuni* and *Campylobacter coli*, is a major cause of foodborne illness, causing human acute bacterial gastroenteritis worldwide (Coker et al., 2002; Samuel et al., 2004). *Campylobacter* spp. have been the focus of growing attention for the past three decades. Their wide host range (ruminants, poultry, wild birds, pets) (Sheppard et al., 2011a; Waldenström et al., 2002), environmental distribution (animals, food, water, beach sand) (Newell, 2001, 2002) and high genetic diversity (Moore et al., 2005; Ogden et al., 2009), contribute to the complex molecular epidemiology of campylobacteriosis in humans. It is possible that a broad host range such as exhibited by pathogens like *Campylobacter*, can increase the risk of disease emergence in novel hosts (Cleaveland et al., 2001). Although some countries have experienced a decline in the rate of campylobacteriosis notification in recent years, the global burden is still high (Ailes et al., 2008; Hermans et al., 2012; Sears et al., 2011).

Campylobacter spp. display extensive genetic variation, which has arisen from intragenomic mechanisms as well as genetic exchange (horizontal gene transfer, HGT) among strains and is well described within (Sheppard et al., 2014; Suerbaum et al., 2001) and between (Sheppard et al., 2008, 2011b) different species and lineages of *Campylobacter* spp.

Sheppard et al. (2011b) applied multilocus sequence typing (MLST) to a large number of *C. jejuni* and *C. coli* isolates from diverse locations and sources, and revealed associations of certain STs with host species. Müllner et al. (2010b) suggested that *Campylobacter* spp. in New Zealand are unique compared to *Campylobacter* spp. populations in other countries, exhibiting a relatively low genetic diversity, with evidence of many niche-associated genotypes.

Campylobacter spp., like other zoonotic bacteria are able to colonise multiple hosts, and hence a number of different niches. When colonising a new niche, bacterial populations become separated from the ancestral gene pool, which can lead to diversification into distinct niche-associated genotypes (Cohan and Koeppel, 2008). Previous attempts to understand the population structure of the genus *Campylobacter* include large-scale population sampling studies combined with nucleotide sequence-based typing of isolates.

The aim of this study is to examine the population structure, host association and genetic admixture within New Zealand *C. jejuni* strains. The isolates were collected from many farm and wild bird species across the North and South Islands of New Zealand. The isolates used were typed by MLST (Dingle et al., 2001) and the model-based techniques BAPS (Bayesian Analysis of Population Structure) (Corander and Marttinen, 2006) and ClonalFrame (Didelot and Falush, 2007) were used to estimate the number of host-associated lineages and the level of admixture within and between *C. jejuni* populations from multiple sources. The level of association between host and sequence type was visualised by applying a multivariate analysis, MDS (non-metric multi-dimensional scaling).

Two hypotheses were tested: that levels of horizontal gene transfer are related to niche specialism; and that *C. jejuni* sequence types can be predictors for host origin in New Zealand's domestic and wild animal populations.

4.2 Methods

4.2.1 Isolates and sequence type (ST) data sets

Nucleotide sequence data from seven defined housekeeping genes of 1957 *C. jejuni* isolates were selected from a database of isolates collected from multiple sources in New Zealand (sampled between 2005-2011). The isolates were sequenced at the seven housekeeping loci (*asp*A, *gln*A, *glt*A, *gly*A, *pgm*, *tkt* and *unc*A) using the multilocus sequence typing (MLST) technique (Dingle et al., 2001; Maiden et al., 1998). A total of 3,309 nucleotides were sequenced per isolate. The data collection and sequence preparation was carried out by students and staff of the mEpiLab team as part of different projects.The isolates originated from eight different sources, namely: chickens (1202 isolates), farmed ducks (52 isolates), cattle (271 isolates), sheep (251 isolates), wild ducks and geese (88 isolates) (Mohan et al., 2013), gulls (21 isolates), rails (13 isolates) (French et al., 2014) and starlings (59 isolates) (Mohan et al., 2013). The non-passerine wild birds were grouped for subsequent analysis due to the low numbers.

clarification who provided the data

4.2.2 Bayesian clustering analysis and admixture

BAPS 5.1, a Bayesian method for inferring admixture events (Corander and Marttinen, 2006) was used for estimating the number of clusters (K) and the level of admixture among the examined isolates. For the calculation of admixture inference, a maximum number (K) of genetically diverged groups has to be supplied *a priori*. K was assumed to lie in the range 10-20 and the obtained marginal likelihoods were compared to identify the optimal number of clusters. The Codon linkage model is an appropriate choice for linkage connections occurring under the second order Markov structure where the denser connections represent a codon like dependence formulation for sequence type data (Corander and Tang, 2007). The Codon linkage model was used as an underlying model as suggested in the BAPS manual (Corander et al., 2006).The BAPS analysis is divided into two steps: a sequential modelling strategy where the number of genetically differentiated sources (the different clusters) contributing to a data set is inferred using a mixture model; followed by admixture events which are identified using a Monte-Carlo simulation-based algorithm (Corander and Marttinen, 2006).

comment 22, Murray

BAPS was used to investigate the admixture events, and based on these results, the gene flow between the identified clusters. A weighted table was constructed to determine the degree of host association for the clusters identified by BAPS.

The degree of association for each host with a cluster defined by BAPS was tested by calculating

a score for each host group per cluster, and then bootstrapping the scores 10,000 times to obtain confidence intervals (CI's). For a given ST k we computed the probability that it arose in host group j, $P(\mathsf{host}\ j|\mathsf{ST}\ k)$. If we assume *a priori* that each host group is equally likely, we can estimate this probability using the model of Van Pelt et al. (1999) which makes use of Bayes' theorem

$$P(\mathsf{host}\ j|\mathsf{ST}\ k) = \frac{P(\mathsf{ST}\ k|\mathsf{host}\ j)}{\sum_j P(\mathsf{ST}\ k|\mathsf{host}\ j)}, \tag{4.1}$$

where $P(\mathsf{ST}\ k|\mathsf{host}\ j)$ may be estimated using the relative prevalence of ST k in host j.
The degree of association of a sequence type k with host j can then be found using

$$P(hostj|STk) - \sum_{h \neq j} P(hosth|STk) \tag{4.2}$$

where larger positive values indicate association with that host, while negative values suggest no association with the host.
For each cluster, we produced an overall score by summing the degrees of associations, weighted by the number of isolates of each ST observed, giving

$$\mathsf{score}_{ij} = \frac{1}{N_i} \sum_{k \in \mathsf{cluster}_i} n_k \left[P(\mathsf{host}\ j|\mathsf{ST}\ k) - \sum_{h \neq j} P(\mathsf{host}\ h|\mathsf{ST}\ k) \right] \tag{4.3}$$

where n_k is the number of isolates of type k, and N_i is the total number of isolates in cluster i.
In addition, the total host associativity across all clusters was evaluated using

$$\mathsf{assoc}_j = \sum_i 1[\mathsf{score}_{ij} > 0]\mathsf{score}_{ij} \tag{4.4}$$

where only those clusters that are host associated are accumulated. The algorithm was run in R (R Development Core Team, 2013).

4.2.3 ST genealogy and association between defined clades and hosts based on ClonalFrame

ClonalFrame (Didelot and Falush, 2007) was used to construct a 75% majority-rule consensus genealogy of recently diverged *Campylobacter* lineages based on concatenated sequences of the seven housekeeping genes. The analysis for the whole dataset was based on 1,000,000 burn-in iterations followed by 1,000,000 data collection iterations. The high number of burn-in and MCMC (Markov chain Monte Carlo) iterations was necessary to obtain convergence between the chains. The convergence was tested by applying the method of Gelman and Rubin (1992) implemented in the ClonalFrame software.

The obtained ClonalFrame tree (75% majority-rule consensus tree) was used to test the association between recently diverged *Campylobacter* lineages. As a high coalescent depth results in long branches, which are from an evolutionary point of view less reliable than short branches (Sánchez-Gracia and Castresana, 2012), we chose the smallest cut-off (≤ 0.04 coalescent units) that still allowed identification of clusters. A clade within the tree was defined as a lineage containing ≥ 2 sequence types that shared a common ancestor at ≤ 0.04 coalescent units (the same approach as used in Sheppard et al. (2011a)).
Evidence for an association between ClonalFrame defined clades and hosts was tested by calculating a score across clades per host, and then permuting the dataset randomly 10,000 times and re-calculating the score. This was done to allow a comparison of the observed score with the expected distribution, if the null hypothesis (that there is no genetic basis for host association) is correct. Similarly, as described before (section 'Bayesian clustering analysis and admixture'), we assumed *a priori* that each host is equally likely, and estimated this probability using the Bayes' theorem (equation 4.1).
Sequence types k associated with host j are those where the probability of the host being j is large relative to the probability of the host being $h \neq j$. Thus

$$\mathsf{score}_{ij} = \sum_{k \in clade_i} \left[P(\mathsf{host}\ j | \mathsf{ST}\ k) - \sum_{h \neq j} P(\mathsf{host}\ j | \mathsf{ST}\ k) \right],$$

is a measure of host-associativity of clade i with host j, with positive scores being strongly host-associated. A measure of overall host-associativity for a host across all clades was then be given by the equation 4.4.

4.2.4 Multidimensional scaling

The objective of the non-metric multidimensional scaling (MDS) algorithm is to find a pattern of points whose relative distances are in the same rank order as the relative distances of the data matrix. The input data for non-metric multi-dimensional scaling is a matrix of dissimilarities or distances:

$$D \qquad = \qquad |\delta_{ij}|$$

where δ_{ij} is the dissimilarity between points i and j. The data δ_{ij} are non-metric, therefore they are interpreted as distance-like. The distance is calculated based on the pairwise comparison of the allelic profile. Each ST is described by seven alleles, therefore two STs can have a maximal distance of seven. The MDS plots were visualised using Primer (Clarke, 1993).

4.3 Results

The Venn diagram, Figure 4.1 shows the number of unique and shared sequence types per defined host group in New Zealand. The 1957 examined *C. jejuni* isolates comprised 165 STs (Figure 4.1). Twenty-three (17%) of the 132 STs that occurred in agricultural sources, occurred in both chickens and ruminants. Of the 32 STs that occurred only in ruminants, eight (25%) STs occurred only in in cattle, seven (22%) only in sheep and the remainder in both. In total, 16 (25%) STs of the 63 that occurred in wild birds also occurred in chickens. Eleven (17%) of the 63 STs were detected in more than one species of wild bird.

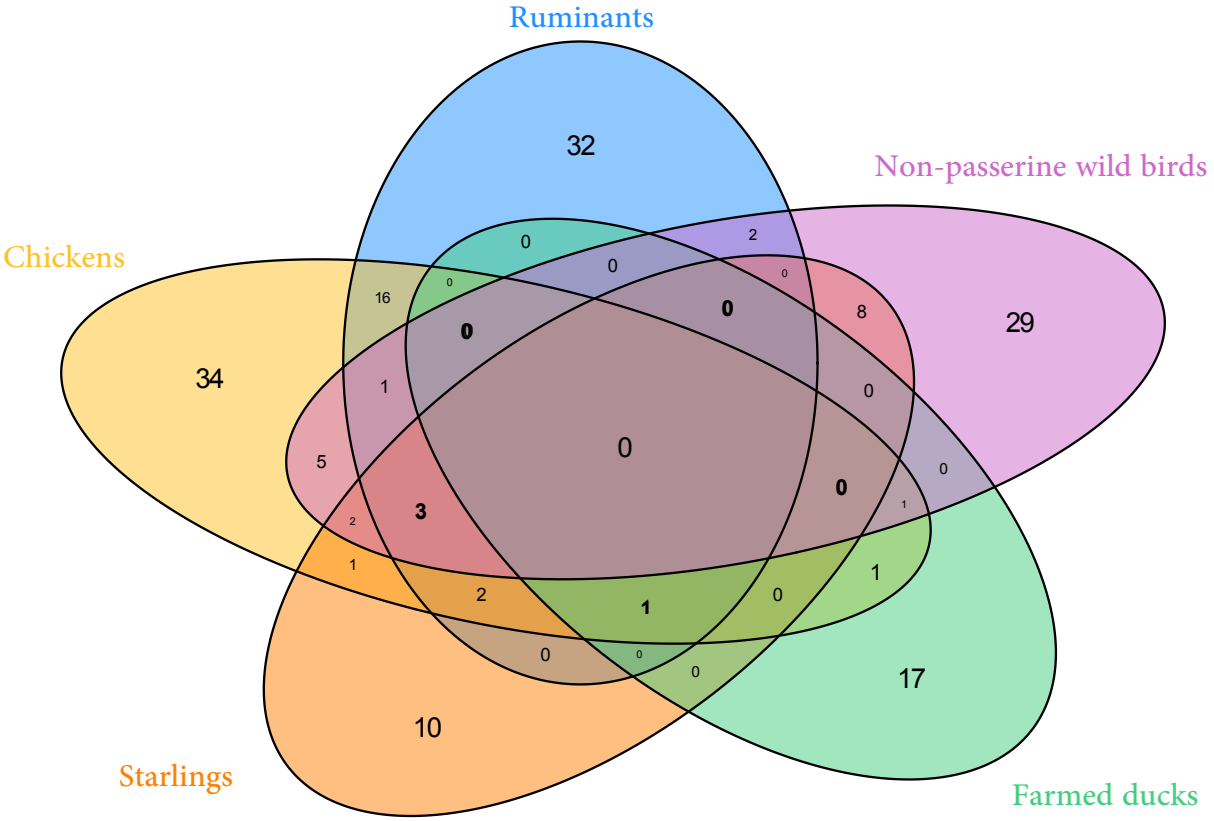


Figure 4.1: The figure shows a Venn diagram of the numbers of unique and shared sequence types for different animal hosts in New Zealand.

It is apparent that, although we can identify a high number of host associated STs, four STs (namely: ST-45, ST-53, ST-137 and ST-583) are common within four out of the five identified host groups (Figure 4.1). The diagram also provides evidence that the farmed ducks represent a niche for *C. jejuni* with highly host associated STs. Only three of their 20 STs are shared with other animal species (one with chickens, one with chickens and non-passerines, one with

ruminants, chickens and starlings).

4.3.1 BAPS and ClonalFrame

The BAPS analysis of the 1957 *Campylobacter* STs identified 17 clusters. Using the weighted table (Table 4.1), most of the clusters were assigned to groups comprising one host species: six clusters (2, 4, 5, 7, 9 and 16) with chickens, three clusters with ruminants (1, 6, 17), one with non-passerine wild birds (cluster 15), one cluster with starlings (10) and six clusters (3, 8, 11, 12, 13, 14) were determined to be generalists not associated with a specific host group.
The ClonalFrame tree (Figure 4.2) is based on concatenated nucleotide sequences and is coloured by host and sequence type (e.g. if the ST was only found in one host species, then it is coloured accordingly). The associated clusters (in brackets) and CCs identified by BAPS are indicated, showing an overall agreement between the two methods.
For example, clusters 4, 5 and 7 which are strongly associated with chickens (Table 4.1) correspond to a chicken associated lineage on the ClonalFrame tree (Figure 4.2), and CC-403 which is strongly associated with ruminants, corresponds to cluster 1 (ruminants) in BAPS. However, there are also noticeable cluster differences between these two methods. Cluster 15 (CC-1275, CC-1332 and unassigned CCs) which is associated with rails and gulls, was split by the ClonalFrame method and clustered on two different parts of the tree, separating the gull lineage from the rail lineage (Figure 4.2). Nevertheless, host associated STs can be identified in each host group; e.g. ST-2381 and ST-4508 in rails, ST-1324 and ST-1304 in starlings, ST-48 and ST-52 in chickens, ST-4343 and ST-4512 in farmed ducks, ST-436 and ST-2026 in ruminants and ST-1223 and ST-2654 in gulls.

Table 4.1: Numbers of isolates in each BAPS- defined cluster that were associated with the six host groups. (Created with the codon linkage model)

	identified Clusters*																
	1(47)	2(54)	3(529)	4(106)	5(55)	6(52)	7(144)	8(419)	9(85)	10(20)	11(24)	12(81)	13(65)	14(96)	15(56)	16(33)	17(91)
CI's based on bootstrap	0.03(0.03-1)	0.69(0.66-0.89)	0	1(1-1)	0.19(0.16-0.25)	0.16(0.08-0.96)	0.2(0.01-0.08)	0	0.58(0.56-0.6)	0.22(0.17-1)	0	0(0-0.03)	0(0-0.33)	0(0-0.02)	0.09(0.06-0.19)	0.14(0.11-1)	10.14(0.09-0.27)
Source																	
Agricultural hosts	**46**	**54**	**526**	**106**	**55**	**52**	**144**	**391**	**85**	**0**	**13**	**78**	**14**	**67**	**21**	**33**	**91**
Chickens	0	49	313	106	33	4	128	373	74	0	9	13	13	19	20	30	18
Farmed ducks	0	0	2	0	0	0	0	1	0	0	1	0	0	48	0	0	0
Ruminants	46	5	211	0	22	48	16	17	11	0	3	65	1	0	1	3	73
Non-agricultural hosts	**1**	**0**	**3**	**0**	**0**	**0**	**0**	**28**	**0**	**20**	**11**	**3**	**51**	**29**	**35**	**0**	**0**
Non-passerine wild birds	1	0	1	0	0	0	0	12	0	2	1	1	46	25	33	0	0
Starlings	0	0	2	0	0	0	0	16	0	18	10	2	5	4	2	0	0
Dominant clonal complex	403	52	21	48	48, 206	21	48	45	354	1304, U/A	177, 677	42	1034, U/A	692, 1034	1275, 1332	257	61

*The number of isolates in each cluster is given in brackets next to the cluster number. The point estimate including lower and upper confidence interval (CI) values is given below the cluster ID. The CI values are based on 10,000 bootstrap iterations. Each cluster shows the number of isolates associated with each host group within. The dominant clonal complexes for each cluster are indicated, U/A is the symbol for an unassigned CC. Clusters associated with a specific host group are coloured in blue.

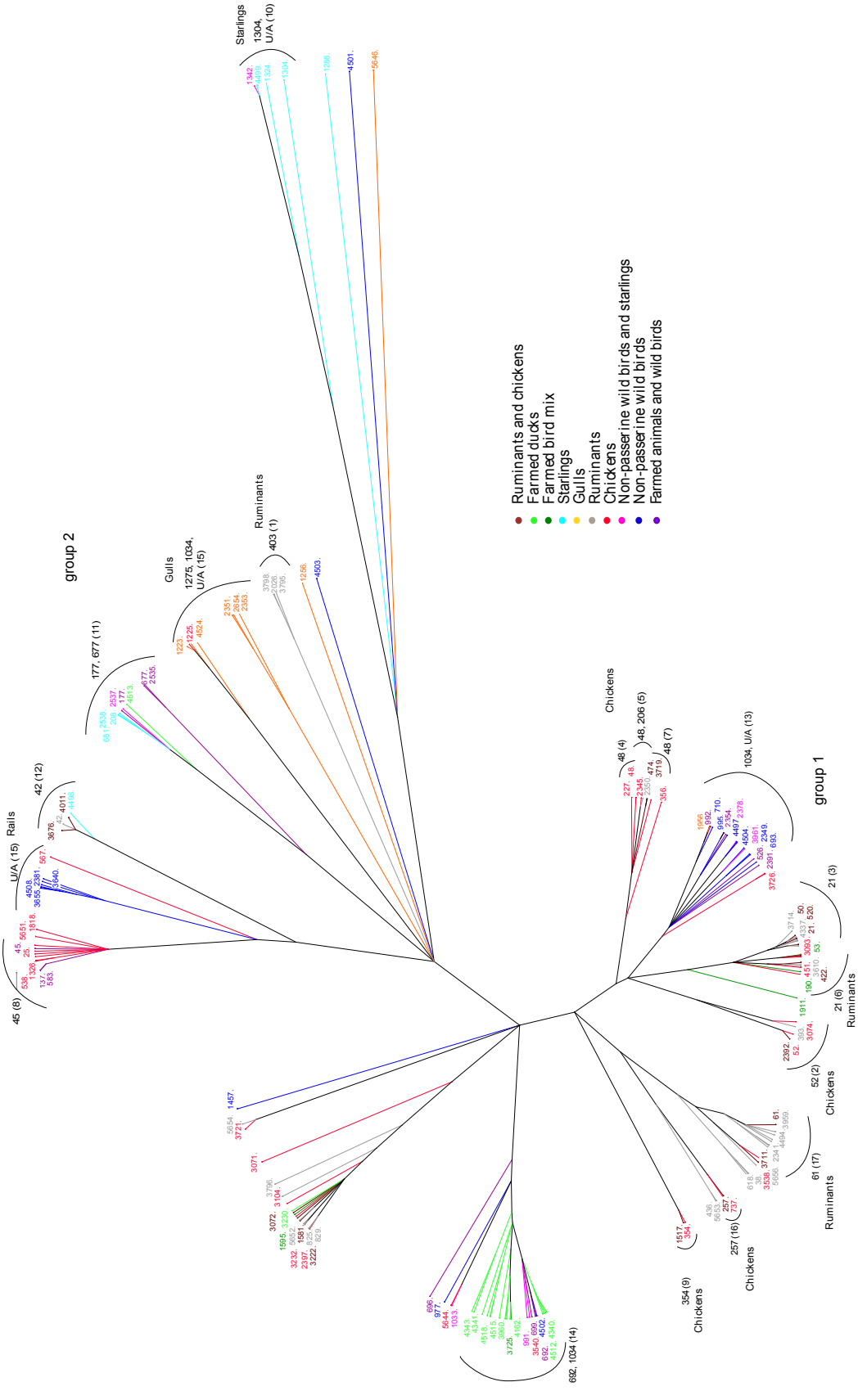


Figure 4.2: ClonalFrame genealogy of *C. jejuni* ST of isolates from New Zealand visualised with SplitTree (Huson and Bryant, 2006). Some of the labels were excluded for the purpose of clarity. The dominant clonal complex and the BAPS cluster (in brackets) associated with each clade were specified.

4.3.2 Estimated admixture and gene flow between host associated clusters estimated by BAPS

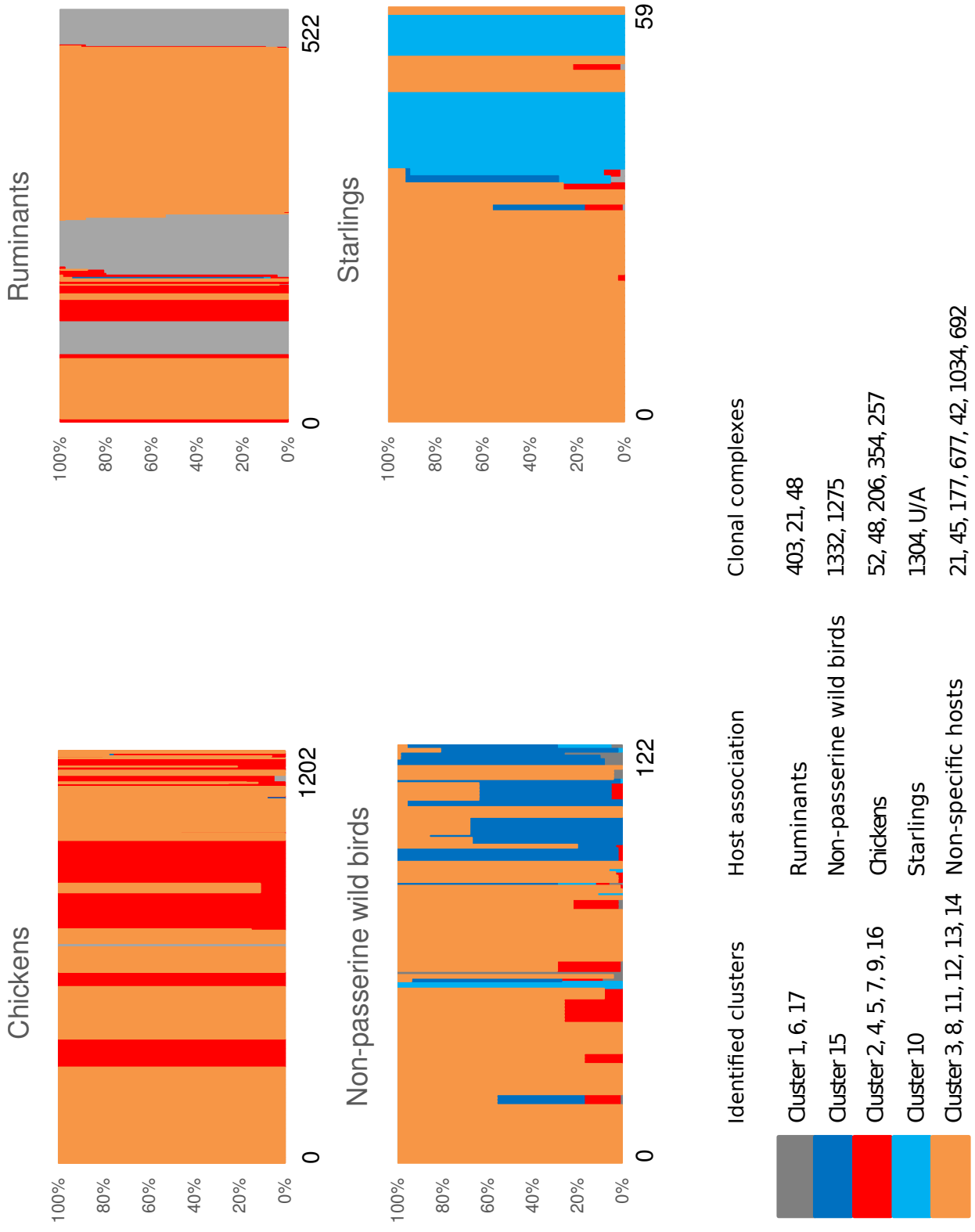


Figure 4.3: The genetic structure of *C. jejuni* genotypes for different host groups. One vertical bar represents one bacterial isolate, showing the probability of the origin from each of the clusters defined by BAPS. The hosts associated with every BAPS cluster are shown by different colours. If genetically distinct populations within each host/niche group (without any admixture or gene flow) were found, then the plots would be uniform in colour. Mosaic alleles (multi-coloured isolates) originate from admixture events (recombination events (HGT) when two or more different strains from different groups of host associated clusters recombine). The colouring representing the hosts is as following: chickens are represented in red, ruminants (grey), non-passerine wild birds (dark blue), starlings (light blue) and generalists in orange. The legend shows the groupings of clusters, the hosts associated with this grouping and the range of clonal complexes within those clusters. U/A stands for unassigned CC.

The admixture plots (Fig. 4.3) show the presence of host associated strains in the agriculture species with a relatively low level of admixture (mosaic alleles) compared to the wild bird species. However it is also apparent that >50% of the isolates in each host-associated group belong to the generalist clusters. This means that they are frequently isolated from more than one host species. It is also noticeable that *C. jejuni* genotypes associated with chickens also appear to be isolated from ruminants, whereas host-associated ruminant isolates are less frequently isolated from chickens. The isolates belonging to the non-passerine wild bird cluster (wild ducks, geese, rails and gulls) show a higher level of admixture than the agricultural host species (chickens and ruminants) (Figure 4.3) and few host-associated genotypes. The opposite is true for the starling cluster which shows a higher number of host-associated isolates. Overall, the admixture plots show the presence of host associated strains in the agriculture species with a relatively low level of admixture compared to the wild bird species.

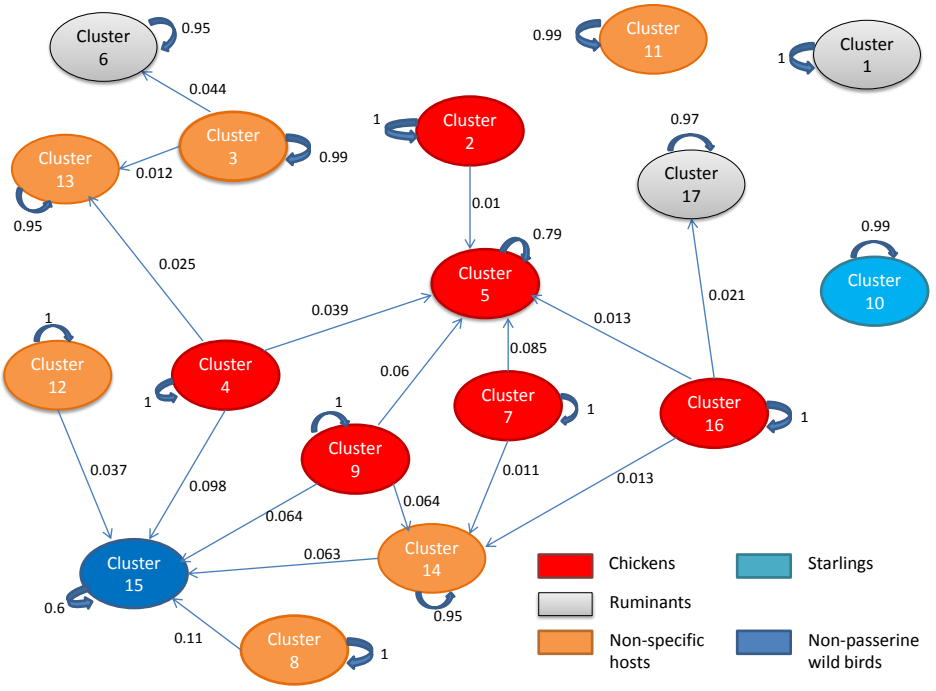


Figure 4.4: This figure visualises a network of clusters where gene flow is indicated by weighted arrows. To investigate the ancestral admixture of a certain population, one can look at all the arrows pointing at this population. The direction and amount of gene flow between the single clusters is indicated by the arrows and the numbers are (proportions) displayed. The self-looping arrows denote the proportion of genes which stay within the population and are not shared with other clusters. Arrows symbolising very low gene flow were pruned to make the plot more readable. The colour of the cluster refers to the host population: red (chicken), grey (ruminants), dark blue (non-passerine wild birds), light blue (starlings) and orange represents the generalist clusters. The number within the cluster is based to the cluster assignment in Table 4.1.

comment 24, Murray

The diagram (Figure 4.4) displays clusters identified by BAPS that are estimated to be involved in the process of gene flow and the direction of the exchange. The clusters can be divided into three main categories: sinks (recipients), donors and clusters which show only few interactions with other clusters. Two main sinks are apparent: cluster 5 and cluster 15. Cluster 5 is a chicken cluster (CC-48 and CC-206) with incoming gene flow from the other chicken associated clusters. Cluster 15, which is a non-passerine wild bird associated cluster (geese, wild ducks, rails, gulls) acts as a recipient, inheriting genetic material from clusters associated with generalists and chickens. This is in agreement with the admixture plots (Figure 4.3) where the non-passerine wild bird *C. jejuni* isolates display high levels of admixture.

4.3.3 Statistical testing of host-association based on ClonalFrame

The density plots in Figure 4.5 show a comparison of the measure of association between *C. jejuni* clades and host-associated STs observed, with that expected by chance. It is apparent that for most of the examined host groups the observed number of host-associated STs is much higher than would be expected by chance. The highest number of host-associated STs was estimated for chickens and ruminants which is in agreement with the admixture plots (Figure 4.3). However, the test also estimated a high number of host-associated STs for the non-passerine wild birds (n=10) which seems to contradict the admixture plots. The host species with the lowest number of host-associated STs were the starlings.

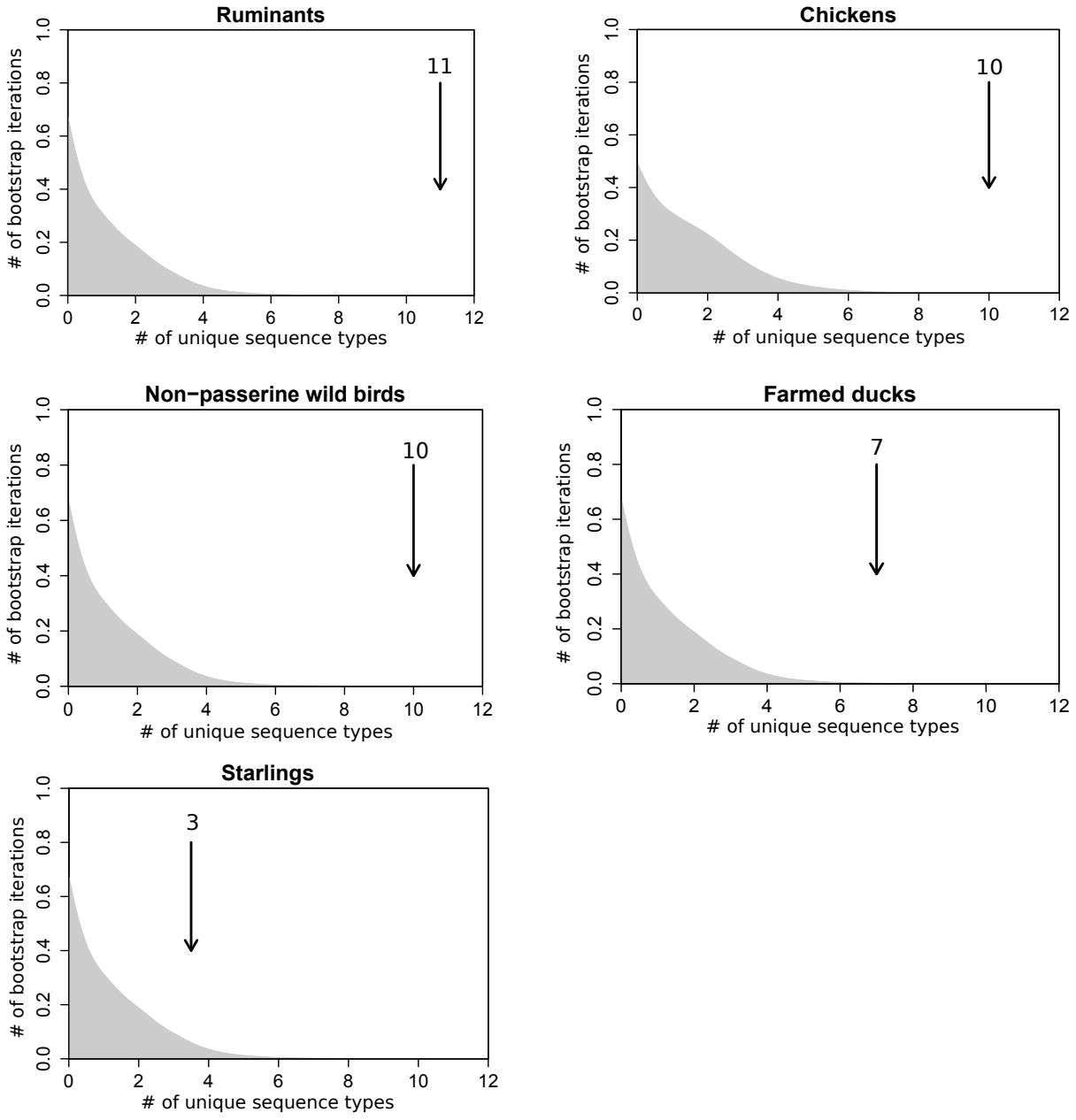


Figure 4.5: The figure shows a comparison of a measure of association between clade as defined by ClonalFrame and isolate host observed with that expected by chance in *C. jejuni* from chickens, ruminants, non-passerine wild birds, farmed ducks and starlings. The arrows indicate the observed measure of association, the densities display the range of results based on 10,000 random permutations of the data set. The graphs are plotted to the same scale.

comment 25, Murray

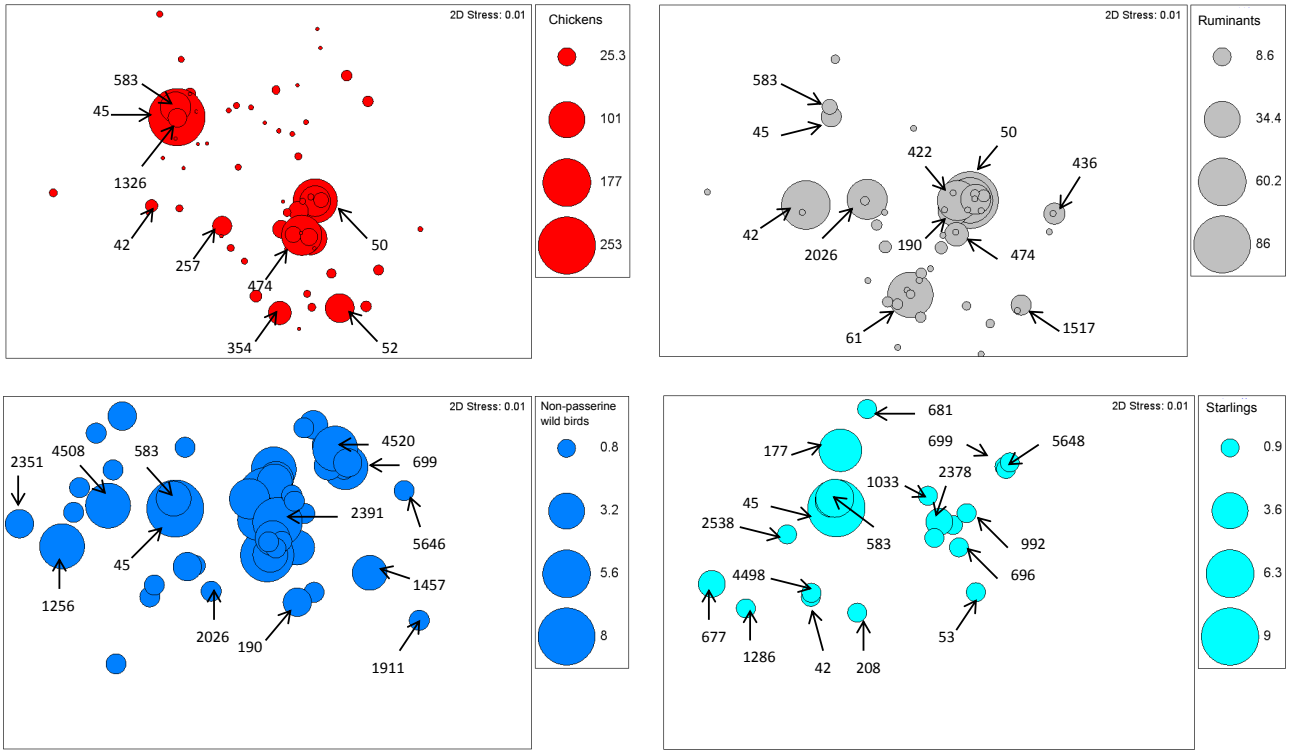


Figure 4.6: Multi-dimensional-scaling bubble plots showing the 3-D *C. jejuni* population structure of agricultural and non-agricultural hosts in New Zealand. Each bubble represents a ST, the distance to the other sequence types is based on a distance matrix generated from their allelic profile. The size of the bubble is scaled according to the number of isolates from each host. ST-2381, which is a ST commonly isolated from rails, has an allelic profile quite different from the other isolates associated with non-passerine wild birds, and is therefore outside the frame of this figure.

The bubble plots in Figure 4.6 show an overview of the population structure of agricultural compared with non-agricultural host populations in NZ. The same scale (distance between the bubbles) is shown for each host group, visualising the differences in the population structure. Each bubble represents a ST and its area is proportional to the number of isolates recovered from each host group. It is evident that, although there are multiple STs shared between the domesticated animals (e.g. ST-474, ST-42, ST-583), some of the STs appear to be host associated as they appear only in one host group, e.g. ST-52 (member of CC-52, group 1 (Figure 4.2), and cluster 2 (Table 4.1)) in chickens. Some STs are present in all host groups, e.g. ST-45, which is considered to be a generalist.

Although the starlings and non-passerine wild birds (rails, gulls, wild ducks and geese) share overlapping habitats, it is noticeable that the different host populations of birds harbour different *C. jejuni* STs with only ST-583 and ST-45 in common. Overall it is apparent that the pattern is more similar, and more STs are shared between the domesticated animals than between the non-passerine wild bird species and starlings.

4.4 Discussion

This study provides insights into the population structure and genetic admixture of *C. jejuni* in agricultural animals and wild birds in New Zealand. The findings of the study indicate that certain *C. jejuni* STs are associated with specific hosts, which is consistent with other studies conducted in Europe (Colles et al., 2008; McCarthy et al., 2007; Sheppard et al., 2011a) and the US (Keller and Shriver, 2014). However, the population of *C. jejuni* STs in New Zealand is different to elsewhere in the world. Although some STs found in both NZ and overseas show similar host associations (e.g. ST-61 with ruminants (Manning et al., 2003), or lack of host association (e.g. ST-45) (Sheppard et al., 2014), others such as CC-692, CC-702, CC-1034 and CC-1332 show different host-associations, being associated with wild geese in the UK (Colles et al., 2008) but wild ducks in NZ. As well as these geographically shared STs, New Zealand has the internationally rare but domestically ubiquitous ST-474 which is strongly host associated (Muellner et al., 2013).

The Venn diagram visualises the number of host-associated and shared STs and indicates the sources of the shared STs (Fig. 4.1). It is apparent that multiple STs found in starlings are also shared with chickens (n=9), ruminants (n=6), non-passerine wild birds (n=11) and farmed ducks (n=1). This result provides further support for the hypothesis that the starlings lifestyle is related to the low number of host-associated STs found. It is apparent from the Venn diagram, that farmed ducks have highly host associated STs. Only three out of 20 STs are shared with other host species. The farmed duck isolates correspond to cluster 14, however, because this cluster also contains several isolates from chickens, non-passerine wild birds and starlings, it was deemed as being not associated with a particular host, even though farmed ducks display a large number of host-associated STs, these STs may be very similar (on the nucleotide level) to STs associated with chickens and other hosts and were therefore clustered together by BAPS. Interestingly, no ST is shared between all five host groups and only four STs (ST-45, ST-53, ST-137, ST-583) are shared between four of the five host groups. ST-45, ST-137 and ST-583 are members of CC-45 whereas ST-53 is member of CC-21. Both of these clonal complexes are generalists (Gripp et al., 2011; Levesque et al., 2008; Sheppard et al., 2014).

The results of the BAPS and ClonalFrame analysis, which were combined and visualised in the neighbour-joining tree (Fig. 4.2), show an overall agreement of the two methods. Generally, STs clustered by ClonalFrame correspond to a cluster identified by BAPS. For example, STs associated with chickens were clustered together by ClonalFrame (group 1) and correspond to the chicken clusters identified by BAPS (clusters 2, 4, 5, 7, 9 and 16). This may not seem surprising as both methods are based on the nucleotide sequence of the seven concatenated housekeeping loci. However, not all the clusters identified by the different methods are identical. The identified cluster 15 (Table 4.1) which consists mainly of isolates from gulls and rails (CC-1275, CC-1034 and ST-2381 (which is not assigned to a CC)), was split by ClonalFrame into two

clusters, separating the STs associated with rails from the STs associated with gulls (Figure 4.2, group 2). The clusters identified by BAPS, are the result obtained from the first step of the analysis, which is a sequential modelling strategy based on the nucleotide sequence similarity of the concatenated MLST alleles modelled with the codon linkage model. ClonalFrame, however, is used to infer bacterial microevolution based on the coalescent theory. Therefore, the cluster identified by BAPS is solely based on the similarity of the sequences within the clusters, whereas ClonalFrame infers the ancestry of the STs. This suggests that the STs in gulls and rails have different most common recent ancestors.

The statistical test of host-association in Fig. 4.5 revealed that the number of host-associated strains, is in general, significantly higher than would be expected by chance alone. However, one group, starlings, showed lower numbers of host associated STs compared to the other hosts (Mohan et al., 2013). The explanation for the observed lower number of host-associated STs for starlings could be as simple as the low number of isolates (n=59) or the environment they live in. Their habitat, lifestyle and mobility exposes starlings to a wide variety of *C. jejuni* STs which might explain the low number of host-associated STs (Mohan et al., 2013). However, due to the low number of *C. jejuni* isolates, the statistical test may not be powerful enough to determine the host associated STs.

The admixture plots (Fig. 4.3) show, that compared to the non-passerine wild birds, there are only few instances of mosaic alleles (admixture) in the agricultural animal and starling *C. jejuni* isolates. The non-passerine wild birds create a larger group containing isolates from rails, wild ducks, wild geese and gulls. These host groups were combined due to the low numbers of isolates from each host group in the database. Extensive exchange of genetic material between *C. jejuni* isolates found in non-passerine wild birds is not surprising due to the shared environment. The greater number of mosaic alleles in the non-passerine wild bird cluster may be partially due to the many bird species combined in this cluster or the greater mobility of the birds. Wild ducks and rails show an annual pattern of dispersion and aggregation (Craig, 1979; Williams et al., 2006) with the aggregation giving greater opportunity for HGT between the *C. jejuni* strains within the birds (due to faecal contamination of the environment). Gulls are scavengers and are commonly found feeding on other potential sources of multiple strains of *Campylobacter*, such as waste in landfill sites.

The admixture plots for starlings (Fig. 4.3) show a large number of host-associated STs, which seems to contradict the results from the statistical test (Figure 4.5. However, only 20 of the 59 starling isolates belong to the BAPS cluster 20, whereas the other isolates are part of the generalist cluster 11. Combining the clusters would show a lower number of host-associated STs as indicated by the statistical test. The less pronounced separation by host species in the agricultural animals, as compared to wild birds, might potentially be explained by movement of lineages between these hosts or recent shared ancestry of these lineages and therefore insufficient genetic variation in the seven housekeeping genes (Sheppard et al., 2011a).

The admixture plots for the agricultural animals show a large number of host associated lineages which seem to be restricted to chickens or ruminants with less indication of admixture within the isolates. However, these isolates only represent a small proportion of the agricultural isolates and the majority of the isolates are combined in the generalist clusters (cluster 3, 8, 11, 12, 13, 14). This means that these isolates are frequently isolated from more than one host group and can therefore not be associated with a particular host (e.g. ST-45 and ST-21) (Sheppard et al., 2014). Farms with free-ranging livestock (cattle and sheep) adjacent to a poultry house pose a risk of contamination to the poultry as, they may excrete *Campylobacter* in high numbers, which may result in contaminated boots, clothing and equipment. In other studies genotypically identical *Campylobacter* strains have been found in cattle kept next to the poultry house and the chickens within the house (Gregory et al., 1997; Newell and Fearnley, 2003). Manipulating the host niche, e.g. the farm environment, by holding genetically distant animals (chickens and ruminants) closely together, could potentially affect lineage associations seen in domestic animals where chickens and ruminants share similar genotypes of *Campylobacter*.

The admixture plots only show the presence of mosaic alleles but not the origin or direction of the exchanged genetic material. However, the direction and amount of the exchange can also be visualised using BAPS (Figure 4.4). The arrows in the plot have been pruned to a specific threshold (≥ 0.01) to increase the readability and to only show the significant amounts of exchange. The main sinks identified by this method are cluster 15 (non-passerine wild bird associated) and cluster 5 (chicken associated). The gene flow plot infers that the isolates in cluster 15 contain genetic material from several generalists and chicken clusters.

An overall picture of the *C. jeuni* population structure in New Zealand agricultural animals and wild bird species can be seen in the multi-dimensional scaling bubble plots (Figure 4.6). The bubble plot represents the shared and unique STs in each host group where the bubbles are scaled according to the numbers of isolates for each ST. It is noticeable that, although the starling and non-passerine wild birds (rails, gulls, wild ducks and geese) share overlapping habitats, the different host populations of birds harbour different *C. jejuni* STs with only ST-583, ST-699 and ST-45 in common. ST-583 and ST-45 belong to CC-45 which is isolated frequently from multiple host species and is therefore referred to as a generalist. These plots strengthen the theory that more shared STs occur between agricultural animals (e.g. ST-583, ST-45, ST-42, ST-474, ST-50) than between wild bird species (McCarthy et al., 2007).

The combined results of the different methods indicate that isolates obtained from wild bird sources have a different population structure to those isolated from agricultural animals. This gives rise to the question as to whether there has been a recent introduction and expansion of certain genotypes in the agriculture environment (Griekspoor et al., 2013; Sheppard et al., 2011a). The limited host range of some of the STs (e.g. ST-1324 and ST-1304 in starlings, ST-2381 and ST-4508 in rails) obtained from wild animals implies host adapted strains. However, there is also evidence of strains that are isolated frequently from multiple sources (e.g. STs

belonging to CC-21, CC-45) (Gripp et al., 2011; Levesque et al., 2008; Sheppard et al., 2014). The two different host ranges (restricted and promiscuous) suggest two different survival strategies for *C. jejuni*. The first is adaptation to and colonisation of one specific host species, and the second one is adaptation to, and colonisation of, several hosts (Griekspoor et al., 2013).

It is still unknown as to how some strains of *C. jejuni* are able to live the life of a generalist as the challenges of colonising organisms with such distinct environments (anatomy, physiology and microflora) is substantial. Sheppard et al. (2014) hypothesised that these generalists could occupy niches different to those occupied by the specialist *Campylobacter* within the same host and therefore avoid selection pressure and out-competition by specialists. Possible factors which have been identified in other pathogens, and which could potentially reduce the cost of adaptation to new environments, are the import of DNA by HGT from already adapted lineages present in each new host species (McCarthy et al., 2007); coordinated genetic regulation of host-specific factors (Gottesman, 1984; Killiny and Almeida, 2011) and genetic change via contingency loci (genes that increase mutation rates in nearby DNA sequences, not across the whole genome) (Kim et al., 2012; Moxon et al., 2006). This study is based on seven housekeeping alleles which are under positive selection pressure and will therefore unlikely reflect genes associated with adaptation.

MLST studies of *C. jejuni* conducted by different groups and in different countries have confirmed the genetic diversity of this bacterium (Cohan and Koeppel, 2008; Coker et al., 2002; Moore et al., 2005; Ogden et al., 2009; Sheppard et al., 2011a). Despite the fact that there is extensive recombination both within and between *Campylobacter* populations (Sheppard et al., 2008, 2011b; Suerbaum et al., 2001), certain groups of related genotypes (clonal complexes), persist over time and geographical spread.

The present study showed that, although some host association is evident at the ST level, this does not necessarily hold at the BAPS cluster scale. It is possible that using higher resolution genotyping, such as ribosomal MLST or whole genome shotgun sequences, would break up some of the generalists clusters and identify more host-associated STs.

New Zealand has a unique geographical location and although the *C. jejuni* population structure appears to be similar to that of other countries, there are some STs highly prevalent in New Zealand that are very rare overseas. One of the highly prevalent STs frequently identified in New Zealand is ST-2381, a sequence type associated with rails, commonly isolated from river water in both the North and South Islands (Carter et al., 2009). This ST has only (as of 2014) been isolated in New Zealand.

Another highly prevalent and important *C. jejuni* ST is ST-474, a member of CC-48. It is strongly associated with poultry and was estimated to account for approximately 25−30% of human campylobacteriosis cases in New Zealand (Muellner et al., 2013; Müllner et al., 2009). ST-474 has only sporadically been isolated outside New Zealand. A few cases have been reported in the literature Clark et al. (2005), Wilson et al. (2008) and Best et al. (2007) and three cases

are recorded on PubMLST (`http://pubmlst.org/campylobacter/`), one is associated with a chicken in the Czech Republic and two isolates were recently identified in the United Kingdom. In conclusion, the applied phylogenetic and population genetic tools in this study have described a detailed picture of the population structure and host associated genotypes within the New Zealand *C. jejuni* population in wild and agricultural animals. The findings show, that the *C. jejuni* STs isolated from non-agricultural animals exhibit a substantially different population structure to those seen in agricultural animals. The *C. jejuni* isolates from non-agricultural animals exhibit a higher number of mosaic alleles and fewer shared genotypes between the host groups, whereas the *C. jeuni* in agricultural animals show a higher number of shared genotypes and fewer occurrences of admixture. This is partially in agreement with a study in the UK (Sheppard et al., 2011a) which also found less separation by genotype in agricultural animals. However, Sheppard et al. (2011a) also found greater admixture among *C. jejuni* isolates in farmed animals, as compared to wild birds, which is in contrast to the results of this study. A reason could be the different cut-offs in the calculation of the host-associated BAPS clusters.

Acknowledgments

AF is funded by the Allan Wilson Centre for Molecular Ecology and Evolution. We are grateful to the staff and students of the mEpiLab team for their support and helpful discussions. Steve Richie for his help with the software Primer and Jukka Corander for helpful comments.

CHAPTER 5

Use of phenotypic microarrays and Genome Wide Association Study (GWAS) to identify phenotype-genotype relationships in New Zealand *Campylobacter jejuni*

Abstract

Campylobacter spp. are human pathogens of international importance readily colonising the intestinal tract of mammalian and avian species. Infections in humans are primarily transmitted through direct contact with livestock or animal-derived food, particularly contaminated poultry. It has been suggested that the different core body temperatures of poultry (42°C) and humans (37°C) trigger potential colonisation or virulence factors and previous studies have demonstrated differential gene expressions at the two temperatures. *Campylobacter* spp. exhibit unique nutritional requirements and have been shown to utilise a limited range of carbon sources for growth. As far as the authors are aware, the present study is the first study to test the ability of a variety of *C. jejuni* isolates from different hosts and STs to oxidise 95 substrates as sole carbon sources and to test the tolerance of the isolates to different osmotic conditions at 38°C and 42°C (simulating the mammalian and poultry GI tracts respectively) using phenotypic microarray (PM) technology. This study aimed to examine the phenotypes associated with different host-associated strains, whether these phenotypes were influenced by different temperatures; and if there was any profound difference between the isolates with and without the *ykg*C gene.

Results indicated that apart from the previously reported differential utilisation of the amino acids L-serine, L-aspartic acid, L-asparagine, and L-glutamic acid, the examined isolates in this study also showed strain-specific utilisation of L-glutamine and citric acid. A genome-wide association study (using the software PhenoLink), combining the phenotypic profiles and whole genome sequences of the isolates, identified genes related to the observed phenotypes. Although no strong, biologically plausible associations between the phenotype and the *ykg*C insertion were detected, the GWAS identified other interesting associations between phenotype expressions and specific genes. This study also identified a type IV secretion system previously unidentified in

C. jejuni which could be involved in the natural transformation of *Campylobacter*. The wider range of potential carbon sources and amino acids utilised by the *C. jejuni* isolates in the present study, and their tolerance to osmolytes and pH, could lead to improvements in culture media for detection and isolation of the bacteria. The different phenotypic profiles of the ST types could also provide new insights into factors associated with colonisation and niche-association.

5.1 Introduction

Campylobacteriosis, mainly caused by the zoonotic bacterial species *Campylobacter jejuni* and *Campylobacter coli*, is a major foodborne illness, causing acute bacterial gastroenteritis in humans worldwide (Coker et al., 2002; Samuel et al., 2004). Its ability to adapt to different animals (ruminants, poultry, wild birds, pets) (Sheppard et al., 2011a; Waldenström et al., 2002) and its wide distribution (animals, food, water, beach sand) (Newell, 2001, 2002) contribute to the complex molecular epidemiology of campylobacteriosis in humans.

It is still unknown how some strains of *C. jejuni* are able to survive as generalists (commonly isolated from multiple sources) as the challenges of colonising animals with such distinct environments (anatomy, physiology and microflora) are substantial. However, the phenotypes of host-associated versus generalist strains, may give a better understanding of metabolic pathways and reveal gene functions associated with adaptation to the environment. Previous studies have demonstrated that different temperatures, for example a chicken's body temperature at 42°C and a human's body temperature of 37°C, trigger expression of potential colonisation or virulence factors which could either lead to commensalism or pathogenesis. These studies demonstrated differential gene expression at the two temperatures (Brás et al., 1999; Pajaniappan et al., 2008; Stintzi, 2003). By taking advantage of fully sequenced bacterial genomes, a relatively new technique, microbial genome-wide association study (GWAS) provides the means to investigate the correlation of genetic variants with phenotypic traits across several bacterial strains enabling the potential discovery of virulence, host-associated or other phenotype-associated genes.

In order to correlate phenotypes and genotypes, for example via the presence/absence of genes, one must be able to consistently measure phenotypic traits. A few studies attempted to correlate phenotype and genotype via comparative genome hybridisation (CGH) (Malloff et al., 2001; Salama et al., 2000; Willenbrock et al., 2006) but one of the limitations is that the genes in the query genome can only be detected in relation to the reference sequence. The cost of genome sequencing has reduced significantly, and the recent development of a phenotype microarray (PM) system has enabled global phenotyping of bacterial strains (Bochner, 2003; Bochner et al., 2001). These developments provide new opportunities for conducting phenotype:genotype comparisons.

C. jejuni is known to be sensitive to environmental stress (temperature, osmolarity and atmospheric oxygen) (Park, 2002), therefore, questions remain how *C. jejuni* are able to adapt to and colonise different hosts. Line et al. (2010) tested the ability of the *C. jejuni* strain NCTC 11168 (GS) (a variant of the wild-type NCTC 11168) to utilise 190 different substrates as sole carbon sources at 38°C and 42°C using phenotypic microarray (PM) technology. Their results indicate that several amino acids, as well as a number of organic acids were utilised, and that the respiration was generally greater at 42°C as compared to 38°C (Line et al., 2010). Tang

et al. (2010) investigated the phenotypic profiles of *C. jejuni* strain ATCC 33560 based on carbon source utilisation and tolerance to osmolytes at 42°C and 30°C using the same phenotypic microarray system. Utilisation was observed for carbon sources from amino acids and carboxylates but not from sugars. No phenotype loss was observed for incubation at 30°C (suboptimal temperature) within the carbon source plate, but a reduction was observed in the tolerance to specific osmolytes and growth at pH 5 (Tang et al., 2010). Previous experiments reported that *C. jejuni* is sensitive to 2% and greater concentrations of NaCl (Cameron et al., 2012; Tang et al., 2010) but showed tolerance to a wide range of food preservatives (sodium lactate, sodium phosphate, sodium benzoate, ammonium sulphate and sodium nitrate) (Tang et al., 2010).
In New Zealand, *C. jejuni* sequence type 474 (ST-474) was estimated to account for approximately 25-30% of human campylobacteriosis cases (Müllner et al., 2009, 2010a), but has been rarely found elsewhere in the world. It was the predominant ST in chicken, but has also occasionally been isolated from red meat sources (Müllner et al., 2009). A recent study (Biggs et al., 2011) comparing a human and a poultry ST-474 isolate, discovered a 12 kb region of non-homologous recombination in one of the genomes. This region includes an insertion between the genes *Cj1069-Cj1070* which has a >99% identity to the *ykg*C gene, a pyridine nucleotidedisulfide-oxidoreductase protein and a conserved hypothetical protein of unknown function. An additional study (Cookson et al., 2011) comparing a broader range of *C. jejuni* strains found that this insertion is over-represented in ruminant-associated STs. Little is known about the function of the *ykg*C insertion but a study in *Escherichia coli* suggested that the gene may be beneficial for survival in seawater (Rozen and Belkin, 2001).
In this study we aimed to examine the phenotypes associated with different host-associated strains and sequence types, whether these phenotypes were influenced by different temperatures; and if there was any profound difference between the isolates with and without the *ykg*C insertion. Using the whole genome sequences of the examined isolates, we aimed to identify genes associated with the observed phenotypes.

5.2 Materials and Methods

5.2.1 Bacteria isolates

C. jejuni isolates in 15% glycerol broth (Nutrient broth No 2. Oxoid, Basingstoke, UK) were sub-cultured on Columbia Horse Blood Agar (BA) (Fort Richard Laboratories, Auckland, New Zealand) and incubated under microaerobic conditions (85% N_2, 5% O_2, 10% CO_2) for 48 hours at 42°C in a VAIN (VA500, Don Whitley Scientific, Yorkshire, UK). A single colony was sub-cultured to a new BA plate for an additional 24 h.
Fifteen *C. jejuni* isolates representing a range of different STs associated with human clinical disease in New Zealand, namely ST-474, ST-42, ST-61 and ST-2026, were chosen based on a

preliminary screening (Cookson et al., 2011) for the presence or absence of the *ykg*C gene. Details of all the isolates used are provided in Table 5.1. The isolates were provided by the mEpiLab team, the OmniLog analysis and preparation for sequencing was carried out by myself.

data clarification

Host	Sequence Type	strain ID	*ykg*C insertion presence/absence
Poultry	474	P694a	+
Poultry	474	P110b	-
Ruminant	474	S168b	+
Human	474	h73020	-
Ruminant	474	S330a	+
Human	474	H22082	+
Ruminant	2026	S22b	-
Human	2026	m28548	-
Ruminant	42	S263b	-
Human	42	H550	-
Ruminant	42	M602b	+
Human	42	H180	+
Ruminant	42	S355b	-
Ruminant	61	S276b	+
Human	61	H450b	+

Table 5.1: Isolates used for phenotypic analysis. The presence/absence of the *ykg*C gene is indicated by a +/- respectively.

5.2.2 Preparation of PM 1 and PM 9 plates

PM 1 plates test for utilisation of 95 different carbon sources. For these plates, *C. jejuni* cells were inoculated into a mixture consisting of 10 mL IF-0a GN/GP (1.2x, Biolog, Hayward, USA), 0.12 mL dye mix D (100x, Biolog, Hayward, USA), 0.2 mL PM additive (12x) and 1.68 mL sterile water. IF0-a buffer is a salt solution for maintaining the cell's viability, without supporting the growth of bacteria cells. The PM additives (12x) consist of 0.6% Bovine Serum Albumin (Sigma, Auckland, New Zealand) and 15 mM $NaHCO_3$ (Sigma, Auckland, New Zealand). A turbiditimeter (Biolog, Hayward, USA) was used to measure a final cell density of 52%, which was used to inoculate the PM 1 plates with 100 μL/well. All steps were carried out in a laminar flow cabinet to ensure a sterile environment. The plates were placed in a microaerobic environment

and inserted without lids into gas impermeable bags (Biolog, Hayward, USA) and heat sealed (Impulse sealer TISH 300C). The PM 1 plates were incubated in the OmniLog incubator (Biolog, Hayward, USA) either at 38°C or 42°C for 42 hours. All assays were repeated at least three times on different days.
Unlike the PM 1 plates where the bacteria use the supplied carbon source for utilisation, the PM 9 plates test for tolerance (survival) of the bacteria to different osmolytes. *C. jejuni* cells were harvested after 24 hours incubation at 42°C and re-suspended into a mixture consisting of 10 mL IF-10a GN/GP (1.2x), 0.12 mL dye mix D (100x), 0.3 mL PM additive (12x) and 1.58 mL sterile water. The PM additives (12x) consist of 0.6% Bovine Serum Albumin, 15 mM $NaHCO_3$ and 12 mM L-lactic acid (Sigma, Auckland, New Zealand). L-lactic acid was identified as the carbon source utilised consistently across all examined isolates and was therefore chosen as the supplement. A turbiditimeter was used to obtain the final cell density of 72%, which was used to inoculate the PM 9 plates with 100 μL/well. From this point on forward, the same procedure as for the PM 1 plates was applied to the PM 9 plates. All assays were repeated at least three times on different days.

5.2.3 Incubation and data recording

The respiration of different assays was recorded spectrophometrically every 15 min by a charged coupled device (CCD) camera over 48 h in the OmniLog. Each well of the 96-well plate is designed to tests the microbe's ability to catabolise a particular metabolite (PM 1) or to measure the tolerance to osmolytes (PM 9). Positive results produce a visible purple colour change, which darkens in conjunction with increasing metabolic response. The camera captures a digital image of the panel and stores the quantitative colour change values in a computer file. These can be accessed with the OmniLog PM software and displayed as kinetic graphs.

5.2.4 Statistical analysis of the kinetic data

The raw data were exported from the OmniLog PM software and further analysis and visualisation were conducted in R (R Development Core Team, 2013). The data were converted into opms files using the R package *opm* (Vaas et al., 2013). The add-on package *lattice* (Sarkar, 2008) with the function *levelplot()* was used to visualise the PM curves as heat maps for a general overview, and then, for a more detailed and comprehensive view the function *xyplot()* to plot the curves of all repeats per isolate in a grid-like structure. All respiration curves were included in the supplementary files A.2. Four different parameters can be estimated from the respiration curves, A (the maximum recorded value), AUC (the area under the curve), λ (the lag phase) and μ (the increase in respiration which corresponds to the "slope") (Vaas et al., 2012). The function *kmeans()* (Hartigan and Wong, 1979) based on the A value was used to group the isolates into two distinct clusters to calculate the means across the repeats for each

isolate. These means were plotted with the package *pheatmap* (Kolde, 2012). The heatmaps are included in the supplementary file A.3.
In order to assess the significance of MLST types on specific wells and the significance of the presence/absence of the *ykg*C gene in the isolates, a linear mixed effects model *lme()* was fitted to the data (Pinheiro et al., 2006). The isolate was treated as a random effect where i is the well, j the presence or absence of the *ykg*C insertion, k is the MLST type, t the temperature and l the isolate.

$$\begin{aligned} \mathsf{A\text{-}value}_{ijktl} &= \mu_i(\mathit{ykg}\mathsf{C}_{ij} + \mathsf{MLST}_{ik} + \mathsf{temp}_{it}) + \mathsf{isolate}_l + \varepsilon_{ijktl} \\ \text{with } \mathsf{isolate}_l &\sim \mathsf{Normal}(0, \sigma^2) \text{ and } \varepsilon_{ijktl} \sim \mathsf{Normal}(0, \sigma^2) \end{aligned}$$

REEMtrees (Random Effects-Expectation Maximization) are a combination of mixed effects models for clustered data with the flexibility of tree-based estimation methods (Sela and Simonoff, 2012). The underlying algorithm for REEMtrees is based on recursive partitioning, where all the data is examined and split into subsets based on one predictor at a time. The data is partitioned recursively to minimise the variability in each node (Sela and Simonoff, 2012). The R package *REEMtree* (Sela and Simonoff, 2010) was used to examine the parameters associated with the greatest variation.

5.2.5 Isolation of DNA and sequencing

DNA was prepared from isolates sub-cultured onto BA and incubated under microaerobic conditions for 48 h at 42°C. A single colony was sub-cultured to a new plate after 48 h. For the nine isolates which have not been sequenced before, DNA was extracted with the QIAamp DNA Mini, QIAamp Blood Mini kit (Qiagen, Auckland, New Zealand) or Wizard Genomic DNA Purification Kit (Promega, Madison, Wisconsin, USA) according to the manufacturer's instructions. The prepared samples were sent to NZGL (New Zealand Genomics Limited, Massey University, Palmerston North, New Zealand) for full-genome sequencing using Next Generation Sequencing Technology (Illumina, MiSeq).
Each of the nine genomic DNA samples were fragmented by nebulisation for 6 minutes at a pressure of 32 psi, purified, then end repaired, A-tailed, adaptor-ligated, fractionated, purified and enriched according to the manufacturer's instructions, using the TruSeq DNA LT Sample Prep Kit v2-Set A (48 reactions with PCR): part number FC-121-2001 and TruSeq DNA LT Sample Prep Kit v2-Set B (48 reactions with PCR): part number FC-121-2002, Illumina Inc. The prepared libraries were normalised to equal molarity, diluted to 2nM and pooled; 20 libraries per pool. A flow cell was prepared for each of the two library pools and sequencing reactions using 9 pmoles of the pooled libraries were performed on an Illumina MiSeq instrument with the MiSeq Reagent Kit v2 (500 cycle): part number MS-102-2003, Illumina Inc (Ilumina, Victoria, Australia), to give approximately 12 to 15 million clusters per run.
The genome characteristics of the 15 isolate are summarised in Table A.1.

5.2.6 Quality control and *de novo* assembly of sequence reads

SolexaQA (Cox et al., 2010) was used for trimming and quality control of the reads, subsequently the *de novo* assembler Velvet (version 1.2.10 (Zerbino and Birney, 2008)) was used for assembling the 2x 250 base pair short reads. The de Bruijn graph-based genome assembly algorithm works in two steps: short reads are broken into small pieces (k-mers) and a de Bruijn graph is constructed from those short pieces; the genome is derived from the paired de Brujin graph (Eulerian path). The sequences were assembled across a range of k-mers (in steps of 10), for the odd numbers between 245 and 55 inclusive. These resulting Velvet contigs were stored in a MySQL database.

5.2.7 Gene prediction and clustering

For each isolate, concatenated contigs generated at each k-mer (an assembly) were ranked based on a score derived from their N50 (a statistical measure of average length of a set of sequences), smallest number of contigs, maximal contig length and assembly length. A subset of three assemblies were chosen and annotated with Prokka (Seemann, 2014) to determine the robustness of the assembly method. Using customised Perl scripts the annotated contigs from each assembly were clustered using OrthoMCL (version 2.09; (Chen et al., 2006a; Li et al., 2003b)) with the default parameters and cluster length range of 20 amino acids.

5.2.8 Gene-trait matching

Genes of relevance were identified using PhenoLink (Bayjanov et al., 2012), a web-based software linking genomic and phenotypic data in order to find underlying associations. PhenoLink uses a random forest approach as its underlying algorithm to determine the importance of the genes in the observed phenotype by building an ensemble of decision trees. The presence/absence matrix of the pangenomes was created using OrthoMCL (Chen et al., 2006b; Li et al., 2003b) and custom Perl scripts. A matrix with the phenotypic profile of the isolates was based on the A value calculated by the function *kmeans()* (Hartigan and Wong, 1979).

5.2.9 Confirmation of growth on salt plates

Tolerance to NaCl was performed as described by On and Holmes (1991), using nutrient broth No 2. (Oxoid, Basingstoke, UK) with 2% agar (Gibco, Life Technologies, Paisley, Scotland) and NaCl (Scharlau, Sentmenat, Spain) to bring the total NaCl concentration to 0.5% (the NaCl concentration in nutrient broth No. 2), 1%, 2% and 4%. *Campylobacter* were grown on BA microaerobically for 24 h at 42°C and suspended in cation-adjusted Mueller-Hinton broth (Fort Richard, Auckland, New Zealand) to a density approximating 10^6 cfu/ml. Ten μl volumes of the suspension were dispensed onto the agar and plates were incubated microaerobically at 42°C for

48 h. *C. jejuni* NCTC 11168 and *C. lari* NCTC 11352 were used as controls in this experiment.

5.3 Results

5.3.1 Utilisation of carbon sources at 38°C and 42°C

The positive utilisation of substrates as sole carbon sources by the examined 15 isolates at 38°C and 42°C is shown in Table 5.2. The means of all repeats at 38°C and 42°C on PM 1 were calculated and visualised as a heatmap in Figure 5.1 and Figure 5.2 respectively. The colours in the heatmap correspond to the average amount of oxidation of the dye due to respiration in the specific well. The heatmaps for PM 1 at 42°C and PM 9 at 38°C and 42°C can be found in the supplementary material A.3 (Figure A.61, Figure A.62 and Figure A.63 respectively). The heatmap in Figure 5.1 shows no apparent association of the respiration of specific carbon sources with the presence or absence of the *ykg*C insertion (H22082, P694a, S168b, S330a, H180, M602b, H450b, S276b), however a clustering relating the ST type (ST-474, ST-42, ST-61 and ST-2026) and the associated phenotype is quite noticeable at 38°C (not at 42°C as much).
Table 5.2 displays the utilisation pattern across the 15 isolates on PM 1 at 38°C and 42°C. The table cells are coloured to show the changed phenotype of the isolates. Green symbolises that the isolate did not utilise the carbon source at 38°C but did at 42°C (e.g. L-aspartic (A07) acid for isolate S22b) and red when the isolate utilised the carbon source at 38°C but not at 42°C (e.g. L-proline (A08) for isolate P110b). Some carbon sources (succinic acid (A05) and L-lactic acid (B09)) are utilised at both temperatures across all isolates, whereas others are primarily associated with one isolate (S263b and glyoxylic acid (F10)) or one ST (gly-glu (G01) mainly ST-474; L-glutamine (E01) mainly ST-42). Some amino acids were utilised as carbon sources by all isolates (L-aspartic acid (A07), L-glutamic acid (B12) and L-asparagine (D01)), whereas others were only utilised across specific isolates (L-proline (A08), L-glutamine (E01) and L-serine (G03)). Overall, an increase in respiration (green cells) was observed from 38°C to 42°C in 16 instances, whereas a decrease in respiration from 38°C to 42°C was observed in 46 instances.
False-positives reactions were observed for the pentose sugars D-xylose (B8), D-ribose (C4), L-Lyxose (H6) and L-arabinose (A2) in PM 1. This also has previously been reported by Line et al. (2010).

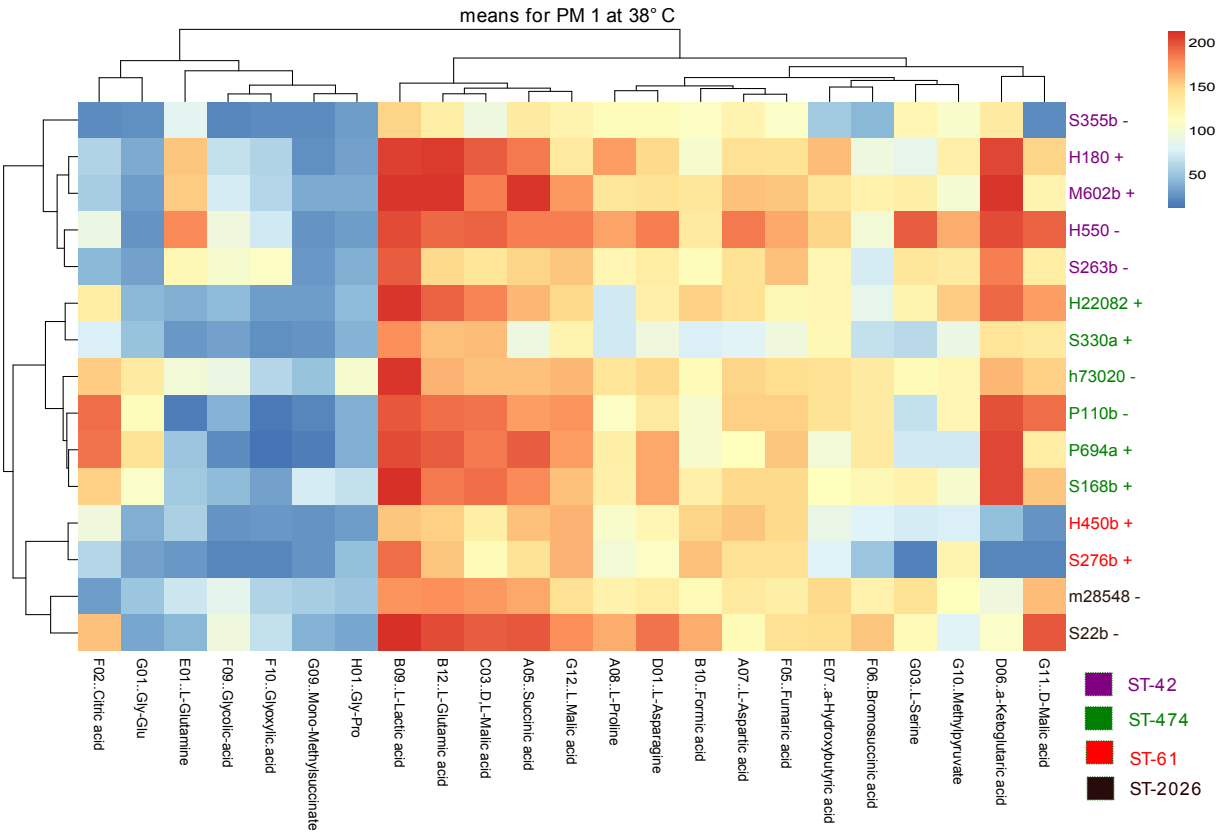


Figure 5.1: Heatmap displaying the utilisation pattern of the 15 examined isolates in PM 1 at 38°C. The plot is based on the calculated means across the repeats per isolate and the scale of the heatmap refers to the mean A value (from no utilisation (blue), to high utilisation (red)). The isolates were coloured according to the corresponding ST (ST-42 in purple, ST-474 in green, ST-61 in red and ST-2026 in brown).

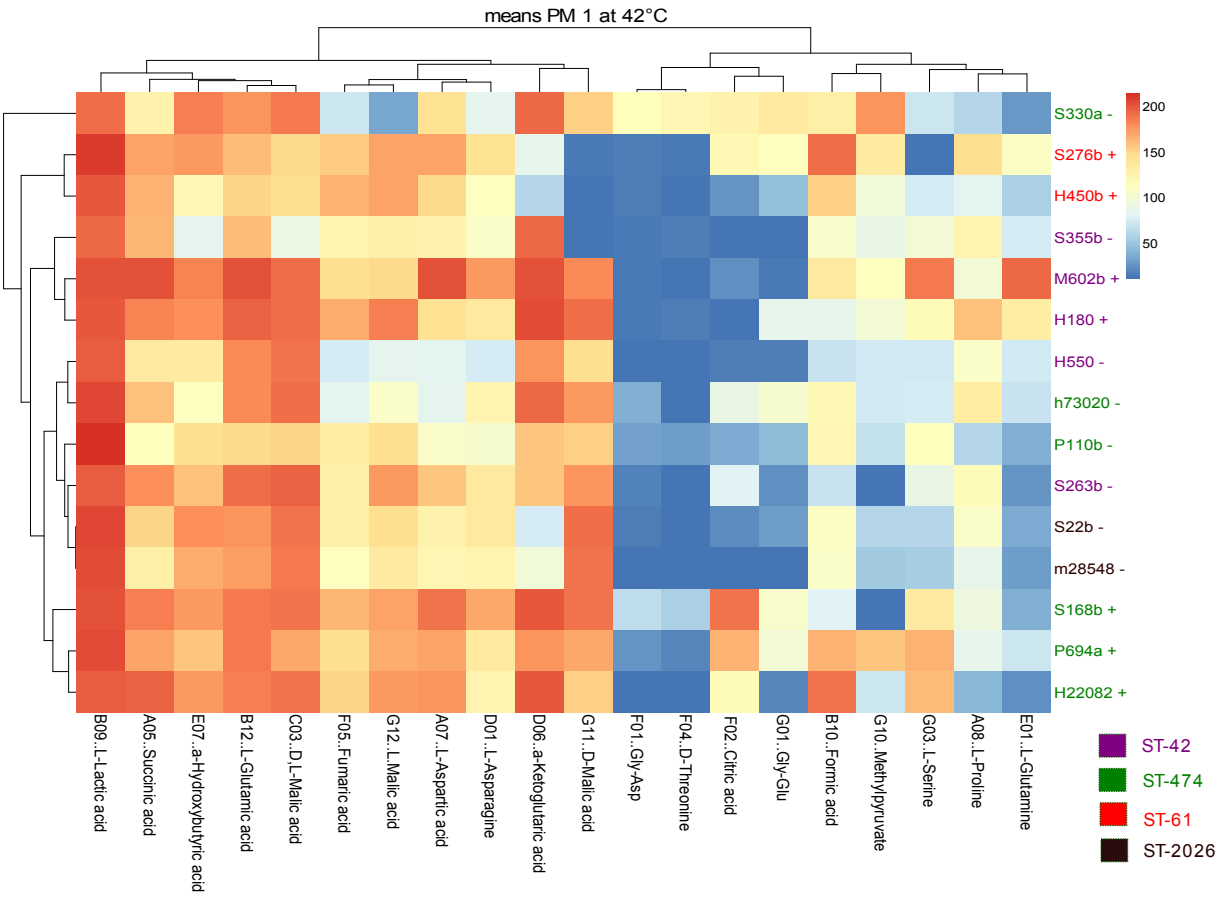


Figure 5.2: Heatmap displaying the utilisation pattern of the 15 examined isolates in PM 1 at 42°C. The plot is based on the calculated means across the repeats per isolate and the scale of the heatmap refers to the mean A value (from no utilisation (blue), to high utilisation (red)). The isolates were coloured according to the corresponding ST (ST-42 in purple, ST-474 in green, ST-61 in red and ST-2026 in brown).

Table 5.2: Utilisation in PM 1 at 38°C and 42°C

	ST-474						ST-2026		ST-42					ST-61	
	P694a(+)	P110b(-)	S168b(+)	h73020(-)	S330a(+)	H22082(+)	S22b(-)	m28548(-)	S263b(-)	H550(-)	M602b(+)	H180(+)	S355b(-)	S276b(+)	H450b(+)
	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42
succinic acid (A05)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+
L-aspartic acid (A07)	+/+	+/+	+/+	+/-	+/+	+/+	-/+	+/+	+/+	+/-	+/+	+/+	+/+	+/+	+/+
L-proline (A08)	+/+	+/-	+/+	+/+	-/-	-/-	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/-
L-lactic acid (B09)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+
formic acid (B10)	+/+	+/+	+/-	+/+	-/+	+/+	+/+	+/+	+/-	+/-	+/+	+/-	+/+	+/+	+/+
L-glutamic acid (B12)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+
D,L-malic acid (C03)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+
L-asparagine (D01)	+/+	+/+	+/+	+/+	+/-	+/+	+/+	+/+	+/+	+/-	+/+	+/+	+/+	+/+	+/+
α-ketoglutaric acid (D06)	+/+	+/+	+/+	+/+	+/+	+/+	-/-	+/+	+/+	+/+	+/+	+/+	+/+	-/-	-/-
L-glutamine (E01)	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	+/-	+/-	+/+	+/+	-/+	-/+	-/-
α-hydroxybutyric acid (E07)	-/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/+	-/+	-/+
citric acid (F02)	+/+	+/-	+/+	+/-	+/+	+/+	+/-	-/-	-/-	-/-	-/-	-/-	-/-	-/+	+/-
D-threonine (F04)	-/-	-/-	-/-	-/-	-/+	-/-	-/-	-/-	-/-	+/-	-/-	-/-	-/-	-/-	-/-
fumaric acid (F05)	+/+	+/+	+/+	+/-	+/-	+/+	+/+	+/+	+/+	+/-	+/+	+/+	+/+	+/+	+/+
bromosuccinic acid (F06)	+/-	+/-	+/-	+/-	-/-	-/-	+/-	+/-	-/-	+/-	+/-	+/-	-/-	-/-	-/-
glycolic acid (F09)	-/-	-/-	-/-	-/-	-/-	-/-	+/-	+/-	+/-	+/-	-/-	-/-	-/-	-/-	-/-
glyoxylic acid (F10)	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	+/-	-/-	-/-	-/-	-/-	-/-	-/-
gly-glu (G01)	+/+	+/-	+/+	+/+	-/+	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/+	-/-
L-serine (G03)	-/+	-/+	+/+	+/-	-/-	+/+	-/-	+/-	+/-	+/-	+/+	-/+	+/+	-/-	-/-
methylpyruvate (G10)	-/+	-/-	-/-	+/-	+/+	+/-	-/-	+/-	+/-	+/-	+/+	+/+	+/+	+/+	+/+
D-malic acid (G11)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/-	-/-	-/-
L-malic acid (G12)	+/+	+/+	+/+	+/+	+/-	+/+	+/+	+/+	+/+	+/-	+/+	+/+	+/+	+/+	+/+

This table shows the overall utilisation pattern of the 15 examined isolates in PM 1 at 38°C and 42°C. Green cells symbolise a phenotype gain (from 38°C to 42°C) and red cell refer to a phenotype loss between the two temperatures.

5.3.2 Tolerance to osmolytes (PM 9) at 38°C and 42°C

The tolerance to osmolytes of the 15 examined isolates at 38°C and 42°C is summarised in Table 5.3. Only the wells in which an isolate showed phenotypic variation were included in the Table 5.3. All isolates showed respiration in wells D06-D12, E01-E09, F01, F02, F04, F05, G01-G04, G09, G10 and H01-H04 across both temperatures. All isolates showed a change in the phenotype expression across the two temperatures. Comparatively, the isolates showed an increased respiration (from 38°C to 42°C) in 53 instances, compared to a decrease in 29 instances. The *C. jejuni* isolates were generally able to respire at concentrations of sodium chloride (NaCl) of up to 3% (A01-A04). All isolates showed tolerance to potassium chloride at levels of 3%-4% (D01 and D02), to sodium sulphate at levels of 2%-8% (D05-D08), to levels of ethylene glycol at 5%-20% (D09-D12), to levels of sodium formate at 1%-6% (E01-E06), to levels of urea at 2%-7% (E07- E12), to levels of sodium lactate at 1%-12% (F01-F12), to levels of sodium phosphate pH 7 (G01 (20mM)- G04 (200mM)), to 20mM sodium benzoate at pH 5.2 (G05), to ammonium sulphate at pH 8 (G09 (10mM)- G12 (100mM)), to sodium nitrate (H01 (10mM)- H06 (100mM)) and to sodium nitrite (H07 (10mM) and H12 (100mM)). Some of the ST-474 isolates (P694a, P110b, h73020 and S330a, H22082) and one ST-42 isolate (H550) showed an increased tolerance to NaCl at 4%. The results related to the increased tolerance at higher concentrations of sodium chloride were examined experimentally as described in On and Holmes (1991). No growth of *C. jejuni* was observed on agar plates for a salt concentration > 1%, which is inconsistent with the results in Table 5.3 but may be due to the fact that within the PM plates respiration of the bacteria can occur independent of cell growth (Bochner et al., 2001).

comment 29, Murray

The heatmaps at 38°C and 42°C (supplementary material A.3) did not show clustering related to the presence or absence of the *ykg*C insertion, temperature or ST type. No false-positive reactions were observed in the PM 9 plate.

Table 5.3: Tolerance to osmolytes in PM 9 at 38°C and 42°C

	ST-474						ST-2026		ST-42					ST-61	
	P694a(+)	P110b(-)	S168b(+)	h73020(-)	S330a(+)	H22082(+)	S22b(-)	m28548(-)	S263b(-)	H550(-)	M602b(+)	H180(+)	S355b(-)	S276b(+)	H450b(+)
	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42	38/42
NaCl 1% (A01)	+/+	+/+	+/+	+/+	+/+	+/+	-/+	+/+	+/+	+/+	+/+	-/+	-/+	+/+	+/+
NaCl 2% (A02)	+/+	-/+	+/+	+/+	+/+	+/+	-/-	+/+	-/+	+/+	+/-	-/+	-/+	+/+	+/+
NaCl 3% (A03)	+/+	-/+	-/+	-/+	+/+	+/+	-/-	-/+	-/-	+/+	-/-	-/+	-/+	+/+	+/-
NaCl 4% (A04)	-/+	-/+	-/-	-/+	+/-	+/+	-/-	-/-	-/-	-/+	-/-	-/-	-/-	-/-	-/-
NaCl 10% (A12)	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/+	-/-
NaCl 6% + L-Carnitine (B12)	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/+	-/-
NaCl 6% + Gluathione (C06)	-/-	-/-	-/-	-/-	+/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-	-/-
potassium chloride 3% (D01)	+/+	+/+	+/+	-/+	+/+	+/+	-/-	-/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+
potassium chloride 4% (D02)	-/-	-/+	-/-	-/-	+/-	-/-	-/-	-/-	-/-	+/+	-/-	+/+	-/-	-/-	+/-
sodium sulphate 2% (D05)	+/+	+/+	+/+	+/+	+/+	+/+	-/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+
urea 5% (E10)	+/+	+/+	+/-	+/+	+/+	+/+	+/-	+/+	+/+	+/+	+/+	+/+	-/+	+/-	+/+
urea 6% (E11)	+/-	+/+	-/-	-/-	+/-	-/-	-/-	+/-	+/-	+/+	+/-	+/-	-/-	-/-	+/-
urea 7% (E12)	+/-	+/-	-/-	-/-	+/-	-/-	-/-	-/-	+/-	+/+	-/-	-/-	-/-	-/-	+/-
sodium lactate 3% (F03)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/-	+/+	+/-
sodium lactate 6% (F06)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/-	+/+	+/-
sodium lactate 7% (F07)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/-	+/+	+/+
sodium lactate 8% (F08)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/-	+/+	+/+
sodium lactate 9% (F09)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/-	+/+	+/-
sodium lactate 10% (F10)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/+	+/+	-/+	+/+	-/-	+/+	+/-
sodium lactate 11% (F11)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/+	+/+	-/+	+/+	-/-	+/+	-/-
sodium lactate 12% (F12)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/+	+/+	-/+	+/+	-/-	+/+	+/+
sodium benzoate pH 5.2 20mM (G05)	+/+	+/+	-/+	-/+	+/+	+/+	-/+	+/+	-/+	-/+	+/+	-/+	-/-	+/+	-/+
ammonium sulphate pH 8 50mM (G11)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/+	+/+	+/+	+/+
ammonium sulphate pH 8 100mM (G12)	+/+	+/+	+/+	+/+	+/+	+/+	-/+	+/+	+/+	+/+	+/+	+/+	-/+	+/+	+/+
sodium nitrate 80mM (H05)	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	-/+	+/+	+/+
sodium nitrate 100mM (H06)	+/+	+/+	+/+	+/+	+/+	+/+	-/+	+/+	+/+	-/+	+/+	-/+	+/+	+/+	+/+
sodium nitrite 10mM (H07)	+/+	-/-	-/+	-/+	+/-	-/+	+/+	+/+	+/+	-/+	-/+	+/-	-/-	+/+	-/+
sodium nitrite 100mM (H12)	-/-	-/-	-/-	-/+	+/+	-/-	-/+	-/-	-/+	-/-	-/+	-/+	-/-	-/-	-/+

This table shows the overall tolerance of the 15 examined isolates in PM 9 at 38°C and 42°C. Green cells symbolise a phenotype gain (from 38°C to 42°C) and red cell refer to a phenotype loss. Only the wells with variation between the isolates are listed in the table. All isolates showed respiration in D06-D12, E01-E09, F01, F02, F04, F05, G01-G04, G09, G10, H01-H04 across both temperatures.

5.3.3 Linear mixed effects models for PM 1 and PM 9

A linear mixed effects model (lme) was used to examine possible relationships between temperature, *ykg*C insertion and ST type and the utilisation and osmotic tolerance profile of the 15 examined isolates. The results are summarised in Table 5.4. The main effect and the standard error (se), which is identical for a given main effect across all wells, is indicated in the header. The first column is the maximum height of the curve in Omnilog units for the given substrate at the baseline (no *ykg*C insertion, temperature at 38°C, ST-474). Subsequent columns are offsets from the intercept, showing an increase or decrease (given in Omnilog units). The numbers in brackets are the p-values, with cells coloured according to the significance level. A blue cell colour represents a positive correlation between the main effect (*ykg*C insertion, temperature, ST-type) and the substrate (increase in Omnilog units), the brown colour represents a negative correlation (decrease in Omnilog units).

The results indicated that the *ykg*C insertion has a positive association with six substrates including succinic acid (A05) showing an increase of 25.52 Omnilog units (P=0.04). All the *C. jejuni* isolates showed significantly higher utilisation in some wells at 42°C compared to 38°C (D,L-malic acid (C03), α-ketoglutaric acid (D06), α-hydroxybutyric acid (E07), bromosuccinic acid (F06), glycolic acid (F09), D-malic acid (G12)) and lower utilisation at higher temperature in citric acid (F02) and gly-glu (G01). ST-42 isolates showed significantly higher utilisation ability in two substrates compared to ST-474 (L-proline (A08), L-glutamine (E01)) and lower utilisation in three substrates (citric acid (F02), gly-glu (G01) and D-malic acid (G11)). One significant difference was observed in L-glutamine (E01, an increase of 89.43 Omnilog units, P< 0.0001), suggesting that ST-474 isolates were less likely to utilise L-glutamine (baseline= 27.35 Omnilg units). The other significant difference was seen in citric acid (F02, -89.92 Omnilog units, P< 0.0001) compared to ST-474 (baseline= 137.26 Omnilog units). This suggests that ST-42 isolates were less likely to utilise citric acid.

Similar results were seen for ST-2026 and ST-61 where the isolates showed higher or lower utilisation of specific substrates compared to ST-474. ST-2026 isolates indicated increased utilisation of five substrates (D,L-malic acid (C03), α-hydroxybutyric acid (E07), bromosuccinic acid (F06), glycolic acid (F09), D-malic acid (G11)) compared to ST-474 and decreased utilisation ability of three substrates (α-ketoglutaric acid (D06), citric acid (F02) and gly-glu (G01)). ST-61 isolates showed overall reduced utilisation compared to ST-474. ST-61 showed greater utilisation ability of only one substrate (formic acid (B10)), compared to nine substrates where it showed less utilisation (L-glutamic acid (B12), D,L-malic acid (C03), α-ketoglutaric acid (D06), α-hydroxybutyric acid (E07), citric acid (F02), bromosuccinic acid (F06), glycolic acid (F09), L-serine (G03) and D-malic acid (G11)) compared to ST-474. The greatest negative difference was seen in substrate D-malic acid (G11), showing a decrease of -152.16 Omnilog units (P< 0.0001). This suggests that the isolates belonging to ST-61 were less likely to utilise D-malic

acid. It is apparent that the substrate L-glutamine (E01) was only utilised by ST-42 and the substrate citric acid (F02) was only utilised by ST-474.

Table 5.4: Results of the linear mixed effects model for PM 1 at 38°C and 42°C

substrate	baseline (12.03)[a]	*ykg*C(+) (11.64)	temp 42°C (9.67)	MLST-42 (11.61)	MLST-2026 (17.22)	MLST-61 (15.57)
succinic acid (A05)	137.73 (0.00)[b]	25.52 (0.04)	-1.06 (0.09)	23.86 (0.78)	29.99 (0.09)	-4.45 (0.78)
L-aspartic acid (A07)	118.52 (0.00)	19.88 (0.10)	7.89 (0.98)	18.82 (0.12)	-0.48 (0.98)	8.42 (0.59)
L-proline (A08)	102.16 (0.00)	0.83 (0.94)	-21.53 (0.07)	41.06 (0.00)	33.07 (0.07)	13.31 (0.40)
L-lactic acid (B09)	195.69 (0.00)	8.10 (0.49)	4.90 (0.63)	-6.40 (0.59)	8.34 (0.63)	-14.68 (0.36)
formic acid (B10)	111.46 (0.00)	14.53 (0.22)	-2.94 (0.19)	-2.98 (0.80)	23.42 (0.19)	40.80 (0.02)
L-glutamic acid (B12)	156.35 (0.00)	32.03 (0.01)	0.03 (0.07)	9.82 (0.41)	32.85 (0.07)	-34.28 (0.04)
D,L-malic acid (C03)	149.85 (0.00)	36.37 (0.00)	4.46 (0.02)	-2.39 (0.84)	41.57 (0.02)	-50.79 (0.00)
L-asparagine (D01)	128.48 (0.00)	14.84 (0.21)	-13.81 (0.15)	7.33 (0.53)	25.89 (0.15)	-16.94 (0.29)
α-ketoglutaric acid (D06)	162.54 (0.00)	26.25 (0.03)	6.34 (0.00)	10.70 (0.37)	-70.84 (0.00)	-141.54 (0.00)
L-glutamine (E01)	27.35 (0.02)	23.94 (0.05)	-5.08 (0.40)	89.43 (0.00)	14.84 (0.40)	10.27 (0.52)
α-hydroxybutyric acid (E07)	99.00 (0.00)	42.77 (0.00)	23.30 (0.01)	-2.47 (0.83)	50.90 (0.01)	-40.86 (0.01)
citric acid (F02)	137.26 (0.00)	7.05 (0.55)	-23.15 (0.00)	-89.92 (0.00)	-70.80 (0.00)	-62.01 (0.00)
D-threonine (F04)	23.65 (0.05)	-0.54 (0.96)	-4.47 (0.81)	-3.22 (0.78)	-4.22 (0.81)	-4.59 (0.77)
fumaric acid (F05)	123.03 (0.00)	16.98 (0.16)	-5.11 (0.68)	12.06 (0.31)	7.30 (0.68)	17.22 (0.28)
bromosuccinic acid (F06)	92.65 (0.00)	20.02 (0.10)	15.37 (0.04)	-8.99 (0.45)	38.21 (0.04)	-44.55 (0.01)
glycolic acid (F09)	45.14 (0.00)	4.13 (0.73)	2.86 (0.02)	19.40 (0.11)	44.73 (0.02)	-33.32 (0.04)
glyoxylic acid (F10)	35.55 (0.00)	-2.00 (0.87)	-13.65 (0.20)	21.79 (0.07)	22.44 (0.20)	-9.72 (0.54)
gly-glu (G01)	85.76 (0.00)	-5.64 (0.63)	-1.78 (0.00)	-52.42 (0.00)	-56.71 (0.00)	-24.23 (0.13)
L-serine (G03)	104.32 (0.00)	8.13 (0.49)	1.34 (0.44)	16.46 (0.17)	-13.38 (0.44)	-72.61 (0.00)
methylpyruvate (G10)	102.21 (0.00)	7.34 (0.53)	-28.96 (0.63)	17.23 (0.15)	-8.31 (0.63)	19.96 (0.21)
D-malic acid (G11)	138.52 (0.00)	27.79 (0.03)	8.68 (0.01)	-29.04 (0.02)	50.08 (0.01)	-152.16 (0.00)
L-malic acid (G12)	147.54 (0.00)	12.21 (0.30)	-5.53 (0.60)	-0.26 (0.98)	9.22 (0.60)	12.15 (0.44)

This table show the results of the lme model for the utilised wells in plate PM 1. The baseline represents ST-474 without *ykg*C insertion at 38°C. The subsequent columns are offsets of the baseline. The cells are coloured according to the significance level ($p<0.05$). Blue represent a significant positive relationship with the main effect on the substrate, brown represent a negative relationship.
[a] standard error, [b] p-value

The same linear mixed effects model (lme) was used to examine possible relationships between temperature, *ykg*C insertion and ST type and the tolerance to osmolytes for PM 9. The results are summarised in Table 5.5 which has the same underlying structure as the table for PM 1 (Table 5.4).
The results indicated that the *ykg*C insertion was negatively associated with three substrates (urea 6% (E11), urea 7% (E12) and sodium lactate 9% (F09)) and no positive associations. The higher temperature at 42°C was positively associated with the tolerance to the effects of 19 of the osmolytes and only the tolerance to one substrate (urea 6% (E11)) was negatively associated with the higher temperature compared to the baseline. Compared to the ST-474 isolates, the ST-42 isolates were significantly negatively associated with 19 substrates (sodium chloride (NaCl) at 1%-4% (A01-A04), potassium chloride at 3% (D01), sodium sulphate at 2% (D05), sodium lactate at 3%-12% (F03-F12), 20mM sodium benzoate at pH 5.2 (G05), 50mM and 100mM ammonium sulphate at pH 8 (G11 and G12) and 80mM and 100mM sodium nitrate (H05 and H06). Isolates belonging to ST-2026 were negatively associated with (compared to ST-474) nine substrates (NaCl at 1%-4% (A01-A04), potassium chloride at 3% (D01) and 4% (D02), sodium sulphate at 2% (D05), 50mN and 100mM ammonium sulphate at pH 8 (G11 and G12) and 100mM sodium nitrate(H06)) and positively associated with one substrate (100mM sodium nitrite, H07). Isolates of ST-61 were significantly negatively associated with two substrates (sodium nitrate 80mM and 100mM, H05 and H06). Overall it is apparent that the tolerance was positively associated with the higher temperature of 42°C and that the *ykg*C insertion had no positive association on the tolerance to osmolytes. Isolates belonging to ST-42, ST-61 and ST-2026 showed in general reduced tolerance to substrates in PM 9, compared to isolates from ST-474.

5.3.4 REEMtrees

The REEMtree in Figure 5.3 is based on the same PM 1 data used for the linear mixed effects model. The main effects are MLST, *ykg*C insertion and temperature and the isolate number is included as a random effect. Within each box, the top number represent the mean of the A-value of the isolates within the split, n equals the number of observations present within each split and the percentage represents the number of isolates present in the split.
The algorithm splits firstly on the explanatory variable (well) and divides it into substrates that are widely utilised among the isolates (A05, A07, A08, B09, B10, B12, C03, D01, D06, E07, F05, F06, G03, G10, G11, G12) and substrates that are less commonly utilised (D07, E01, E05, E06, F01, F02, F04, F09, F10, G01, G09, H01, H02, H05). No initial division on the presence or absence of the *ykg*C insertion, temperature or ST type is apparent. However, there are several interesting sub-splits present in the tree. One such split is in the left part of the tree (Figure 5.3). Following the left branch down, and taking the right branch at the intersection (split 5), the algorithm starts to split on the presence/absence of the *ykg*C insertion (split 47)

Table 5.5: Results of the linear mixed effects model for PM 9 at 38°C and 42°C

substrate	baseline (9.25)[a]	*ykg*C(+) (9.28)	temp 42°C (8.05)	MLST-42 (9.25)	MLST-2026 (14.40)	MLST-61 (13.27)
NaCl 1% (A01)	180.59	3.41	35.99	-65.01	-70.48	-2.05
	(0.00)	(0.72)	(0.00)	(0.00)	(0.00)	(0.88)
NaCl 2% (A02)	122.74	-5.31	42.94	-35.35	-41.28	-0.69
	(0.00)	(0.57)	(0.00)	(0.00)	(0.01)	(0.96)
NaCl 3% (A03)	106.87	-5.08	33.16	-39.05	-53.99	-18.42
	(0.00)	(0.59)	(0.00)	(0.00)	(0.00)	(0.18)
NaCl 4% (A04)	88.24	-4.20	13.54	-36.76	-37.40	-25.69
	(0.00)	(0.65)	(0.10)	(0.00)	(0.02)	(0.06)
NaCl 10% (A12)	64.85	-0.84	9.99	-6.26	-5.80	-1.70
	(0.00)	(0.93)	(0.23)	(0.50)	(0.69)	(0.90)
NaCl 6% + L-Carnitine (B12)	50.65	-2.36	20.84	-7.11	-3.03	-1.23
	(0.00)	(0.80)	(0.02)	(0.45)	(0.83)	(0.93)
NaCl 6% + Gluathione (C06)	76.49	-7.86	-9.76	-11.47	-23.42	-0.18
	(0.00)	(0.41)	(0.24)	(0.23)	(0.12)	(0.99)
potassium chloride 3% (D01)	120.55	-5.66	39.59	-22.69	-48.34	12.32
	(0.00)	(0.55)	(0.00)	(0.02)	(0.00)	(0.36)
potassium chloride 4% (D02)	92.27	-6.50	1.11	-14.22	-30.66	4.48
	(0.00)	(0.49)	(0.89)	(0.14)	(0.04)	(0.74)
sodium sulphate 2% (D05)	180.17	-1.54	24.50	-47.55	-61.09	13.55
	(0.00)	(0.87)	(0.01)	(0.00)	(0.00)	(0.32)
urea 5% (E10)	135.30	-16.23	-12.29	-16.79	-22.02	-6.96
	(0.00)	(0.09)	(0.14)	(0.08)	(0.14)	(0.60)
urea 6% (E11)	107.71	-21.74	-17.33	-9.13	-26.25	-6.55
	(0.00)	(0.03)	(0.04)	(0.33)	(0.08)	(0.63)
urea 7% (E12)	95.17	-19.69	-8.73	-5.61	-26.00	-2.27
	(0.00)	(0.04)	(0.29)	(0.55)	(0.08)	(0.87)
sodium lactate 3% (F03)	174.13	-3.14	-1.88	-23.24	-5.73	-20.70
	(0.00)	(0.74)	(0.82)	(0.02)	(0.69)	(0.13)
sodium lactate 6% (F06)	163.48	-12.72	20.02	-25.60	2.95	-26.82
	(0.00)	(0.18)	(0.02)	(0.01)	(0.84)	(0.05)
sodium lactate 7% (F07)	156.46	-14.74	24.83	-27.12	-2.29	-18.60
	(0.00)	(0.12)	(0.00)	(0.01)	(0.88)	(0.17)
sodium lactate 8% (F08)	159.42	-13.44	17.81	-22.77	-1.18	-8.02
	(0.00)	(0.16)	(0.04)	(0.02)	(0.94)	(0.55)
sodium lactate 9% (F09)	150.23	-21.77	18.83	-29.52	-1.84	-20.28
	(0.00)	(0.03)	(0.03)	(0.00)	(0.90)	(0.14)
sodium lactate 10% (F10)	133.68	-15.31	24.40	-34.46	-2.51	-21.57
	(0.00)	(0.11)	(0.01)	(0.00)	(0.86)	(0.12)
sodium lactate 11% (F11)	117.11	-12.62	24.59	-34.01	3.94	-24.63
	(0.00)	(0.19)	(0.01)	(0.00)	(0.79)	(0.08)
sodium lactate 12% (F12)	131.18	-16.09	39.96	-33.80	5.61	-19.25
	(0.00)	(0.10)	(0.00)	(0.00)	(0.70)	(0.16)
sodium benzoate pH 5.2 20mM (G05)	92.70	0.45	62.65	-32.27	-15.96	-1.13
	(0.00)	(0.96)	(0.00)	(0.00)	(0.28)	(0.93)
ammonium sulphate pH 8 50mM (G11)	164.56	-0.88	22.69	-33.53	-33.66	10.49
	(0.00)	(0.93)	(0.01)	(0.00)	(0.03)	(0.44)
ammonium sulphate pH 8 100mM (G12)	164.92	1.00	21.27	-33.28	-50.75	9.82
	(0.00)	(0.91)	(0.01)	(0.00)	(0.00)	(0.47)
sodium nitrate 80mM (H05)	164.72	8.36	24.90	-39.65	-18.18	-28.50
	(0.00)	(0.38)	(0.00)	(0.00)	(0.22)	(0.04)
sodium nitrate 100mM (H06)	156.28	5.58	34.76	-50.78	-39.35	-33.58
	(0.00)	(0.55)	(0.00)	(0.00)	(0.01)	(0.02)
sodium nitrite 10mM (H07)	94.54	-1.90	15.68	-9.82	31.80	27.84
	(0.00)	(0.84)	(0.06)	(0.30)	(0.04)	(0.05)
sodium nitrite 100mM (H12)	60.13	-0.58	25.14	2.28	2.81	-1.36
	(0.00)	(0.95)	(0.00)	(0.81)	(0.85)	(0.92)

This table show the results of the lme model for the tolerance of the *C. jejuni* isolates to osmolytes in plate PM 9. The baseline represents ST-474 without the *ykg*C insertion at 38°C. The subsequent columns are offsets of the baseline. The cells are coloured according to the significance level. Blue represent a positive effect of the main effect on the substrate, brown represent a negative effect.
[a] standard error, [b] p-value

and further down on the ability of ST-474 and ST-42 to utilise citric acid (F02) and L-glutamine (E01) (see enlargement in Figure 5.4). ST-474 isolates that are *ykg*C +ve are able to utilise citric acid but not L-glutamine, whereas the opposite is true for the isolates belonging to ST-42. It is important to keep in mind that each step is splitting the remaining data, minimising the number of observations in each split. Therefore the splits 380-383 only represent the isolates belonging to ST-42 and ST-474 that are *ykg*C positive and are able to utilise either citric acid or L-glutamine. These are only 1.4% of the overall observations.

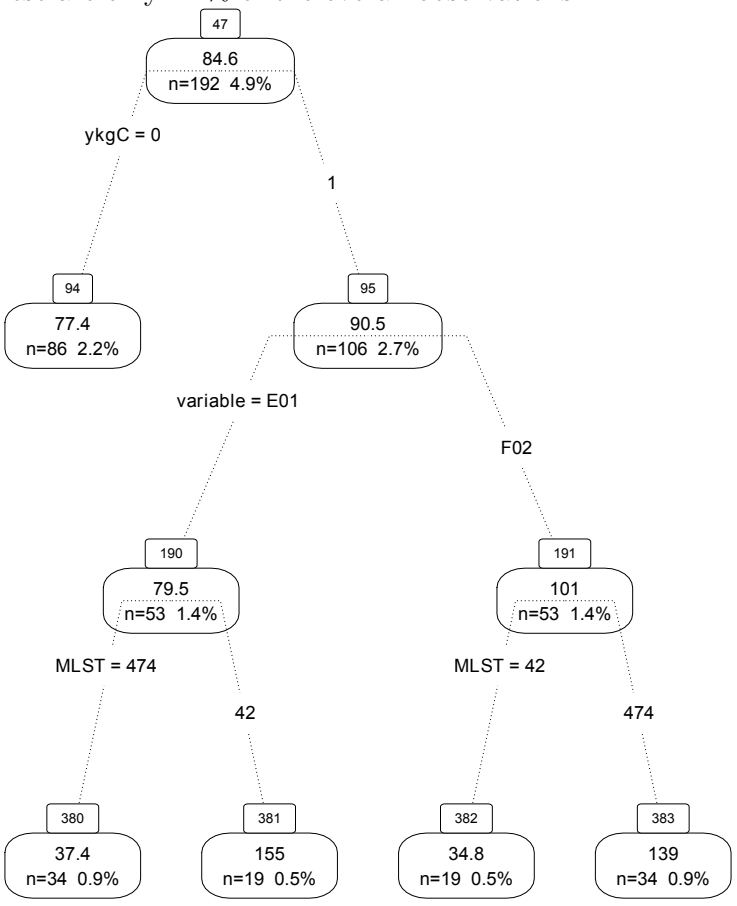


Figure 5.4: Enlargement of the partial REEMtree showing the split on the presence/absence of the *ykg*C insertion and the split based on the ability of ST-42 and ST-474 to utilise L-glutamine (E01) and citric acid (F02). This part of the tree considers the *ykg*C positive isolates belonging to ST-474 and ST-42 only.

5.3.5 Gene-trait matching

The following algorithm was employed to identify relevant genes and to assign their importance:

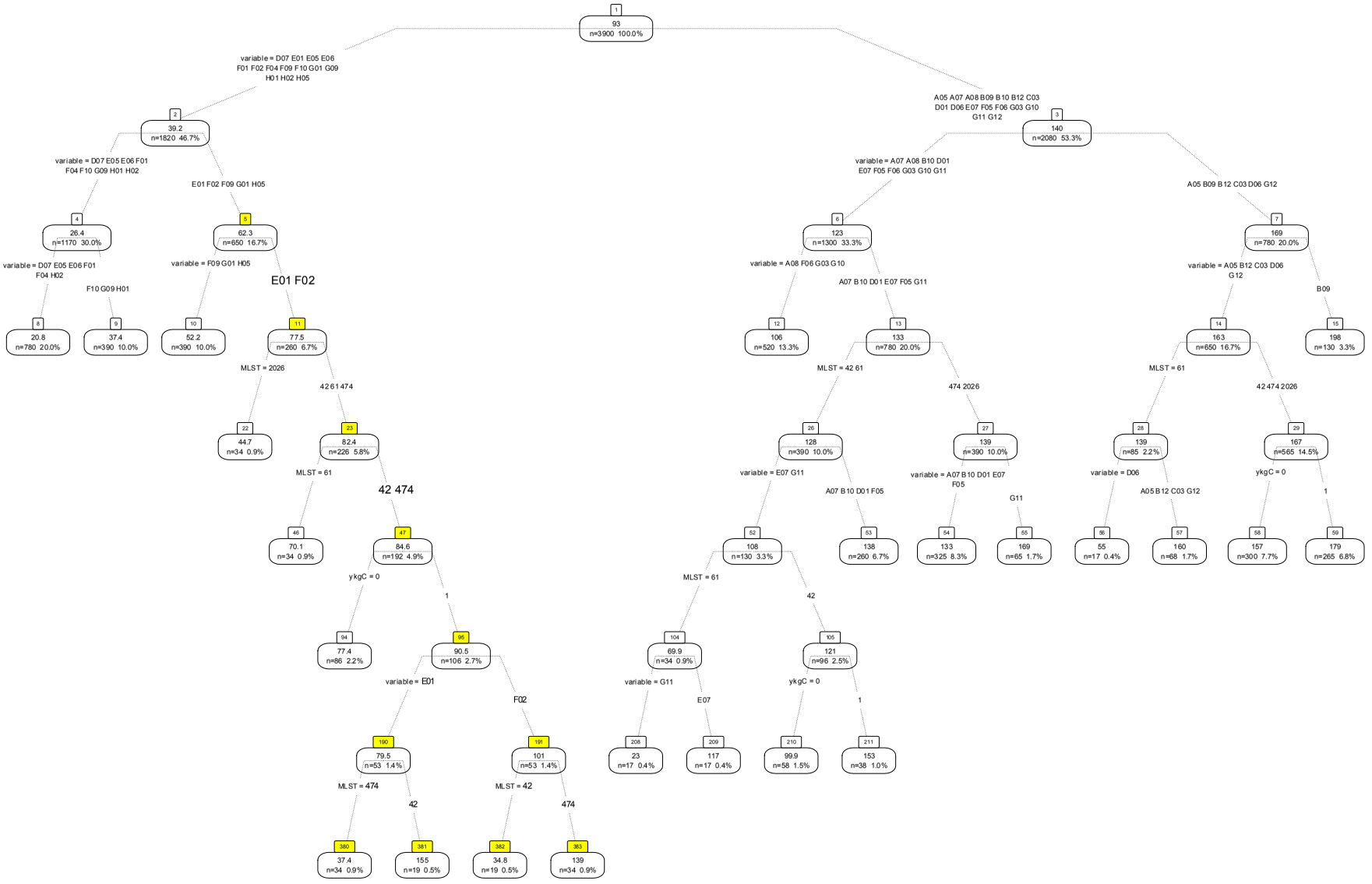


Figure 5.3: REEMtree estimated based on the A value of all repeats on the PM 1 plate. Within each box, the top number represents the mean within the split, n equals the number of observations present within each split and the percentage represents the number of isolates present in the split.

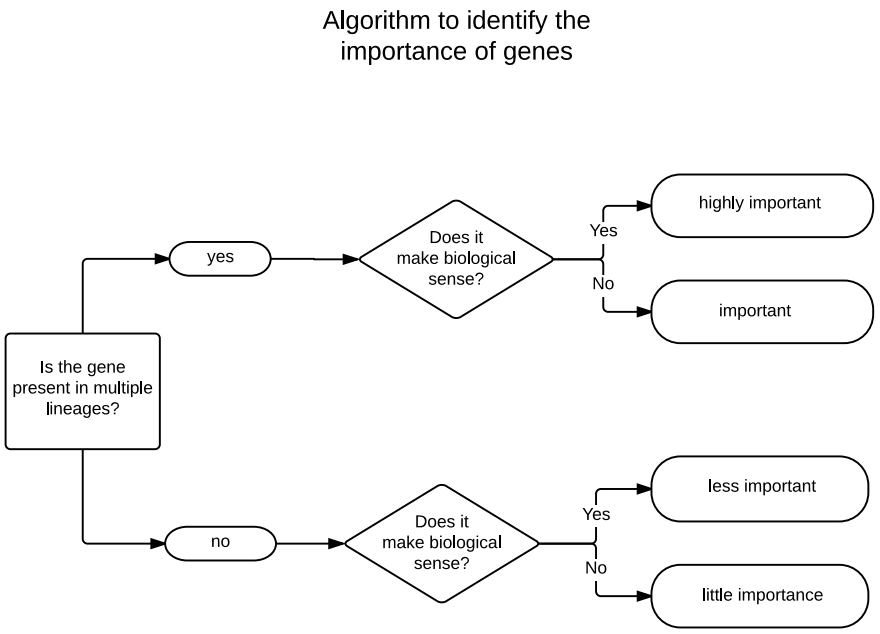


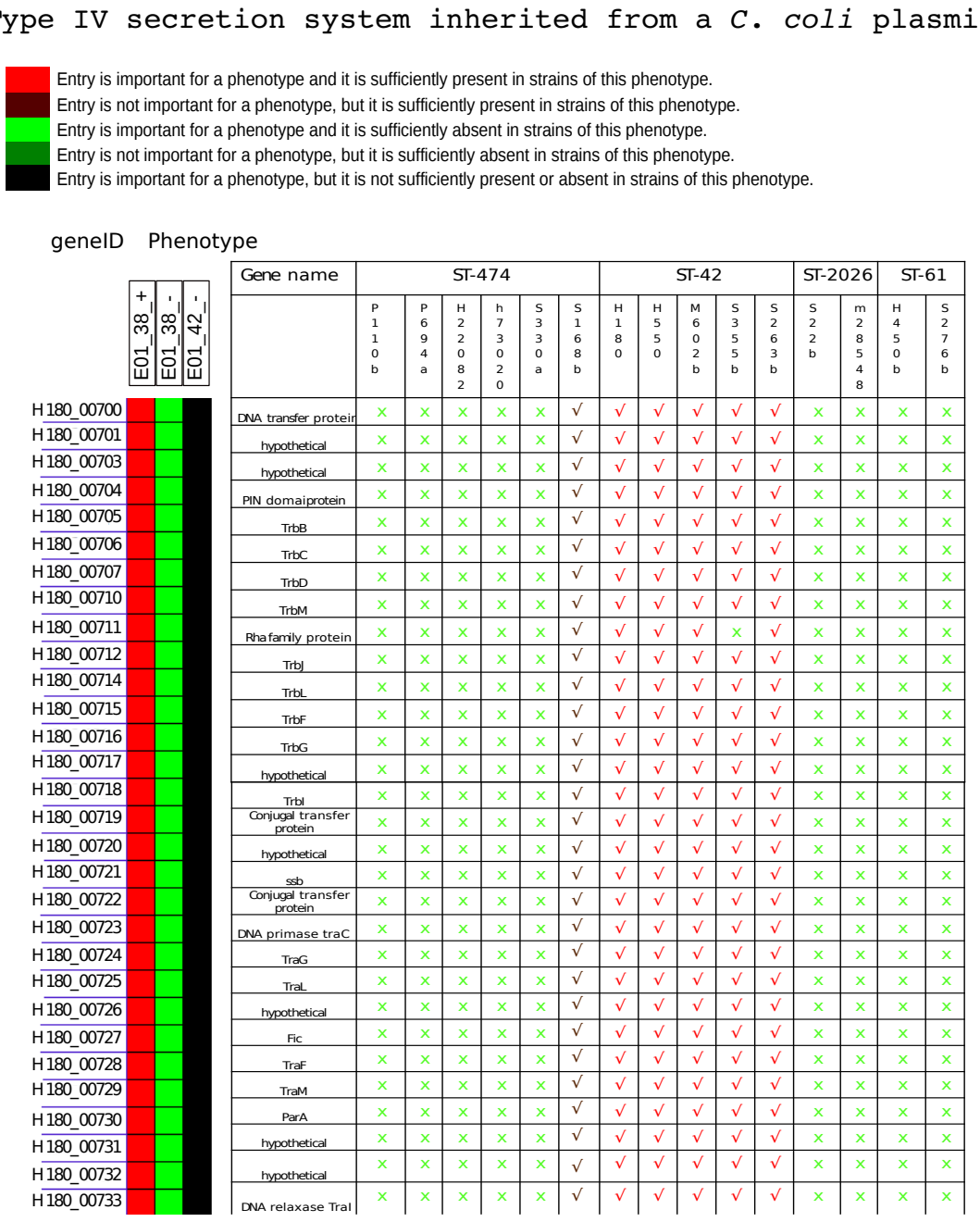
Figure 5.5: Algorithm to identify whether an identified gene or gene cassette is important for an observed phenotype.

Based on the results obtained in the first part of the study (utilisation of carbon sources and linear mixed effects model), relevant wells were investigated further. The linear mixed effects model, the REEMtree analysis and the phenotypic profile (heatmap) based on the OmniLog data suggest that, comparing ST-42 and ST-474, the well E01 (L-glutamine) was mainly utilised by the isolates belonging to ST-42 and the well F02 (citric acid) was mainly utilised by the isolates belonging to ST-474 (some isolates of ST-61 and ST-2026 constituted an exception). These two wells appeared to be the most interesting and were therefore investigated further with a genome wide association analysis.

The random forest algorithm implemented in PhenoLink (Bayjanov et al., 2012) was used to identify correlations between the genotypes (the presence or absence of genes for a given isolate) and phenotypes (the utilisation of specific carbon sources based on the results obtained from OmniLog). A gene whose presence or absence appears to be associated with a given phenotype was considered for further analysis. 'Present' means that the gene is present in at least p percent (75% in this study) of the strains for a given phenotype (Bayjanov et al., 2012). The same principle applies to the absence of a gene. Examples of some interesting gene cassettes identified by PhenoLink can be found in Figure 5.6 and Figure 5.9 for the phenotype 'utilisation of L-glutamine as a carbon source' as represented by E01.

Figure 5.6 and Figure 5.8 show a partial bacterial type IV secretion system which had not previously been identified in *C. jejuni*. The NCBI BLAST of this area revealed a partial hit (∼90%) to the *C. coli* plasmid pCC42yr (length: 26,269bp, sequence ID: gb|CP006703.1). PhenoLink identified its presence as being important for the isolates belonging to the ST-42 lineage but not for the isolate S168b which belongs to the ST-474 lineage. This region is present across a few lineages but not relevant to the observed phenotype (utilisation of L-glutamine) which is associated with isolates of ST-42. A BLASTN search on the pubmlst database (`http://pubmlst.org/campylobacter/`) identified the type IV secretion system in several isolates belonging to ST-42 and CC-828 (*C. coli*). However, it has previously not been identified in *C. jejuni* on GenBank or documented in the literature. 5.7 shows the mapped regions of the *C. coli* isolate CP006703 and the *C. jejuni* isolate H180 showing evidence that the whole region is one single gene cluster.

comment 32, Murray



Gene name	ST-474						ST-42					ST-2026		ST-61	
	P110b	P694a	H20082	h73020	S330a	S168b	H180	H550	M6022b	S3558b	S2638b	S22b	m28548	H4508b	S2766b
DNA transfer protein	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
hypothetical	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
hypothetical	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
PIN domaiprotein	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TrbB	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TrbC	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TrbD	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TrbM	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
Rha family protein	×	×	×	×	×	√	√	√	√	×	√	×	×	×	×
TrbJ	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TrbL	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TrbF	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TrbG	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
hypothetical	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TrbI	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
Conjugal transfer protein	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
hypothetical	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
ssb	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
Conjugal transfer protein	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
DNA primase traC	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TraG	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TraL	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
hypothetical	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
Fic	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TraF	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
TraM	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
ParA	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
hypothetical	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
hypothetical	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×
DNA relaxase TraI	×	×	×	×	×	√	√	√	√	√	√	×	×	×	×

Figure 5.6: Gene cassette showing all the genes that are part of the type IV secretion system. The colours in the table are linked to the importance of the gene as defined by PhenoLink. It shows the presence of the type IV secretion system in the isolates belonging to the ST-42 lineage and one ST-474 isolate. However, the secretion system is only deemed to be important for the ST-42 isolates but not for the ST-474 isolate.

Another relevant region identified by PhenoLink is shown in Figure 5.9 and Figure 5.10. This gene cassette is comprised of a cytochrome C gene and a gamma-glutamyltransferase (*ggt*) gene, which may potentially be involved in the utilisation of L-glutamine. PhenoLink identified this cassette as being important for the observed phenotype. The cassette is only present in

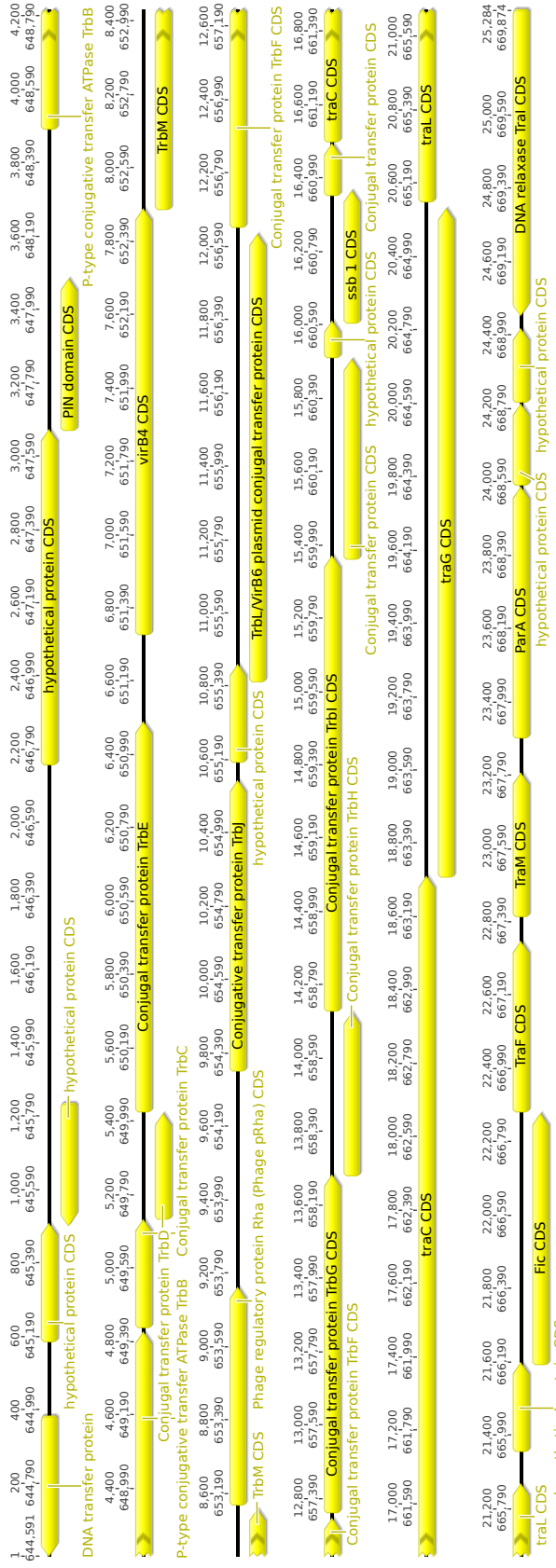


Figure 5.7: This figure shows the mapped 26kB region of type IV secretion system of the *C. jejuni* isolate H180 and the *C. coli* isolate CP006703 (extracted with Geneious version 7.1.7).

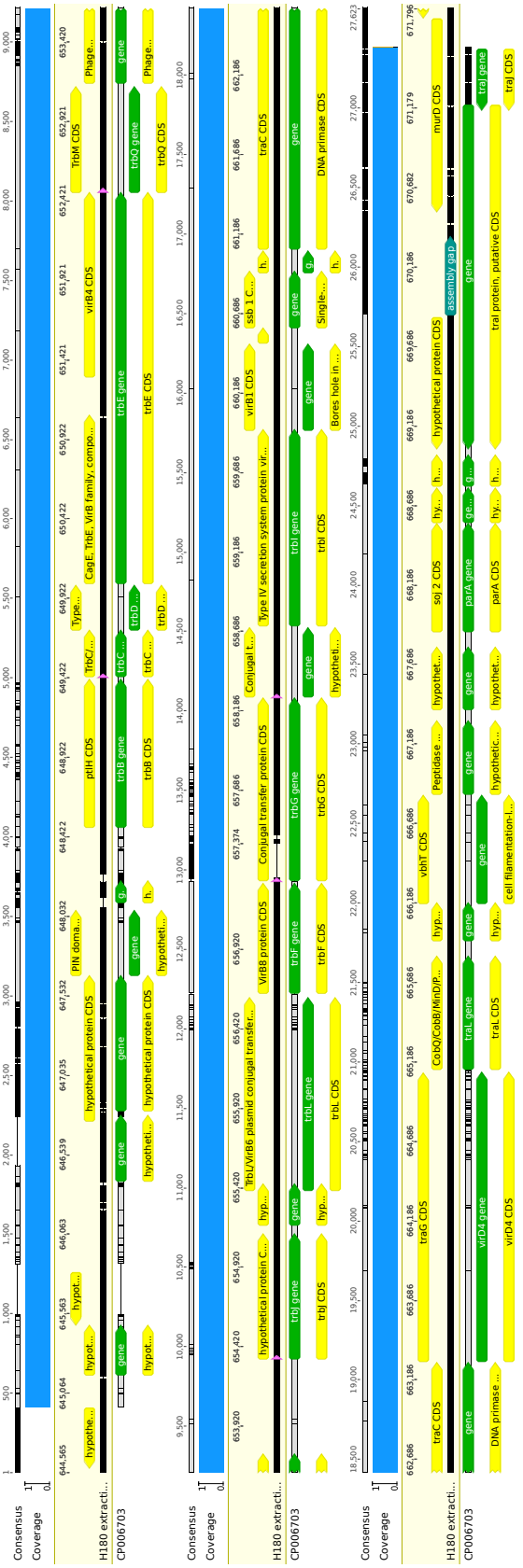
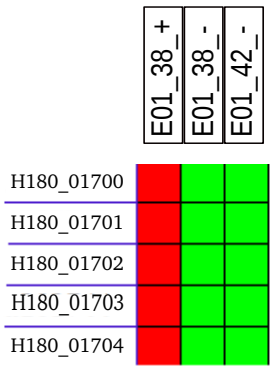


Figure 5.8: This figure shows the extracted 26kB region of type IV secretion system of the *C. jejuni* isolate H180 (extracted with Geneious version 7.1.7).

the ST-42 lineage, the lineage that predominately displays the L-glutamine utilisation phenotype.



Gene name	ST-474						ST-42					ST-2026		ST-61	
	P110b	P694a	H220822	h730220	S330a	S168b	H180	H550	M602b	S355b	S263b	S22b	m28548	H450b	S276b
Cytochrome C	x	x	x	x	x	x	√	√	√	√	√	x	x	x	x
Membrane protein, putative	x	x	x	x	x	x	√	√	√	√	√	x	x	x	x
hypothetical	x	x	x	x	x	x	√	√	√	√	√	x	x	x	x
Cytochrome C biogenesis protein	x	x	x	x	x	x	√	√	√	√	√	x	x	x	x
ggt	x	x	x	x	x	x	√	√	√	√	√	x	x	x	x

Figure 5.9: Gene cassette including the genes gamma-glutamyltransferase (*ggt*) gene and cytochrome C gene.

The same approach as for L-glutamine was applied to citric acid. An example of a gene cassette identified by PhenoLink for citric acid can be found in Figure 5.11 and Figure 5.12. The L-fucose genomic island has no obvious biochemical correlation to the utilisation of citric acid but has been shown to be advantageous for the colonisation of chickens (Muraoka and Zhang, 2011; Stahl et al., 2011). This island is only present in the ST-474 lineages, interestingly it appears that its absence in the other three lineages is of differing importance to the phenotype at the tested temperatures. Figure 5.11 indicates that for some of the non-ST-474 isolates (H550, S263b, m28548 and H450b) the absence of the genomic island is important at 38°C but not at 42°C, whereas for the other non-ST-474 isolates it is either consistently important (H180, M602b, S355b) or not important (S22b, S276b) across the two temperatures. A mapping of the ST-474 genomic island to NCTC 11168 revealed SNPs in the ST-474 *fuc*P gene.

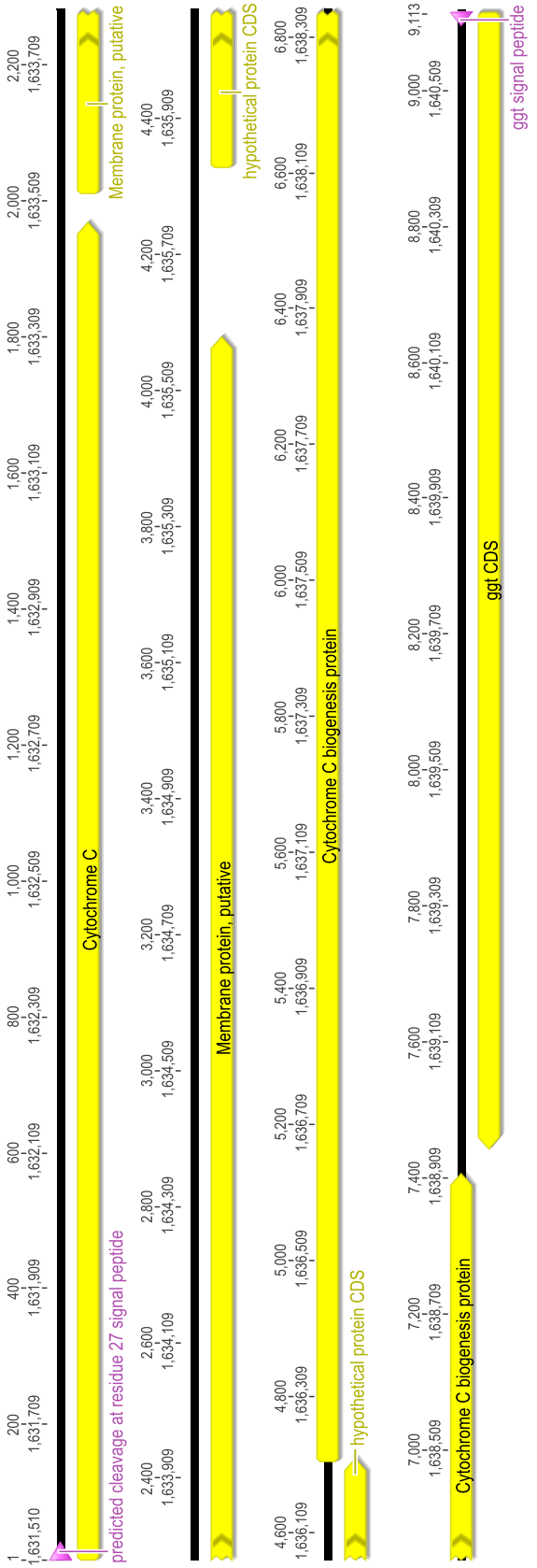


Figure 5.10: Geneious sequence extraction for the gamma-glutamyltransferase cassette.

	Gene name	ST-474						ST-42					ST-2026		ST-61	
		P110b	P694a	H2208	h7302	S330a	S168b	H180	H550	M602b	S355b	S263b	S22b	m28554	H450b	S276b
H22082_00831	ArsR	√	√	√	√	√	√	x	x / x	x	x	x / x	x	x / x	x / x	x
H22082_00830	Dihydrodipicolinate synthase	√	√	√	√	√	√	x	x / x	x	x	x / x	x	x / x	x / x	x
H22082_00829	Altronate hydrolase	√	√	√	√	√	√	x	x / x	x	x	x / x	x	x / x	x / x	x
H22082_00828	Putative altronate hydrolase	√	√	√	√	√	√	x	x / x	x	x	x / x	x	x / x	x / x	x
H22082_00827	Tartrate transporter	√	√	√	√	√	√	x	x / x	x	x	x / x	x	x / x	x / x	x
H22082_00826	3-oxoacyl-[acyl-carrier protein] reductase	√	√	√	√	√	√	x	x / x	x	x	x / x	x	x / x	x / x	x
H22082_00825	L-fucose permease	√	√	√	√	√	√	x	x / x	x	x	x / x	x	x / x	x / x	x

Figure 5.11: Gene cassette showing all the genes that are part of the L-fucose genomic island (*cj0480c-cj0490*) which is present across the ST-474 lineage but not the other examined *C. jejuni* lineages. For most isolates the island appears to be important at 38°C (except for the isolates S22b and S276b). However, it is not important for the isolates H550, S263b, m28548 and H450b at 42°C (indicated by the split cells).

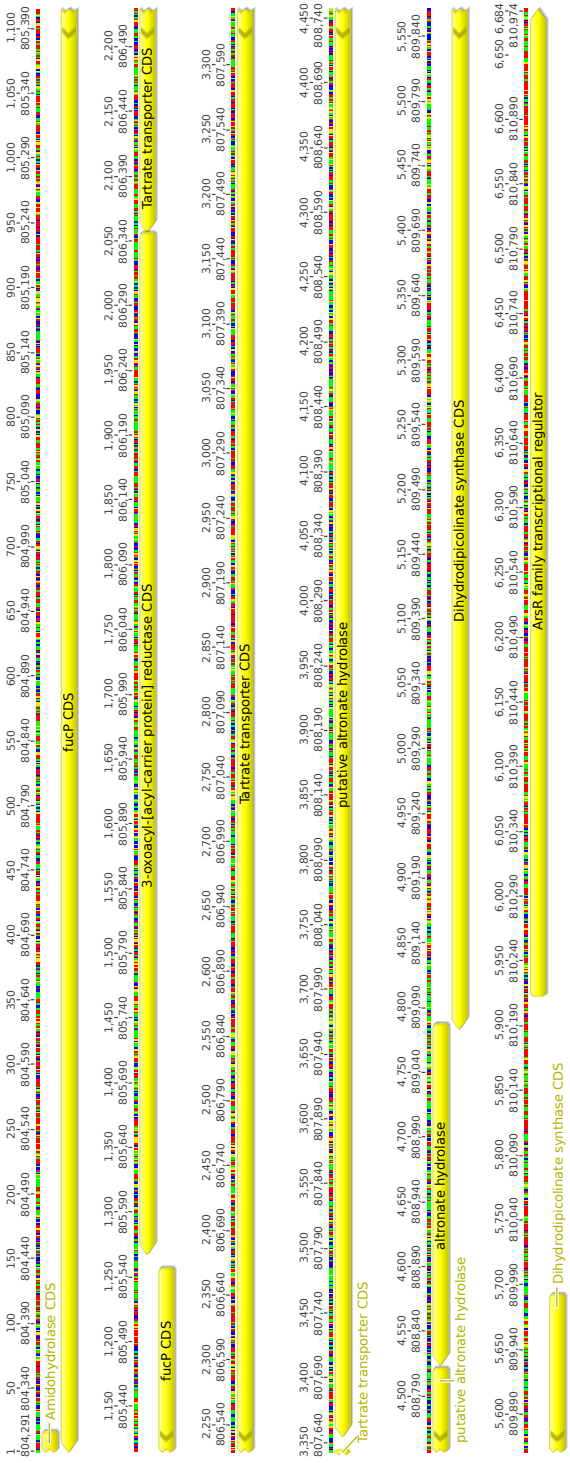


Figure 5.12: Gene cassette showing all the genes that are part of the L-fucose genomic island (*cj0480c-cj0490*) which is present across the ST-474 lineage but not the other examined *C. jejuni* lineages. For most isolates the island appears to be important at 38°C (except for the isolates S22b and S276b). However, it is not important for the isolates H550, S263b, m28548 and H450b at 42°C (indicated by the split cells).

5.4 Discussion

The present study gives a detailed view of the correlations between specific genotypes and their phenotypes related to utilisation of carbon sources (PM 1) and tolerance to osmolytes (PM 9) through the use of the phenotypic microarray system. Here we examined 15 isolates based on eight genes (seven housekeeping genes (MLST) and the *ykg*C gene) from four sequence types, hosts and temperatures (Table 5.1). These were related to their phenotypic expression to examine the relationships between phenotype and multilocus genotype (seven core genes and one accessory gene). We then using a GWAS approach aimed to identify genes that may be related to the observed phenotypes.

Other studies examined the phenotypic profiles of NCTC 11168 (GS) (Line et al., 2010) related to carbon sources (PM 1 and PM 2) and ATCC 33560 (Tang et al., 2010) related to carbon sources and tolerance to osmolarity. However, these studies only examined one isolate and did not identify the genes that may be associated with the observed phenotypic profile.

The heatmap in Figure 5.1 displays a strong clustering by ST corresponding to the phenotypic utilisation profile of carbon sources at 38°C. Isolates belonging to one lineage show a different utilisation pattern for some carbon sources as compared to isolates from other lineages (for example well E01, L-glutamine was only utilised by the lineage ST-42, whereas well F02, citric acid, was only utilised by the isolates of ST-474). A clustering by lineages may not seem surprising, however, no apparent clustering was observed on the same plate at 42°C or on the PM 9 plate at 38°C or 42°C (the plots can be found in the supplementary material Figure A.61, Figure A.62 and Figure A.63 respectively). The phenotypic utilisation profile differs for some isolates at 42°C (e.g. the isolates belonging to ST-42 lose the ability to utilise L-glutamine) which may explain the absence of clustering by ST. The different temperatures symbolise different core body temperatures of avian (42°C) and mammalian (38°C) species and therefore, the differences in the phenotypic profiles may be related to the different colonisation and/or virulence factors. ST-42 is a ruminant associated strain and the loss of its ability to utilise L-glutamine at 42°C is potentially associated with the differences between poultry and ruminant core body temperatures.

The linear mixed effects model (Table 5.4) revealed that the presence of the *ykg*C insertion was significantly correlated with an increased respiration of the sole carbon sources succinic acid, L-glutamic acid, D,L-malic acid, α-ketoglutaric acid, α-hydroxybutyric acid and D-malic acid. However, the significant increase of respiration in these substrates (25-42 Omnilog units), does not change the phenotypic expression (e.g. a given isolate would utilise the carbon source without the *ykg*C insertion). As the *ykg*C insertion consists of two genes (a hypothetical protein and the *ykg*C), it may also be possible that the increase is due to the protein with unknown function, stem-loop structures or methylation. Another significant finding of the model, which is supported by the heatmap (Figure 5.1), is that L-glutamine is mainly utilised by the ST-42 lineage and citric acid is mainly utilised by the ST-474 lineage.

The model results for the PM 9 plate summarised in Table 5.5 indicate that the ST-474 lineage showed in general increased tolerance to multiple osmolytes compared to the other ST types. Overall all STs showed increased tolerance at 42°C as compared to 38°C. The greatest increase in tolerance, for the ST-474 isolates at 42°C, was seen for the osmolyte sodium benzoate (pH 5.2 20mM). Sodium benzoate is, apart from being a food preservative, also commonly found in nature in fruits, grains and cereals (Chipley, 2010). The ST-474 lineage is mainly chicken-associated and 42°C resembles the core body temperature of avian species. Assuming that chickens pick up sodium benzoate from the environment through grains, the increased tolerance at the higher temperature may provide the chicken-associated isolates with an advantage compared to ruminant-associated *C. jejuni* isolates.

Having identified relationships using conventional regression modelling the data were examined using regression trees to further examine the complex interactions between the response variable and the covariates. The REEMtree algorithm examined the data and split on the greatest difference, minimising the variation of the data within each split. The linear mixed effects model indicated that the temperature and *ykg*C insertion were only associated with some of the substrates. Therefore it is not surprising that the first split divides the wells that are utilised by all isolates from the ones that are only utilised by a subset (Figure 5.3). However, further down the tree, the algorithm starts to split on STs, indicating a specific phenotypic difference between the lineages. The three methods (heatmap, linear mixed effects model and REEMtree) associated the ST-42 lineage with the utilisation of L-glutamine and the ST-474 lineage with the utilisation of citric acid, strengthening the finding and its importance.

Line et al. (2010) used the phenotypic microarray system to evaluate the carbon source utilisation of NCTC 11168(GS) and determined that, although genomic analysis suggests that the mechanisms for catabolism of L-glutamine and proline are present, they did not observe utilisation in the PM 1 plate. However, in the present study L-glutamine was utilised by the ST-42 lineage and L-proline by all examined isolates. Although NCTC 11168 is a ruminant associated strain (ST-53, CC-21), NCTC 11168(GS) is a variant of the wild-type NCTC 11168 that could have adapted to the laboratory and lost its ability to utilise these carbon sources. Additionally, the gene *ggt* is absent in wild-type NCTC 11168 and therefore it is unlikely to utilise L-glutamine (de Haan et al., 2012).

Having identified associations between MLST type and two carbon sources, this study used a GWAS approach to potentially identify biologically plausible associations between the phenotypic expression and the accessory genome. For the first carbon source of interest (L-glutamine, E01) two interesting regions in the accessory genome were identified by PhenoLink. One region, involving a gene cassette, is comprised of a cytochrome C gene and a γ-glutamyltransferase (also called γ-glutamyltranspeptidase *ggt*) gene, which has been shown to be involved in the utilisation of L-glutamine (Hofreuter et al., 2008). Gamma-glutamyltranspeptidases are enzymes catalysing the hydrolysis of γ-glutamyl bonds in glutamine and glutathione. They also transfer

the released γ-glutamyl group to amino acids or short peptides (Castellano and Merlino, 2012). In *Helicobacter pylori*, a close relative to *Campylobacter*, the gamma-glutamyltranspeptidase enables bacterial cells to use extracellular glutamine and glutathione as a source of glutamate (Schmees et al., 2007; Shibayama et al., 2007). Previous research has shown an association of the *ggt* with human and poultry *C. jejuni* isolates (de Haan et al., 2012; Zautner et al., 2011) and it has also been shown to play a role in the persistent colonisation of the avian gut (Barnes et al., 2007) by *C. jejuni*. Despite the fact that in New Zealand, ST-42 is a mainly ruminant-associated lineage (Muellner, 2009), the presence of the *ggt* may under certain conditions, or in certain environments offer competitive advantage to the isolates (Hofreuter et al., 2006, 2008; Muraoka and Zhang, 2011; Stahl et al., 2011). It is still unclear whether the *ggt* may provide competitive advantage in bovine gut colonisation, virulence in humans or stress response in environmental isolates.

The other interesting region identified by PhenoLink is a type IV secretion system potentially inserted by the *C. coli* plasmid pCC42yr. The plasmid is part of the *C. coli* genome 15-537360 (GenBank: CP006702) which was identified as ST-855 (CC-828). The CC-828 belongs to *C. coli* clade 1 which has been shown to frequently recombine with *C. jejuni* (Sheppard et al., 2008). Type IV secretion systems in bacteria are related to genetic exchange (natural transformation) and the delivery of DNA and proteins to eukaryotic target cells during bacterial infections (Cascales and Christie, 2003). *C. jejuni* is one of the more than 40 naturally competent bacterial species able to import DNA from the environment and to successfully incorporate it into their genomes (Wiesner et al., 2003). However, a study examining the type IV secretion system of the pVir plasmid in *C. jejuni* 81-176 (Bacon et al., 2000) knocked out two of the pVir-encoded genes, *vir*B11 and *com*B3, and showed that while the *com*B3 mutant had a transformation efficiency approximately five times lower than the wild-type, the *vir*B11 mutant showed the same efficiency as the wild-type (Wiesner et al., 2003). As Wiesner et al. (2003) did not supply evidence in showing the natural transformation of *C. jejuni*, it is possible that the type IV secretion system plays a different role in *C. jejuni*, however, the knock-out was only performed on a limited number of genes and natural DNA uptake may be different to in vitro DNA uptake. CRISPRs (clustered regularly interspaced short-palindromic-repeat) are a bacterial defence mechanism against foreign genetic elements such as plasmids and phages, however, it has been shown that ST-42 is CRISPR negative (Price et al., 2007; Schouls et al., 2003) which may explain the presence of the type IV secretion system. Even though the type IV secretion system is only present in the ST-42 and one ST-474 isolate, it is unlikely to be involved in the utilisation of L-glutamine. Interestingly, even though ST-42 is a ruminant associated lineage, the presence of the gene cassette involving the *ggt* and the type IV secretion system (natural transformation) may suggest that it is more capable than other lineages to incorporate foreign DNA and to adapt to different hosts.

No specific genes or gene cassettes (with a function which would make biological sense) were

identified to be associated with the utilisation of citric acid. Biological sense refers to the linkage between the phenotype and the biological function of the observed gene (e.g. gamma-glutamyltransferase and the utilisation of L-glutamine). This may either be due to the fact that PhenoLink removed more than 50% of the genes before the analysis (removal of homogeneous features) or that there are more than one gene (complex gene cassettes) involved in the utilisation of citric acid not being correlated with the specific phenotype. *C. jejuni* encodes all the enzymes required for a complete oxidative tricarboxylic acid cycle (TCA) (Line et al., 2010), and all the isolates examined in our study had the genes for the TCA, therefore it is surprising that only ST-474 was able to utilise citric acid.

comment 31, Murray

Even though PhenoLink did not identify genes biologically associated with the utilisation of citric acid, it identified another interesting region, namely the L-fucose genomic island (Figure 5.11) which, across the examined isolates, was only present in the ST-474 lineage. Evidence for the L-fucose pathway in certain strains of *C. jejuni* was provided genotypically and phenotypically (Muraoka and Zhang, 2011; Stahl et al., 2011) and has also been shown to be advantageous for the colonisation of birds (Muraoka and Zhang, 2011). *C. jejuni* obtains L-fucose from the host during colonisation of the intestine (Stahl et al., 2011). ST-474 is an internationally rare, but in New Zealand, strongly chicken-associated lineage (Muellner et al., 2013). The presence of the L-fucose pathway may provide a significant advantage for chicken colonisation and explain the strong host-association of the ST-474 lineage. However, although the genomic island was present, and *C. jejuni* has been shown to utilise L-fucose (Muraoka and Zhang, 2011; Stahl et al., 2011) it did not support the active respiration in this experiment (PM 1, well B04). A more detailed analysis of the *fuc*P region identified SNPs in the *fuc*P gene (both synonymous and nonsynonymous) in each of the isolates. *Fuc*P has been determined to be essential for the utilisation of L-fucose (Muraoka and Zhang, 2011). Both Line et al. (2010) and Muraoka and Zhang (2011) used a variant of NCTC 11168 and the phenotypic microarray (PM) system, but whereas Muraoka and Zhang (2011) used the wild-type of NCTC 11168 and reported the utilisation of L-fucose, Line et al. (2010) used a variant (NCTC 11168 (GS)) and did not report utilisation. It is possible that the variant (NCTC 11168 (GS)) adapted to laboratory conditions and therefore lost the ability to utilise L-fucose. It has yet to be determined if this could have happened to the isolates used in this study.

Previous studies examining the *C. jejuni* genomes NCTC 11168 (GS) (ST-53, CC-21) (Line et al., 2010) and ATCC 33560 (ST-403, CC-403) (Tang et al., 2010) showed that *C. jejuni* had a limited amino acid utilisation and a preference for specific amino acids. However, the present study shows that the examined isolates utilised a wider range of carbon sources and amino acids. In addition to the commonly utilised carbon sources aspartate, asparagine, glutamate and serine (Line et al., 2010), all of the examined lineages were able to utilise L-proline and ST-42 was able to utilise L-glutamine. Line et al. (2010) also reported that respiration was generally greater at 42°C as compared to 38°C, however the results of this study indicate a decrease of respiration

at 42°C (Table 5.2) for 45 instances compared to higher respiration at 42°C for 16 instances. One possible explanation could be that most of the isolates used in this study were ruminant-associated and would therefore be adapted to a mammalian core body temperature of 38°C.
Previous experiments reported that *C. jejuni* is sensitive to 2% and greater concentrations of NaCl (Cameron et al., 2012; Tang et al., 2010) but showed tolerance to a wide range of food preservatives (sodium lactate, sodium phosphate, sodium benzoate, ammonium sulphate and sodium nitrate) (Tang et al., 2010). We observed respiration of *C. jejuni* in concentrations of NaCl of up to 4% (Table 5.3), mainly associated with the ST-474 lineage and 42°C. However, a confirmatory experiment on salt plates was not in agreement with the OmniLog results. This could have two explanations: the first one is that the confirmatory test was not the appropriate experimental analogy, it could be repeated with broth to give a closer replication of the OmniLog milieu. The second explanation could be that, while cells must respire to grow, utilisation can occur independent of cell growth (Bochner et al., 2001). Therefore, the observed tolerance could be evidence of the respiration in NaCl by the bacteria but not actual cell growth on the substrate (Line et al., 2010). Several studies have confirmed the tolerance of *C. jejuni* to higher concentrations of sodium chloride (Doyle and Roman, 1982; On, 1996; Tang et al., 2010) and therefore the second explanation is more plausible. Care should also be taken with the other results obtained solely from the OmniLog analysis and always be validated experimentally.

comment 34, Murray

When interpreting the results from the PhenoLink random forest analysis, care should be taken to consider the effect of confounding by sequence type. For example, if the type IV secretion system which is present in one isolate of ST-474 and all isolates of ST-42 had expressed the same phenotype across both lineages it would have provided stronger evidence of a true association between the phenotype and the presence of the operon. However as the insertion is only deemed important in the ST-42 lineage and not in the ST-474 lineage the association is more likely to be the result of confounding.
As far as the authors are aware, the present study is the first to test the ability of a wide variety of *C. jejuni* isolates from different hosts and STs to respire in 95 substrates as sole carbon sources and to test the tolerance of the isolates to different osmotic conditions at 38°C and 42°C using phenotypic microarray (PM) technology. It is also the first study to correlate the phenotypic expressions of the examined isolates with their whole genome sequences to identify genes that may be related to the observed phenotypes. We have identified a type IV secretion system, previously not identified in *C. jejuni*, which could be involved in the natural transformation of the bacteria. The wider range of potential carbon sources and amino acids utilised by *C. jejuni*, and its tolerance to osmolytes identified in the present study, could lead to improvements in culture media for detection and isolation of the bacteria. The different phenotypic profiles of the ST types could also provide new insights into factors associated with colonisation and niche-association.

CHAPTER 6

Short-term evolutionary dynamics of *Campylobacter jejuni* in chickens

Abstract

Chickens are an important source of human infection with *Campylobacter jejuni* via contaminated food and, although many studies have examined the epidemiology and evolution of this bacterium, few have considered the short-term evolutionary dynamics following inoculation with fully characterised strains. The present study was nested within a larger randomised trial of the effect of an in-feed additive on colonisation. In total 176 chickens were challenged with two *C. jejuni* genotypes and the present study used multilocus and full genome sequencing to examine the evolution of the microbial population over 43 days. The birds were inoculated with an equal mixture of an isolate from each of two sequence types, ST-474 and ST-45. One hundred-sixty-eight isolates collected at the different kill days were multilocus sequence typed, but only isolates with the sequence type ST-474 were recovered.

Twenty-seven of the 168 isolates cultured from the caeca of 12 birds and the two inoculum strains were whole genome sequenced and analysed. No evidence of recombination between the ST-45 and ST-474 strains was observed and therefore all subsequent comparisons were between the isolated ST-474 strains and the ST-474 inoculum strain. We observed 15 core single nucleotide polymorphisms (SNPs) and three non-core SNPs across the ST-474 isolates (compared to the inoculum ST-474 isolate). Fourteen of the core SNPs were non-synonymous point mutations confined to nine genes which were all associated with cell shape, chemotaxis or motility of the bacteria. We identified six independent SNPs in a single gene, *mre*B, encoding a homologue of actin, two of which have also been identified in Oxford isolates on PubMLST potentially suggesting parallel evolution. We identified alterations in the motility of the bacteria which could be associated with the detected SNPs. No recombination events were observed.

6.1 Introduction

Campylobacter spp. are the most common cause of acute gastroenteritis in the developed world (Adak et al., 2005; Coker et al., 2002; Friedman et al., 2000; Samuel et al., 2004) and have been isolated from a wide variety of domestic livestock, including poultry (Eberhart-Phillips et al., 1997; Müllner et al., 2009, 2010b), ruminants (Stanley and Jones, 2003) and pigs (Brown et al., 2004; Nesbakken et al., 2003). The majority of human campylobacteriosis incidences are caused by *C. jejuni* (90%) and *C. coli* (10%) where most of the *C. jejuni* infections are believed to be caused through the consumption of contaminated food (Carter et al., 2009; Eberhart-Phillips et al., 1997; Gillespie et al., 2002; Lee and Newell, 2006; Nachamkin et al., 1992).

Campylobacter spp. display extensive genetic variation, which has arisen from intragenomic mechanisms (such as point mutations, duplications, inversions and translocations) as well as genetic exchange (horizontal gene transfer, HGT) among strains. HGT, primarily involving the incorporation of genetic material from closely related lineages, is well known in *Campylobacter* spp., both within (Dingle et al., 2001; Suerbaum et al., 2001) and between microbiological species (Sheppard et al., 2008, 2011b, 2013a). Wilson et al. (2009) estimated the rate of cross-species gene flow to be less frequent between *C. jejuni* and *C. coli* than within-species gene flow. This suggests that recombination is not only responsible for molecular changes in *C. jejuni* but also enables cross-species gene flow, which can potentially lead to long-term adaptation (Wilson et al., 2009). Wilson et al. (2009) suggested that *C. jejuni* is subject to purifying selection that eliminates approximately 60% of novel variation.

Previous studies using data from multilocus sequence typing (MLST) estimated the range of the relative contributions of recombination and point mutation events to genetic diversity to be approximately 1:1 (Fearnhead et al., 2005; Feil et al., 1999, 2000). Yu et al. (2012), however, estimated recombination to contribute to the genetic diversity between 2.97 and 8.91 times more than mutation events (depending on the MLST locus examined). Other studies attempted to estimate the relative rate of recombination to mutation based on the whole genome level (Biggs et al., 2011; Falush et al., 2001) and identified a ratio of approximately 1:1 per event.

One study examining the short-term evolutionary dynamics of *Campylobacter* spp. used whole-genome sequencing (WGS) to assemble short-reads from 379 *Campylobacter* clinical isolates which were collected during a period of four month of disease surveillance in Oxford (Cody et al., 2013). Cody et al. (2013) identified a great diversity amongst the *Campylobacter* isolates recovered in Oxfordshire and that the small subsets of closely related isolates originated mainly from repeated sampling of the same patients. Despite the close relatedness of the within-patient isolates, Cody et al. (2013) identified genetic variations which were largely restricted to phase variable genes.

Previous studies involving passage of a bacterial strain through a host attempted to identify *in vivo* genetic variation (de Boer et al., 2002; Espedido et al., 2013; Hänel et al., 2009; Koskiniemi

et al., 2013; Thomas et al., 2014; Woodall et al., 2005), but only a few identified these variations at the nucleotide level (Espedido et al., 2013; Koskiniemi et al., 2013; Thomas et al., 2014) as opposed to comparing PFGE banding patterns (de Boer et al., 2002; Hänel et al., 2009; Woodall et al., 2005).

C. jejuni is a commensal bacteria in chickens and infected birds carry a high load of the bacteria (10^4-10^7 cells per g) in the gastrointestinal tract (GI), especially the caeca (Beery et al., 1988). This makes the chicken an ideal model to study the colonisation ability of different *C. jejuni* isolates and their resulting phenotype after successful passage through the chicken. They are also a prime vehicle for *Campylobacter* dissemination to humans. Several studies examining the colonisation potential of different *Campylobacter* spp. strains concluded that, based on the isolation source (human, bovine or chickens), the strains have a different potential to colonise the GI tract ranging from immediately sustained colonisation to non-colonising strains (Hänel et al., 2009; Korolik et al., 1998; Ringoir and Korolik, 2003). Chaloner et al. (2014) examined dual infections of chickens with a ST-21 and ST-137 isolate and determined different colonisation strategies of the isolates. Whereas the ST-137 isolate showed classical infection biology (rapid colonisation and shedding largely associated with the caeca), ST-21 was slower in colonisation but more able to colonise the upper GI tract.

After ingestion, the bacteria reach the caecum and start to multiply, establishing colonisation within 24 h (Coward et al., 2008). Models of colonisation in orally challenged chickens enabled the identification of some colonisation mechanisms including multi-drug and bile resistance (Lin et al., 2002, 2003), chemotaxis (Kanungpean et al., 2011; Stephens et al., 2006; Vegge et al., 2009) and motility (Champion et al., 2005; Howard et al., 2009).

Ring et al. (2005) showed that many chicken flocks are dominated by a single *C. jejuni* genotype, however, other studies have also shown that two or more genotypes can be isolated from the same flock (Berndtson et al., 1996; Hiett et al., 2002; Jacobs-Reitsma et al., 1995; Thomas et al., 1997). Generally, when more than one genotype had colonised a flock, the individual genotypes were found to coexist over time rather than exclude each other (Höök et al., 2005).

A randomised controlled trial conducted for the New Zealand Food Safety Authority in 2010 examined the effect of in-feed caprylic acid on *C. jejuni* concentrations in broiler chicken caeca. The objectives of the trial were to investigate the changes in *C. jejuni* concentrations as the birds aged and to examine the possible competition between two *C. jejuni* sequence types (ST-474 and ST-45). Both ST-474 and ST-45 sequence types are commonly found in poultry, however, ST-474, which has rarely been isolated outside New Zealand belongs to a poultry associated clonal complex (CC-48) (Muellner et al., 2013), whereas ST-45, belonging to CC-45 is a ubiquitous clonal complex, being frequently isolated from poultry, livestock, the environment and wildlife (Gripp et al., 2011; Sheppard et al., 2011a, 2014).

This study investigated the genotypic and phenotypic changes of two *C. jejuni* isolates after passage through the caeca. The degree of mutation and recombination and its effect on the

phenotype of the bacteria was investigated by whole genome sequencing and SNP analysis. The information generated in this study provides valuable experimental information on microevolution over a short term in experimentally infected birds and mutation rates in the natural habitat of the bacteria.

6.2 Materials and methods

6.2.1 Bacterial strains and challenge

The two *C. jejuni* isolates P179a and P197b were isolated from retail chicken carcasses and stored at -80°C in 15% glycerol broth. The isolates were revived from glycerol broth and sub-cultured on Columbia Horse Blood Agar (BA) (Fort Richard Laboratories, Auckland, New Zealand). The plates were incubated under microaerobic conditions (85% N_2, 5% O_2, 10% CO_2) for 48 h at 42°C in a VAIN (VA500, Don Whitley Scientific, Yorkshire, UK). Bacterial colonies were collected from a 48 h culture plate and re-suspended in 10 ml of Phosphate Buffered Saline (PBS), pH 7.3 to a density equivalent to a 0.5 McFarland standard. Two ml of the PBS suspension were added to 198 ml of PBS to give a suspension of approximately 1 x 10^6 cells/ml and the two 200 ml suspensions were immediately dispatched for use in the challenge.
Serial ten-fold dilutions of the 10 ml cell suspensions were made in peptone diluent (Fort Richard Laboratories, Auckland, New Zealand) and 100 μl volumes were spiral plated onto BA with microaerobic incubation for 48 h to ascertain the viable cell counts.
After challenge, the remaining suspension was returned to the laboratory, serial ten-fold dilutions were made in peptone diluent and 100 μl volumes were spiral plated onto modified Charcoal Cefoperazone Deoxycholate Agar (mCCDA) (Fort Richard Laboratories, Auckland, New Zealand) to ascertain the post-challenge viable cell counts.
The challenge preparation and sampling was carried out by Dr. Anne C. Midwinter and members of the mEpiLab team, whereas the oral gavaging of the chickens was carried out by the team of Professor V. Ravindran (IFNHH, Institute of Food, Nutrition and Human Health, Massey University, Palmerston North, New Zealand).

data clarification

6.2.2 Experimental infection

Male broiler chicks were obtained as 1 day-olds and randomly assigned to eight floor pens on wood shavings in an environmentally controlled room. Each treatment was allocated randomly to four pens. Each pen held 22 birds, giving a total of 88 birds per treatment group. On day 21, all birds were orally gavaged with *C. jejuni* P179a (ST-474) and P197b (ST-45) at a dose of 1 x 10^6 cfu each. At the ages of 33, 39 and 43 days, seven birds per replicate pen were killed and caecal samples were collected. This resulted in a total of 56 samples at each age (7 birds/ replicate x 4 replicates/ treatment x 2 treatments) and a total of 168 samples.

6.2.3 Motility agars

Motility phenotypes of isolates were tested in Brucella media with 0.4% agar (BD, Sparks, Maryland, USA). One ml of sterile 2% 2,3,5 triphenyl tetrazolium chloride (per 200 ml agar) were added to the agar for visualisation of the bacterial growth. Bacterial isolates were grown on BA for 48 h under microaerobic conditions and one colony was sub-cultured to BA for an additional 24 h. Bacterial cells were harvested and re-suspended in Mueller-Hinton broth (Fort Richardson, Auckland, New Zealand) at a density of approximately 10^8 cfu/ml. The centres of the motility agar plates were stabbed with 1 μl of the bacterial suspension at a density of approximately 10^8 cfu/ml and incubated under microaerobic conditions at 42°C for 24 h. The motility assay examined 14 isolates, 12 of the experimentally obtained cultures which were selected based on their SNP profile, and the two inoculum strains. The motility was tested three times for each isolate on three different days which resulted in nine measurements per examined isolate. The diameter of the ring of growth was measured with callipers. This work was mainly carried out by Dr. Anne C. Midwinter.

6.2.4 Isolation of DNA and sequencing

Sections of chicken caeca were received in the laboratory in individual test-tubes, on ice, as soon as practicable after dissection. 1.0 +/- 0.2 gram aliquots were added to 9.0 ml volumes of peptone diluent and vortexed well to suspend the caecal contents evenly. Sterile glass beads were added when required to aid homogenisation. Hundred-fold dilutions in peptone diluent were made from the primary dilution and 100 μl volumes of the 10^{-1}, 10^{-3} and 10^{-5} dilutions were spiral plated onto mCCDA followed by microaerobic incubation at 42°C for 48 h. A 1 ml aliquot of the primary 10^{-1} dilution was added to 20 ml of Boltons Campylobacter enrichment broth (Lab M, Bury, England) which was also incubated microaerobically at 42°C for 48 h.
After incubation colonies resembling *Campylobacter* were counted and Boltons broth was sub-cultured to mCCDA if no *Campylobacter*-like colonies were seen. At least one colony from every caecum was sub-cultured to BA and frozen in glycerol broth at -80°C.
Twenty-seven of the 168 isolates cultured from the caeca of the 12 birds were randomly selected (from all 3 kill days) and with the two inoculum strains, sub-cultured on BA and incubated under microaerobic conditions for 48 h at 42°C. A single colony was sub-cultured to a new plate after 48 h. DNA from the 27 isolates were extracted with the QIAamp DNA Mini, QIAamp Blood Mini kit (Qiagen, Auckland, New Zealand) or Wizard Genomic DNA Purification Kit (Promega, Madison, Wisconsin, USA) according to the manufacturer's instructions. The prepared samples were sent to New Zealand Genomics Limited (NZGL) for full-genome sequencing using a MiSeq (Illumina).
Each of the 27 genomic DNA samples were fragmented by nebulisation for 6 minutes at a pressure of 32 psi, purified, then end repaired, A-tailed, adaptor-ligated, fractionated, purified

and enriched according to the manufacturer's instructions, using the TruSeq DNA LT Sample Prep Kit v2-Set A (48 reactions with PCR): part number FC-121-2001 and TruSeq DNA LT Sample Prep Kit v2-Set B (48rxn with PCR): part number FC-121-2002, Illumina Inc. The prepared libraries were normalized to equal molarity, diluted to 2 nM and pooled. A flow cell was prepared for each of the 2 library pools and sequencing reactions using 9 pmoles of the pooled libraries were performed on an Illumina MiSeq instrument with the MiSeq Reagent Kit v2 (500 cycle): part number MS-102-2003, Illumina Inc (Victoria, Australia), to give approximately 12 to 15 million clusters per run.
Figure 6.1 shows the workflow for the quality control, assembly and analysis. The single steps are described in the corresponding method sections. The DNA extraction and analysis was carried out by myself.

Quality control and *de novo* assembly of sequence reads

SolexaQA (Cox et al., 2010) was used for trimming and quality control of the reads (Q30, expected probability of incorrect base calling of 1:1000), subsequently the *de novo* assembler Velvet (version 1.2.10 (Zerbino and Birney, 2008)) was used for assembling the 2x250 base short reads. The de Bruijn graph-based genome assembly algorithm works in two steps: short reads are broken into small pieces (k-mers) and a de Bruijn graph is constructed from those short pieces; the genome is then derived from the paired de Brujin graph (Eulerian path). The sequences were assembled across a range of k-mers in steps of 10, for the odd numbers between 245 and 55 inclusive. These resulting Velvet contigs were then stored in a MySQL database.

Gene prediction and clustering

For each isolate, concatenated contigs generated at each k-mer (an assembly) were ranked based on a score derived from their N50 (a statistical measure of average length of a set of sequences), smallest number of contigs, maximal contig length and assembly length. A subset of three assemblies were chosen and annotated with Prokka (Seemann, 2014) to determine the robustness of the assembly method. Using customised Perl scripts the annotated contigs from each assembly were clustered using OrthoMCL (version 2.09; (Chen et al., 2006a; Li et al., 2003b)) with the default parameters and cluster length range of 20 amino acids.
From this point on OrthoMCL was used to calculate a pan genome across the experimentally obtained genomes and the two inoculum strains. However, the ST-45 inoculum strain did not share any genes with the 27 experimentally obtained isolates that were not present in the ST-474 inoculum strain. Therefore the pan genome matrix was re-calculated using the experimentally obtained isolates and the ST-474 inoculum strain only. The matrix indicated 48 instances of potential recombination events (missing or inserted genes). The coordinates of the genes were

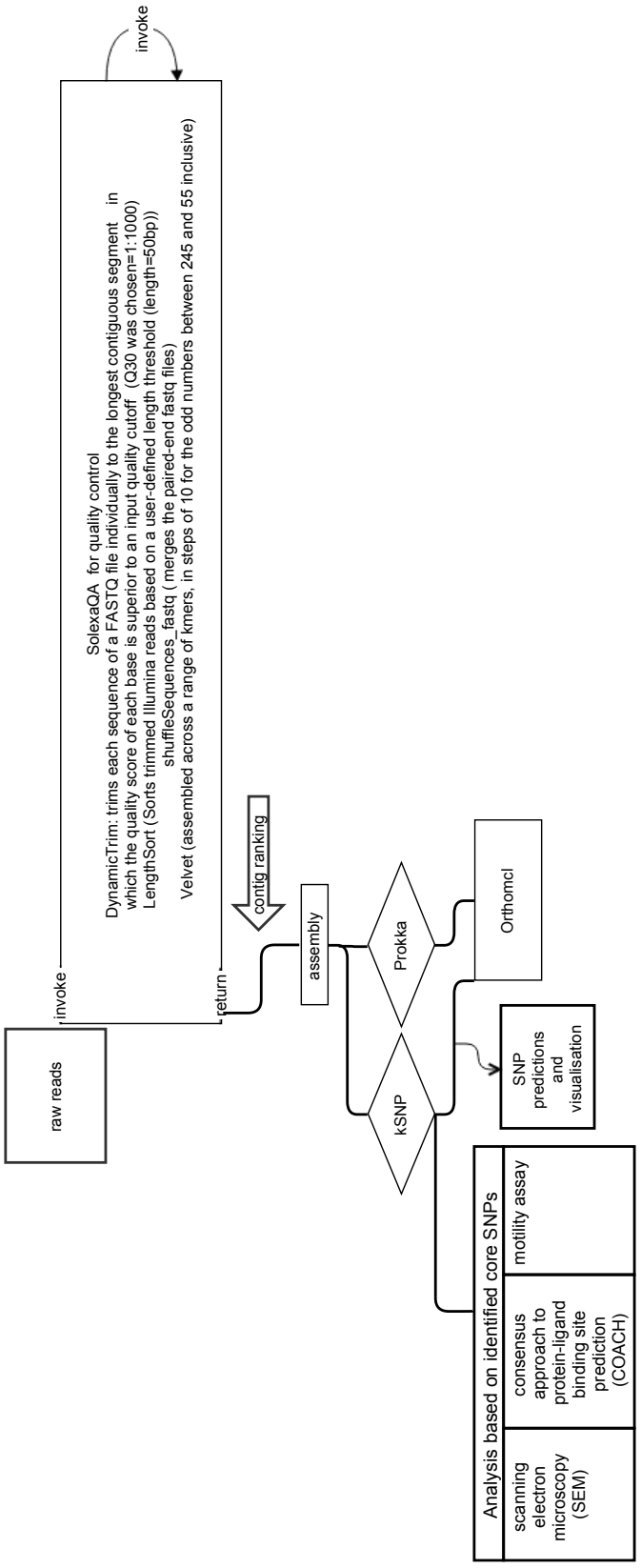


Figure 6.1: This figure shows the details of the quality control, assembly strategy and different steps of the analysis.

extracted from the MySQL database, and for each instance, a gene of interest, in a given gene cluster, was extracted and BLASTN was performed against all genomes to find the adjacent genes. One of the matched genes was extracted with 1kb flanking at each side and Muscle (Edgar, 2004) was used to generate an alignment which was visualised in Geneious (version 7.1.7, Kearse et al. (2012)). Each gene was then mapped back to the raw reads to determine whether it is truly present or absent or if the presence/absence call was based on an assembly error.

6.2.5 SNP detection methods

kSNP

The program kSNP (Gardner and Hall, 2013; Gardner and Slezak, 2010; Hysom et al., 2012; Kurtz et al., 2004; Marçais and Kingsford, 2011; Price et al., 2010; Stamatakis, 2006) identifies the pangenome SNPs for a set of given genome sequences and generates a phylogenetic tree based on the identified SNPs. The program also provides multiple alignment files with all the SNP positions and files that provide the exact location of the SNP in the genome. The program automatically analyses the core SNPs to generate a multiple alignment file and trees based on the identified core SNPs.
The k-mer size defines the length of an oligonucleotide (k-mer) that is identified by kSNP in all of the given sequences and subsequently detects SNPs based on this k-mer. Therefore it does not require a reference genome for the identification of SNPs. The optimal k-mer was determined through a sub-program of the software called Kchooser. kSNP was run for the 27 experimental chicken isolates and the two inoculum strains with the identified k-mer K=21.

Mapping of SNPs

Bowtie2 (version 2.1.0, Langmead and Salzberg (2012)) a tool for aligning sequencing reads to long reference sequences, and SAMtools (Li et al., 2009a) were used to map the region with the identified SNP back to the raw reads and the ST-474 inoculum strain to verify the SNPs which were visualise in Geneious. Additionally, OrthoMCL was used to cluster the genomes and identify the genes with same length but variable sequence which were also used to validate the SNPs identified by kSNP.

6.2.6 The probability of SNP co-locations

The probability for a given number of SNPs (h) in one gene, given r SNPs in n genes is given by:

$$\frac{n\binom{n+r-h-1}{r-h}}{\binom{n+r-1}{r}}$$

Where the denominator gives the number of ways of picking r SNPs in n genes allowing multiple SNPs per gene (sampling with replacement) and the dominator represents the number of ways to choose a gene with at least h SNPs where the remaining $r - h$ SNPs are distributed across the n genes with replacement.

6.3 Results

The challenge suspension was counted before and after the challenge to ensure the *C. jejuni* isolates were alive. Before the challenge each ST had a count of approximately 1.3 x10^6, cfu/ml, after the challenge the (composite) suspension contained approximately 6.0x10^5 cfu/ml. This implies that the *C. jejuni* were alive when the chickens were challenged. Even though the chickens were inoculated with two STs (ST-474 and ST-45) commonly isolated from chickens, and the isolates had each been isolated from chicken carcasses, only one ST (ST-474) was recovered after 33, 39 and 43 days from the caeca of 168 chickens. Out of the 168 isolates, 27 isolates were randomly chosen from all three kill days and sequenced using a MiSeq, Illumina.

The algorithms SolexaQA (Cox et al., 2010), Dynamic Trim and LengthSort were used for quality control (Q30, expected probability of incorrect base calling of 1:1000) and trimming of the raw reads. Subsequently the *de novo* assembler Velvet (Zerbino and Birney, 2008) was used for the assembly of the 2x250 base short reads. The sequences were assembled across a range of k-mers in steps of 10, for the odd numbers between 245 and 55 inclusive. The resulting contigs were ranked based on a score derived from their N50, number of contigs, maximal contig length and overall assembly length and annotated using Prokka. The annotated contigs were all very similar in genome length and number of genes and therefore one contig (for each genome) was chosen for the downstream analysis. The assemblies chosen occurred when the k-mer was approximately two-thirds of the 250 base read length. Using customised Perl scripts the annotated contigs from each genome were clustered using OrthoMCL (version 2.09; (Chen et al., 2006a; Li et al., 2003b)) with the default parameters and cluster length range of 20 amino acids.

Table 6.1 shows a summary of some key characteristics of the 29 isolates used in this study. The ST-45 isolate contains on average 80 fewer coding sequences (CDS) compared to the ST-474 isolates and is about 60 kb shorter. All the ST-474 isolates are very similar in their genome size, their number of signal peptides and genes.

The gene presence absence matrix for the 29 isolates did not identify genes that were present in the ST-45 inoculum strain (but not the ST-474 inoculum strain) and any of the 27 experimentally obtained isolates. A subsequent kSNP analysis across the 29 genomes did not identify SNPs that were found in the ST-45 challenge strain and the experimentally obtained isolates only and not in the ST-474 challenge strain. Given these results and the fact that all the sequence typed isolates (from 168 isolates) were all ST-474, we concluded that the ST-45 inoculum strain did not recombine with the ST-474 but failed to persistently colonise the chickens. Therefore it was

isolate	bases	sig_peptides	CDS	rRNA	tRNA	GC content	no of contigs
P197b	1626298	111	1609	4	36	30.2	28
P179a	1688575	109	1691	6	36	30.2	23
116b	1685535	109	1690	4	36	30.2	15
116c	1688127	109	1693	6	36	30.2	18
116d	1685535	109	1694	6	36	30.2	18
117a	1687839	109	1686	5	35	30.1	21
117c	1687189	109	1691	6	36	30.2	19
117e	1689717	109	1687	6	32	30.1	19
121a	1688069	109	1691	5	40	30.1	23
121c	1684628	109	1693	6	36	30.2	15
122a	1687470	109	1687	6	36	30.2	17
122d	1687484	109	1691	6	36	30.2	23
122f	1685455	109	1689	6	36	30.2	15
131a	1685932	109	1692	5	36	30.2	16
131b	1686478	109	1691	6	35	30.2	20
131c	1688515	108	1686	6	36	30.1	18
132c	1683094	109	1689	5	36	30.2	17
231a	1685080	109	1694	6	36	30.2	18
231c	1685867	109	1691	6	36	30.2	15
232a	1685802	109	1691	5	36	30.2	17
232c	1687128	109	1687	6	36	30.2	18
316a	1687694	109	1691	6	36	30.1	20
316d	1687649	109	1688	6	36	30.2	19
317b	1687775	109	1689	6	36	30.2	22
321a	1684657	109	1693	6	36	30.2	14
321b	1687514	109	1690	6	36	30.2	18
321c	1685447	109	1689	6	36	30.2	16
322a	1686270	109	1689	6	36	30.2	15
322c	1683582	109	1688	6	36	30.2	16

Table 6.1: Summary of genome characteristics of the inoculum strains (P179a (ST-474) and P197b (ST-45)) and the experimental isolates used in this study (all ST-474).

excluded from the downstream analysis.

Figure 6.2 shows a comparison of the two inoculum strains and one of the experimental genomes. The experimental *C. jejuni* isolate 117e (kill day 33) was mapped back to its raw reads to obtain the mapping coverage. BRIG (Blast Ring Image Generator) (Alikhan et al., 2011) was then used to visualise the two inoculum strains and the mapping coverage of the isolate 117e. The innermost ring displays the mapping coverage based on the sam file of the raw reads, followed by the isolate P197b (ST-45), the isolate P179a (ST-474) and the coding sequences. As the mapping coverage was based on the raw reads of the isolate 117e and did not show any drops in coverage, the gaps in the genome of P197b symbolise missing genes. There are no genes shared between P179b and the isolate 117e that are not present in P179a.

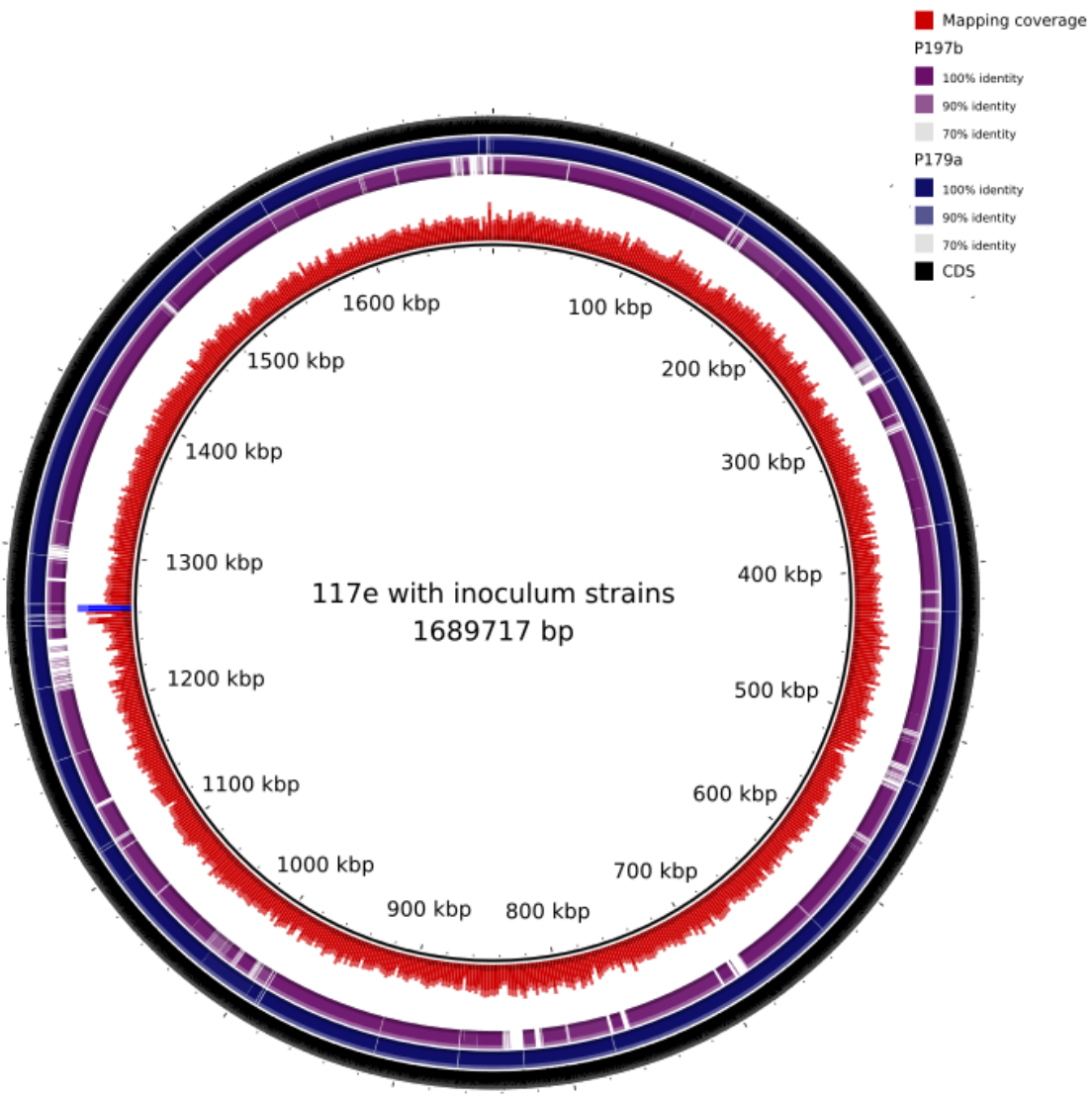


Figure 6.2: This BRIG plot shows a comparison of the two inoculum strains (P179a (purple) and P197b (blue)) to the mapping coverage of the experimental isolates (117e). No drops in the coverage were observed, therefore the gaps in the genome of P197b symbolise missing genes. The apparently missing genes in the P179a isolate compared to the 117e isolate are due to the blast identities estimated by BRIG and annotation/ assembly errors.

6.3.1 Testing for recombination events between the ST-474 isolates

The alignments of genes of interest created with Muscle were visualised in Geneious to identify whether the apparent absence or presence of the gene could be attributed to recombination or if it was an artefact (due to assembly or annotation error). For the cases with an apparently missing gene in a single genome it was due to a truncated gene in the genome which caused

it to be detected as being unique. These genes were mapped back to the raw reads to verify the presence of the full gene. Other instances, where a gene was missing in one or two of the genomes, were attributable to frameshifts or endings of contigs causing the gene to be truncated. For all the examined instances, no recombination was observed.

6.3.2 Identification of single nucleotide polymorphisms

The software kSNP was employed to detect and identify possible SNPs within the 28 examined chicken isolates. The neighbor-joining tree below (Fig 6.3) is based on the identified core SNPs within the 27 isolates in relation to one of the inoculum strains (ST-474). Overall kSNP identified 15 core SNPs and three non-core SNPs. Fourteen of the core SNPs are non-synonymous point mutations confined to nine genes and six of these non-synonymous point mutations occurred in one gene. These genes are listed in table 6.2. The probability of six SNPs occurring in one gene is 4.1×10^{-9}.

Although the isolates were derived from chickens belonging either to the treatment (treated with caprylic acid) or the control group, and were isolated at different days, there is no apparent clustering by group or kill day. It is also interesting to notice that isolates from the same chicken (e.g. 321a-c) show different SNPs.

The identified SNPs were mapped to the raw reads with Bowtie2 to validate the SNPs. Additionally, OrthoMCL was used to cluster the predicted genes from the genomes and identify the genes with the same length but a variable sequence which were also used to validate the SNPs identified by kSNP.

For the 14 non-synonymous SNPs the InterProScan plugin in Geneious was used to identify associated PFAM (protein family) domains (Finn et al., 2013). No domain was identified for *ptm*G and the SNPs for *cut*E and the 'putative MCP-type signal transduction protein' are located outside the domain. The remaining SNPs were located in highly conserved regions of the genes. The three non-core SNPs were located in the motility accessory factor gene 6 (*maf*-6) and a hypothetical protein.

The associated PFAM domain identified for the *mre*B gene is MreB_Mbl and all the identified SNPs are located within this domain. The amino acid identities of the point mutations were compared to the *C. jejuni* sequence collection on PFAM and it was observed that the SNPs were located in highly conserved regions. The SNPs present in *mre*B were also compared to ∼2000 *C. jejuni* isolates in the PubMLST database (`http://pubmlst.org/campylobacter/`) and we identified the two most common SNPs (*mre*B 2 and *mre*B 3) in our dataset to be present also in the Oxford isolates (0.69% and 2.76% respectively).

To further study the SNPs in *mre*B, scanning electron microscopy (SEM) was used to examine possible phenotypic changes of the mutants (the 27 chicken isolates) compared to the wild type (ST-474 inoculum strain). Two examples showing the ST-474 inoculum strain and one of the experimental isolates can be seen in Figure 6.5.

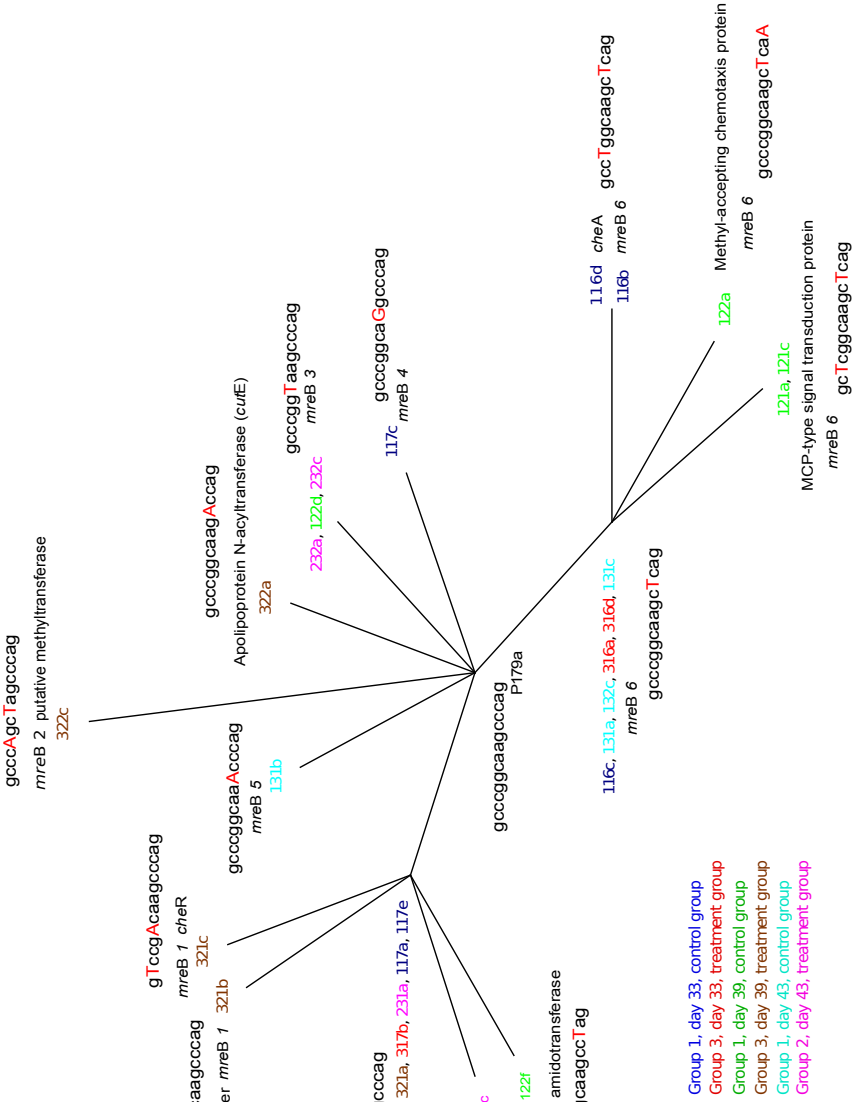
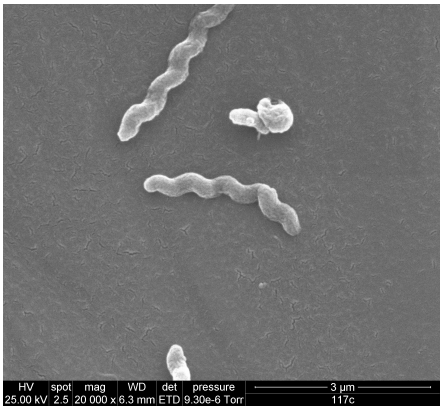


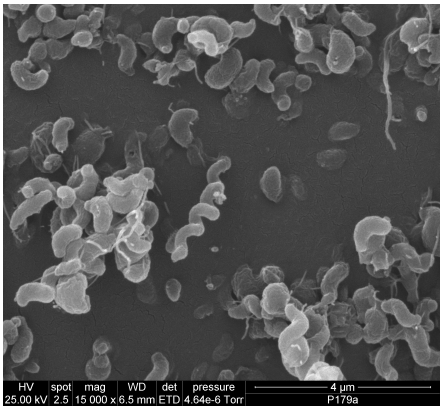
Figure 6.3: The figure shows a neighbour-joining tree based on the detected SNPs within the 27 isolates and one of the inoculum isolates (P179a). The network was generated by kSNP and visualised with SplitsTree (Huson, 1998). The colours of the isolates correspond to the kill day and whether they belonged to the treatment or control group.

identified gene	synonyms	SNP coverage	aa position of SNP	aa change	number of isolates	PFAM domain
*mre*B	Cell shape-determining protein MreB	coverage of 247-537, variant frequency of 98.0%-100.0%	40 107 158 207 212 329	V→M R→C P→S N→S E→K T→I	1 9 11 1 3 1	MreB_Mbl
N-acetyl sugar amidotransferase	*ptm*G, Cj1324	coverage of 500, variant frequency of 99.8%	73	G→R	2	-
*che*A	-	coverage of 425, variant frequency of 100%	548	V→I	2	CheW
*che*R	-	coverage of 193, variant frequency of 99.5%	243	P→S	1	CheR
Methyl-accepting chemotaxis protein	mcp	coverage of 383, variant frequency of 98.7%	213	G→D	1	Pas_9
Apolipoprotein N-acyltransferase	lnt,*cut*E, ALP N-acyltransferase	coverage of 292, variant frequency of 99.7%	160	D→Y	1	-
Malonyl CoA-acyl carrier protein transacylase	*fab*D	coverage of 236, variant frequency of 91.1%	26	I→V	1	$Acyl\_trasf\_1$
putative Methyltransferase	-	coverage of 428, variant frequency of 99.5%	53	S→T	1	Methyltransf_23
Putative MCP-type signal transduction protein	-	coverage of 257, variant frequency of 100.0%	160	A→T	1	-

Table 6.2: The table shows a summary of the identified genes with non-synonymous SNPs, their synonyms, the nucleotide coverage in the raw reads, the amino acid position, the point mutation, the number of isolates the SNP occurred in and the associated PFAM domain.



(a) *C. jejuni* isolate 117c



(b) *C. jejuni* isolate P179a

Figure 6.5: Scanning electron microscopy (SEM) images showing the *C. jejuni* isolates P179a (ST-474) and 117c (ST-474).

No obvious changes in the shape of the bacteria were observed. We used the software COACH (Yang et al., 2013a,b) to inferthe 3-D protein structure of *mre*B and Chimera (Pettersen et al., 2004) to visualise the binding sites of the protein and the non-synonymous SNPs (Figure 6.6). The binding sites were visualised in red, the SNPs were visualised in green. None of the SNPs were located within the binding sites of the protein.

comment 37, Murray

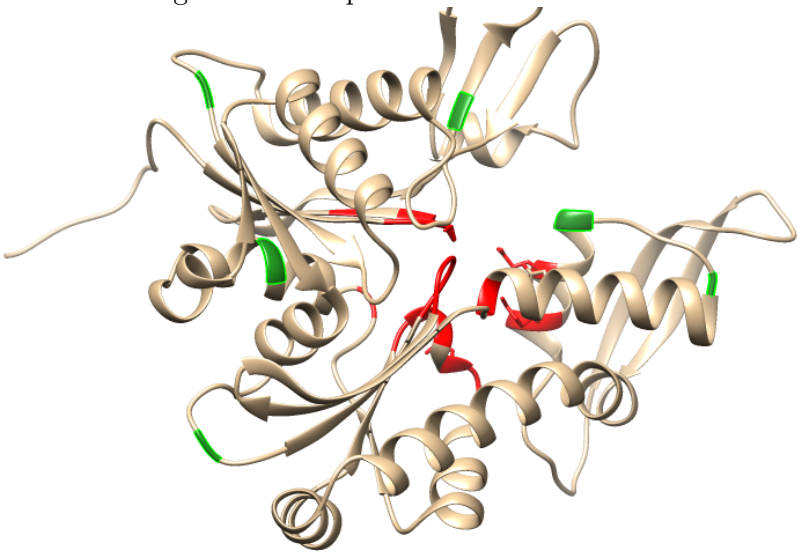


Figure 6.6: 3-D protein structure of the protein MreB. The binding sites are visualised in green, the identified SNPs are visualised in red.

Another test for phenotypic differences between the inoculum strain and the experimental genomes was based on motility in soft agar. The isolates tested were selected based on their SNP profile (such as isolate 322c which has a non-core SNP in the *maf* 6 (motility accessory factor 6)

Isolates	motility in mm (95 % CI[a])	Pr(> |t|)
baseline (P179a)	22.42 (20.94-23.91)	$< 2^{-16} * **$
116b	20.69 (18.60-22.79)	0.18
116c	20.24 (18.15-22.34)	0.10
116d	19.73 (17.64-21.83)	0.04*
117a	23.93 (21.84-26.03)	0.25
117c	21.99 (19.90-24.09)	0.74
122a	23.21 (21.11-25.31)	0.54
122d	24.94 (22.85-27.04)	0.054
122f	9.22 (7.12-11.32)	$< 2^{-16} * **$
131b	15.20 (13.10-17.30)	$< 1.6^{-7} * **$
321c	24.00 (21.90-26.10)	0.23
322a	26.98 (24.88-29.08)	$< 6.3^{-5} * **$
322c	4.6 (2.50-6.70)	$< 2^{-16} * **$
P197b	7.01 (4.91-9.11)	$< 2^{-16} * **$

Table 6.3: Results of the linear model displaying the isolates that differ significantly in motility from the inoculum strain P179a (baseline).
Significance codes: 0 ***' 0.001 **' 0.01 *'; [a] 95% confidence interval

gene, 122a which has a SNP in a methyl-accepting chemotaxis protein and *mre*B 6, 122f which has a SNP in mreB 1 and N-acetyl sugar amidotransferase). Examples of motility plates for the two inoculum strains and two of the experimental isolates can be seen in Figure 6.7. The growth rings of the isolates were measured with callipers and recorded (in mm). A linear model based on the measurements for each replica was used to identify the isolates that showed a significantly different motility from P179a. The results of the model are summarised in Table 6.3. Out of the 14 examined isolates only P197b, 322c, 322a, 131b and 122f appear to be significantly different in their motility compared to the inoculum ST-474 isolate P179a (baseline). The second column displays the mean value for the motility per isolate and the 95% confidence interval.

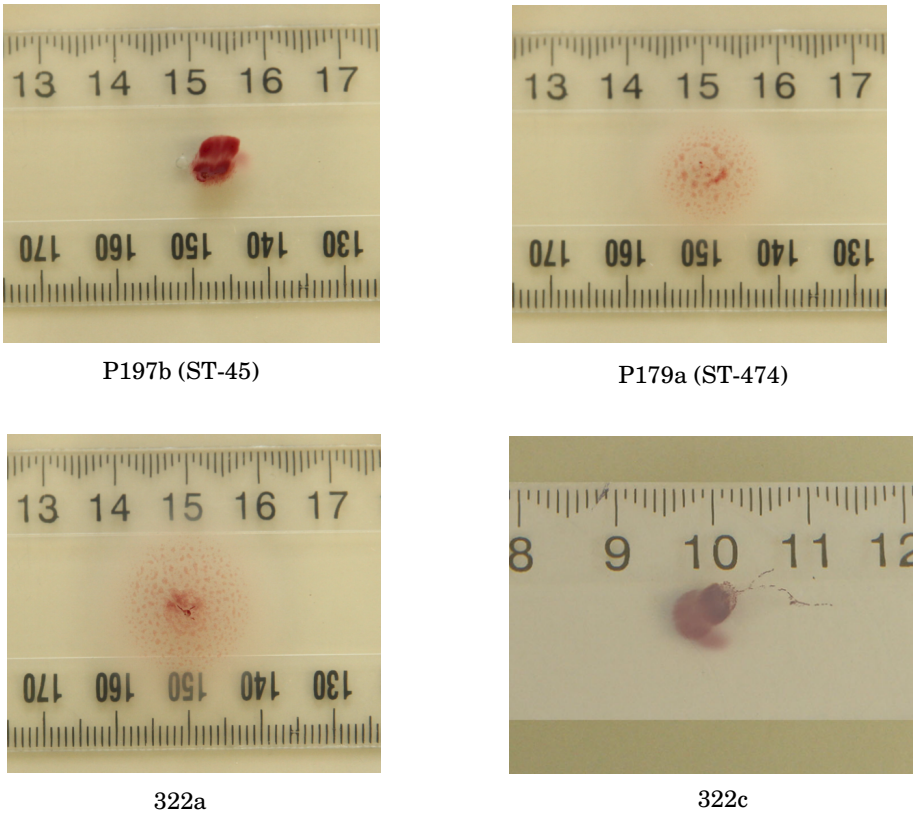


Figure 6.7: Different isolates were, based on their SNP profile, selected for the motility assays. This figure shows an examples of the resulting motility phenotype for the two inoculum strains P179a and P197b and two experimental isolates. The reduction of the triphenyl tetrazoloium chloride giving a pink colour indicates the motility of the bacteria after 24 h. Although the isolates 322a (SNP in Apolipoprotein N-acyltransferase (*cut*E) and 322c (SNP in *maf* 6, *mre*B 2 and a putative methyltransferase)) are from the same chicken sample, they show differences in their motility.

Relatively few mutations were observed in genes related to chemotaxis as opposed to genes related to motility or cell shape. Therefore, (and due to time limitations), only a motility test and examinations of cell shape were performed. Further work investigating this could involve chemotaxis test as described in Hazeleger et al. (1998).

comment 38, Murray

6.4 Discussion

The present study used data from a dual challenge with two *C. jejuni* isolates in chickens to examine short-term evolutionary dynamics and distinct strain-to-strain variations. Previous studies examining colonisation potential of different *C. jejuni* isolates in chickens discovered variances in their dynamics of infection and gut ecology (Chaloner et al., 2014; Hänel et al., 2009; Korolik et al., 1998; Ringoir and Korolik, 2003). Here we showed that although the chickens were inoculated with two isolates of *C. jejuni* (ST-45 and ST-474), out of 168 isolates, only ST-474 was

recovered after the trial. This could be due to two possibilities: either ST-45 did not colonise the chickens at all, or it did not manage to colonise the chickens persistently. Both STs are frequently isolated from chickens, but whereas ST-474 is, in New Zealand, a chicken-associated lineage (Muellner et al., 2013), ST-45 is a ubiquitous sequence type, frequently isolated from multiple sources (French et al., 2009; Gripp et al., 2011; Sheppard et al., 2014). Konkel et al. (2007) has found that a more motile *C. jejuni* strain could inhibit a second, less motile *C. jejuni* strain from binding to chicken epithelial cells. It is possible that this has happened during this experiment, especially as the ST-474 inoculum strain showed greater motility compared to the ST-45 inoculum strain (Figure 6.7).

None of the 27 sequenced experimental isolates showed evidence of recombination, strengthening the theory that the ST-45 isolate was rapidly out-competed by the ST-474 isolate. This is an interesting finding considering that past research estimated the ratio of mutation to recombination to be at least 1:1 (Biggs et al., 2011; Falush et al., 2001). However, the Biggs et al. (2011) was a macroevolutionary study and therefore based on more diverse genomes. This microevolutionary study is examining an almost clonal population of ST-474 isolates. It is possible that within the chicken caecum were no other diverse *C. jejuni* isolates to recombine with, or that the isolates were under positive selection pressure adapting to the chicken caecum. No genes were identified that were shared between the ST-45 inoculum strain and any of the 27 experimental genomes that were not shared with the ST-474 inoculum strain. Figure 6.2 shows a mapping coverage plot of one experimental genome (117e) and the two inoculum strains. No drop in coverage was observed, therefore the gaps in the genome of P197b symbolise missing genes (compared to the experimental genome). The examined instances of missing or inserted genes did not show evidence of recombination but were due to assembly or annotation errors. However, not all of the instances were examined.

comment 36, Murray

The software kSNP was applied to determine potential SNPs between the 27 experimental isolates and the ST-474 inoculum strain. Fifteen core SNPs and three non-core SNPs were observed, 14 of the 15 core SNPs and the three non-core SNPs were non-synonymous point mutations. The high K_a to K_s ratio indicates that the genes are evolving under positive selection. Although housekeeping genes are under purifying selection, this is not necessarily true for the accessory genome which includes the genes responsible for the adaptation of the organisms to the environment and will therefore be under positive selection. The 15 core SNPs and three non-core SNPs were confined to 12 genes, 11 of which were associated with motility or chemotaxis. The remaining gene has an unknown function. The 15 core SNPs were visualised in the SplitsTree in Figure 6.3. No apparent clustering was observed based on kill day or treatment/control group. Furthermore, isolates from the same chicken (e.g. 321a-c) were clustered at different branches of the tree.

comment 35, Murray

One identified protein, MreB (mecillinam-resistance protein B) had six parallel mutations across

26 of the 27 experimental genomes. MreB is a rod-shape determining protein that has been identified as a homologue of actin (van den Ent et al., 2001). The homology was inferred because MreB had been shown to be associated with the maintenance and regulation of cell shape, as rod-like bacteria became spherical when the *mre*B gene was knocked out (Doi et al., 1988). Furthermore, MreB is associated with membrane proteins such as MreC, which help to anchor MreB to the cell wall (Egelman, 2003). To date it is only known that MreB and its associated proteins (such as MreC) are important in either establishing or maintaining the cell diameter, however, the molecular mechanisms of this action are still unknown. Other studies have also suggested that MreB may be involved in the bacterial adherence to host cells (Tu et al., 2008) and general stress response (Garénaux et al., 2008a).

The MreB associated PFAM domain was identified as MreB_Mbl and all the SNPs were located in highly conserved regions of the gene. However, comparing the SNPs present in the *mre*B gene to ∼2000 clinical *C. jejuni* isolates from Oxford in the PubMLST database, we identified two of the SNPs in our dataset as being also present in the Oxford isolates (data not shown). This could suggest parallel evolution or selection pressure to adapt to the host. This theory of adaptation would also be strengthened by the finding of Jerome et al. (2011) who, when examining contingency loci, identified a mutation in the mouse adapted strain compared to the original NCTC 11168 in *mre*B (Jerome et al. (2011), table S2).

An attempt to link the identified genotypic changes to the phenotypes of the isolates was made by using scanning electron microscopy (SEM) images. However, despite the inferred homology to actin and MreB's role in establishing and maintaining the cell diameter, no obvious phenotypic changes were observed (Figure 6.4b). One explanation could be that the changes were too subtle to be observed on the SEM images, or that the SNPs, despite being non-synonymous did not affect the cell shape of the bacteria. In order to identify whether the SNPs could have potentially altered the function of the protein, a 3-D structure was determined for MreB and the binding sites and SNPs were visualised (Figure 6.6). None of the SNPs were located within the binding sites of the protein which could imply that the function was not altered and the protein was still functional.

As some of the genes with SNPs were associated with motility, we used a motility assay to determine whether the SNPs had an effect on the motility of the isolates. An example of four motility plates can be seen in Figure 6.7. Overall we observed a number of different phenotypes, especially comparing between the inoculum strains and the experimental genomes. The observed difference in motility between the ST-474 and ST-45 inoculum strains, with the ST-474 isolate showing a greater motility (Figure 6.7), could explain the different colonisation abilities of the strains. Konkel et al. (2007) showed that one *C. jejuni* isolate could inhibit another strain from binding to chicken epithelial cells. The greater motility of ST-474 could potentially have decreased the ability of ST-45 to bind to the epithelial cells of the caecum and therefore decreased its capabilities to colonise the chickens.

The isolate 322c has, compared to the isolate 322a, a non-synonymous SNP in the motility accessory factor 6, which could explain the different motility observed in the motility assay (Figure 6.7). *Maf*6 is part of a 37 kb cluster of seven closely related genes which were described as the motility accessory factor (*maf*) family of flagellin associated proteins (Karlyshev et al., 2002). Karlyshev et al. (2002) demonstrated that the genes of the *maf* family are involved in the variation of motility via a slipped-strand mispairing mechanism. However, Karlyshev et al. (2002) did not identify the function of the genes and therefore they were designated as 'genes of unknown function'. However, as the isolate 322c has a SNP in *maf*6 and demonstrated reduced motility compared to isolates without the mutation, the *maf*6 gene may be essential for the motility of the isolate.

SNPs located in genes related to chemotaxis were found in *che*A, *che*R a methyl-accepting chemotaxis protein (MCP), a putative methyltransferase and a putative MCP-type signal transduction protein. Chemotaxis is the movement of bacteria as a response to chemical changes in the environment. *C. jejuni* isolates have a single polar unsheathed flagellum at one or both ends of the cell (Vandamme et al., 1991). These flagella can rotate counter clock-wise which enables the bacterium to swim in a straight line, or clockwise causing the bacterium to tumble in place and gives it the opportunity to re-orientate itself (Eisenbach, 2011). Whole genome sequencing has revealed CheB and CheR as being putative adaptation proteins, however, although these proteins influence chemotaxis, they do not seem to be essential for the behaviour (Kanungpean et al., 2011). The response regulatory domains are encoded by the genes *che*Y, *che*A and *che*V. The histidine kinase CheA is activated by coupling to a methyl-accepting chemotaxis protein (MCP), which then phosphorylates itself and transfers a phosphate group to CheY (Kanungpean et al., 2011). The activity of CheA and CheY change the rotation of the flagella, enabling a sudden change in bacterial swimming direction (Kanungpean et al., 2011).

Chemotaxis is an important determinant for colonisation of chickens. Previous research showed that strains with mutations in CheA or CheY were unable to colonise mice, chickens or ferrets (Chang and Miller, 2006; Hendrixson and DiRita, 2004; Yao et al., 1997). We did not perform a chemotaxis test, therefore we do not know whether the SNPs had a significant effect on the chemotactic behaviour of the isolates, however, all the isolates were extracted from the chicken caecum and therefore it is unlikely that the function of the genes were altered. One way to determine this would be to apply the methods described in Hugdahl et al. (1988) and test the ability of the isolates to respond to chicken-associated chemotactic stimulants.

A number of possible processes can generate variations in the assembled genome (such as SNPs generated *in vitro*, indels, sequencing errors and assembly errors) (Croucher et al., 2013). However, as all of the identified SNPs in this study were mapped back to the raw reads, we can exclude the assembly errors. Although the observed genetic variations could potentially have been created *in vitro*, the probability is very small. *Campylobacter* have a generation time of approximately two hours (based on NCTC 11168) (Vegge et al., 2012) and assuming that the

mutations would have occurred within the first ten generations, we would expect the occurrence of a SNP to be at most $2^{-10} \sim 0.001$.

This study showed the out-competition of a generalist *C. jejuni* isolate (ST-45, P197b) by a specialist isolate (ST-474, P179a). Although the birds were orally gavaged with equal amounts of the *C. jejuni* isolates, only ST-474 was recovered after the experiment. We observed 15 core (SNPs) and three non-core SNPs across the ST-474 isolates (including the inoculum isolate). Fourteen of the core SNPs were non-synonymous point mutations confined to nine genes which were all associated with cell shape, chemotaxis or motility of the bacteria. We identified alterations in the motility of the bacteria which could be associated with the detected SNPs. No recombination events were observed. This could potentially influence our knowledge of short-term evolution of *C. jejuni* in natural hosts.

Acknowledgments

AF is funded by the Allan Wilson Centre for Molecular Ecology and Evolution. We are grateful to the staff and students of the mEpiLab team for their support and helpful discussions and Professor V. Ravindran and colleagues for the animal work. We would also like to thank Mauro Tuglio for his expertise in biostructural analysis and his support with the 3-D protein structure of the MreB protein.

CHAPTER 7

Discussion

7.1 General discussion

The term 'phenotype' is used to describe the observable characteristics of an organism (e.g. seasonality, host association, metabolic activities or colonisation ability), whereas the 'genotype' denotes its genetic constituents. The genotype can be defined at different discriminatory levels (clonal complex, multi locus sequence type, whole genome sequence). In the context of evolutionary biology we need to understand how the genotype and its phenotype relate to each other; what constitutes a phenotype, which genes are involved and how do they provide an efficient response to environmental change. Phenotypic characteristics are determined by a combination of environmental and genetic factors. Even if the genotype is well adapted to its environments it can be influenced by environmental conditions and therefore the same genotype (such as a sequence type (ST)) may yield different phenotypes (such as pathogenicity) in different environments (hosts). Natural selection pressure acts on an organism's phenotype but evolution occurs through genetic changes in the genotype. Therefore the organism's ability to survive and colonise specific niches is affected by selection pressure.

The aim of this thesis was to investigate the phenotype:genotype relationships of *C. jejuni* and how they influence the micro-evolution of these bacteria in New Zealand. *Campylobacter* spp. are the most common cause of acute gastroenteritis in the developed world (Adak et al., 2005; Coker et al., 2002; Friedman et al., 2000; Samuel et al., 2004) and are frequently isolated from a wide variety of domestic livestock, including poultry (Eberhart-Phillips et al., 1997; Müllner et al., 2009, 2010b), ruminants (Stanley and Jones, 2003) and pigs (Brown et al., 2004; Nesbakken et al., 2003). The majority of human campylobacteriosis cases are caused by *C. jejuni* (90%) and *C. coli* (10%) where most of the *C. jejuni* infections are believed to be caused through the consumption of contaminated food (Carter et al., 2009; Eberhart-Phillips et al., 1997; Gillespie et al., 2002; Lee and Newell, 2006; Nachamkin et al., 1992).

Although the research conducted over the last few decades has increased our knowledge of *Campylobacter* spp. ecology, evolution and population biology, questions remain about the genetic basis and ecology of host specificity and adaptation. Several studies applying multilocus

sequence typing (Colles et al., 2003; Dingle et al., 2001; French et al., 2005; Karenlampi et al., 2007; Manning et al., 2003; McCarthy et al., 2007; Sails et al., 2003; Sheppard et al., 2009; Taboada et al., 2008; Wilson et al., 2008) aimed to identify the prevalence of specific types among *C. jejuni* isolates from diverse sources (such as agriculture, wild birds and environment). These studies observed a large variation between the host distributions of MLST clonal complexes (CC). A clonal complex is a cluster of closely related bacterial strains that group around a founder strain (Dingle et al., 2002).

Overseas studies examining CCs related to *C. jejuni* infections in humans have identified seasonal patterns for specific CCs in many temperate countries (Cody et al., 2012; McCarthy et al., 2007; Sopwith et al., 2008). Studies have been conducted to analyse the seasonality of *C. jejuni* in New Zealand and have identified marked differences between rural and urban areas, and the North and South Island of New Zealand (Hearnden et al., 2003; Spencer et al., 2012). However, as these studies were based on clinically reported case numbers, they did not determine the underlying clonal complexes causing the observed seasonality. Chapter 3 used a nine year time series of genotyped human campylobacteriosis cases from a sentinel surveillance site in the Manawatu region of New Zealand to investigate if the CCs identified in human cases in this region showed a seasonal pattern, and if so, how these CCs compared with seasonality-associated CCs that had been identified in other countries.

CC-45 was identified as a prevalent clonal complex showing a consistent summer peak over nine years. This was in contrast to a previous study (McCarthy et al., 2012) that found no apparent seasonality of CC-45 in New Zealand. However, the McCarthy et al. (2012) study used a small-scale dataset from New Zealand which was focused on a prolonged outbreak in the winter of one year (McTavish et al., 2008), instead of one of the larger published datasets (Müllner et al., 2009). CC-45 is an internationally well-known CC being associated with a consistent summer peak in several temperate countries (Cody et al., 2012; McCarthy et al., 2012; Sopwith et al., 2008). It is also termed a 'generalist' which means it is frequently isolated from multiple sources (e.g. livestock, environmental water, wild birds) (Colles et al., 2003; French et al., 2009; Sheppard et al., 2014; Sopwith et al., 2008). These findings suggest that CC-45 is adapted to survival outside the host and makes this adaptation a key driver of transmission between livestock, environment and humans (Sopwith et al., 2008). The findings that only a limited number of CCs (e.g. CC-45, CC-283, CC-42, CC-353 and CC-403) (Cody et al., 2012; McCarthy et al., 2012) are associated with human campylobacteriosis suggests specific exposure, transmission routes and virulence not commonly found in other CCs.

Prior to 2006, New Zealand had a high and steadily increasing number of campylobacteriosis cases with a peak of >380 per 100,000 population in 2006 (The Institute of Environmental Science and Research Ltd., 2007). After an intervention in the poultry industry (New Zealand Food Safety Authority, 2008; Sears et al., 2011) cases linked to the poultry-associated ST-474

(CC-48) dropped to < 5% (Muellner et al., 2013), which has also been confirmed in the statistical analysis in chapter 3. As *Campylobacter* infections in humans are primarily transmitted through livestock or animal-derived food, particularly contaminated poultry, and CC-48 is a strongly poultry associated cluster, one would assume a seasonal pattern of CC-48 associated with human campylobacteriosis. However, the data for this study were collected over nine years and the fact that CC-48 was influenced strongly by the intervention, may have obscured or eliminated existing patterns.

The intervention reduced the risk of poultry associated human campylobacteriosis cases but source attribution studies have estimated that there has been a relative increase in the importance of ruminant-associated strains (French et al., 2010; Muellner et al., 2011). New strategies have to be developed to control campylobacteriosis cases acquired from both poultry- and non-poultry sources (New Zealand Food Safety Authority, 2008). *Campylobacter* spp. are able to colonise multiple hosts, and hence a number of different niches. Identifying the population structure of these bacteria and strains that may be associated with specific niches may inform future source attribution and seasonality studies.

In chapter 4 phylogenetic and population genetic tools were applied to describe a detailed picture of the population structure and host associated genotypes within the New Zealand *C. jejuni* population in wild and agricultural animals. The findings show that the *C. jejuni* STs isolated from non-agricultural animals exhibit a substantially different population structure to those seen in agricultural animals. The *C. jejuni* isolates from non-agricultural animals exhibit a higher proportion of mosaic alleles (due to admixture) and fewer shared genotypes between the host groups, whereas the *C. jeuni* in agricultural animals show a higher proportion of shared genotypes and fewer occurrences of admixture. The variation between *C. jejuni* STs appears to be less pronounced within those isolated from domesticated animals as compared to those isolated from non-agricultural sources. This could suggest that the farm environment constitutes a specific niche for *Campylobacter* and that keeping different agricultural species in close proximity to each other may influence the population structure of *C. jejuni*.

The ability of *Campylobacter* to colonise a diversity of hosts and to survive in a huge range of environments is evidence of their ability to detect, adapt and evolve in response to such environments (Lee and Newell, 2006). Habib et al. (2010) compared phenotypic characteristics of the two major generalist lineages (CC-21 and CC-45) and found that CC-45 had more resilience to oxidative and freezing stress but a poorer survival response to heat or chilling compared to CC-21. One way to identify the reason for the wide distribution and prevalence of CCs such as CC-45 is to identify genes that are present in the generalist isolate but not in the specialists ones. Evolutionary forces drive bacteria to grow in and adapt to as many niches as possible (Bochner, 2009). The adapting bacteria have to use the basic nutrients (sulphur, nitrogen, oxygen, carbon) in that environment which may differ between hosts. Therefore, another way to determine why some genotypes show a higher degree of host-association than others is to

determine the specific phenotype of generalist and specialist genotypes based on biochemical requirements. Previous studies (Brás et al., 1999; Pajaniappan et al., 2008; Stintzi, 2003) have demonstrated that different temperatures, for example a chicken's body temperature at 42°C and a human's body temperature of 37°C, trigger expression of potential colonisation or virulence factors which could either lead to commensalism or pathogenesis.

Chapter 5 tested the ability of a variety of *C. jejuni* isolates to oxidise 95 substrates as sole carbon sources and tested the tolerance of the isolates to different osmotic conditions at 38°C and 42°C using phenotypic microarray (PM) technology. As far as the authors are aware, it was the first study to compare these parameters amongst different STs and host derived isolates and the first study to correlate the phenotypic expressions of the examined isolates with their genome sequences to identify genes that may be related to the observed phenotypes. The isolates belonging to ST-42 showed a unique phenotypic feature as they were able to utilise L-glutamine which was not commonly utilised by the other examined STs. In contrast, isolates belonging to ST-474 were the only ones to consistently utilise citric acid. Using the software PhenoLink and the whole genome sequences of the isolates, the gene gamma-glutamyltransferase (*ggt*) was identified as being essential for the utilisation of L-glutamine. This study also led to the identification of a type IV secretion system, previously not reported in *C. jejuni*, which could be involved in the natural transformation of the bacteria. The wider range of potential carbon sources and amino acids utilised by *C. jejuni*, and its tolerance to osmolytes identified in chapter 5, could lead to improvements in culture media for detection and isolation of the bacteria from diverse environments, and could assist in developing potential substances useful for intervention in colonisation of avian or human hosts. The different phenotypic profiles of the ST types could also provide new insights into genes associated with colonisation and niche-association.

Passages through an animal model give the opportunity to monitor changes in genes of interest while the isolates adapt to a new environment. Previous studies involving passage of a bacterial strain through a host attempted to identify *in vivo* genetic variation (de Boer et al., 2002; Espedido et al., 2013; Hänel et al., 2009; Koskiniemi et al., 2013; Thomas et al., 2014; Woodall et al., 2005). Only three of these studies identified the genetic variations based at the nucleotide level (Espedido et al., 2013; Koskiniemi et al., 2013; Thomas et al., 2014) as opposed to comparing PFGE band pattern (de Boer et al., 2002; Hänel et al., 2009; Woodall et al., 2005). Passages of *C. jejuni* isolates belonging to either generalist clusters or specialist clusters through an animal model can provide information on the pressure acting on the isolates and identify changes in genes (pre- and post passage) that were involved in the adaptation process.

The last research chapter made use of data from a previously performed chicken feed trial experiment by colleagues in IFNHH (Institute of Food, Nutrition and Human Health, Massey University, Palmerston North, New Zealand) in 2010 using caprylic acid as an in-feed additive.

Several studies examining the colonisation potential of different *Campylobacter* spp. strains concluded that, based on the isolation source (human, bovine or chickens), the strains had a different potential to colonise the gastrointestinal tract ranging from immediately sustained colonisation to non-colonising strains (Hänel et al., 2009; Korolik et al., 1998; Ringoir and Korolik, 2003). Chapter 4 showed that in New Zealand ST-45 is frequently isolated from a wide range of hosts, whereas ST-474 is strongly poultry associated. Although the isolates used for the dual challenge were both poultry derived, the generalist isolate ST-45 was out-competed by the specialist isolate ST-474 and did not persistently colonise the chickens. After the passage through the chickens we identified genetic variations of 15 core SNPs in the experimental isolates compared to the ST-474 inoculum strain. Fourteen of the 15 core SNPs translated into non-synonymous point mutations. These SNPs were confined to nine genes all of which were associated with cell shape, chemotaxis or motility of the bacteria. No evidence of recombination events between the ST-45 and ST-474 or amongst the ST-474 isolates were observed. This is an very interesting finding considering that past research based on whole genome sequences estimated the ratio of mutation to recombination to be at least 1:1 (Biggs et al., 2011; Fearnhead et al., 2005). This could suggest the presence of a strong selection pressure acting on *C. jejuni* during *in vivo* establishment in the host. These findings have influenced our knowledge of short-term evolution of *C. jejuni* in natural hosts and their adaptation to chickens as a host.

7.2 Limitations and future directions

The seasonality study was the first MLST-based study to examine the seasonal patterns of human campylobacteriosis in New Zealand. However, this was based on only one region and did not differentiate between rural and urban areas. Therefore, the first recommendation would be to extend the seasonality study conducted in the Manawatu region of New Zealand, to the entire North and South Island of New Zealand. It would be very interesting to see whether the differences between rural and urban areas that have been observed in previous studies (Hearnden et al., 2003; Kovats et al., 2005; Nylen et al., 2002; Spencer et al., 2012) are associated with specific MLST types or clonal complexes. There are two possible ways to achieve this: firstly, to rely on the existing laboratory-based notification system and collect the samples from the medical laboratories, or secondly, to conduct a prospective epidemiological study and sample individuals from the community independently of the medical notification system. Although campylobacteriosis is a notifiable disease, it is believed to be underreported at two levels, from the community (not all ill people seek healthcare) and from the health-care system (failure to sufficiently report symptomatic cases that have sought medical advice) (Gibbons et al., 2014).
It would also be interesting to obtain the meteorological data for the examined regions to identify potential correlations between climate and human campylobacteriosis cases. The correlation of these have been discussed extensively and showed a wide range of results (Bi et al., 2008; Febriani

et al., 2010; Lal et al., 2013; Louis et al., 2005; Patrick et al., 2004) potentially dependent on the geographical location.

Chapter 4 described a detailed picture of the *C. jejuni* population structure in New Zealand and the existence of STs belonging to the 'specialist' or 'generalist' clusters. However, as this study was based on the MLST genes, the separation between isolates is limited and it is possible that using higher discriminations of genotyping, such as ribosomal MLST (rMLST) or whole genome sequencing (WGS), would break up some of the generalists clusters and identify more host-associated STs. The higher discrimination of rMLST or WGS may also enable the identification of markers associated with specific hosts, genes associated with specific niches or virulence genes (Sheppard et al., 2013b). Chapter 4 established an excellent foundation for identifying potential isolates that can be sequenced and examined further using genome wide association studies and feature frequency profiles (Sims et al., 2009) (comparison tool for phylogenetic analysis to identify frequencies of words (genes) across multiple genomes).

In order to compare specific phenotypes amongst different STs, the isolates have to be grown and monitored under constant, repeatable experimental conditions. The phenotypic microarray (PM) system has enabled global phenotyping of bacterial strains and the recording via a CCD camera of the changing phenotypes. Using this system we have discovered some limitations, especially for microaerophilic bacteria such as *Campylobacter* that require the use of special bags to keep the atmosphere constant. Some of the limitations are: the not sufficient insulation of the machine, 'floaters' in front of the camera that affect the reading and jammed bags. All of these limitations in conjunction with the biological difficulties to obtain repeatable results (phase variation) make it difficult to obtain consistent readings for biological replicates and multiple runs are necessary to obtain confidence in the results as they vary with each run.

The isolates used for the phenotypic study were based on a previous screening of different STs for the absence or presence of the *ykg*C insertion (Cookson et al., 2011). A study in *E. coli* suggested that the *ykg*C gene may be associated with the survival in seawater which is the reason for selecting the PM 9 plate (to test the isolates for tolerance to specific osmolytes). No association between the *ykg*C insertion and the PM 1 (carbon sources) or PM 9 plate were observed. We are collaborating with Professor Mike Konkel's laboratory in Washington to create a *ykg*C knock-out. The study can then be repeated focusing on the two isolates (the wild type and the mutant) but potentially across a greater range of PM plates screening for tolerance to different pH and antimicrobials.

The function of the identified type IV secretion system needs biological validation. An experiment as described in Wiesner et al. (2003) or Wang and Taylor (1990) and Gaasbeek et al. (2009) and performed by Sheppard et al. (2014) could be performed to test the isolates ability to transform naturally *in vitro*.

Chapter 6 used data from a dual challenge with two *C. jejuni* isolates in chickens to investigate distinct strain-to-strain variations and short-term evolutionary dynamics *in vivo*. Interestingly,

one of the inoculum strains has been out-competed and we did not identify evidence of recombination events. Jerome et al. (2011) suggested that the genetic variation in the hypervariable regions (contingency loci) in *C. jejuni* plays an important role during the adaptation of the isolate to a new environment. It would be interesting to identify contingency loci in the ST-474 inoculum strain and compare these regions to the experimentally obtained isolates. Additionally, Thomas et al. (2014) showed that two SNPs that were acquired during the passage of NCTC 11168-GSv (a variant of the wild-type NCTC 11168) through a human were reverted when an isolate from the infected human was passaged through mice. All the observed SNPs in this study were identified in genes related to chemotaxis, motility and cell shape. It would be interesting to see if the genetic variation that potentially resulted from the passage through the chickens is stable or will be reverted during the passage in another animal (such as mouse, ferret, wax-moth caterpillar).

To date, there are only a small number of direct estimates of mutation rates from experimental data, leading to large uncertainty around estimates of coalescent times to common ancestors in *Campylobacter* genomes. This study could be used for estimating the molecular clock based on *in vivo* experimental data.

A better understanding of the ecology and genetic epidemiology of *C. jejuni* will be advantageous in attempts to design interventions to reduce the disease burden. This thesis has furthered our understanding of the seasonality of human campylobacteriosis in New Zealand, the existing population structure of *C. jejuni*, its biochemical requirements and tolerance to osmolytes and novel insights into short-term evolutionary dynamics *in vivo*. Based on these findings and the recommendations for future directions, our enhanced knowledge could lead to a greater understanding of host-association and new intervention strategies.

APPENDIX A

Supplementary Material for chapter 5

A.1 Genome characteristics for the 15 isolates

isolate	bases	sig_peptides	CDS	rRNA	tRNA	GC content	no of contigs
P694a	1826042	119	1894	9	42	30.4	20
P110b	1656194	110	1668	9	42	30.3	28
S168b*	1762522	127	1836	4	36	30.1	17
h73020	1668906	108	1764	3	33	30.0	40
S330a*	1649981	109	1672	6	36	30.2	18
H22082	1659123	111	1677	9	43	30.3	28
S22b	1647149	113	1696	6	37	29.9	53
m28548	1698828	111	1761	6	43	30.1	37
S263b*	1661374	110	1719	3	36	30.3	21
H550*	1627085	110	1663	3	36	30.3	22
M602b*	1672453	114	1724	3	35	30.2	21
H180*	1634383	112	1669	5	38	30.2	25
S355b*	1672095	111	1722	6	38	30.3	17
S276b*	1700503	122	1783	4	36	30.1	10
H450b*	1615775	110	1670	6	35	30.3	11

Table A.1: Summary of specific genome characteristics of the 15 isolates used in this study. The isolates with a * were newly sequenced for this study, the other genomes were retrieved from the database.

A.2 Respiration curves for PM 1 and PM 9 at 38°C and 42°C

A.2.1 Isolate m28548 (ST-2026) in PM 1 at 38°C and 42°C

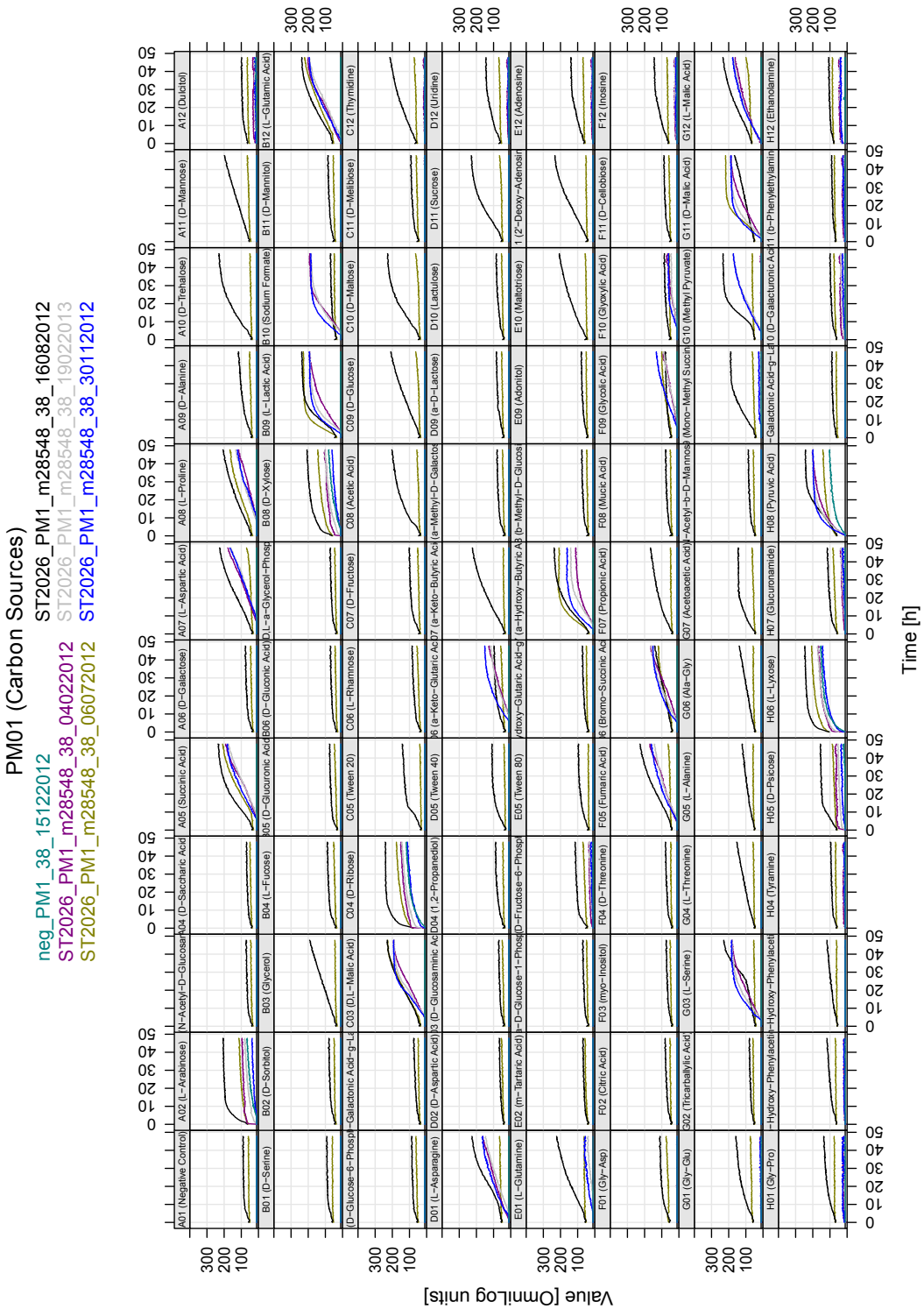


Figure A.1: xy plot for the *C. jejuni* isolate m28548 in PM 1 at 38°C

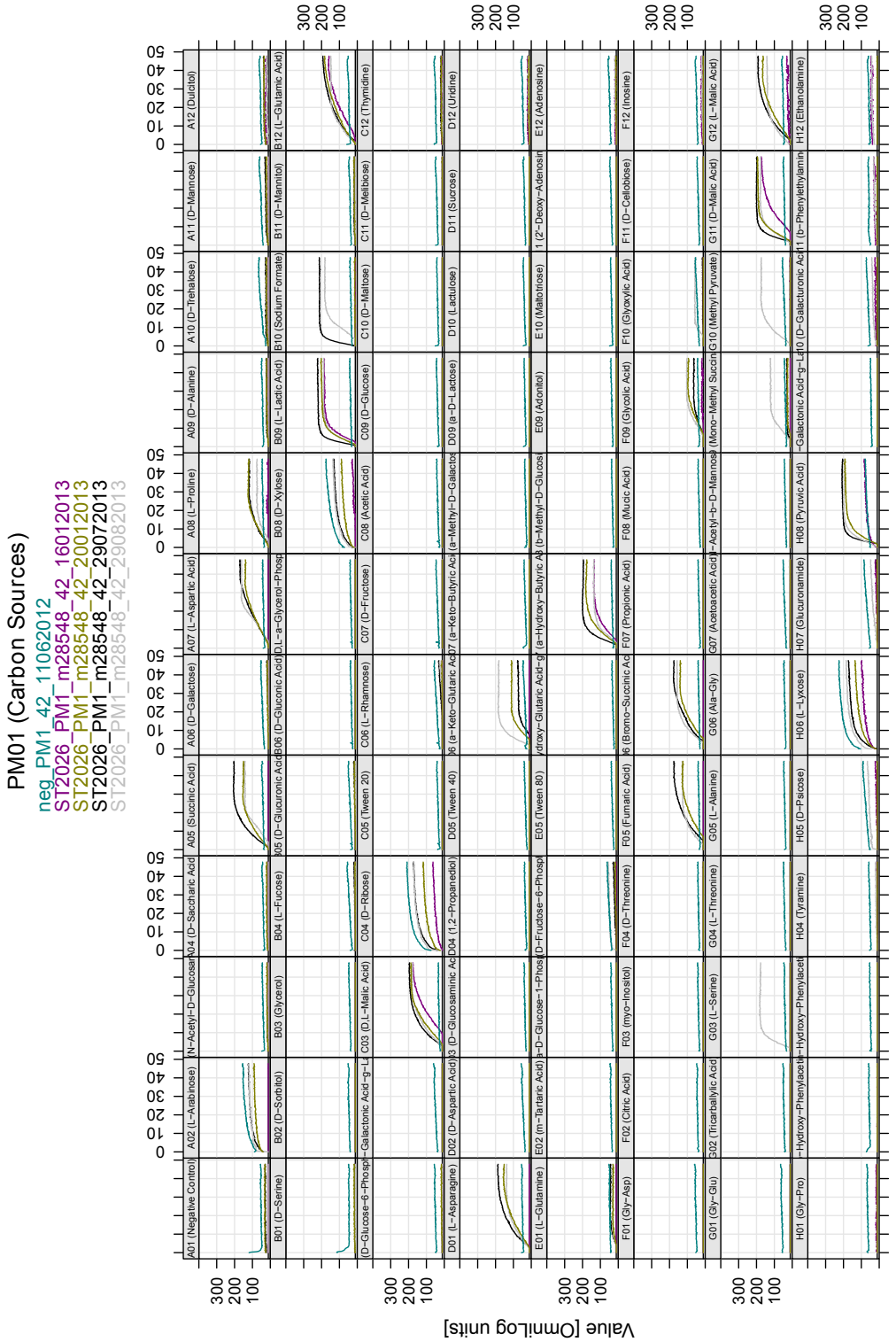


Figure A.2: xy plot for the *C. jejuni* isolate m28548 in PM 1 at 42°C

A.2.2 Isolate m28548 (ST-2026) in PM 9 at 38°C and 42°C

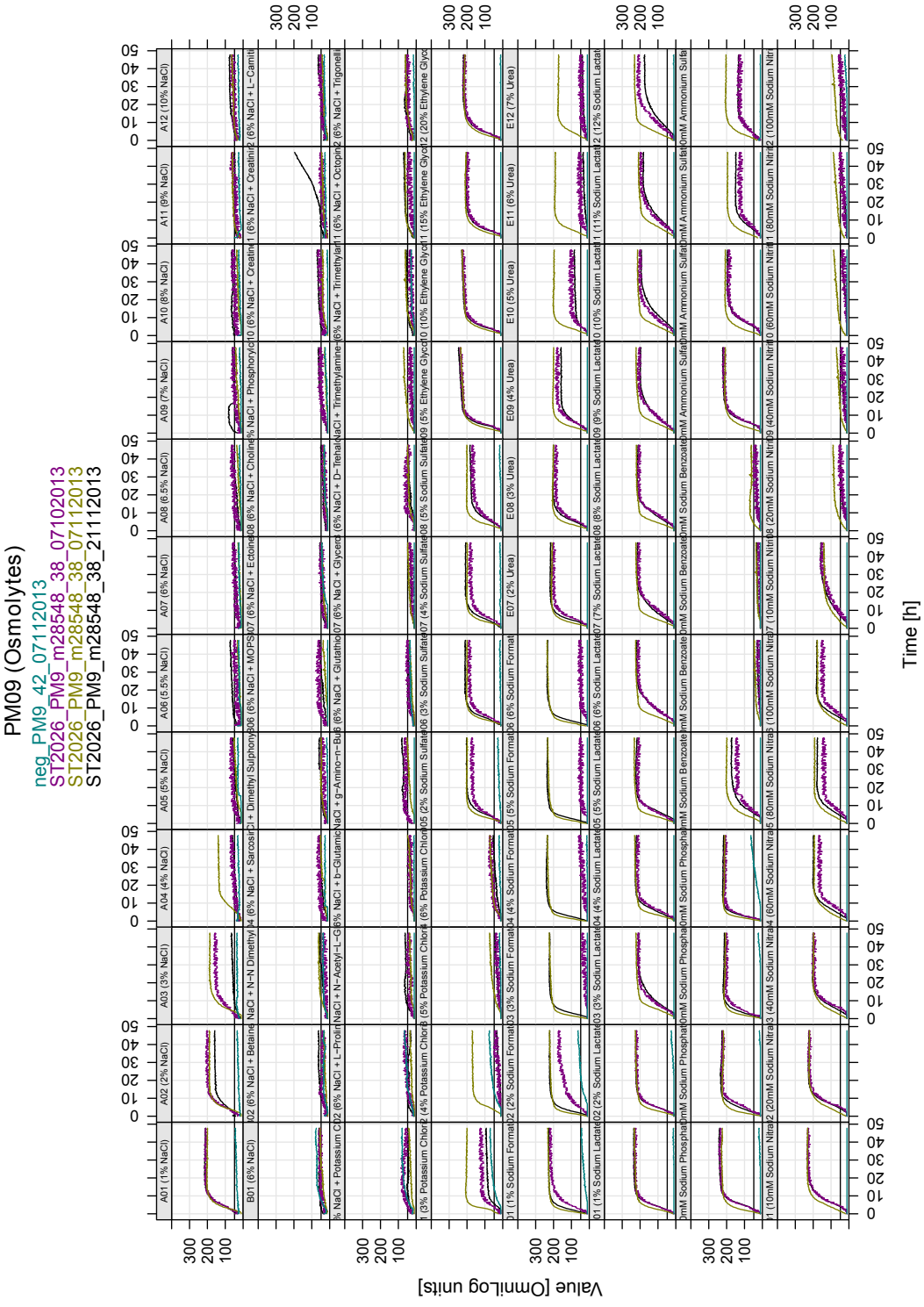


Figure A.3: xy plot for the *C. jejuni* isolate m28548 in PM 9 at 38°C

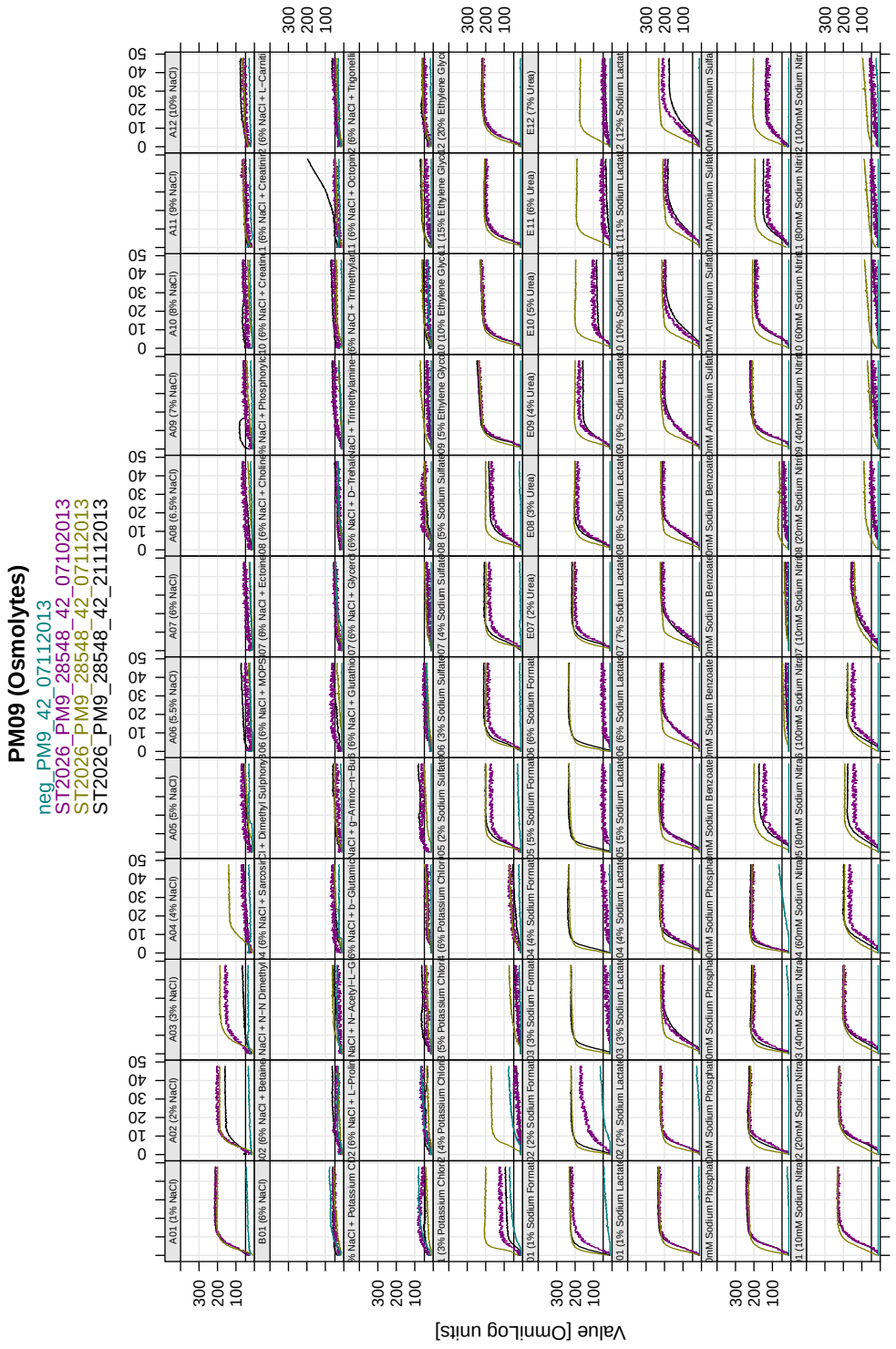


Figure A.4: xy plot for the *C. jejuni* isolate m28548 in PM 9 at 42°C

A.2.3 Isolate h73020 (ST-474) in PM 1 at 38°C and 42°C

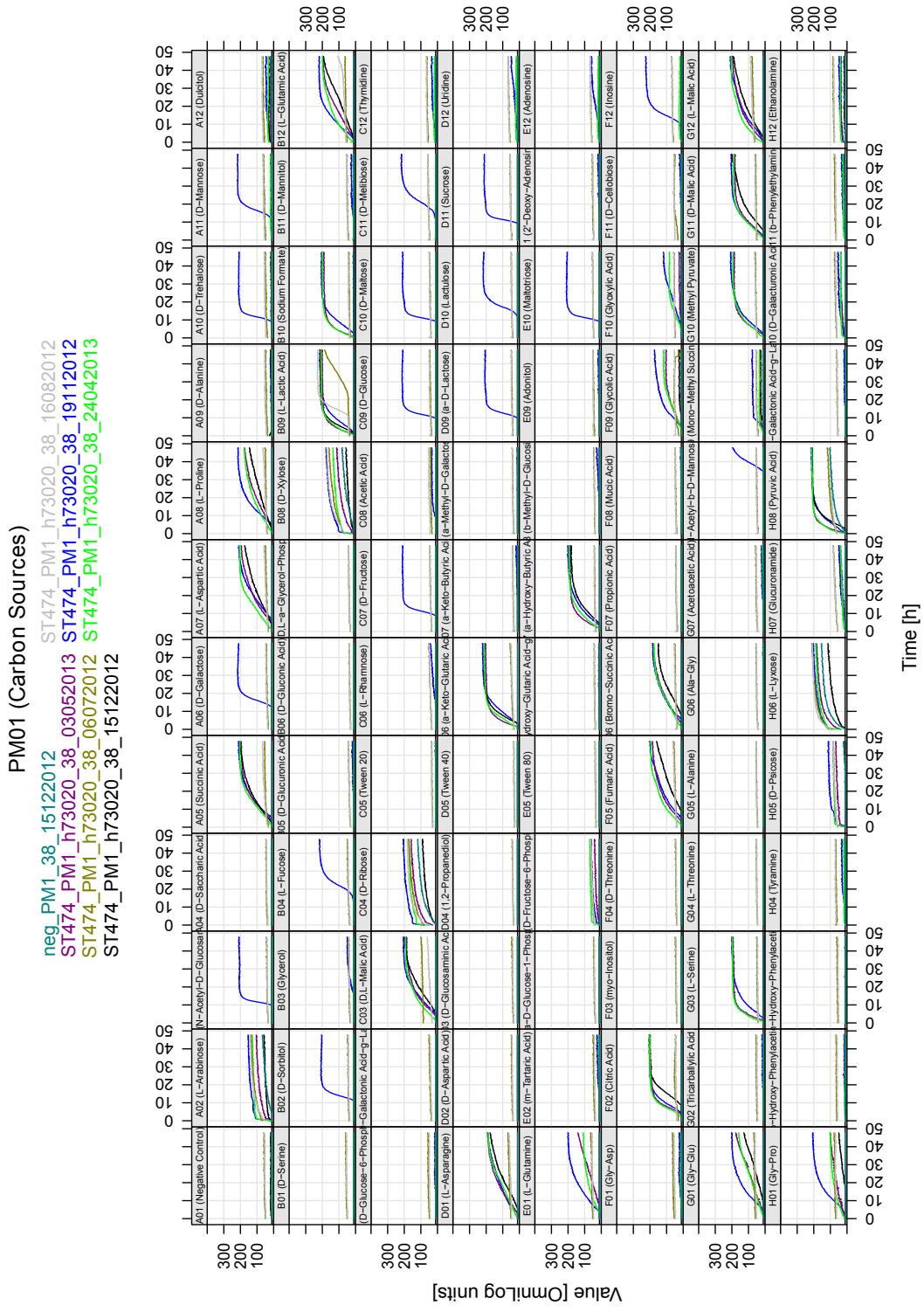


Figure A.5: xy plot for the *C. jejuni* isolate h73020 in PM 1 at 38°C

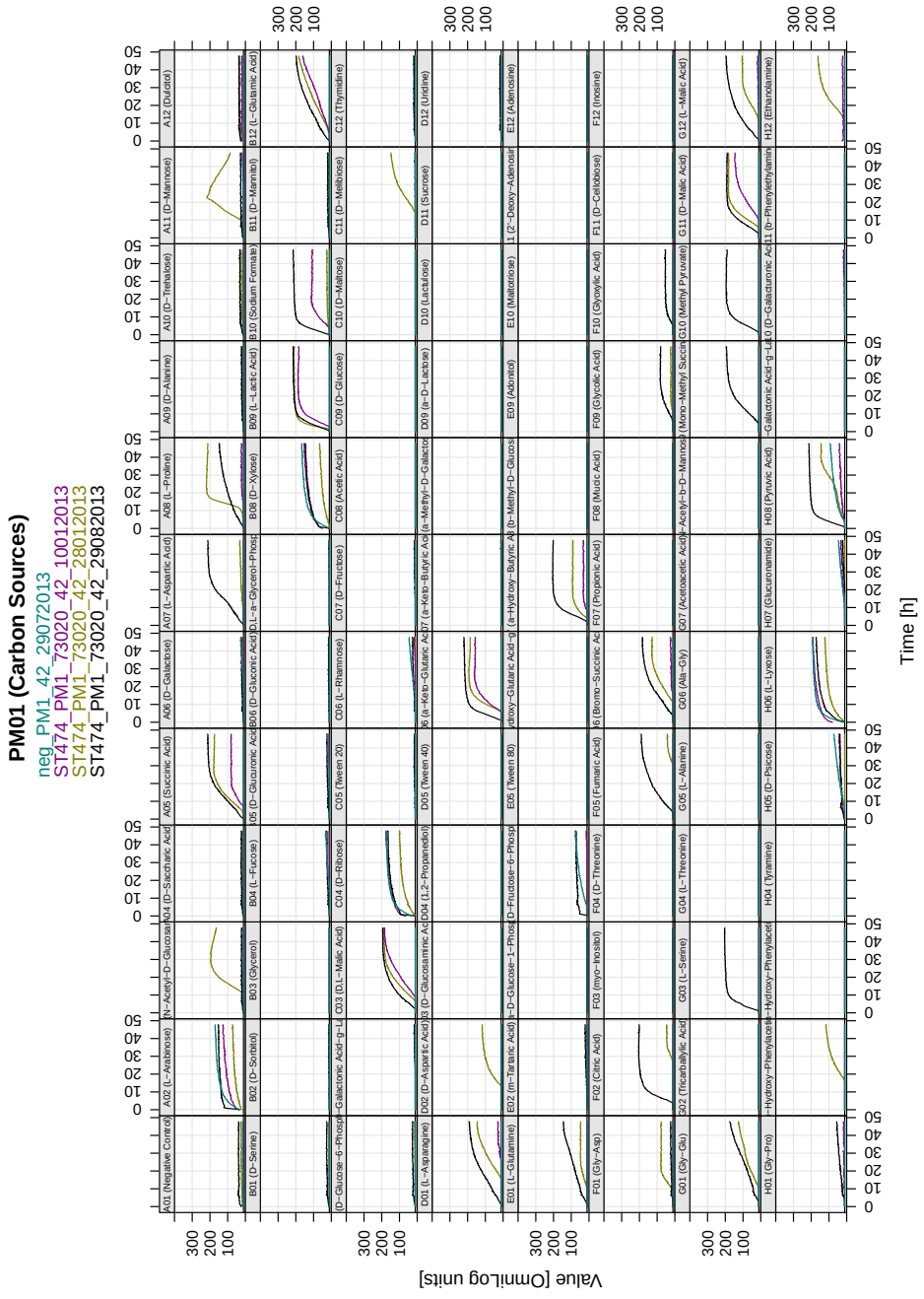


Figure A.6: xy plot for the *C. jejuni* isolate h73020 in PM 1 at 42°C

A.2.4 Isolate h73020 (ST-474) in PM 9 at 38°C and 42°C

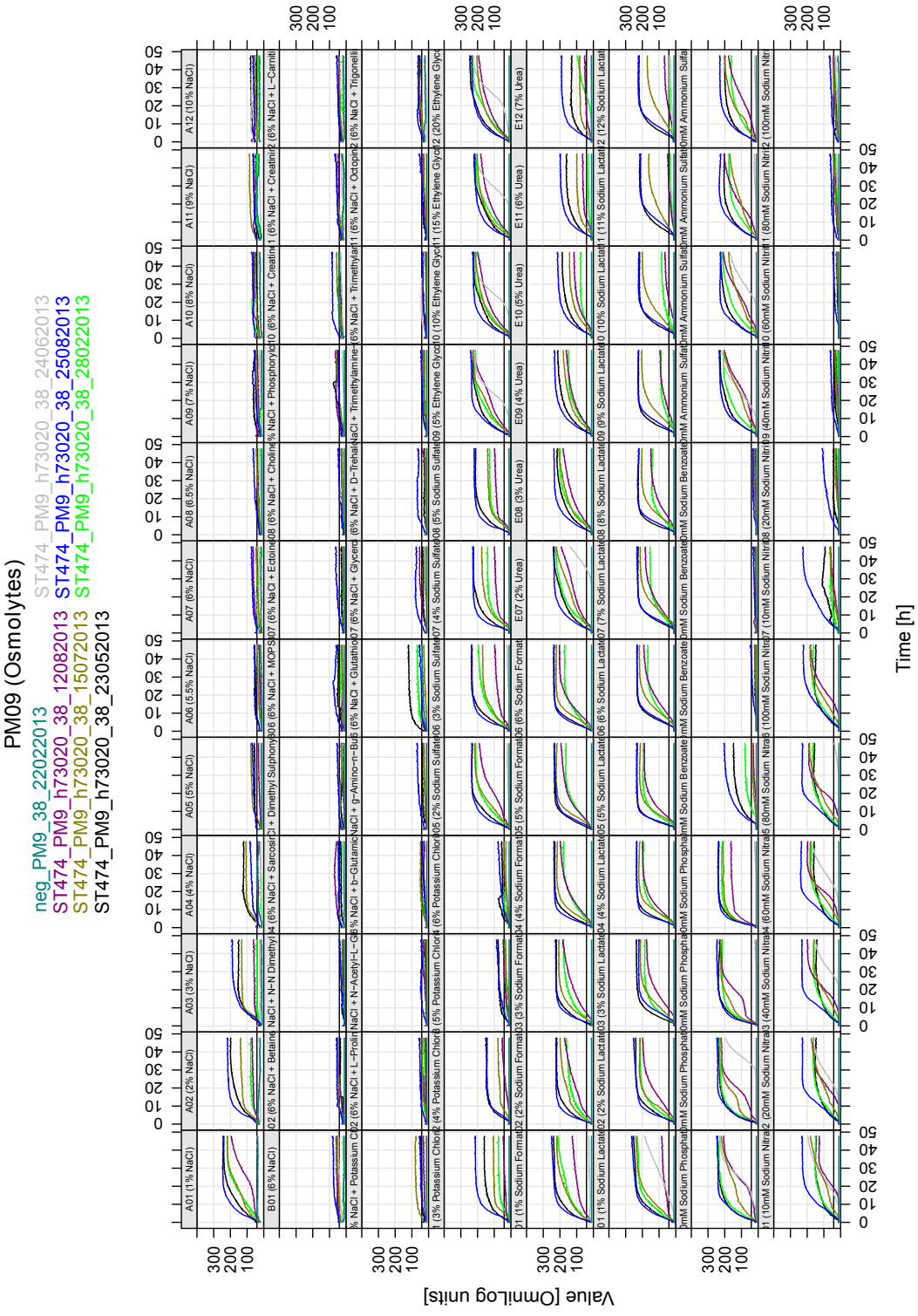


Figure A.7: xy plot for the *C. jejuni* isolate h73020 in PM 9 at 38°C

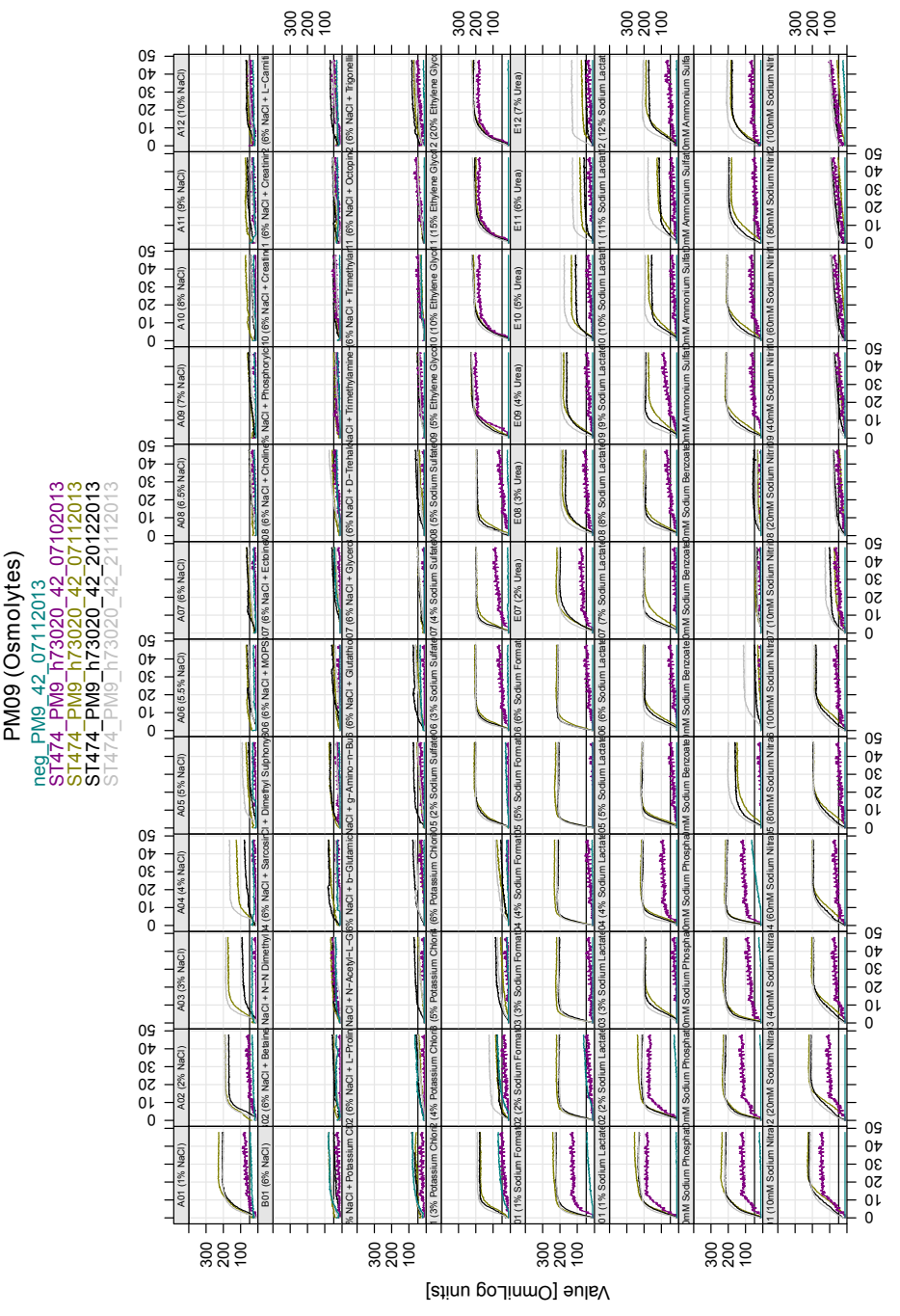


Figure A.8: xy plot for the *C. jejuni* isolate h73020 in PM 9 at 42°C

A.2.5 Isolate S355b (ST-42) in PM 1 at 38°C and 42°C

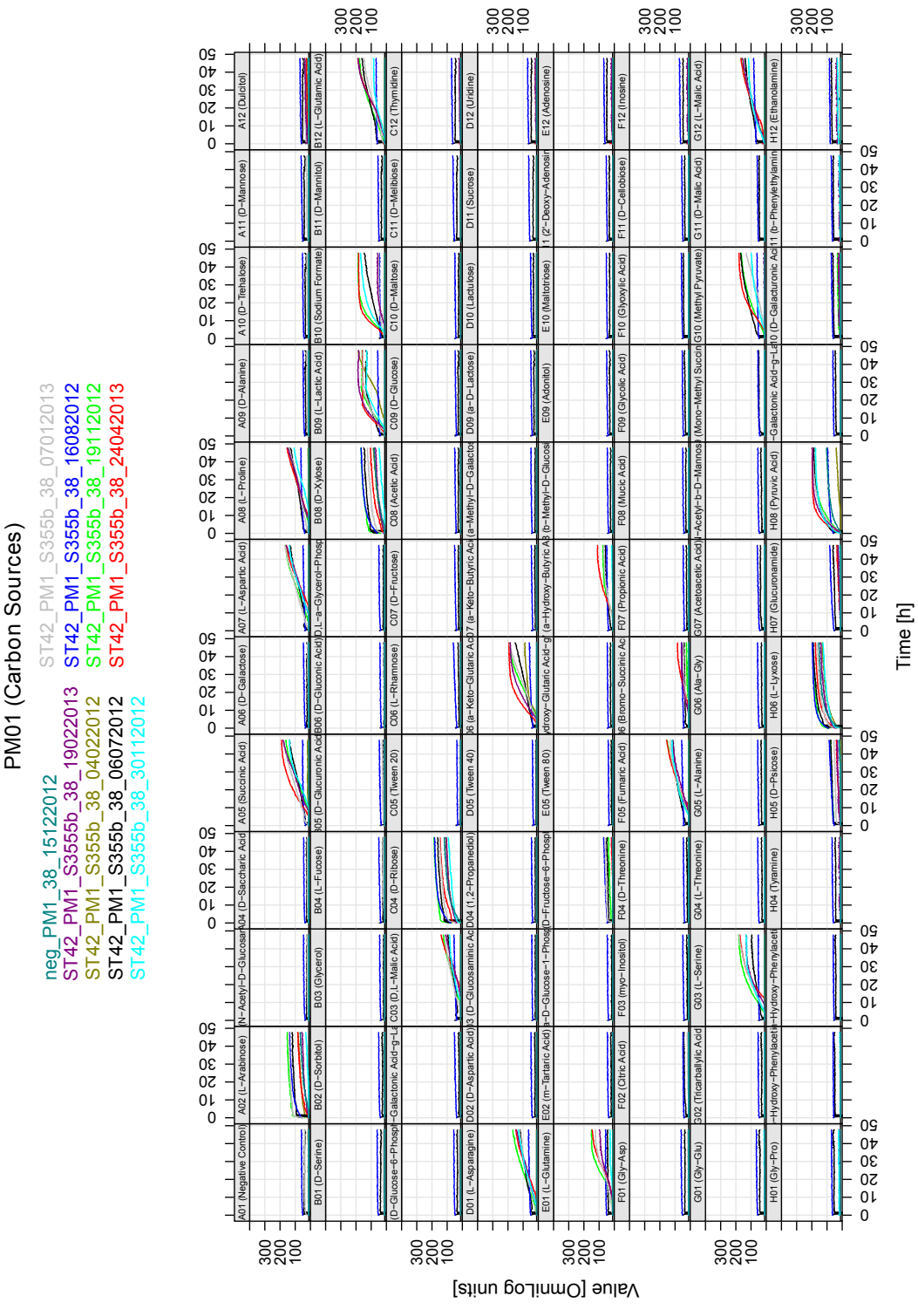


Figure A.9: xy plot for the *C. jejuni* isolate S355b in PM 1 at 38°C

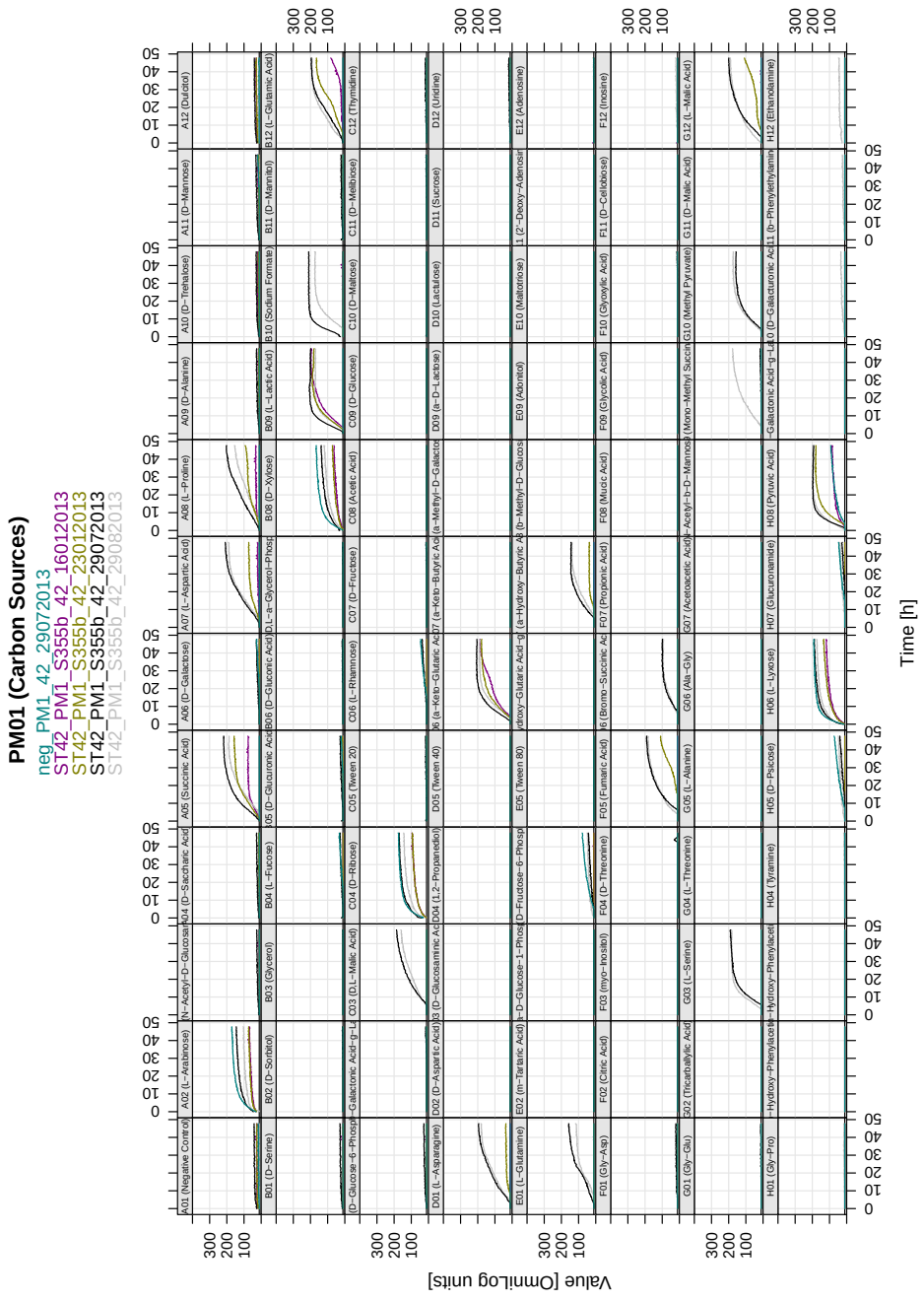


Figure A.10: xy plot for the *C. jejuni* isolate S355b in PM 1 at 42°C

A.2.6 Isolate S355b (ST-42) in PM 9 at 38°C and 42°C

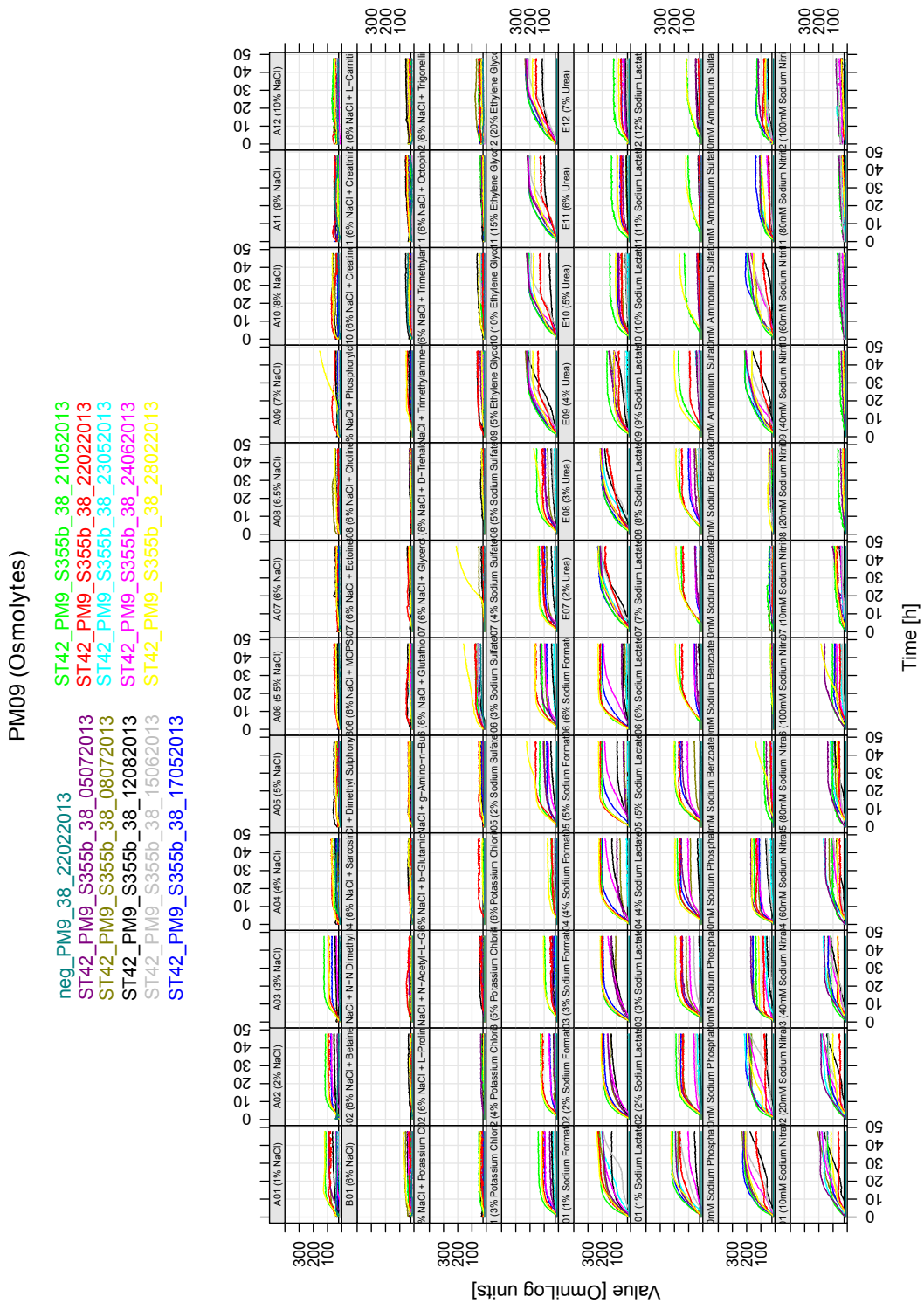


Figure A.11: xy plot for the *C. jejuni* isolate S355b in PM 9 at 38°C

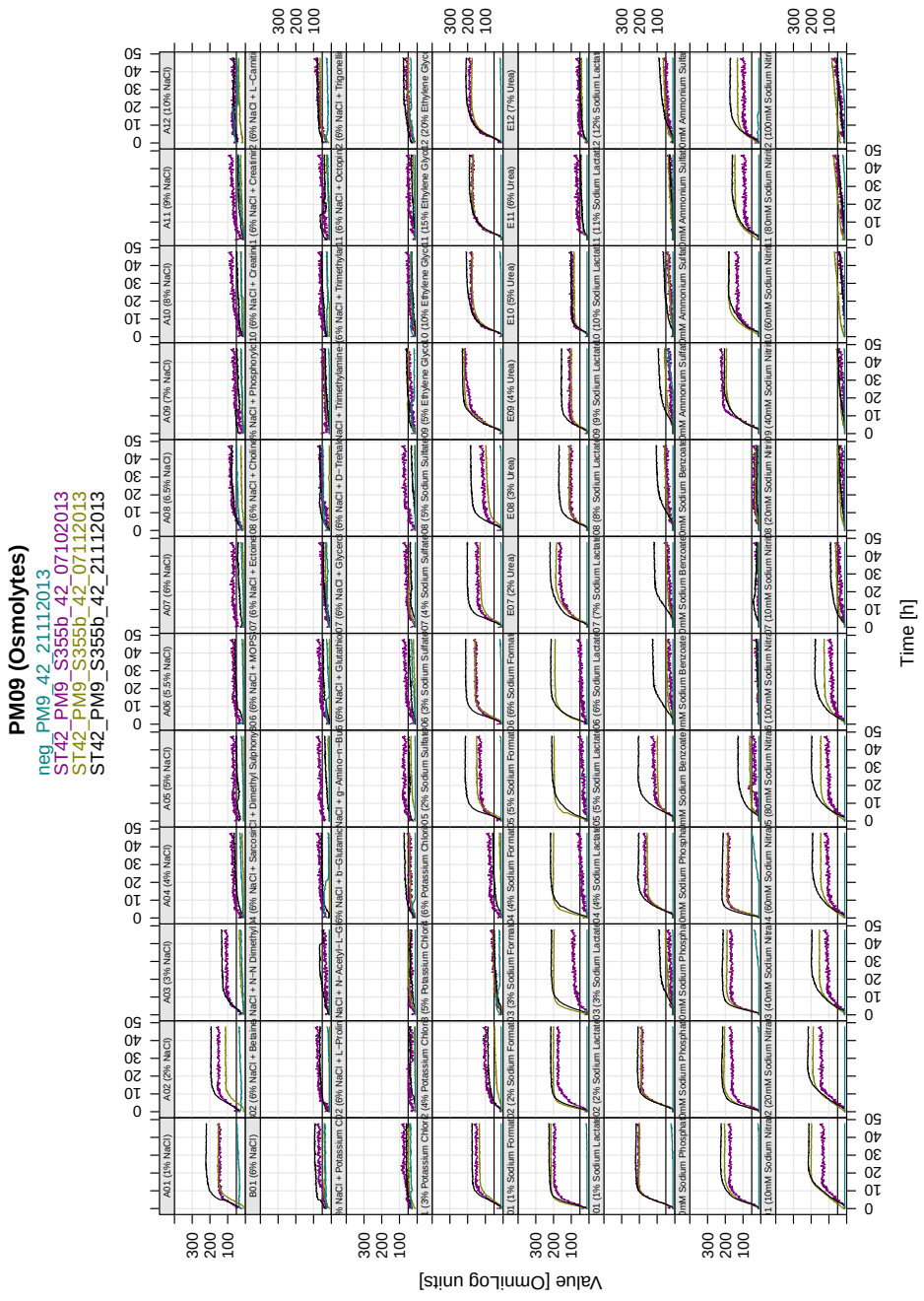


Figure A.12: xy plot for the *C. jejuni* isolate S355b in PM 9 at 42°C

A.2.7 Isolate S330a (ST-474) in PM 1 at 38°C and 42°C

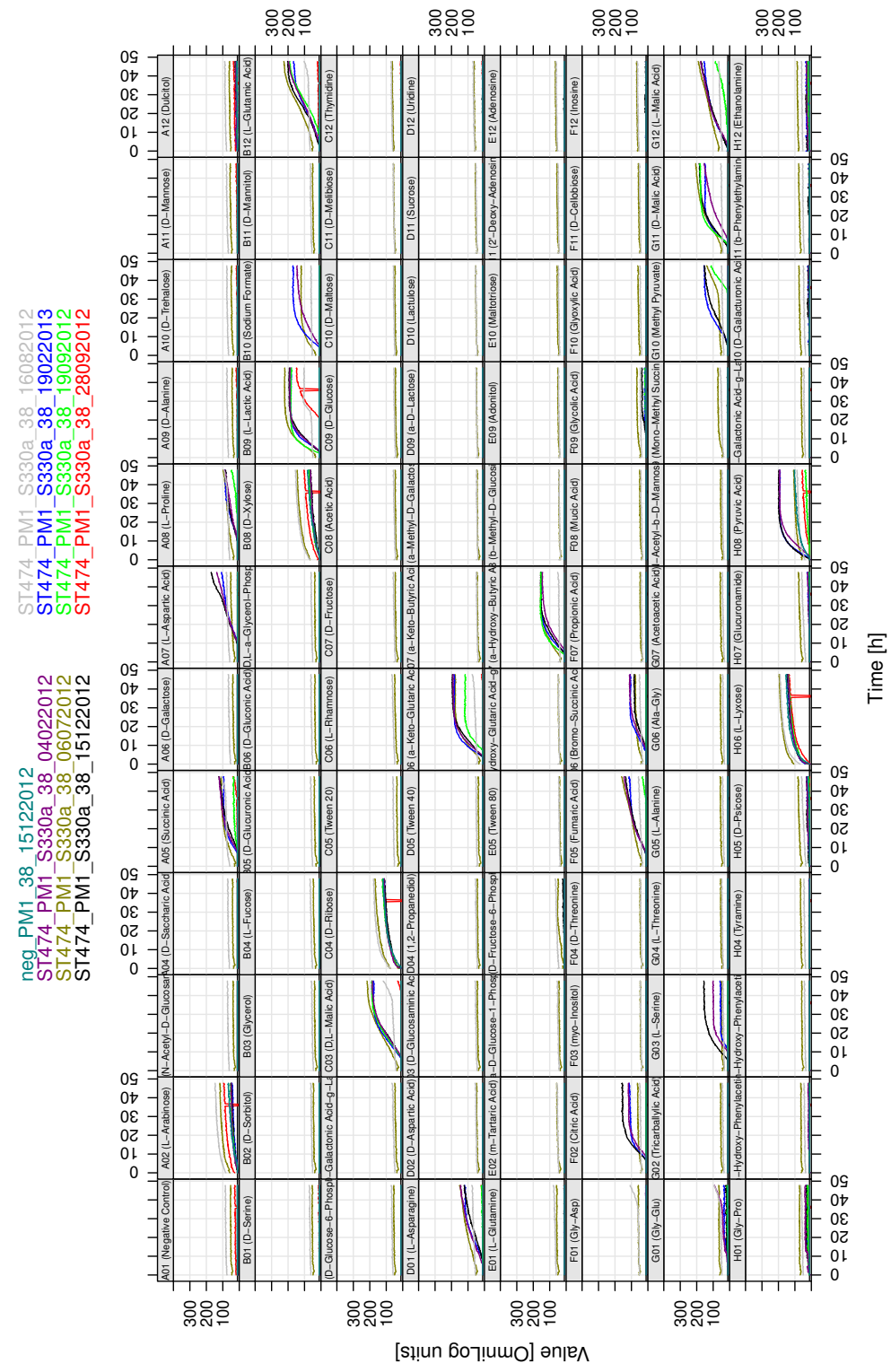


Figure A.13: xy plot for the *C. jejuni* isolate S330a in PM 1 at 38°C

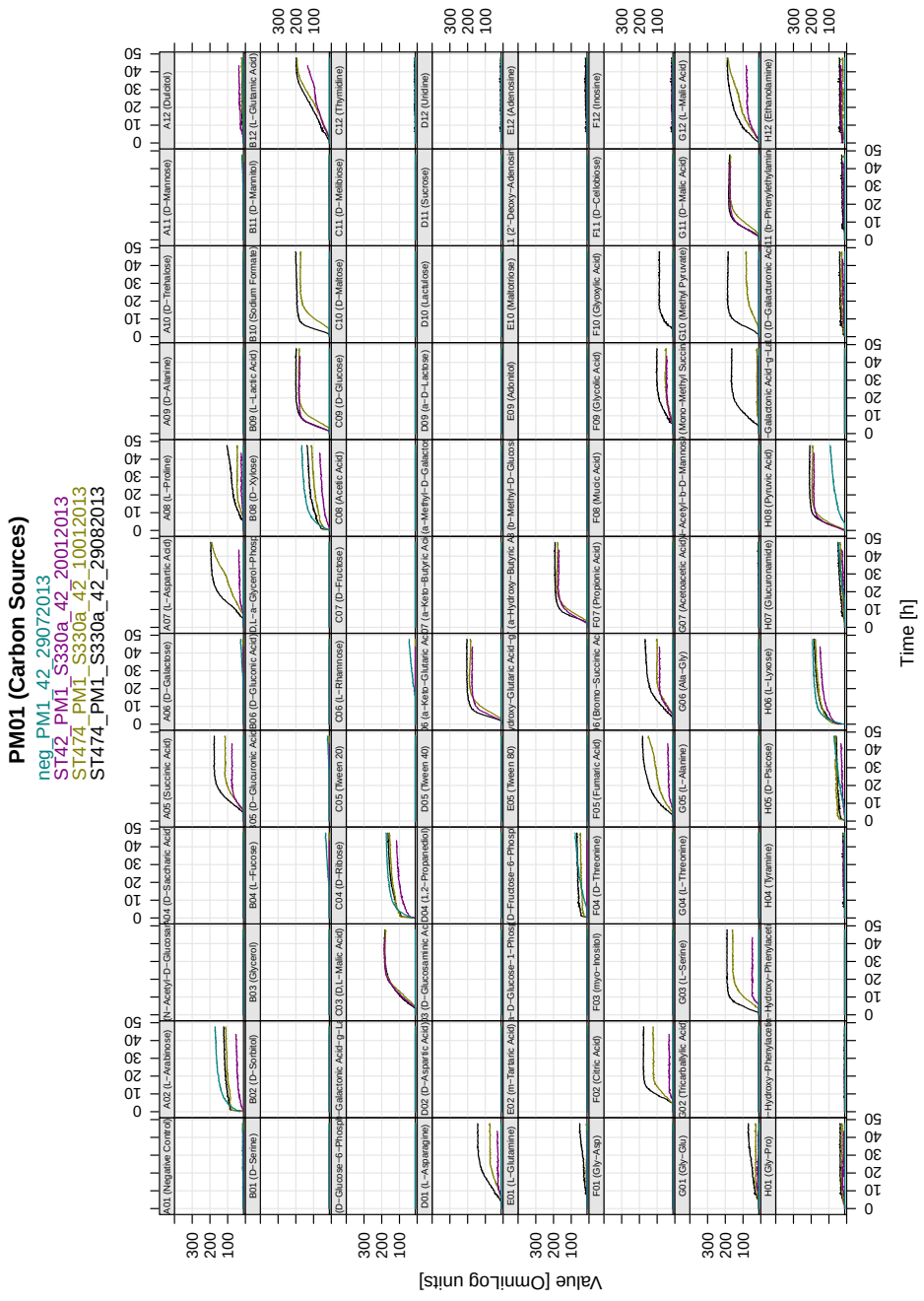


Figure A.14: xy plot for the *C. jejuni* isolate S330a in PM 1 at 42°C

A.2.8 Isolate S330a (ST-474) in PM 9 at 38°C and 42°C

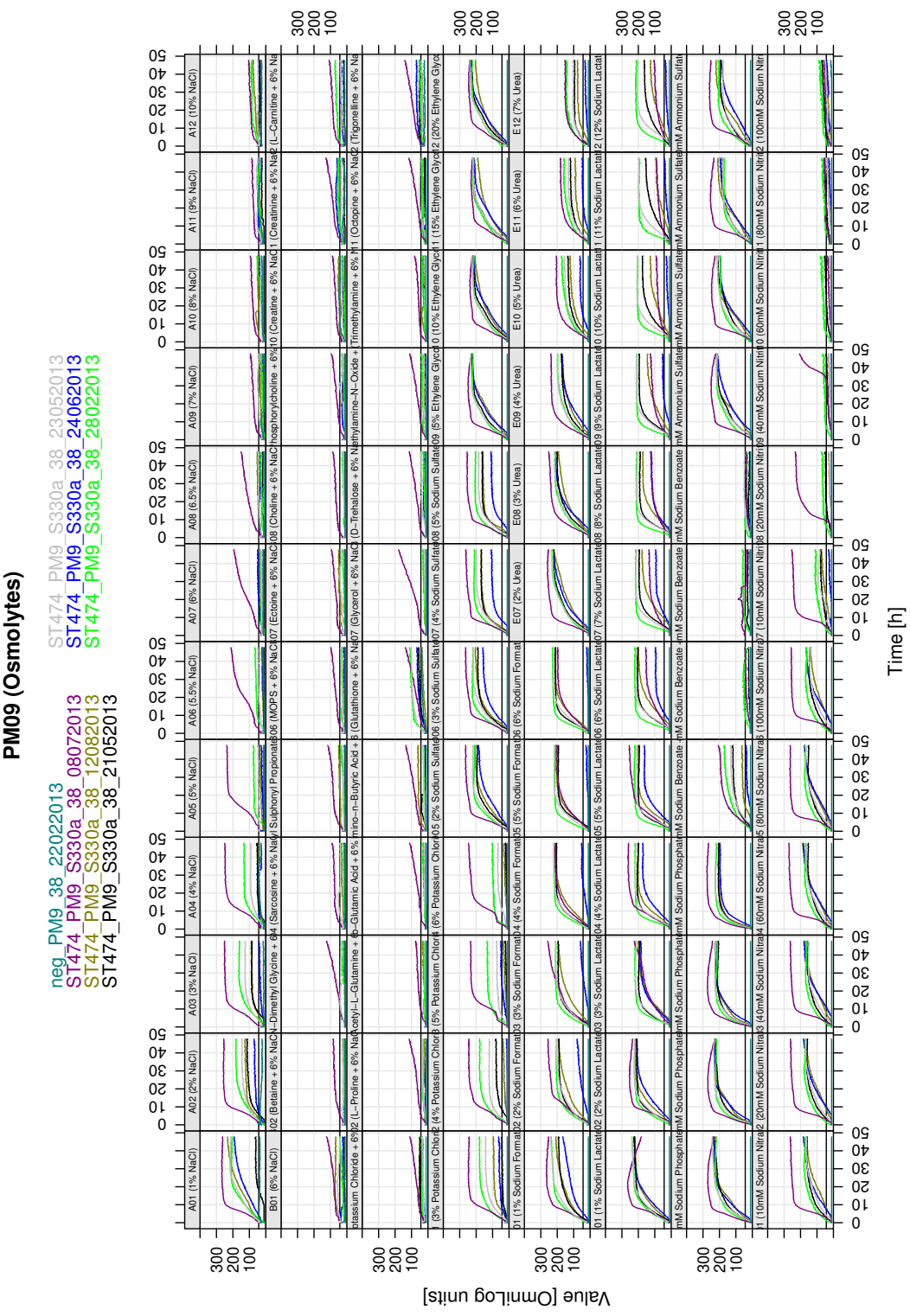


Figure A.15: xy plot for the *C. jejuni* isolate S330a in PM 9 at 38°C

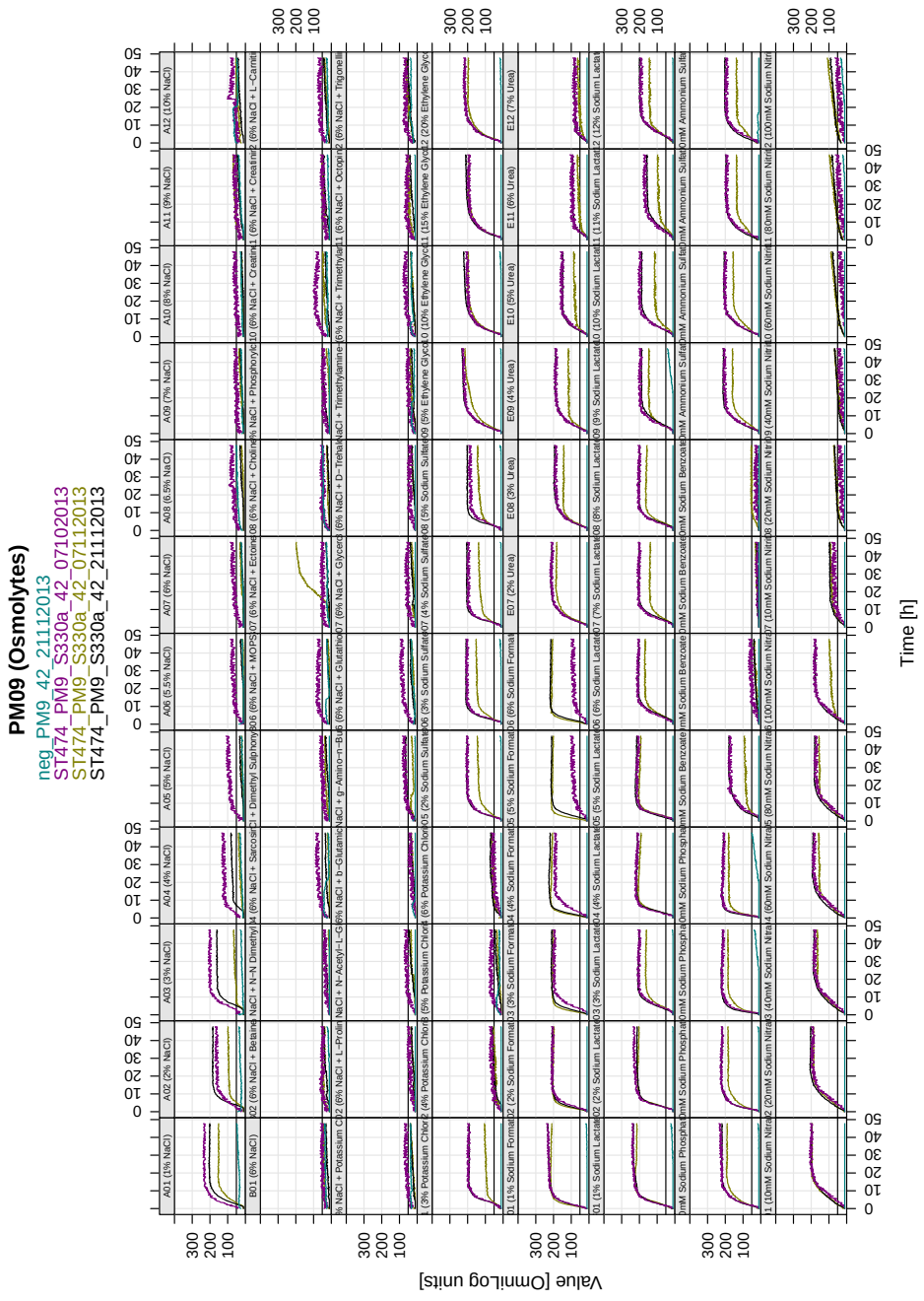


Figure A.16: xy plot for the *C. jejuni* isolate S330a in PM 9 at 42°C

A.2.9 Isolate H22082 (ST-474) in PM 1 at 38°C and 42°C

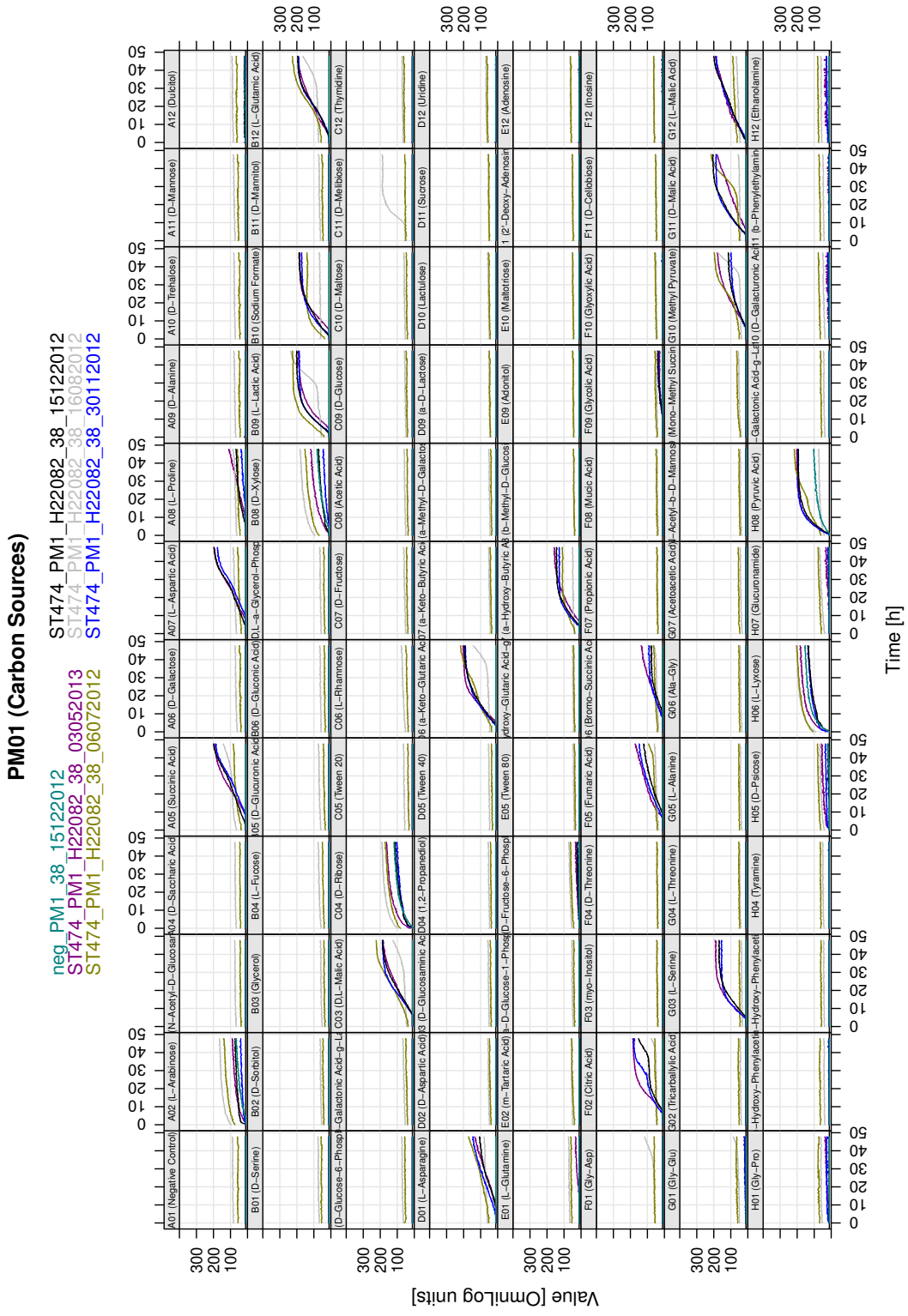


Figure A.17: xy plot for the *C. jejuni* isolate H22082 in PM 1 at 38°C

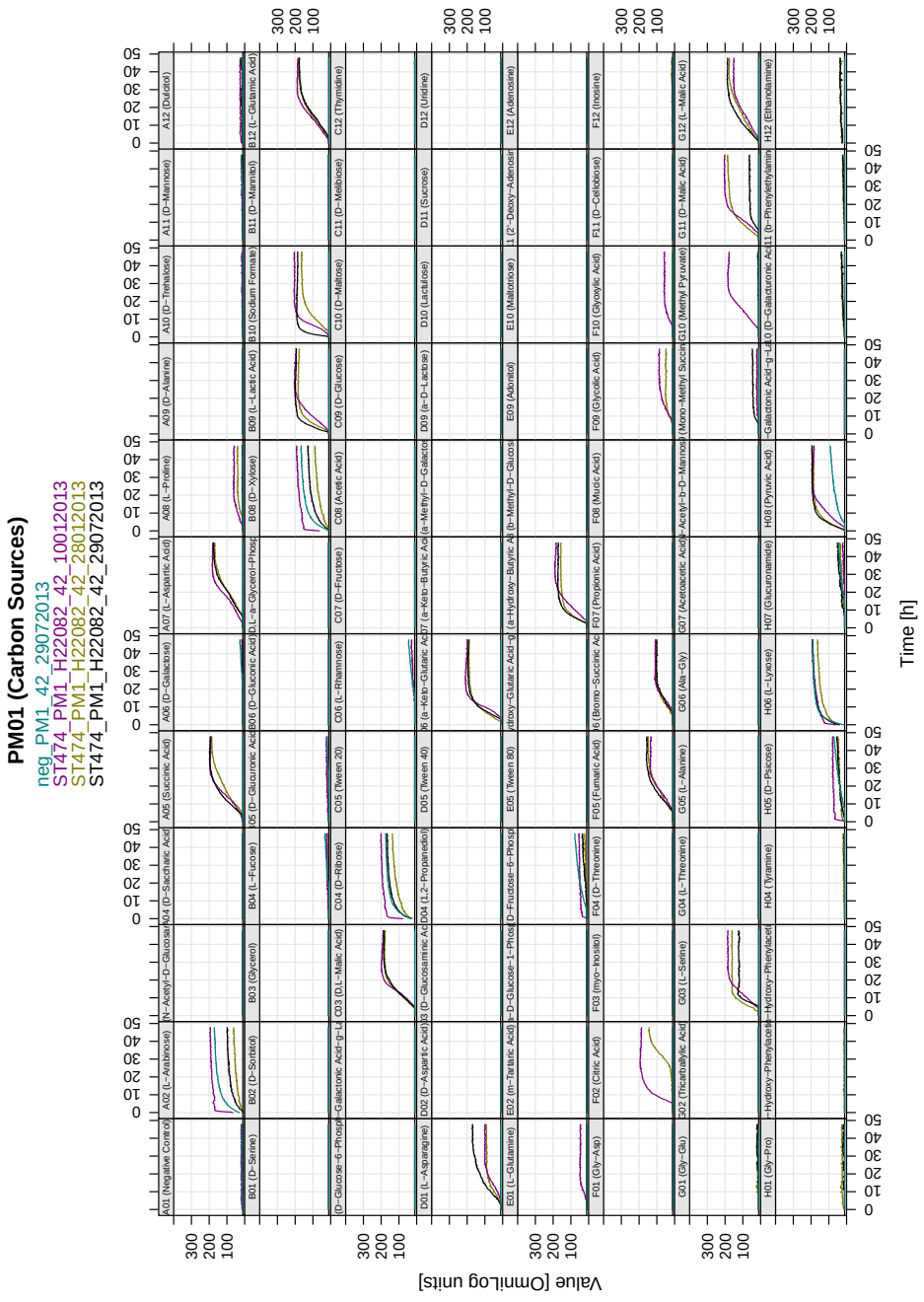


Figure A.18: xy plot for the *C. jejuni* isolate H22082 in PM 1 at 42°C

A.2.10 Isolate H22082 (ST-474) in PM 9 at 38°C and 42°C

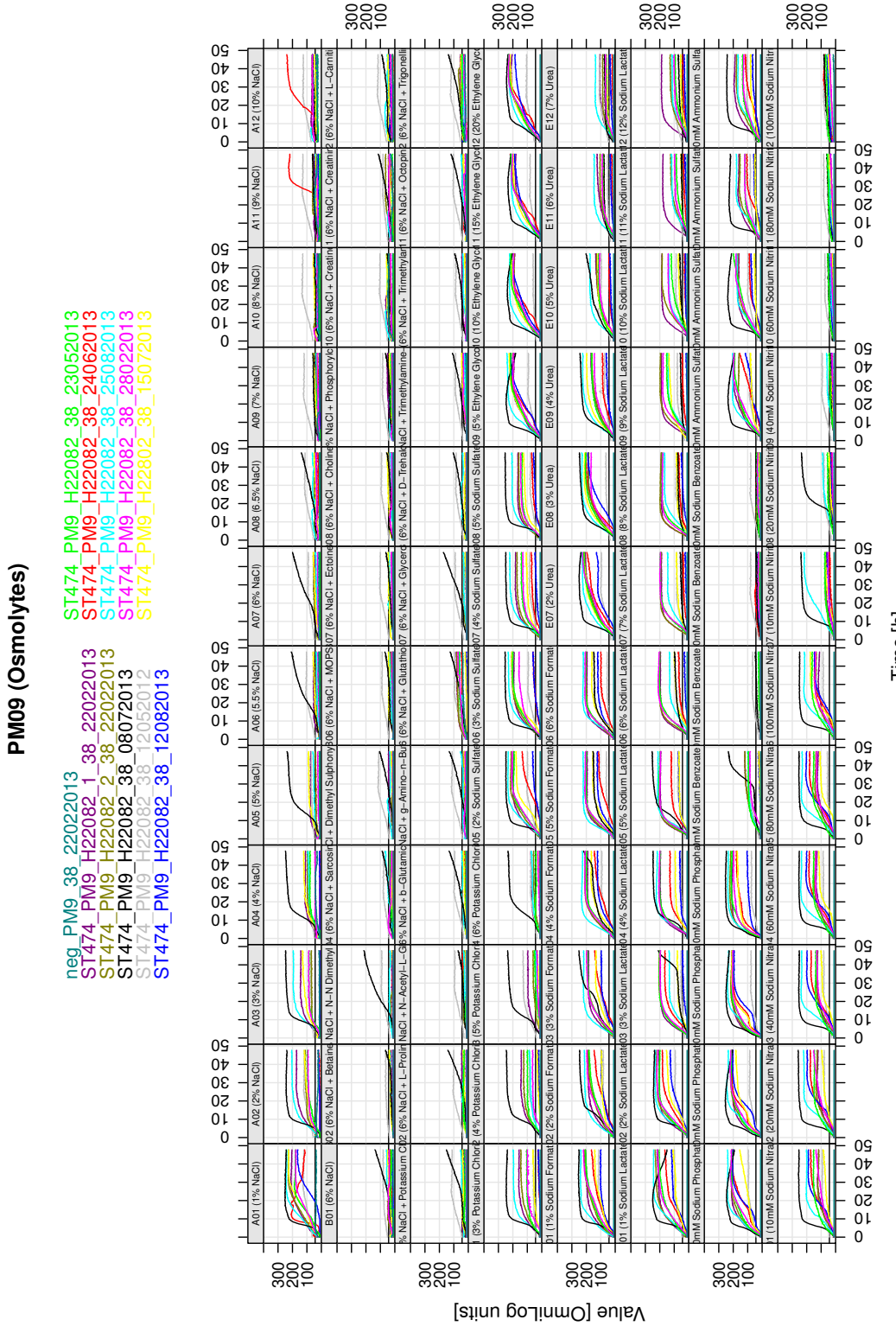


Figure A.19: xy plot for the *C. jejuni* isolate H22082 in PM 9 at 38°C

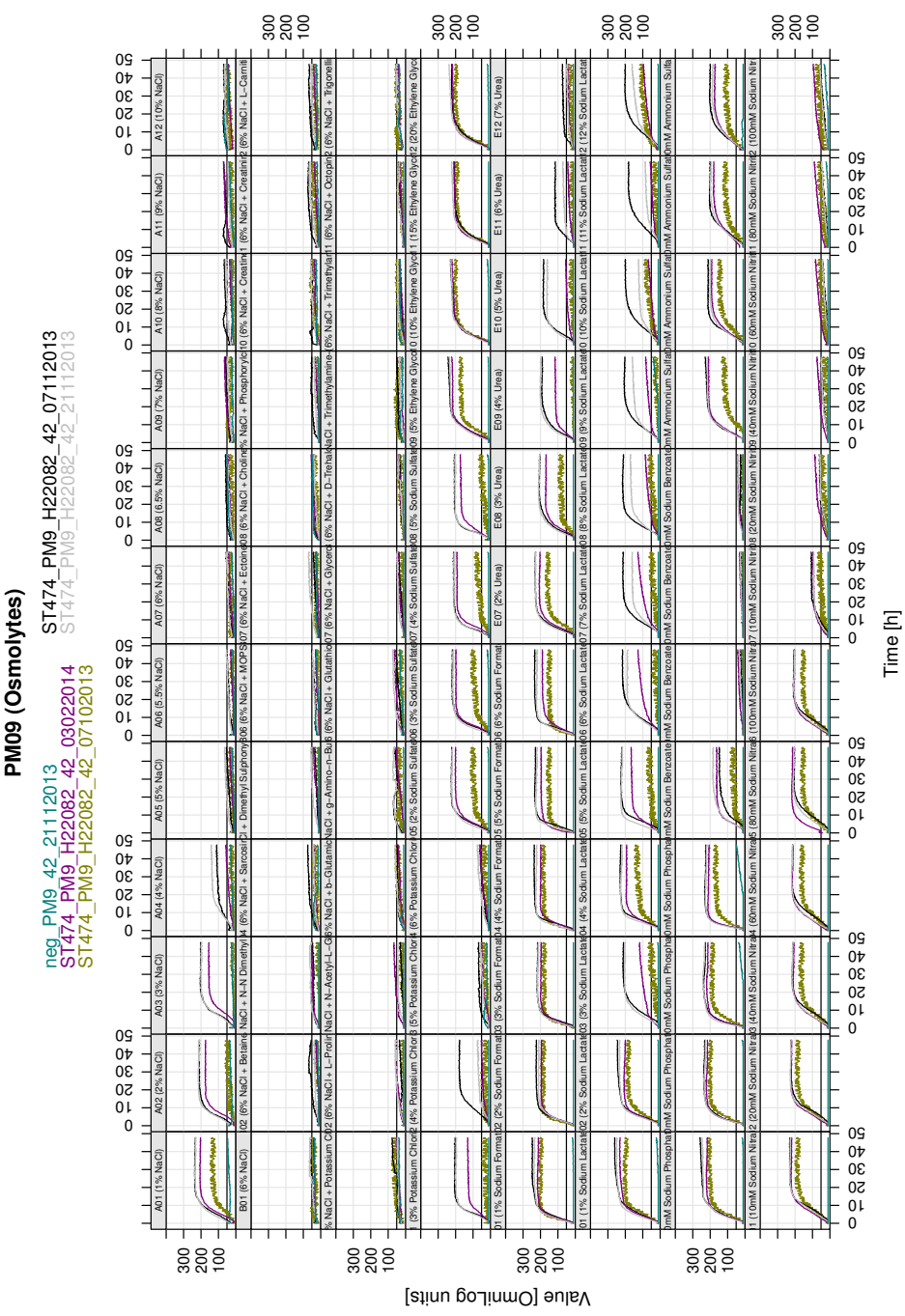


Figure A.20: xy plot for the *C. jejuni* isolate H22082 in PM 9 at 42°C

A.2.11 Isolate H450b (ST-61) in PM 1 at 38°C and 42°C

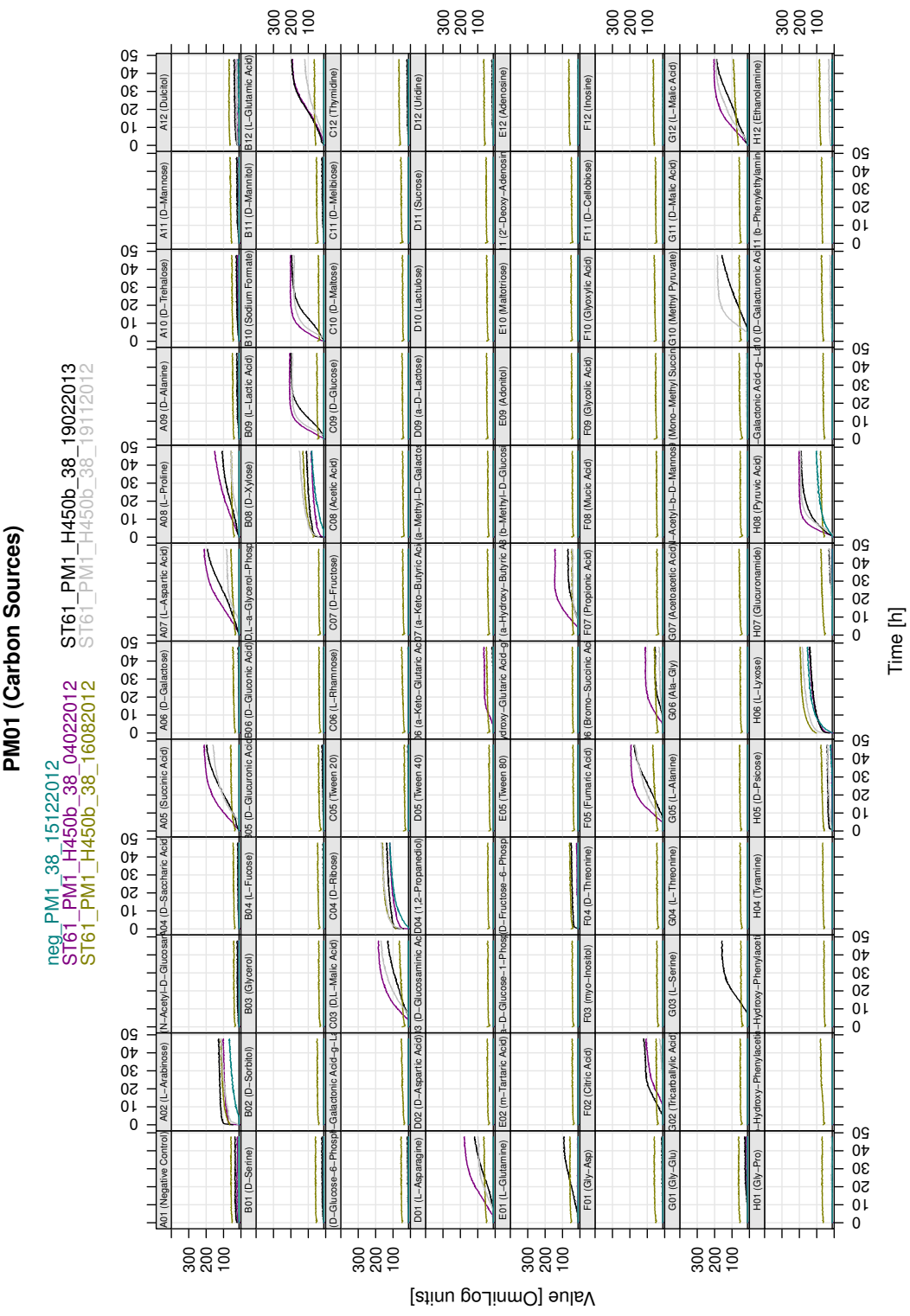


Figure A.21: xy plot for the *C. jejuni* isolate H450b in PM 1 at 38°C

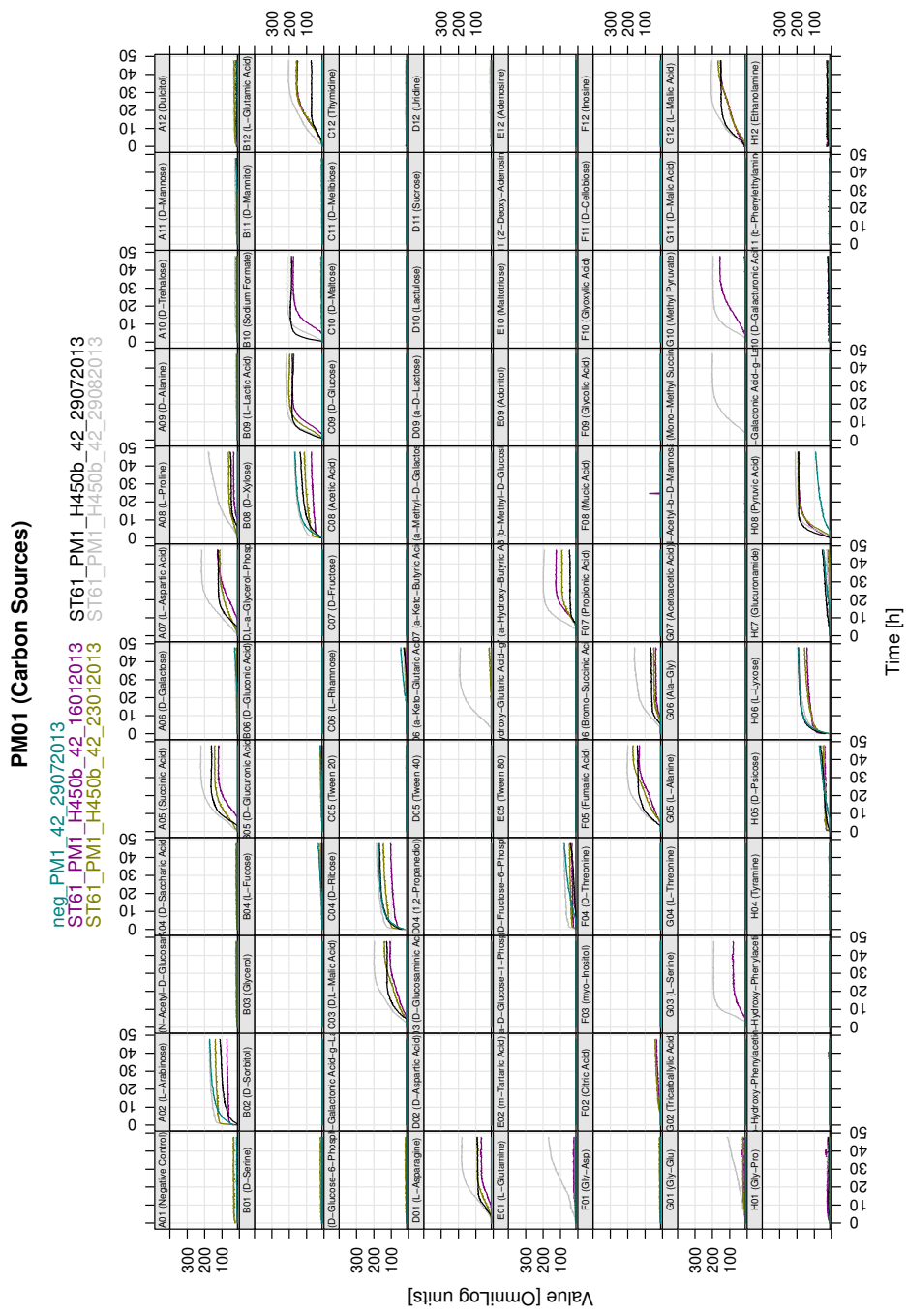


Figure A.22: xy plot for the *C. jejuni* isolate H450b in PM 1 at 42°C

A.2.12 Isolate H450b (ST-61) in PM 9 at 38°C and 42°C

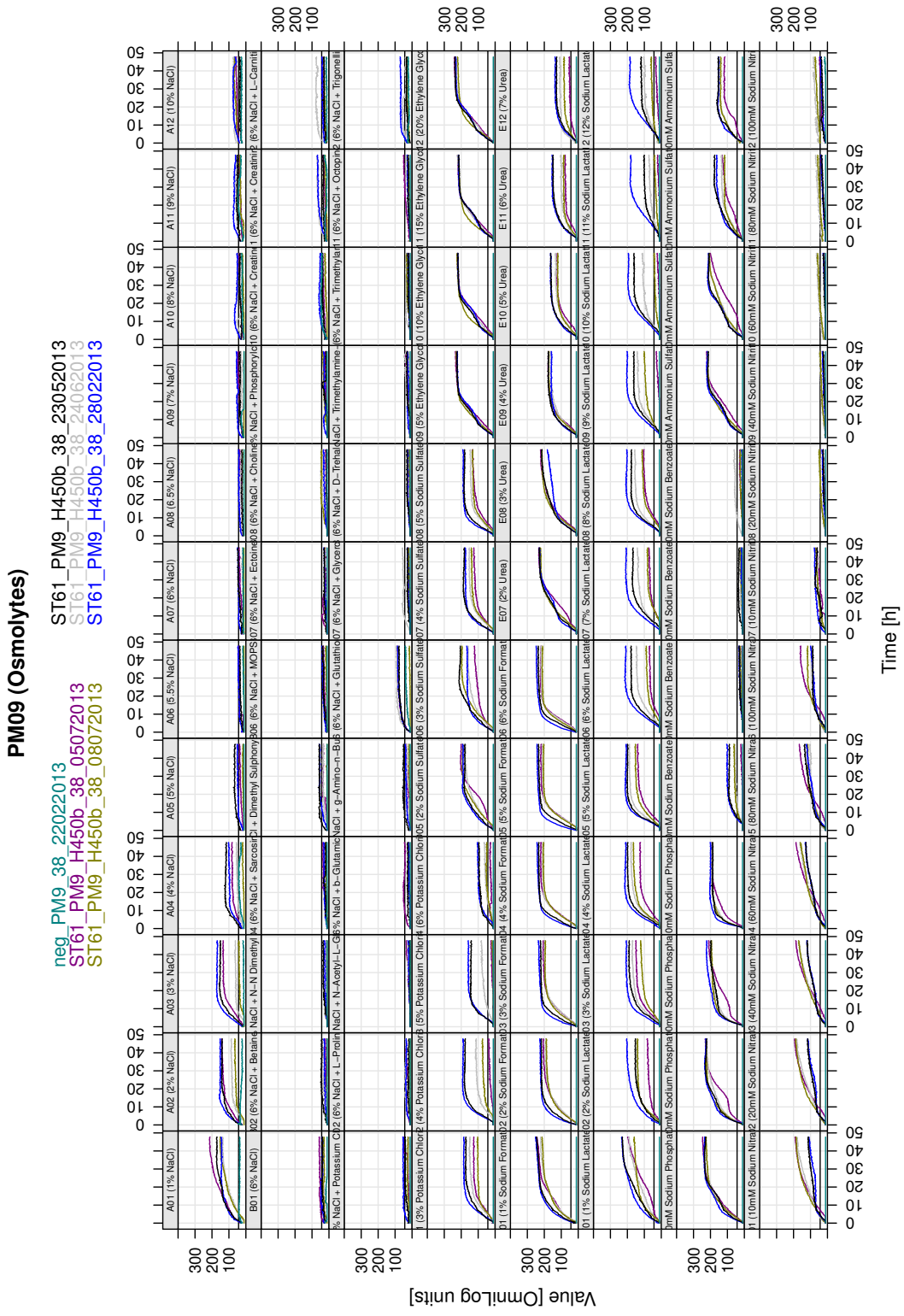


Figure A.23: xy plot for the *C. jejuni* isolate H450b in PM 9 at 38°C

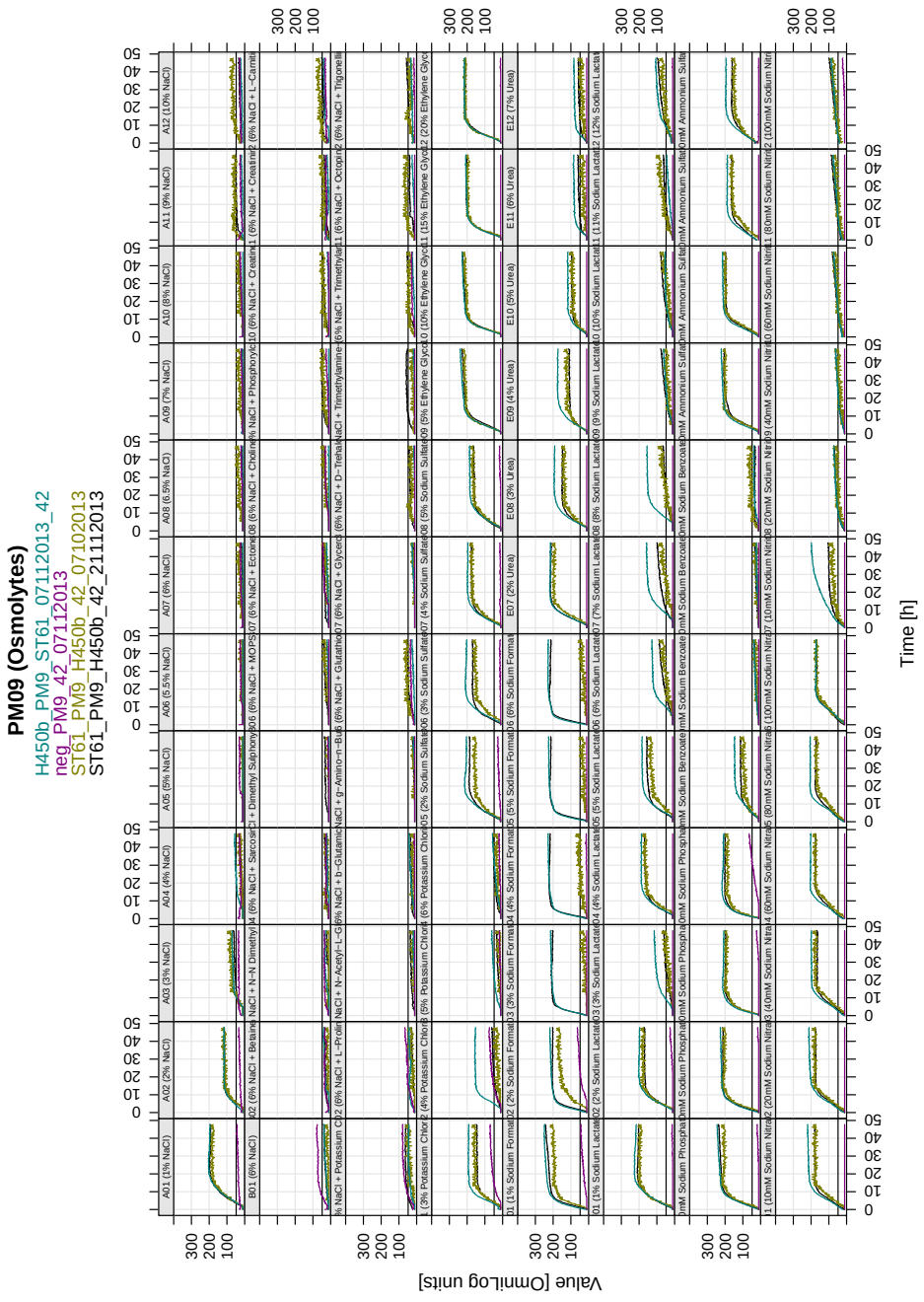


Figure A.24: xy plot for the *C. jejuni* isolate H450b in PM 9 at 42°C

A.2.13 Isolate H550 (ST-42) in PM 1 at 38°C and 42°C

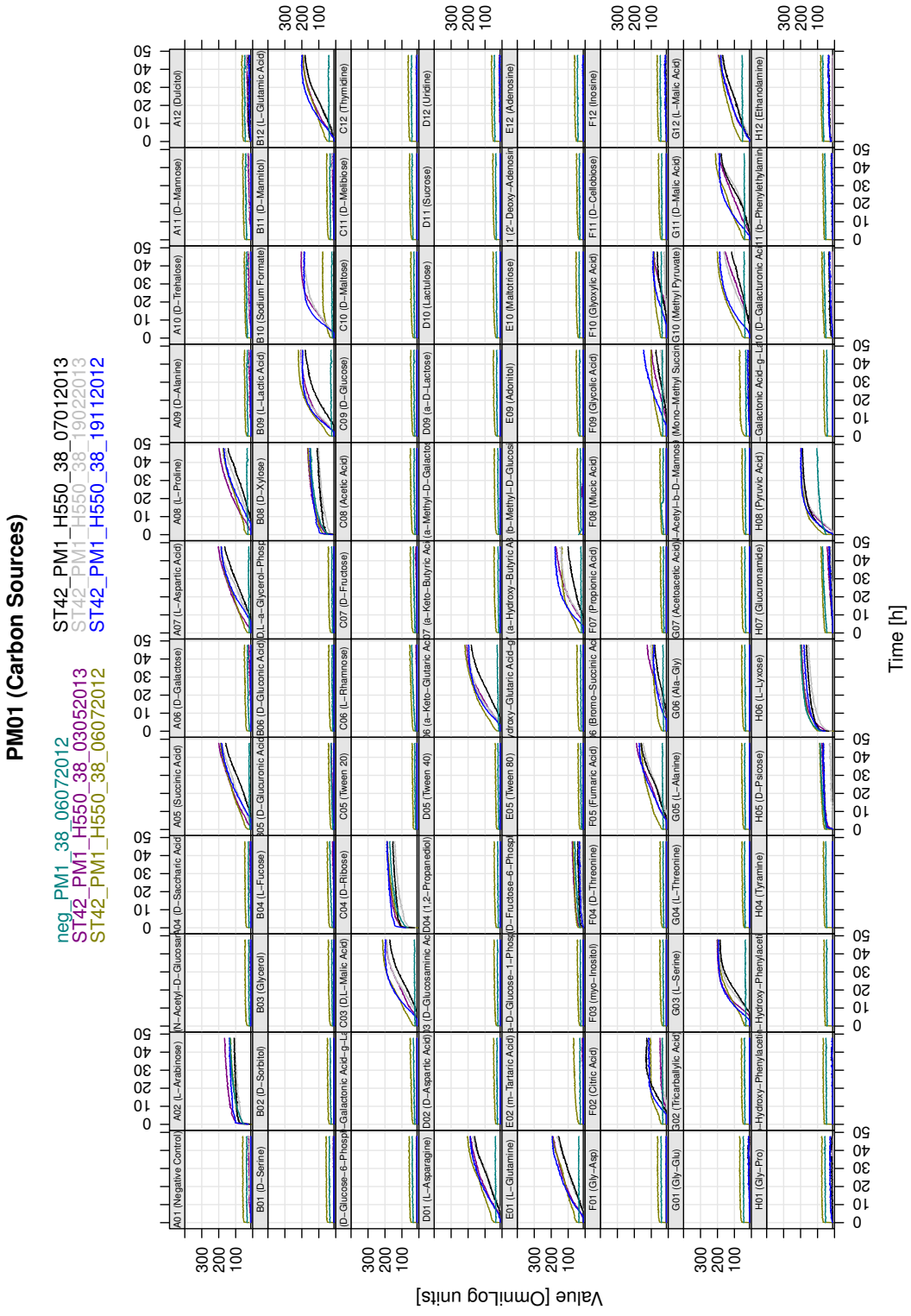


Figure A.25: xy plot for the *C. jejuni* isolate H550 in PM 1 at 38°C

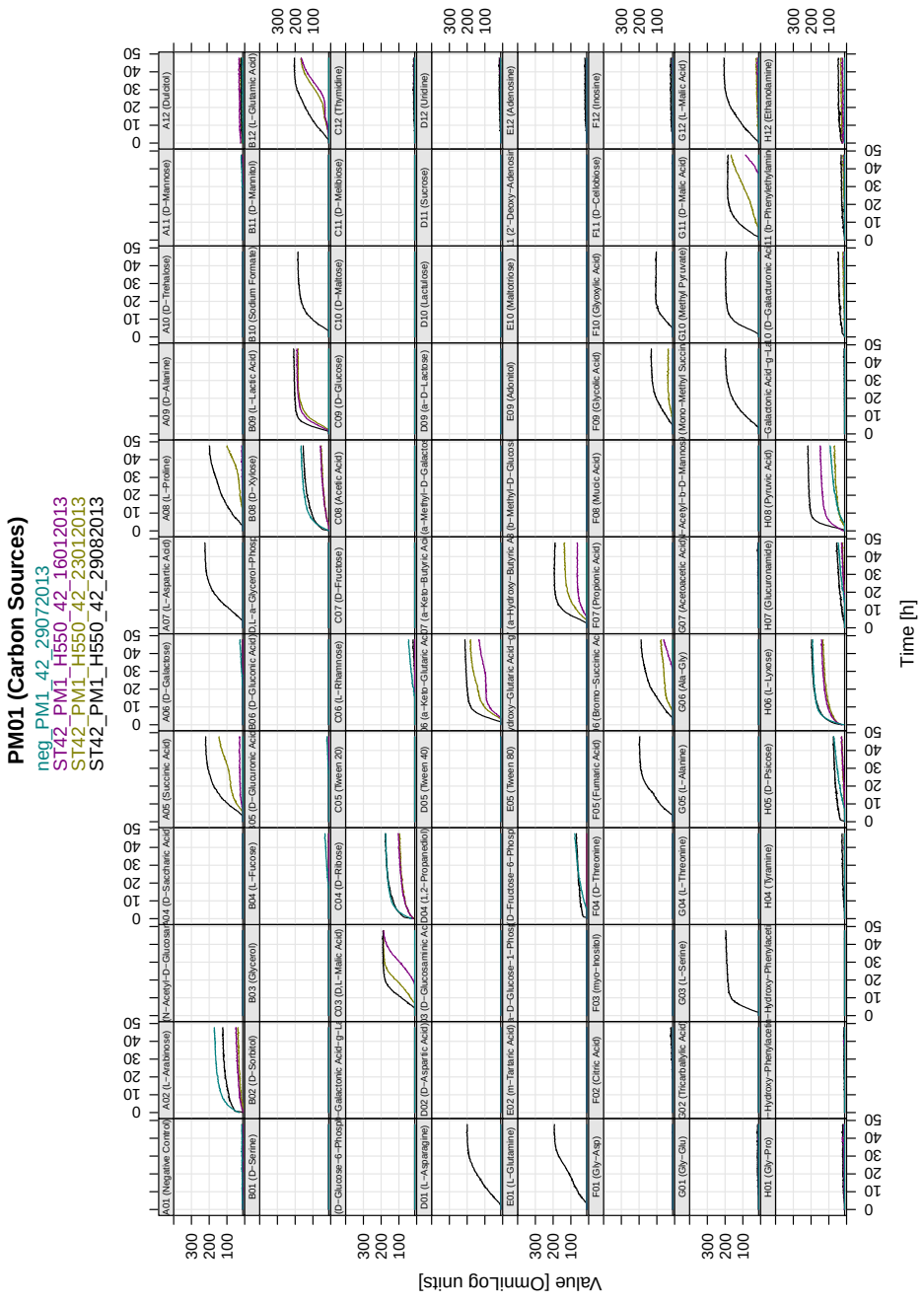


Figure A.26: xy plot for the *C. jejuni* isolate H550 in PM 1 at 42°C

A.2.14 Isolate H550 (ST-42) in PM 9 at 38°C and 42°C

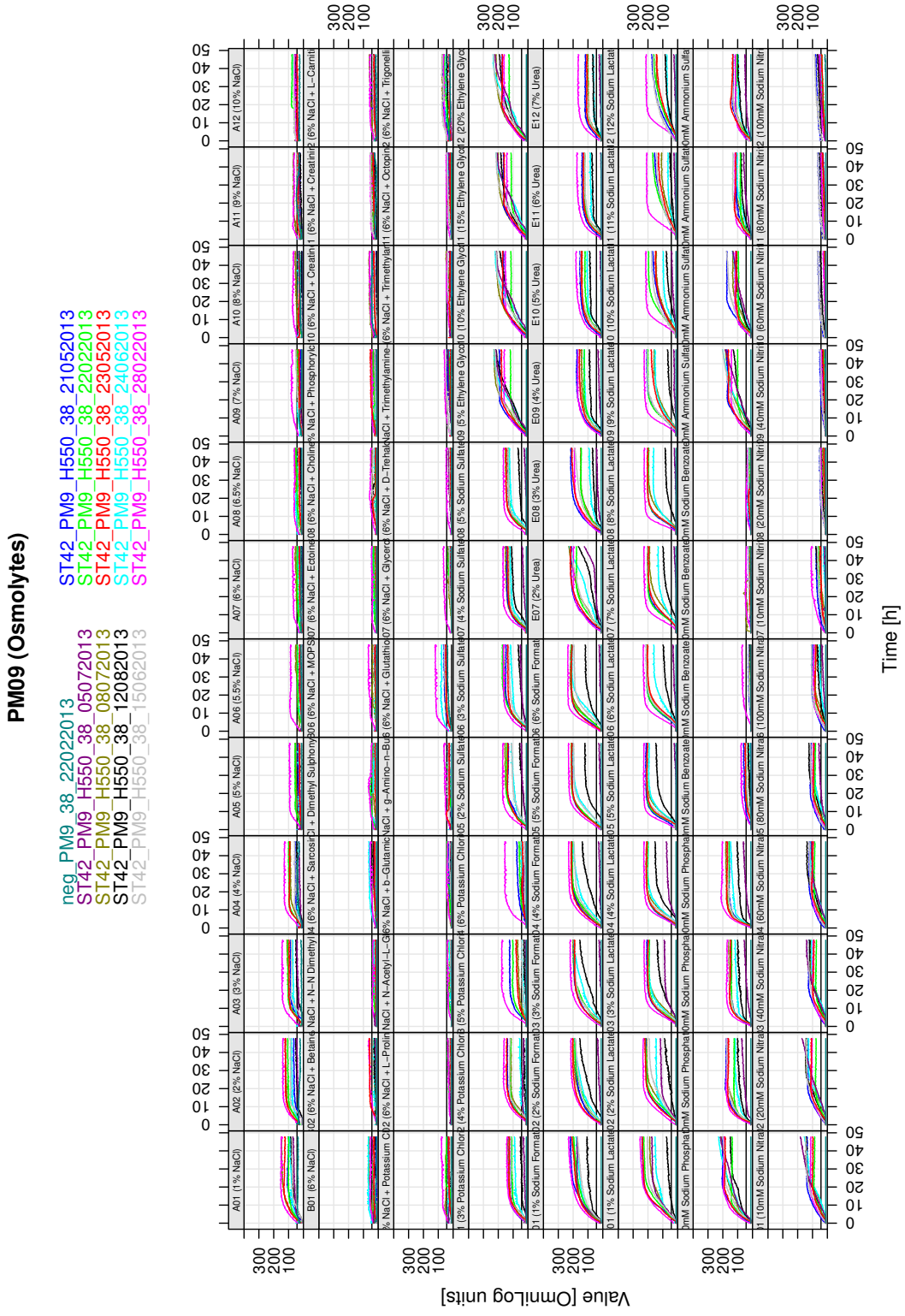


Figure A.27: xy plot for the *C. jejuni* isolate H550 in PM 9 at 38°C

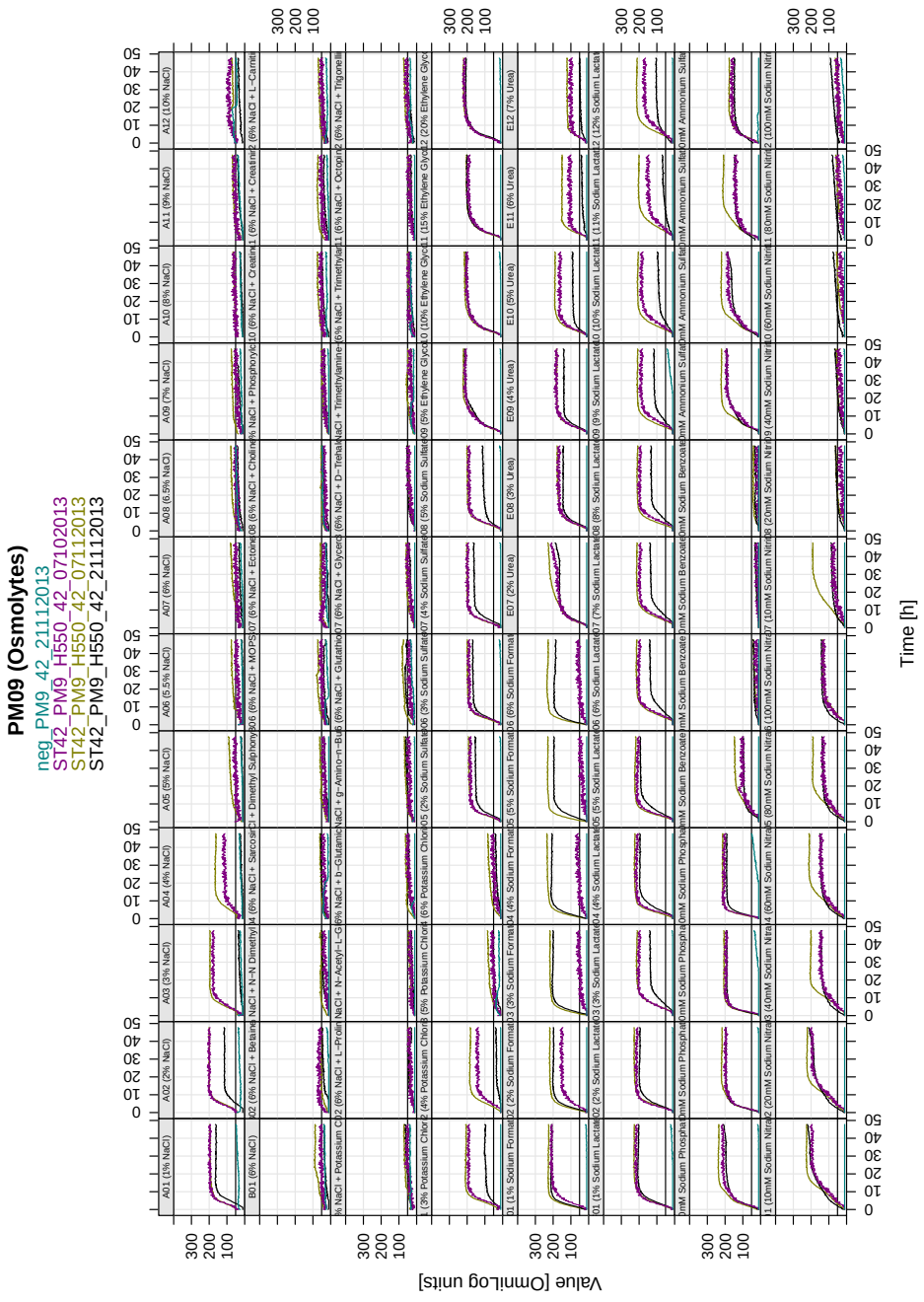


Figure A.28: xy plot for the *C. jejuni* isolate H550 in PM 9 at 42°C

A.2.15 Isolate M602b (ST-42) in PM 1 at 38°C and 42°C

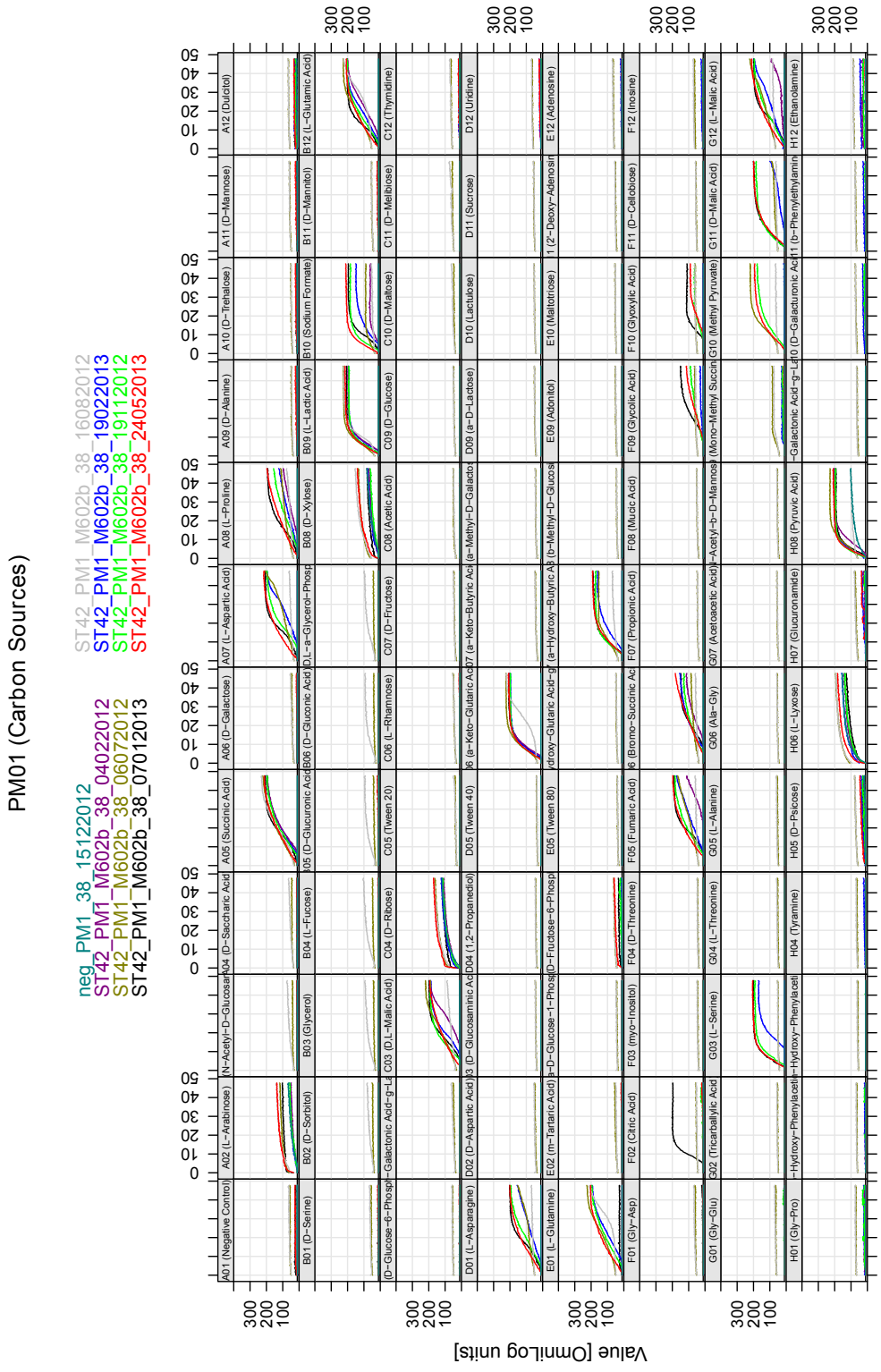


Figure A.29: xy plot for the *C. jejuni* isolate M602b in PM 1 at 38°C

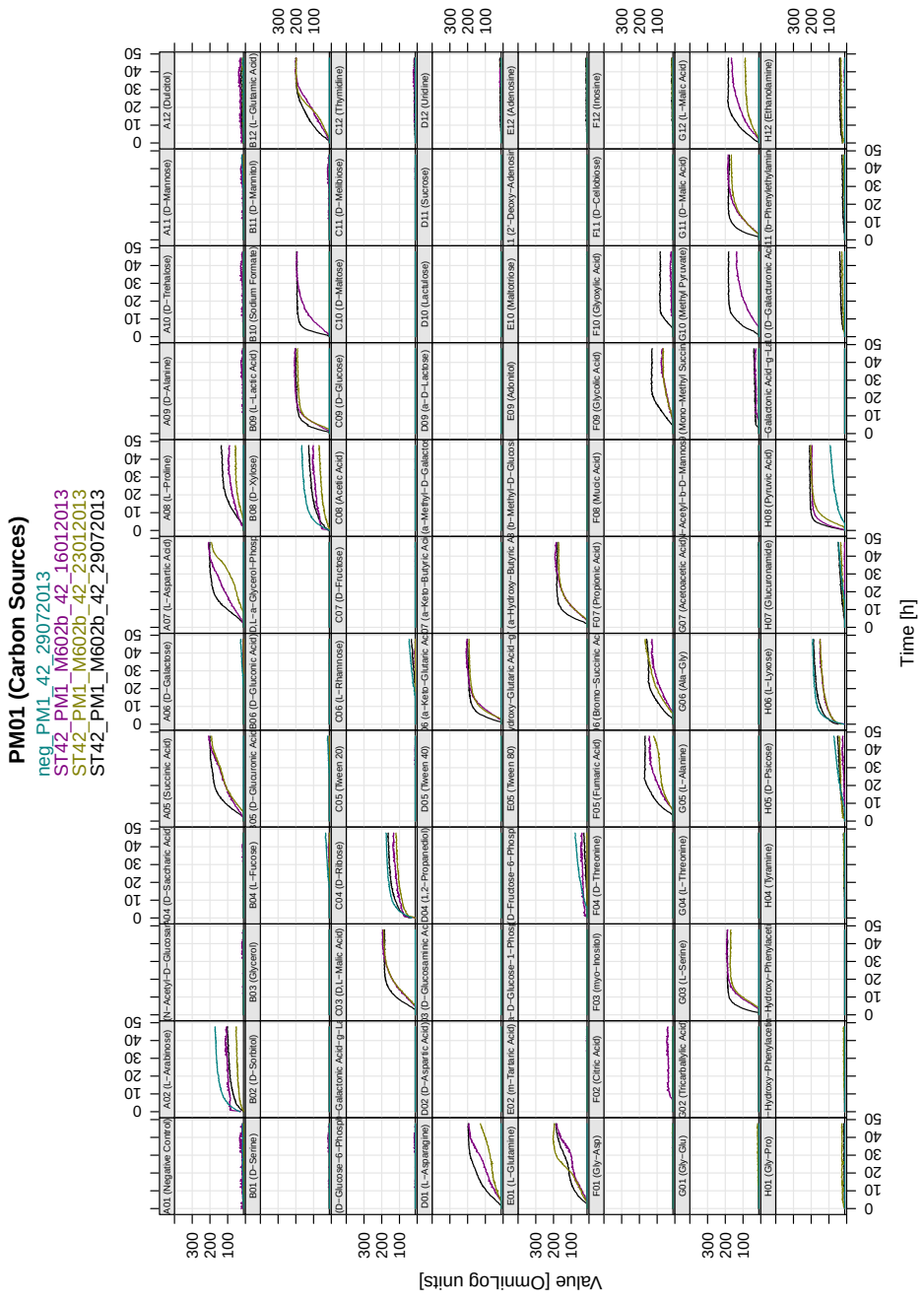


Figure A.30: xy plot for the *C. jejuni* isolate M602b in PM 1 at 42°C

A.2.16 Isolate M602b (ST-42) in PM 9 at 38°C and 42°C

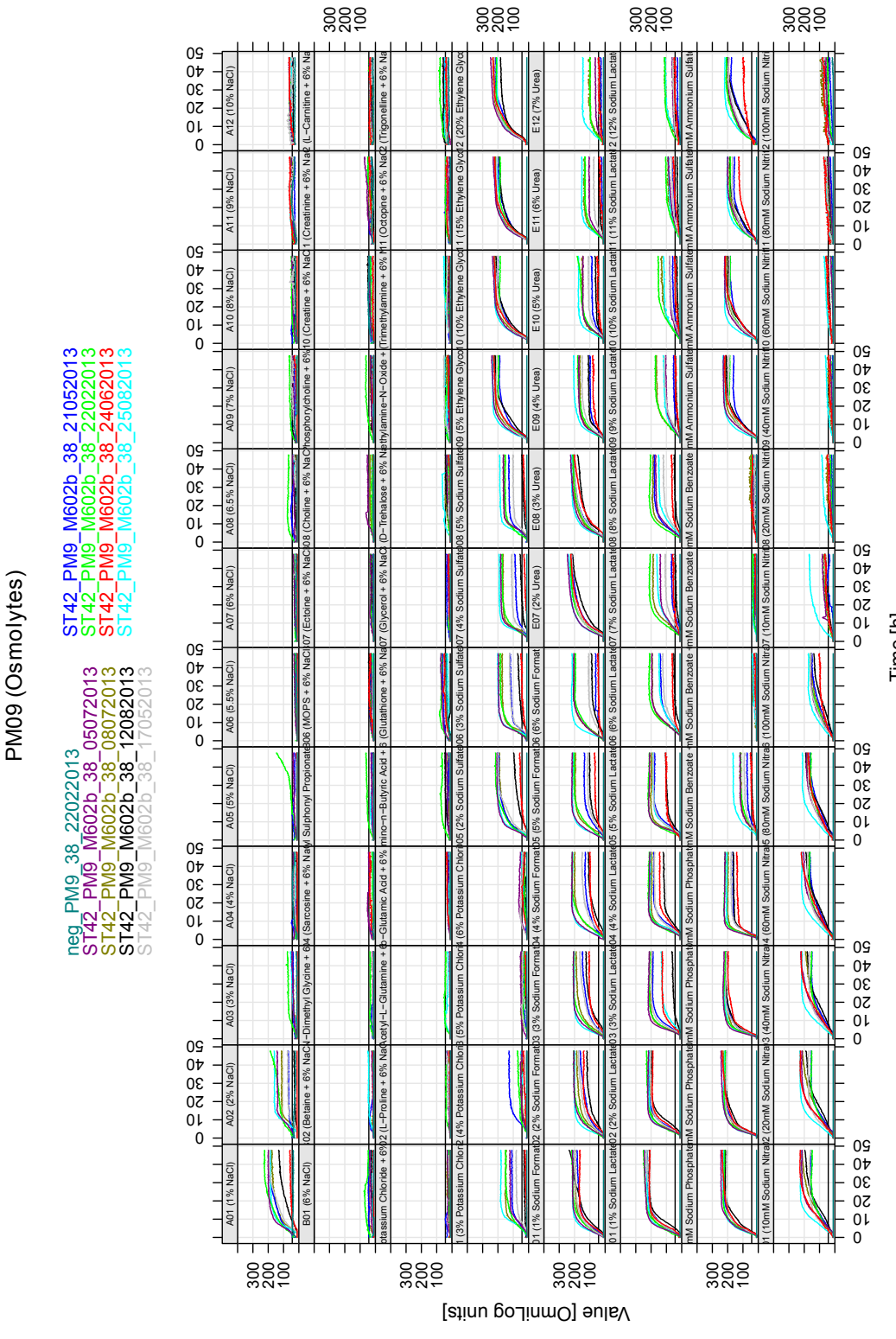


Figure A.31: xy plot for the *C. jejuni* isolate M602b in PM 9 at 38°C

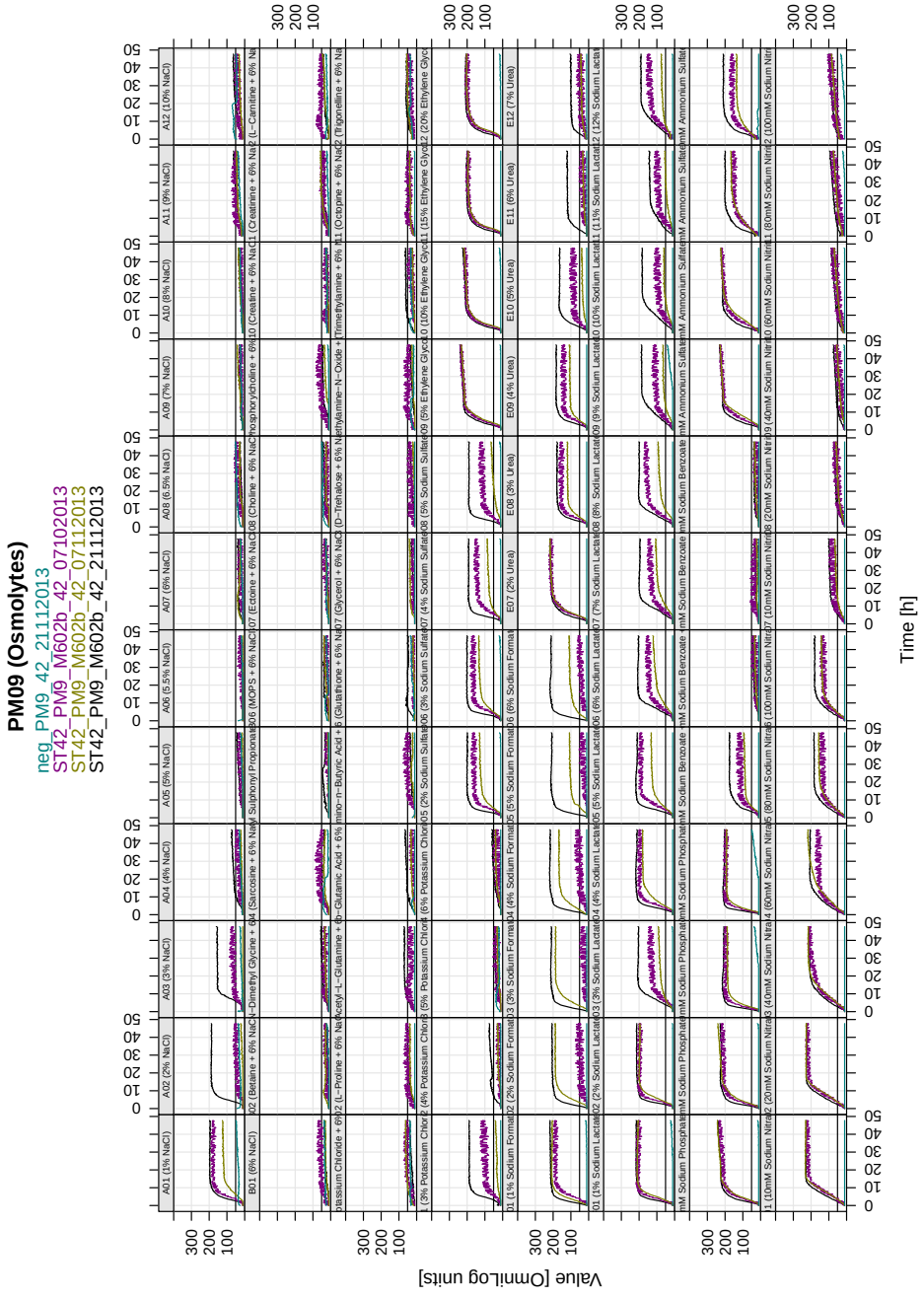


Figure A.32: xy plot for the *C. jejuni* isolate M602b in PM 9 at 42°C

A.2.17 Isolate P110b (ST-474) in PM 1 at 38°C and 42°C

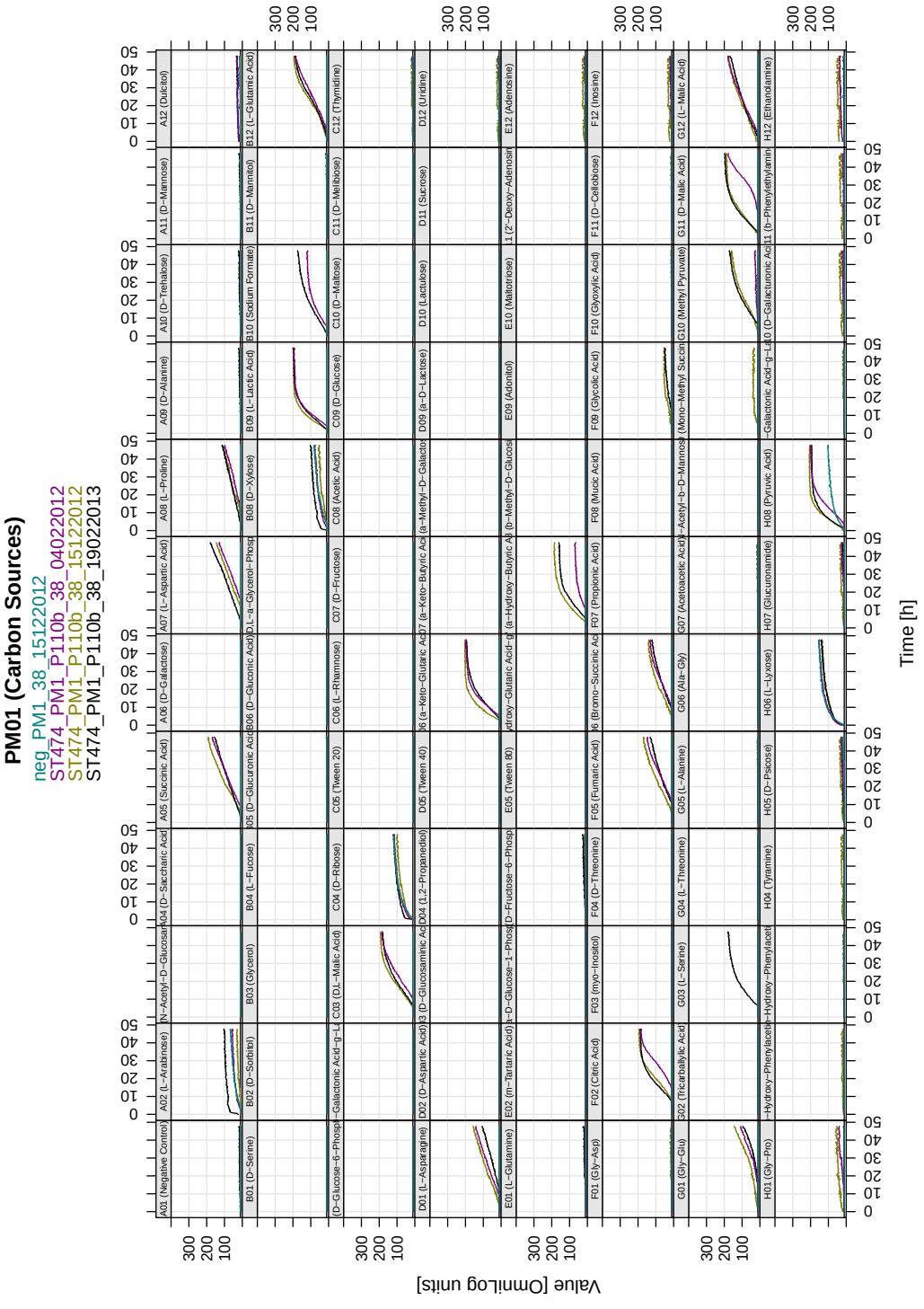


Figure A.33: xy plot for the *C. jejuni* isolate P110b in PM 1 at 38°C

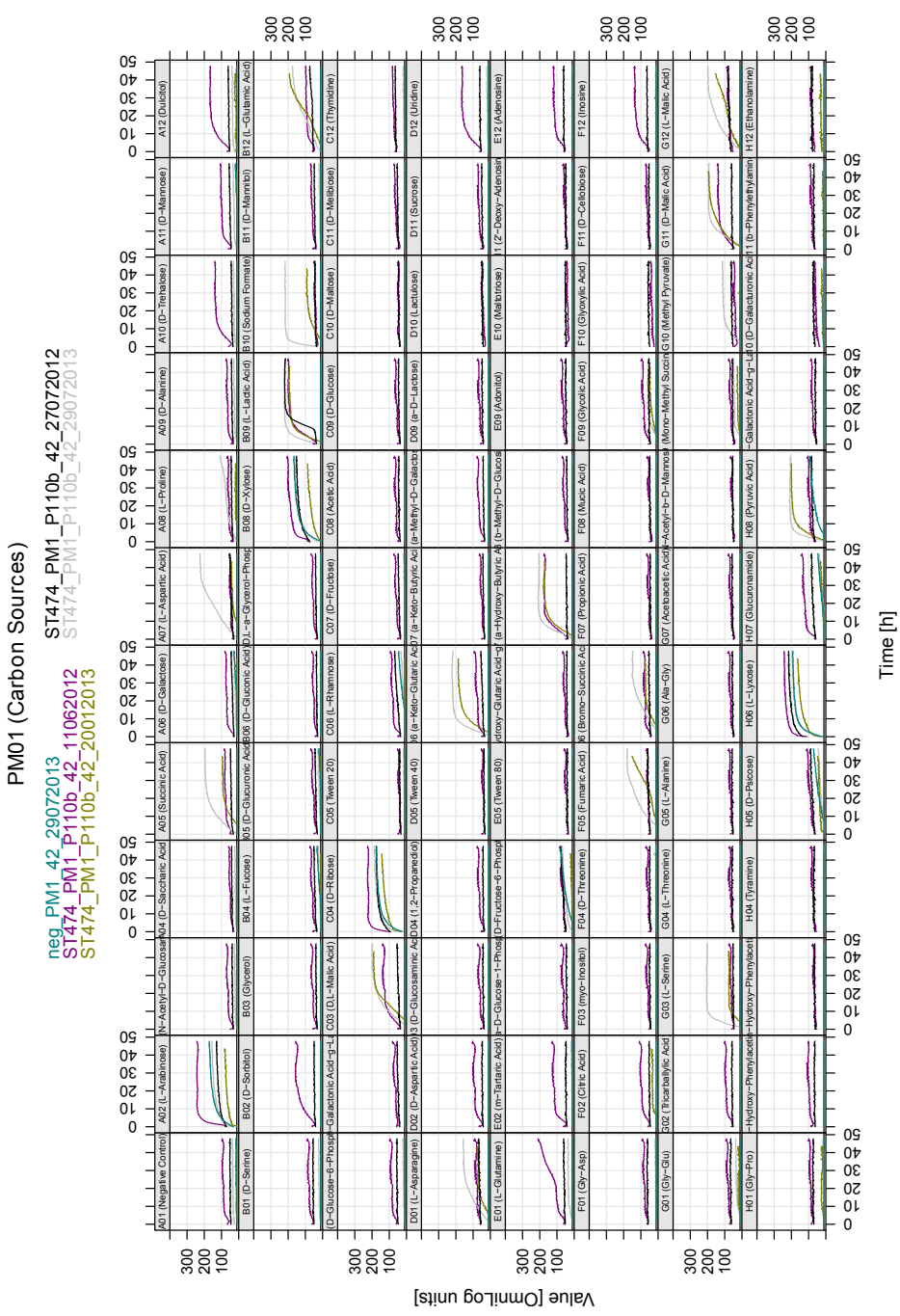


Figure A.34: xy plot for the *C. jejuni* isolate P110b in PM 1 at 42°C

A.2.18 Isolate P110b (ST-474) in PM 9 at 38°C and 42°C

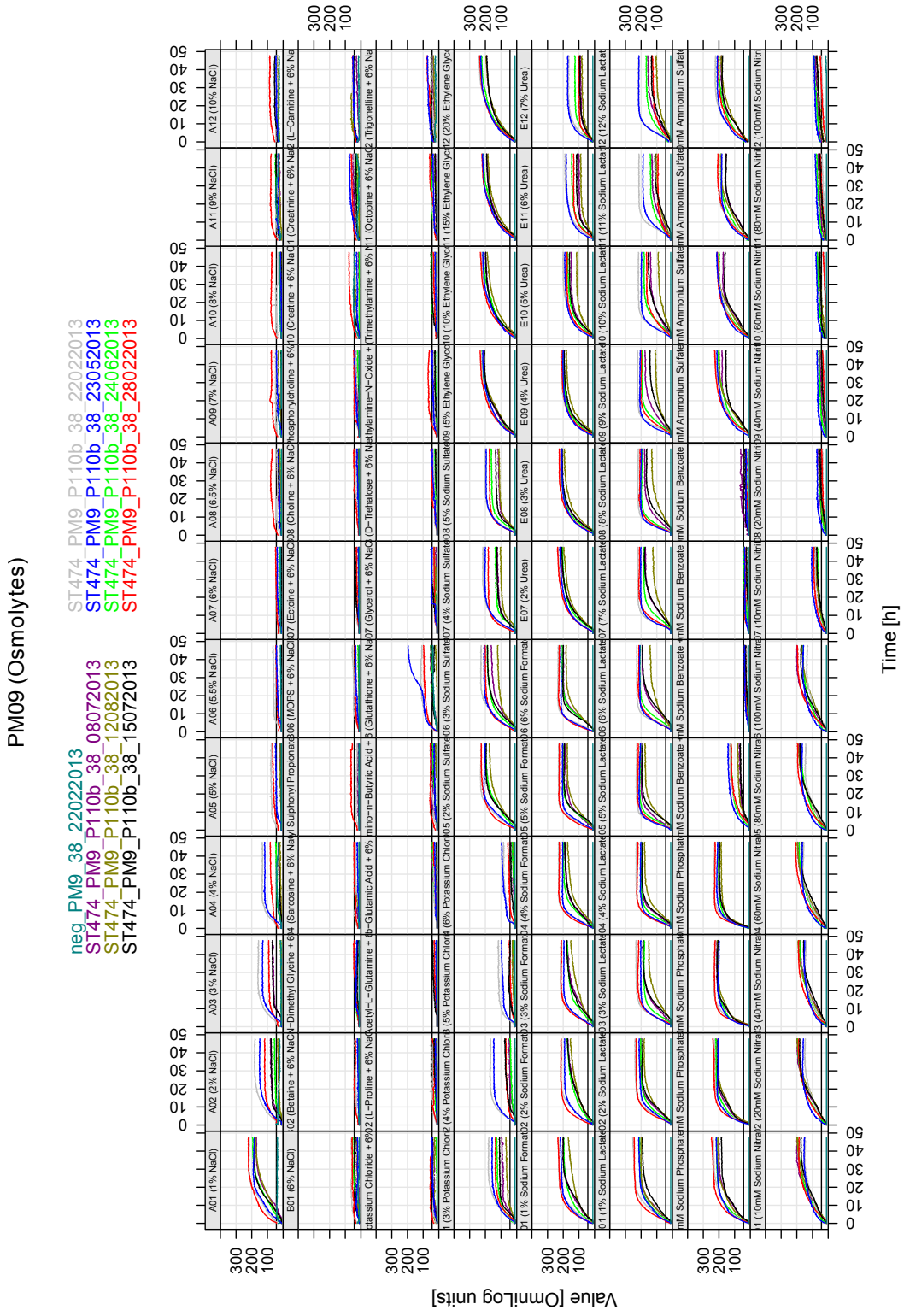


Figure A.35: xy plot for the *C. jejuni* isolate P110b in PM 9 at 38°C

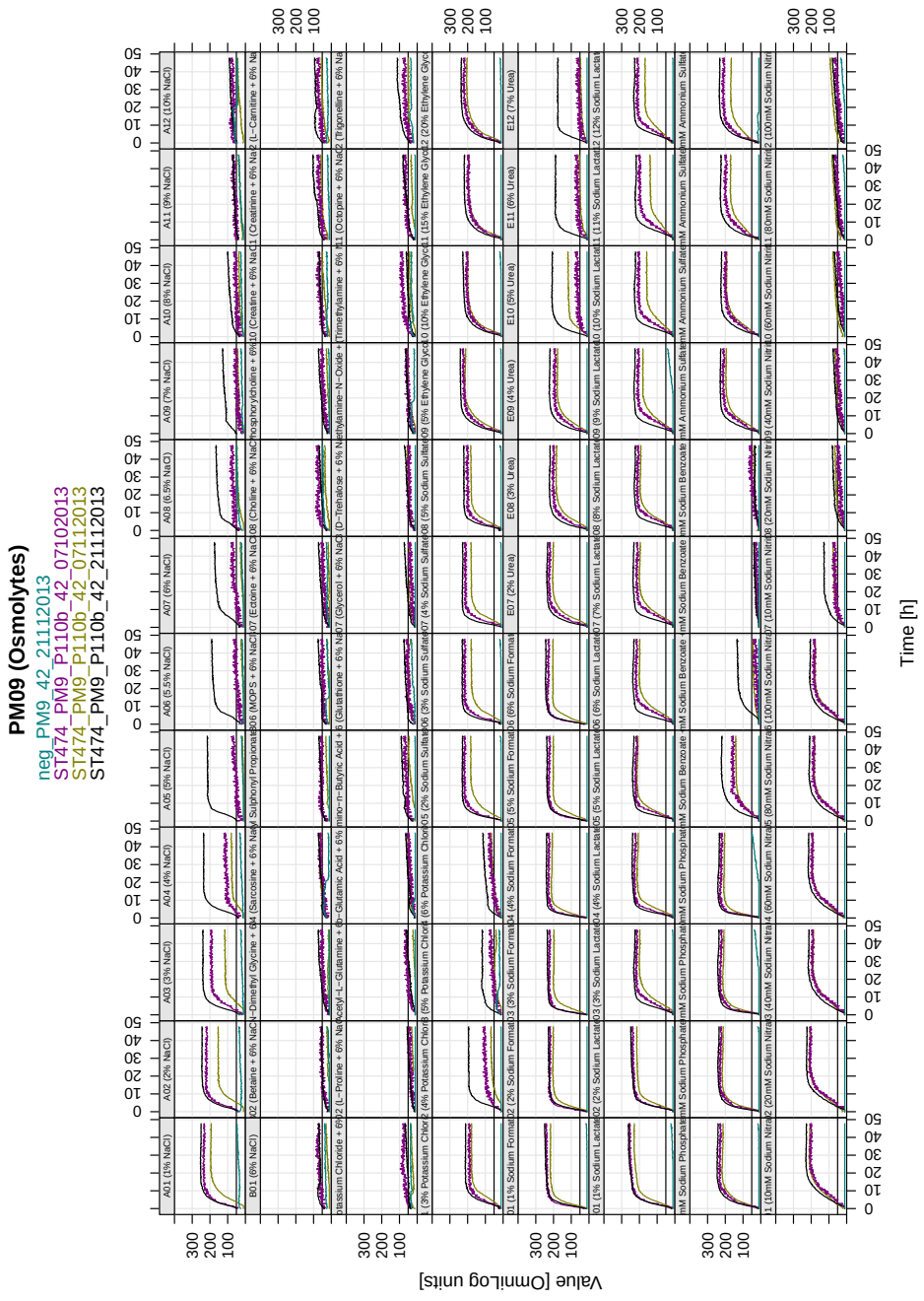


Figure A.36: *xy* plot for the *C. jejuni* isolate P110b in PM 9 at 42°C

A.2.19 Isolate P694a (ST-474) in PM 1 at 38°C and 42°C

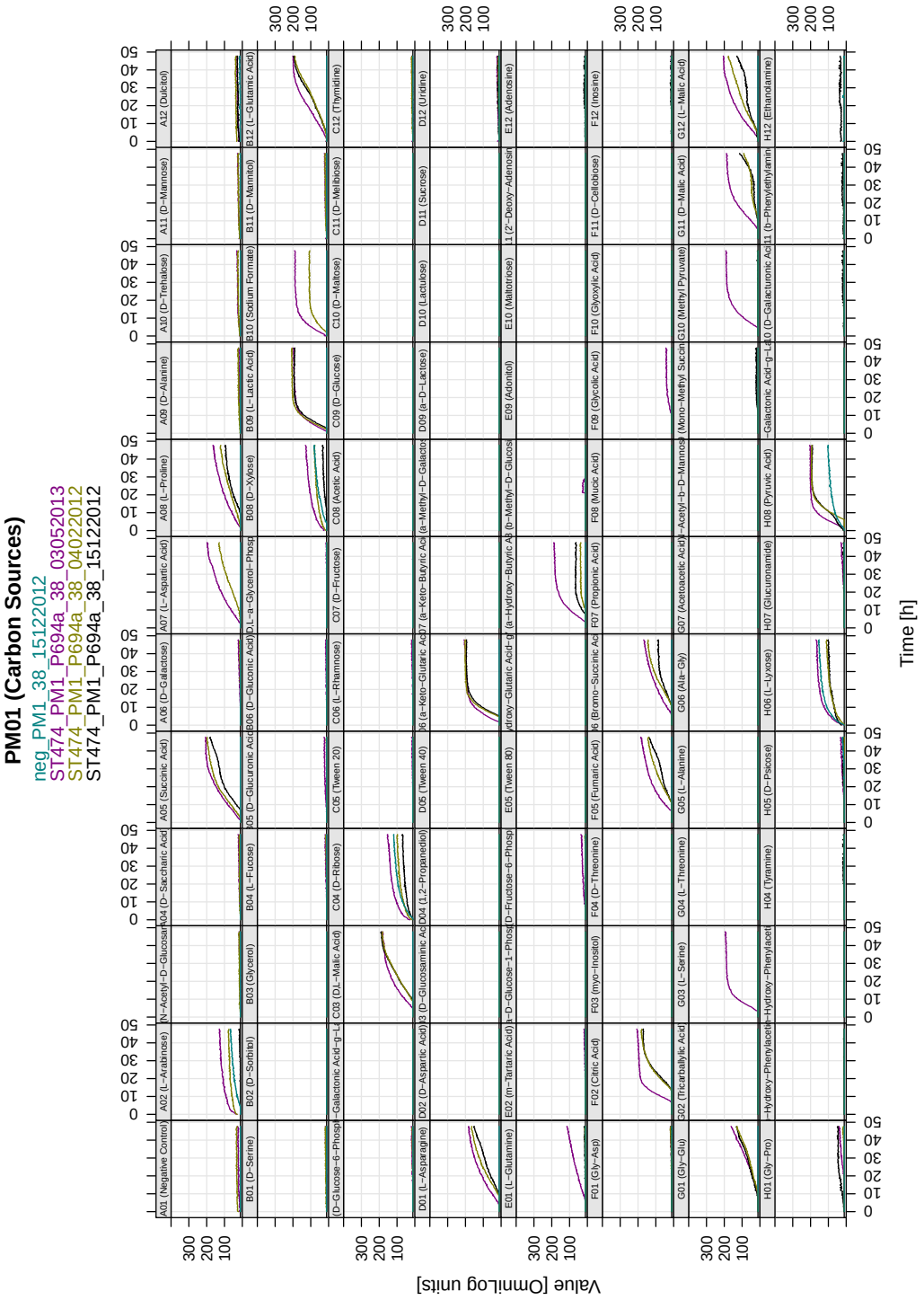


Figure A.37: xy plot for the *C. jejuni* isolate P694a in PM 1 at 38°C

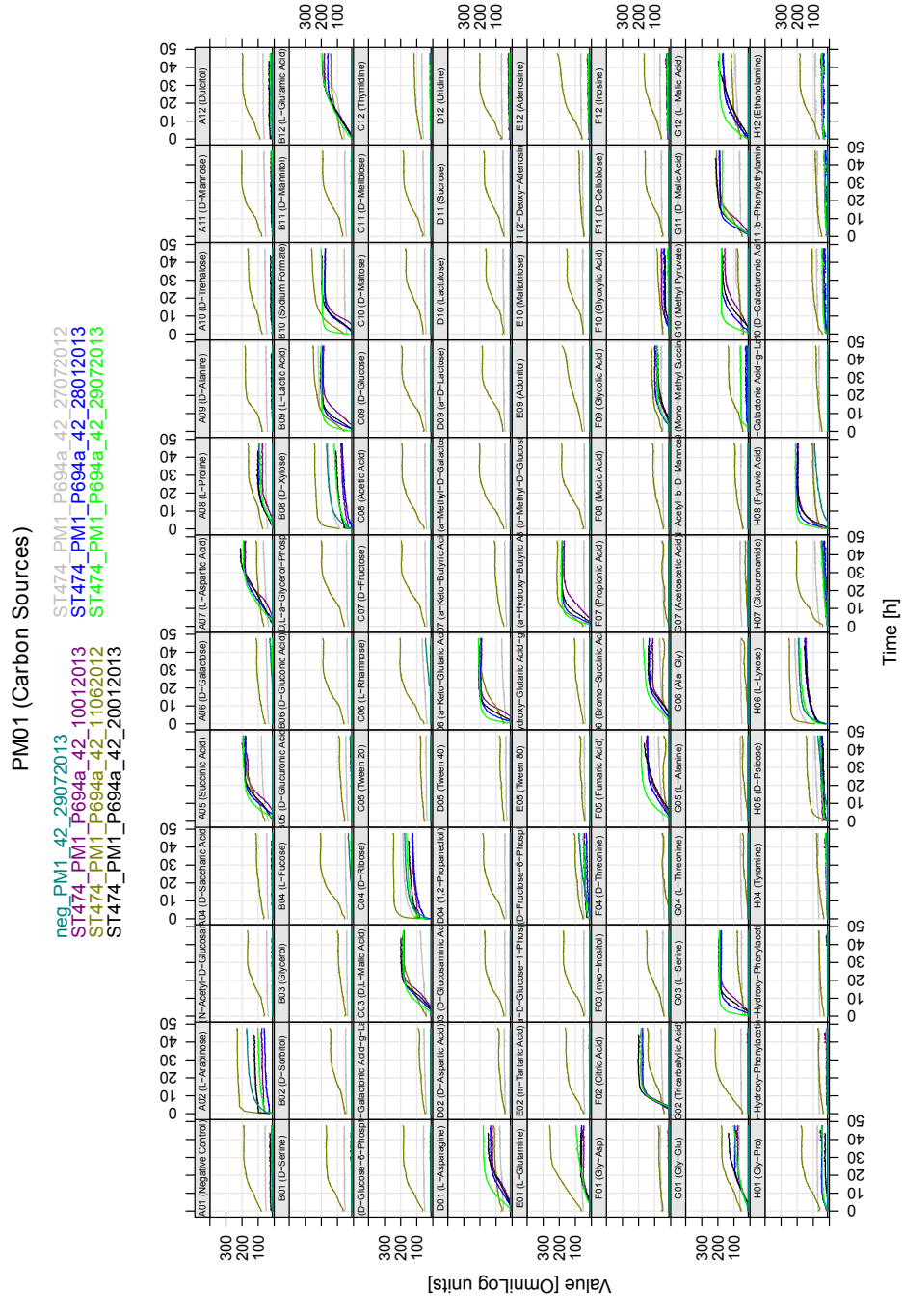


Figure A.38: xy plot for the *C. jejuni* isolate P694a in PM 1 at 42°C

A.2.20 Isolate M602b (ST-474) in PM 9 at 38°C and 42°C

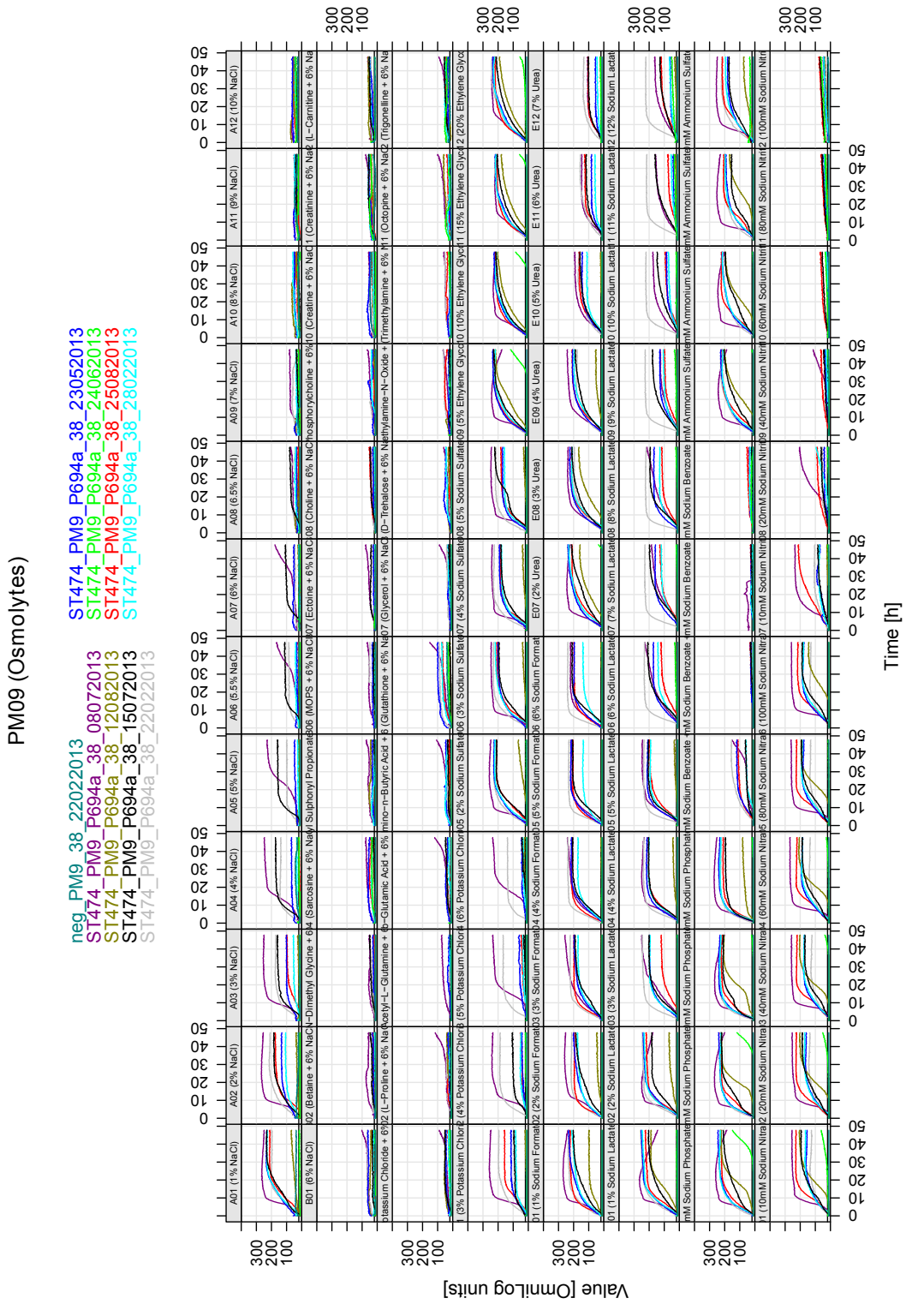


Figure A.39: xy plot for the *C. jejuni* isolate P694a in PM 9 at 38°C

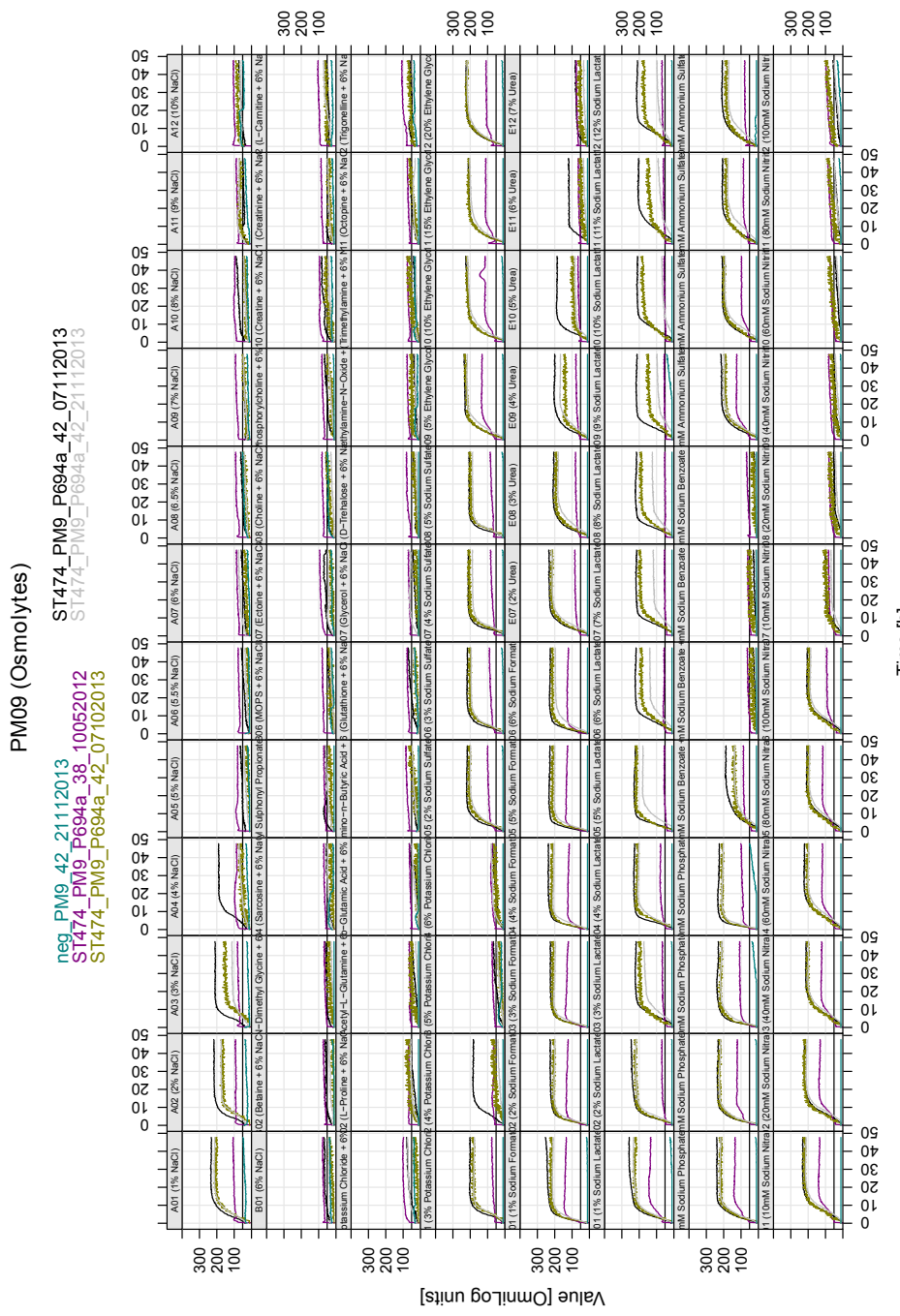


Figure A.40: xy plot for the *C. jejuni* isolate P694a in PM 9 at 42°C

A.2.21 Isolate S168b (ST-474) in PM 1 at 38°C and 42°C

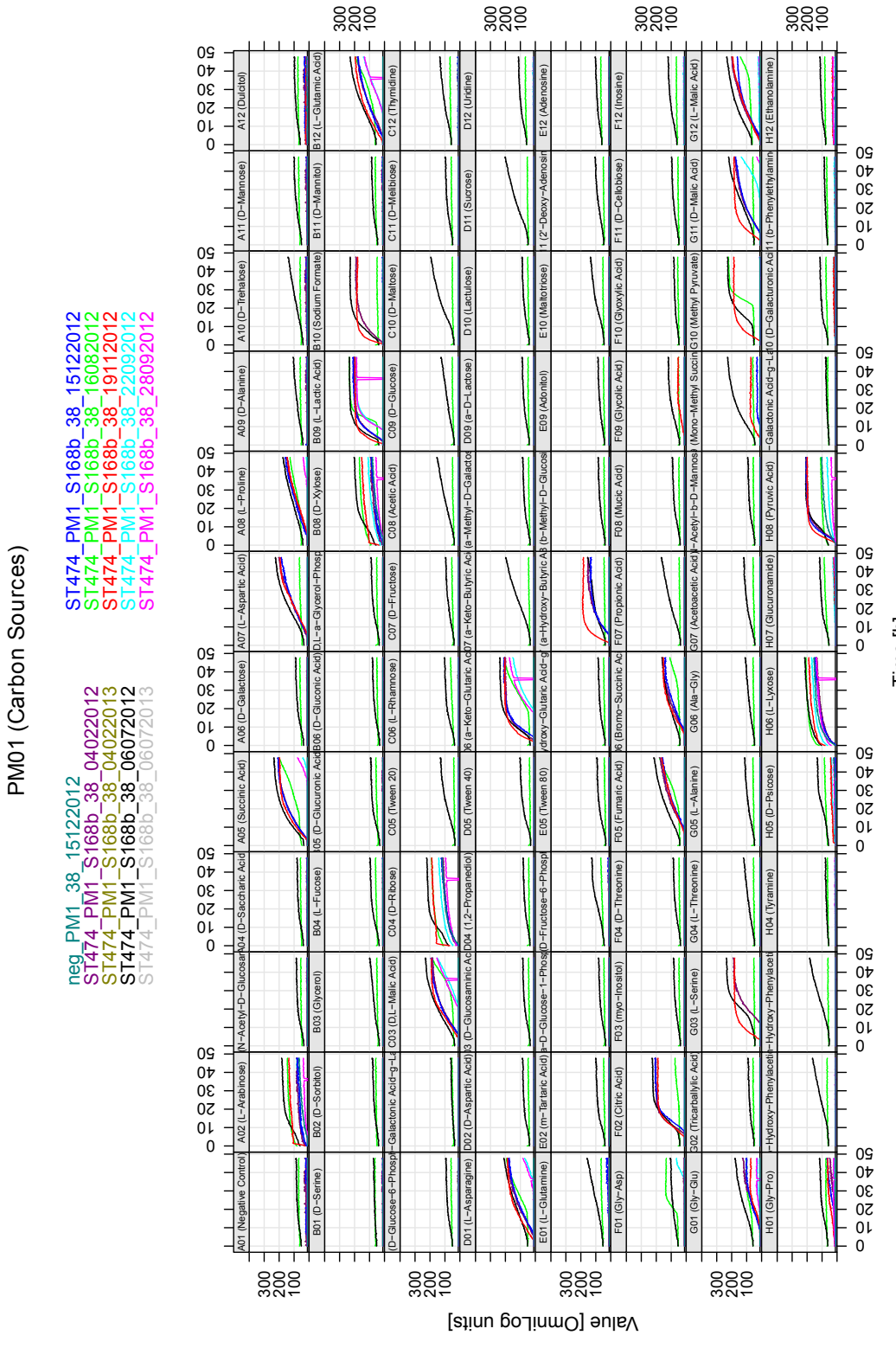


Figure A.41: xy plot for the *C. jejuni* isolate S168b in PM 1 at 38°C

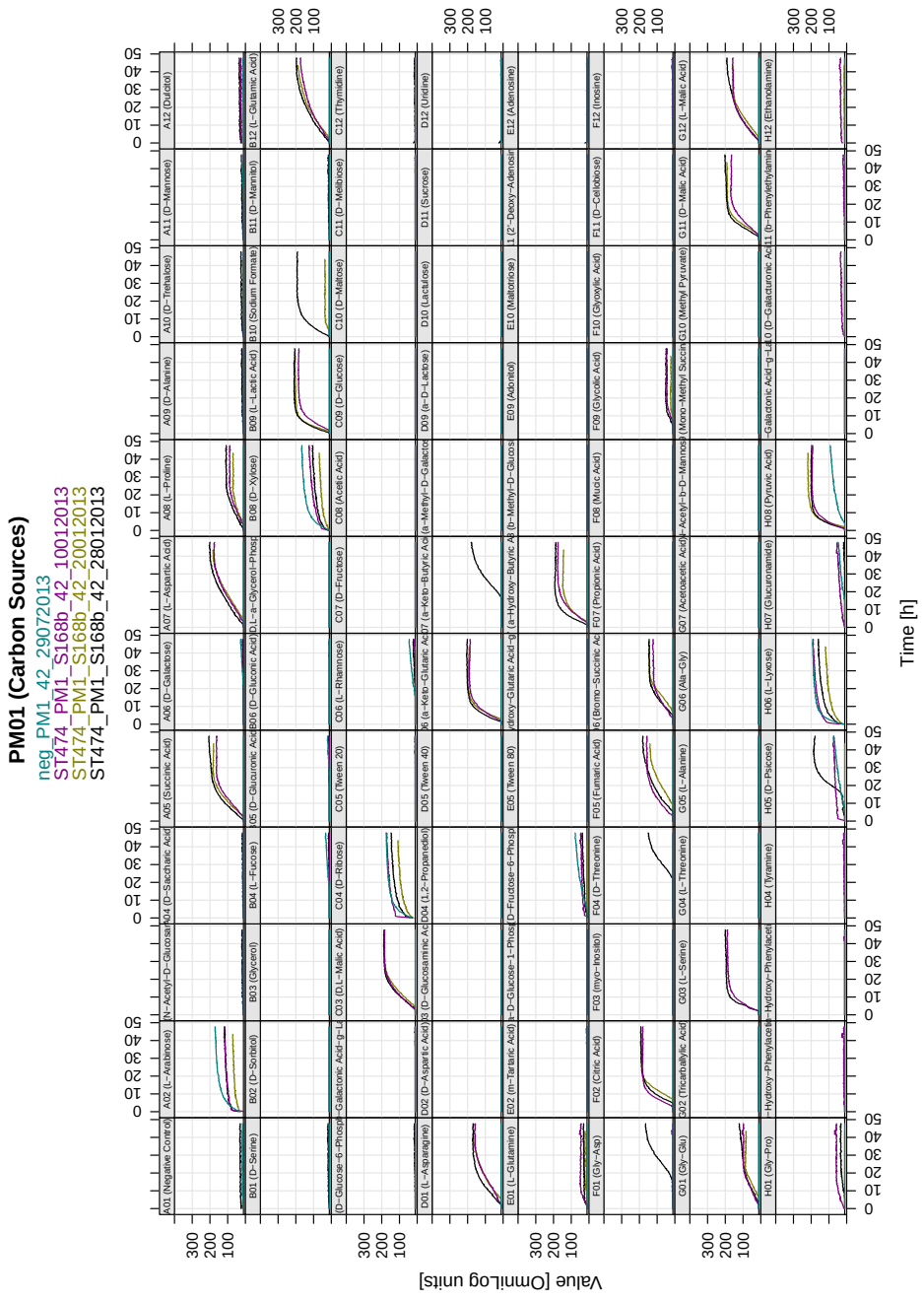


Figure A.42: xy plot for the *C. jejuni* isolate S168b in PM 1 at 42°C

A.2.22 Isolate S168b (ST-474) in PM 9 at 38°C and 42°C

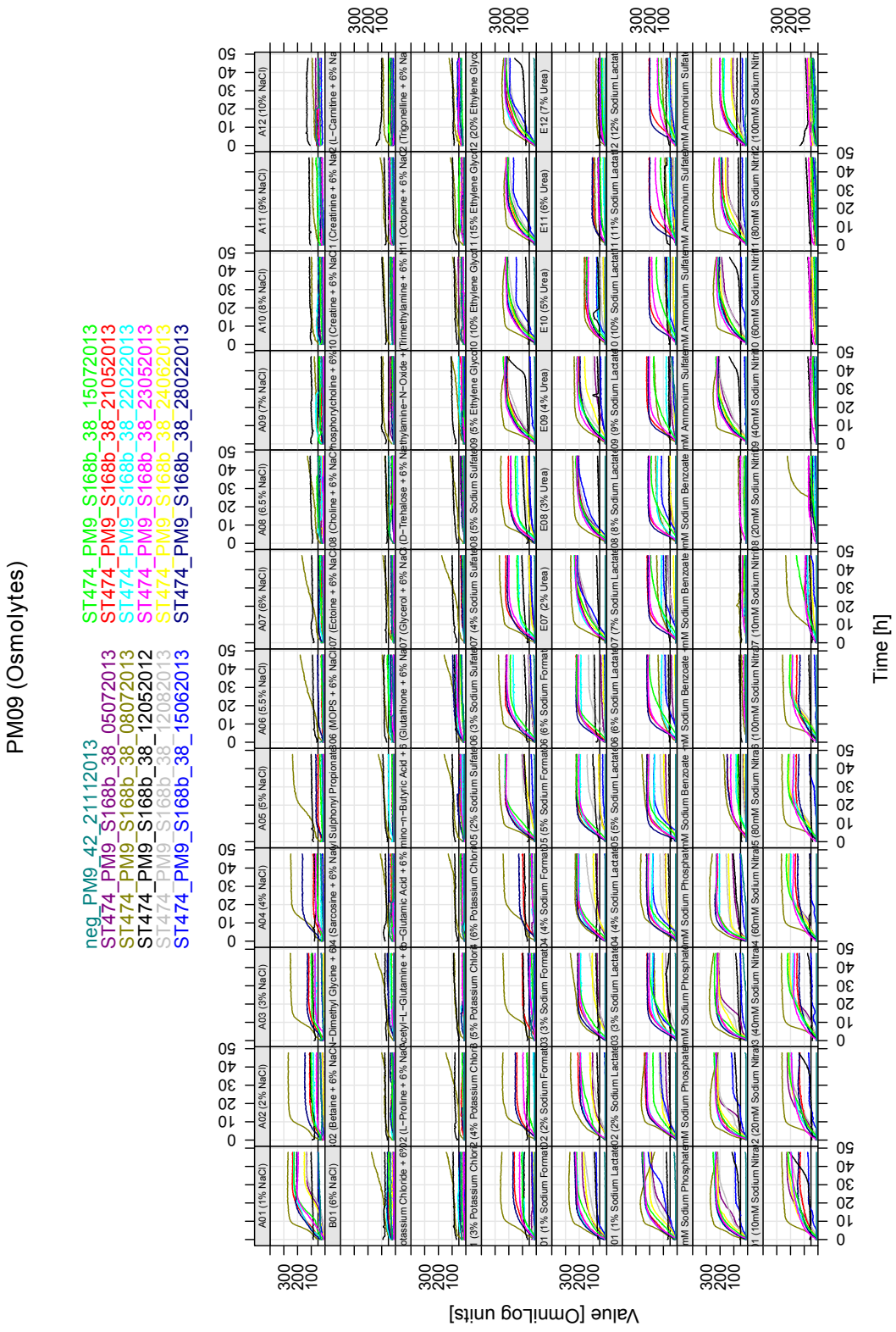


Figure A.43: xy plot for the *C. jejuni* isolate S168b in PM 9 at 38°C

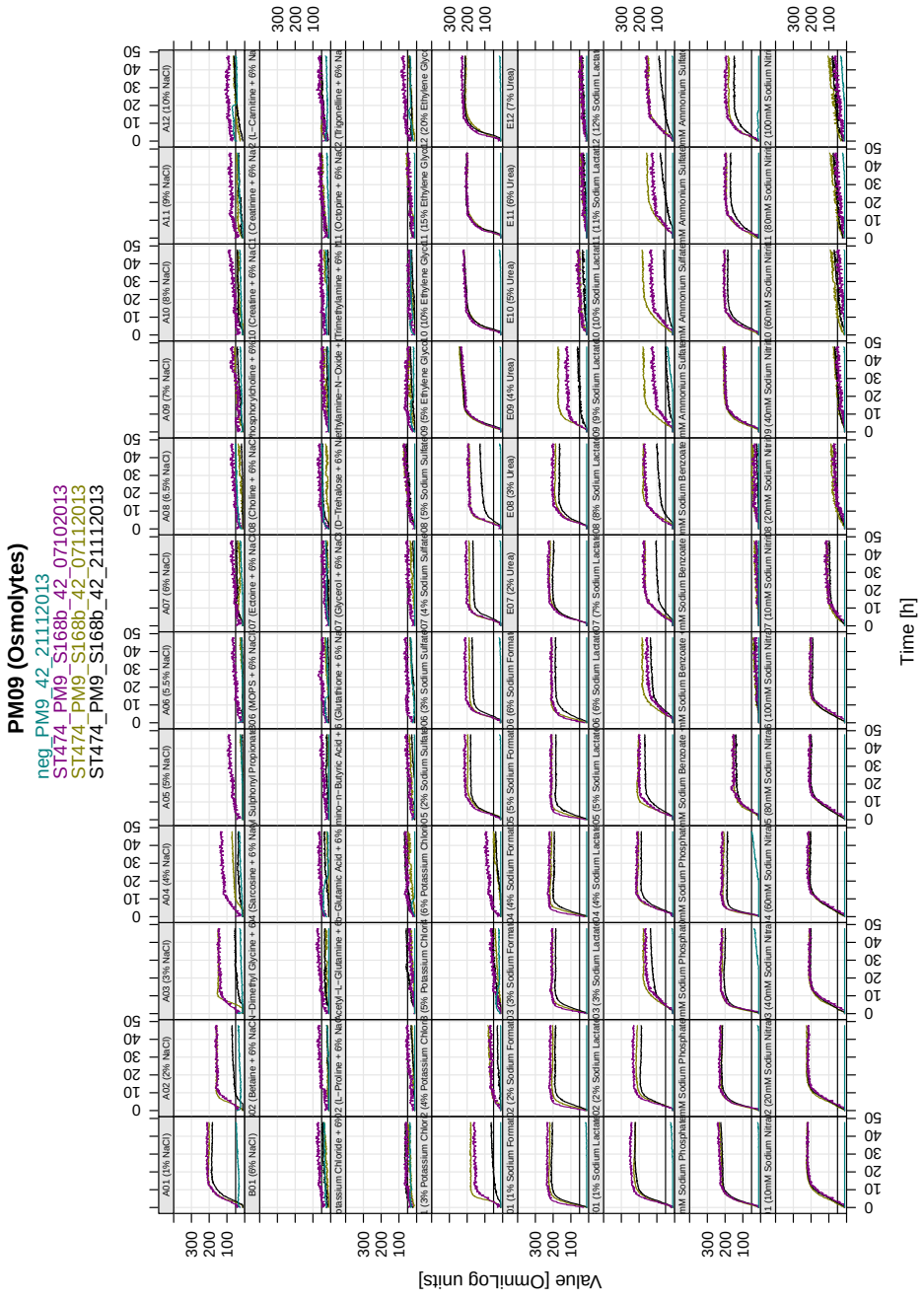


Figure A.44: xy plot for the *C. jejuni* isolate S168b in PM 9 at 42°C

A.2.23 Isolate S22b (ST-2026) in PM 1 at 38°C and 42°C

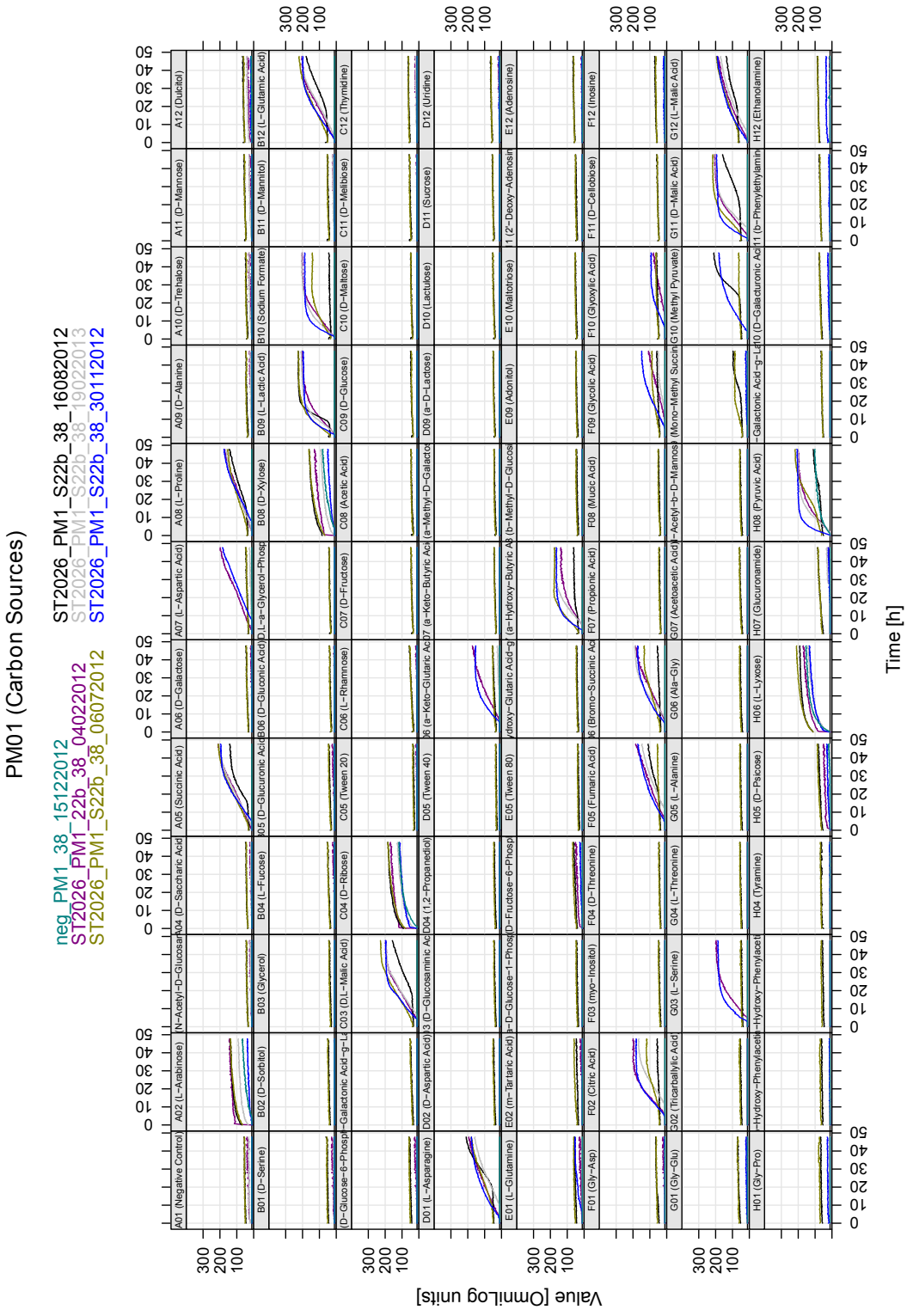


Figure A.45: xy plot for the *C. jejuni* isolate S22b in PM 1 at 38°C

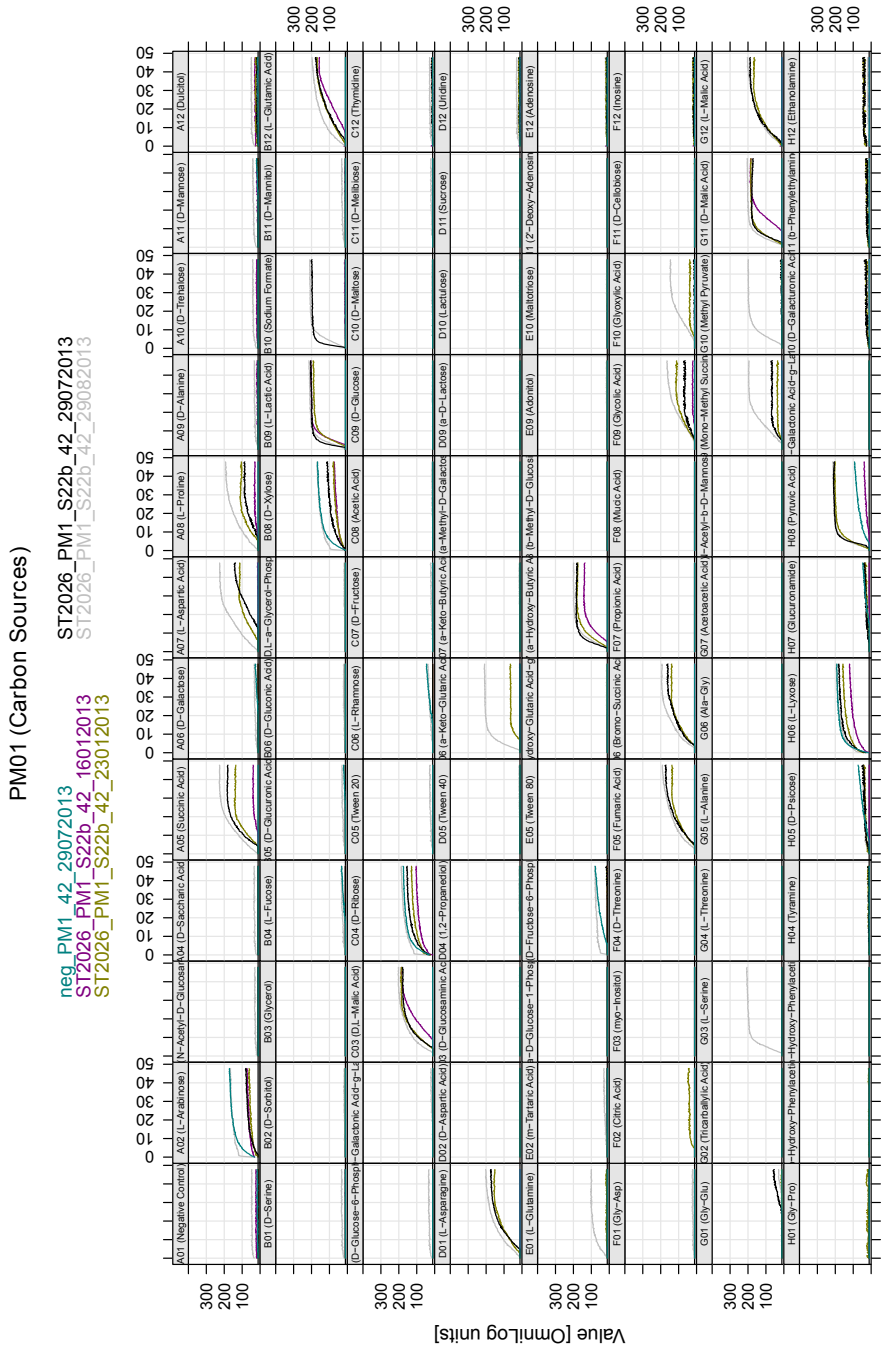


Figure A.46: xy plot for the *C. jejuni* isolate S22b in PM 1 at 42°C

A.2.24 Isolate S22b (ST-2026) in PM 9 at 38°C and 42°C

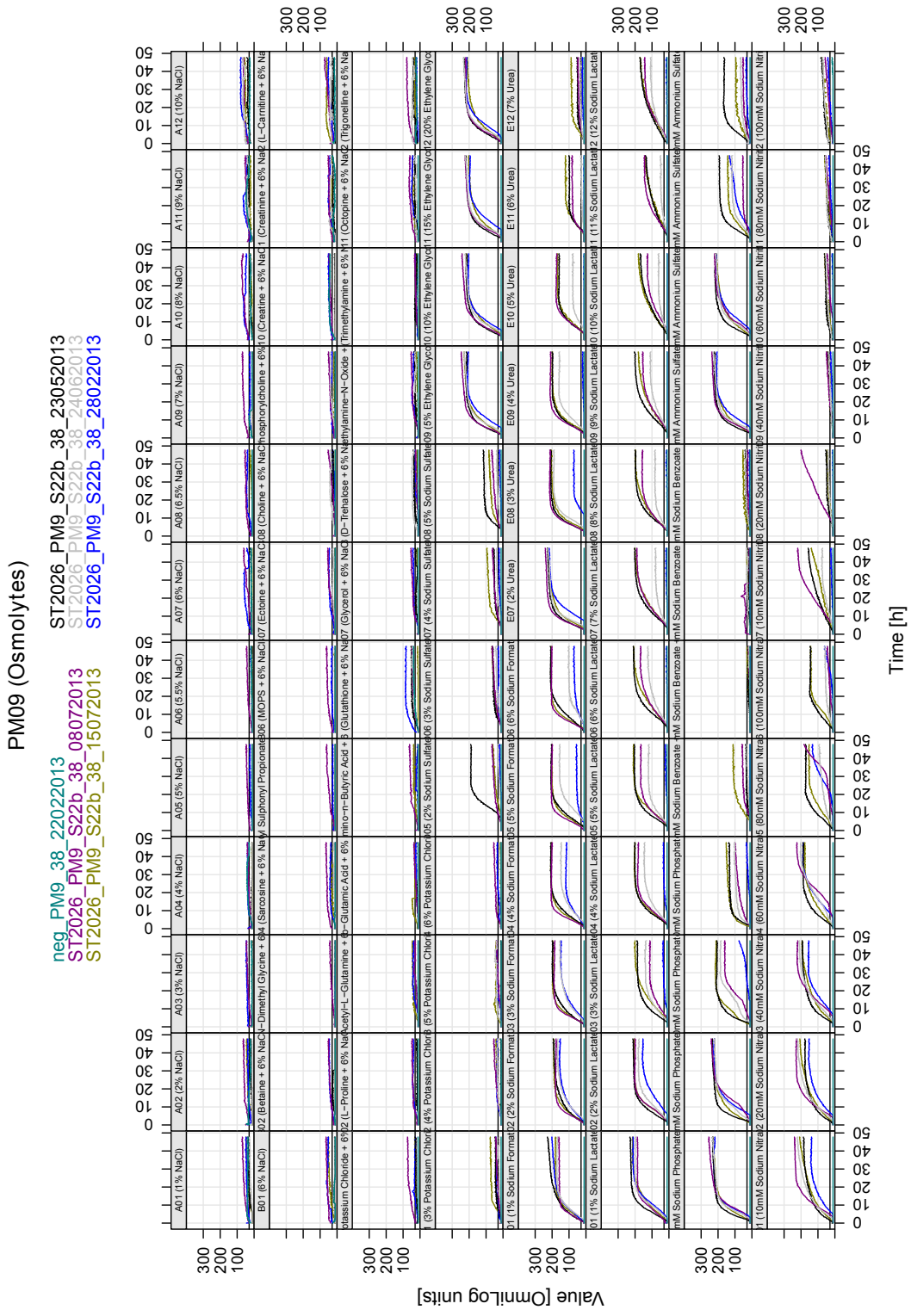


Figure A.47: xy plot for the *C. jejuni* isolate S22b in PM 9 at 38°C

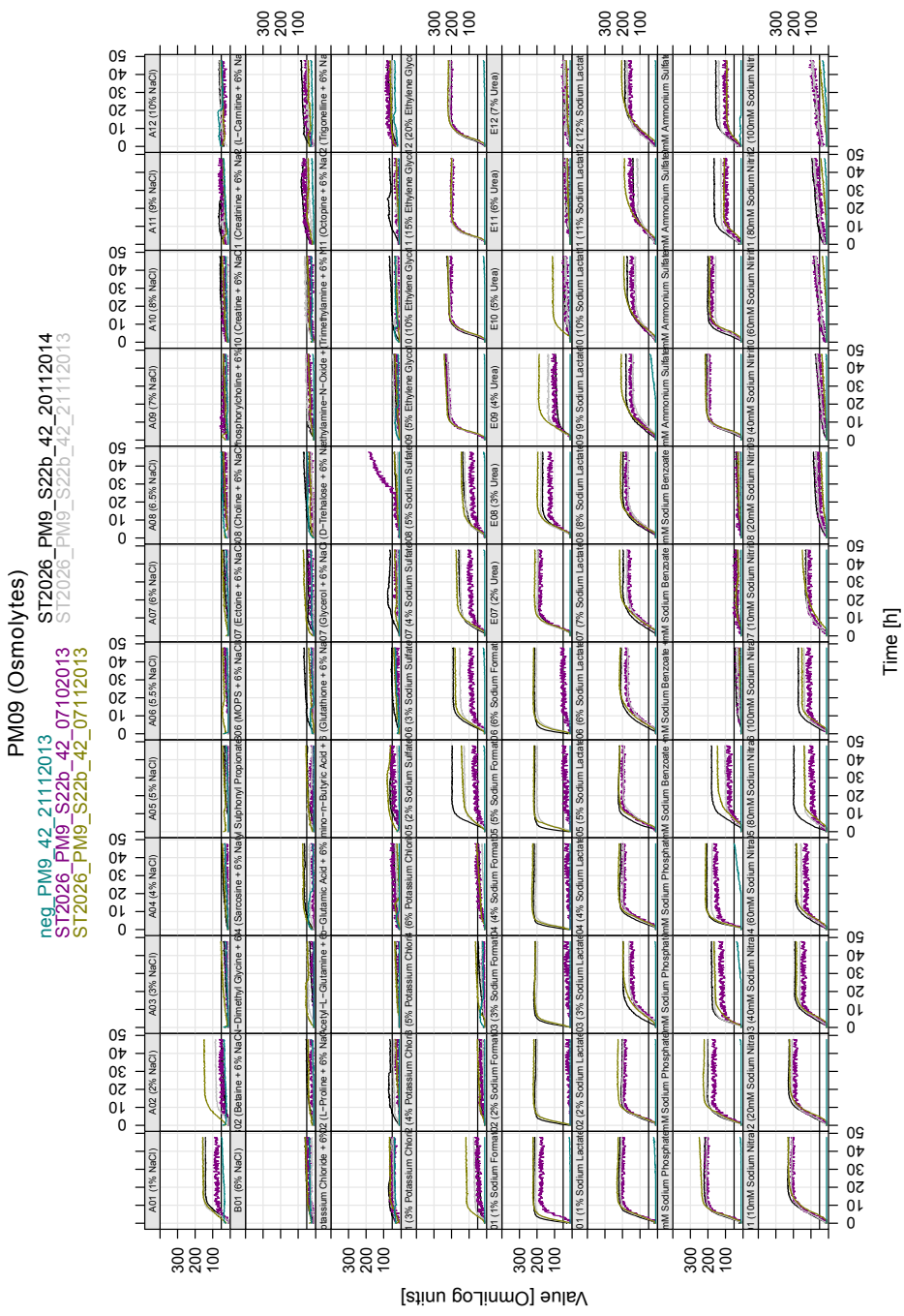


Figure A.48: xy plot for the *C. jejuni* isolate S22b in PM 9 at 42°C

A.2.25 Isolate S276b (ST-61) in PM 1 at 38°C and 42°C

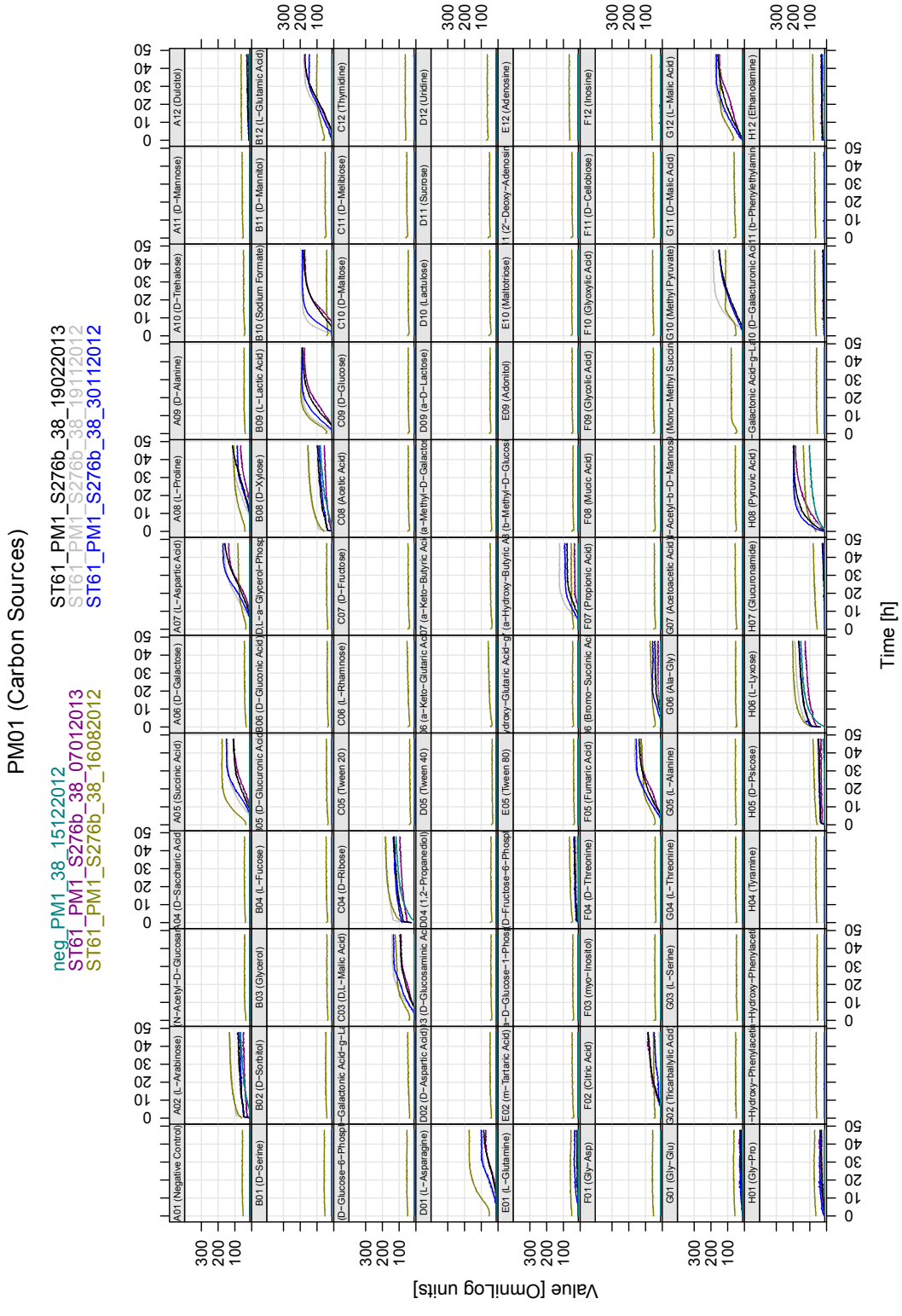


Figure A.49: xy plot for the *C. jejuni* isolate S276b in PM 1 at 38°C

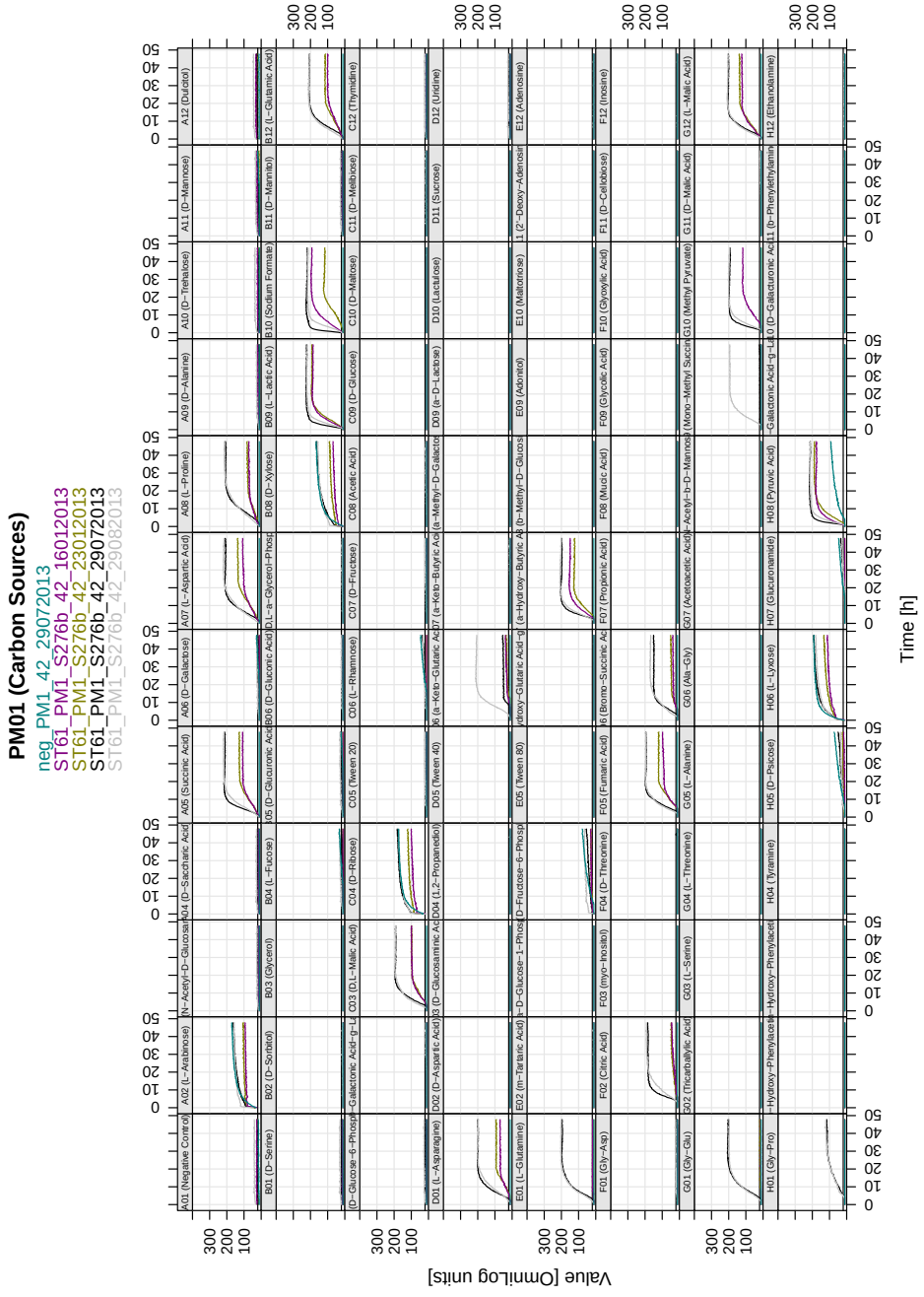


Figure A.50: xy plot for the *C. jejuni* isolate S276b in PM 1 at 42°C

A.2.26 Isolate S276b (ST-61) in PM 9 at 38°C and 42°C

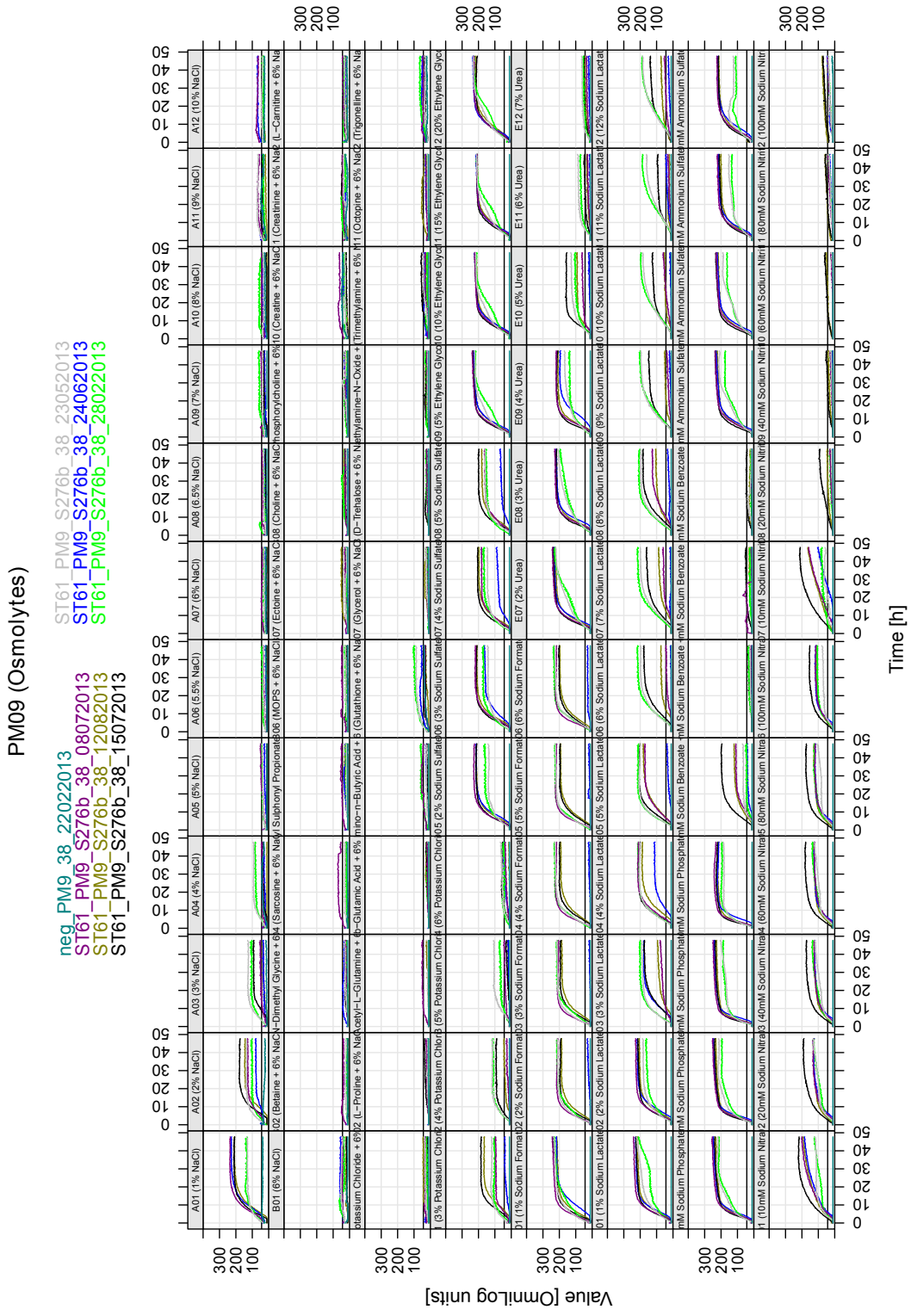


Figure A.51: xy plot for the *C. jejuni* isolate S276b in PM 9 at 38°C

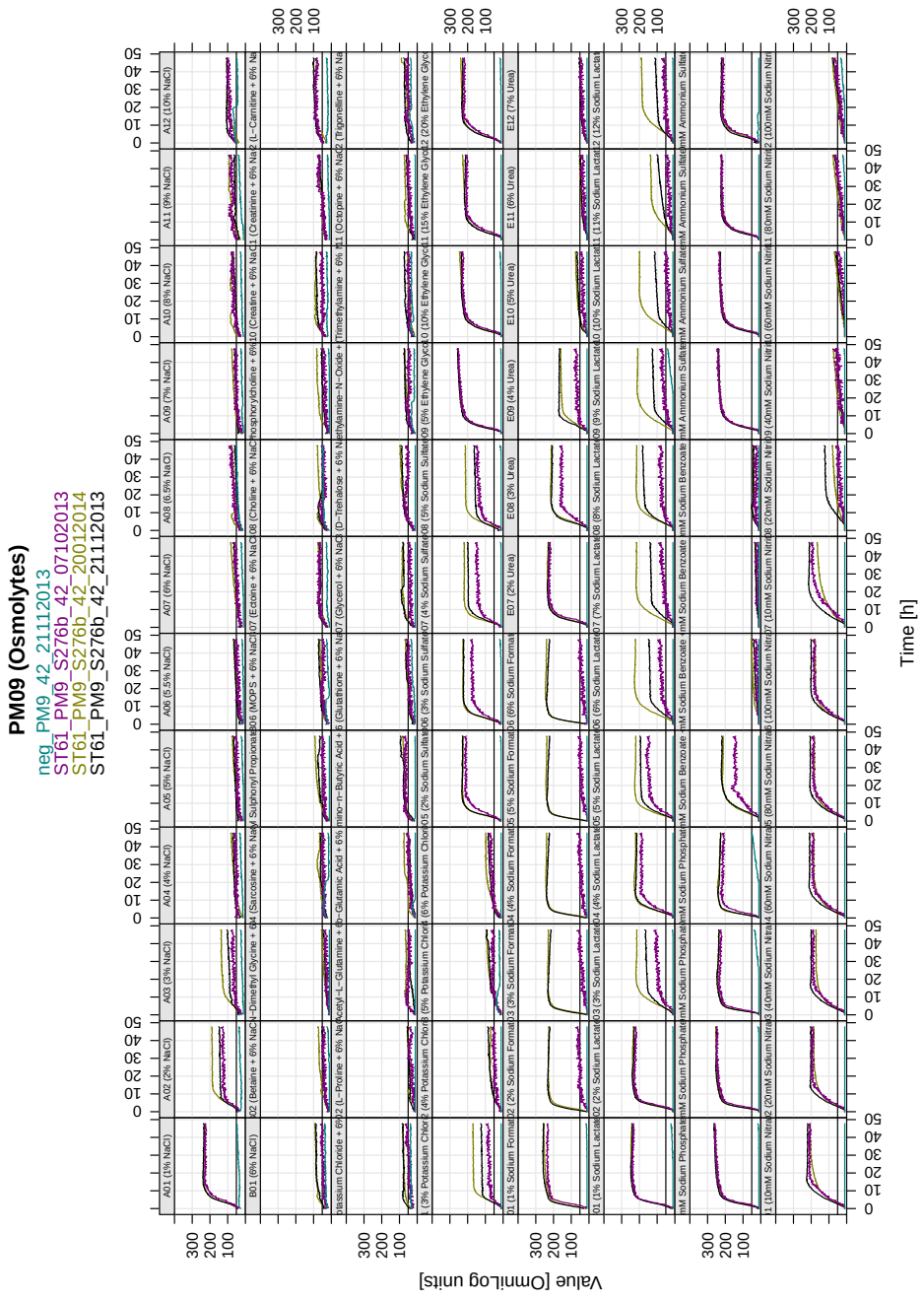


Figure A.52: xy plot for the *C. jejuni* isolate S276b in PM 9 at 42°C

A.2.27 Isolate S263b (ST-61) in PM 1 at 38°C and 42°C

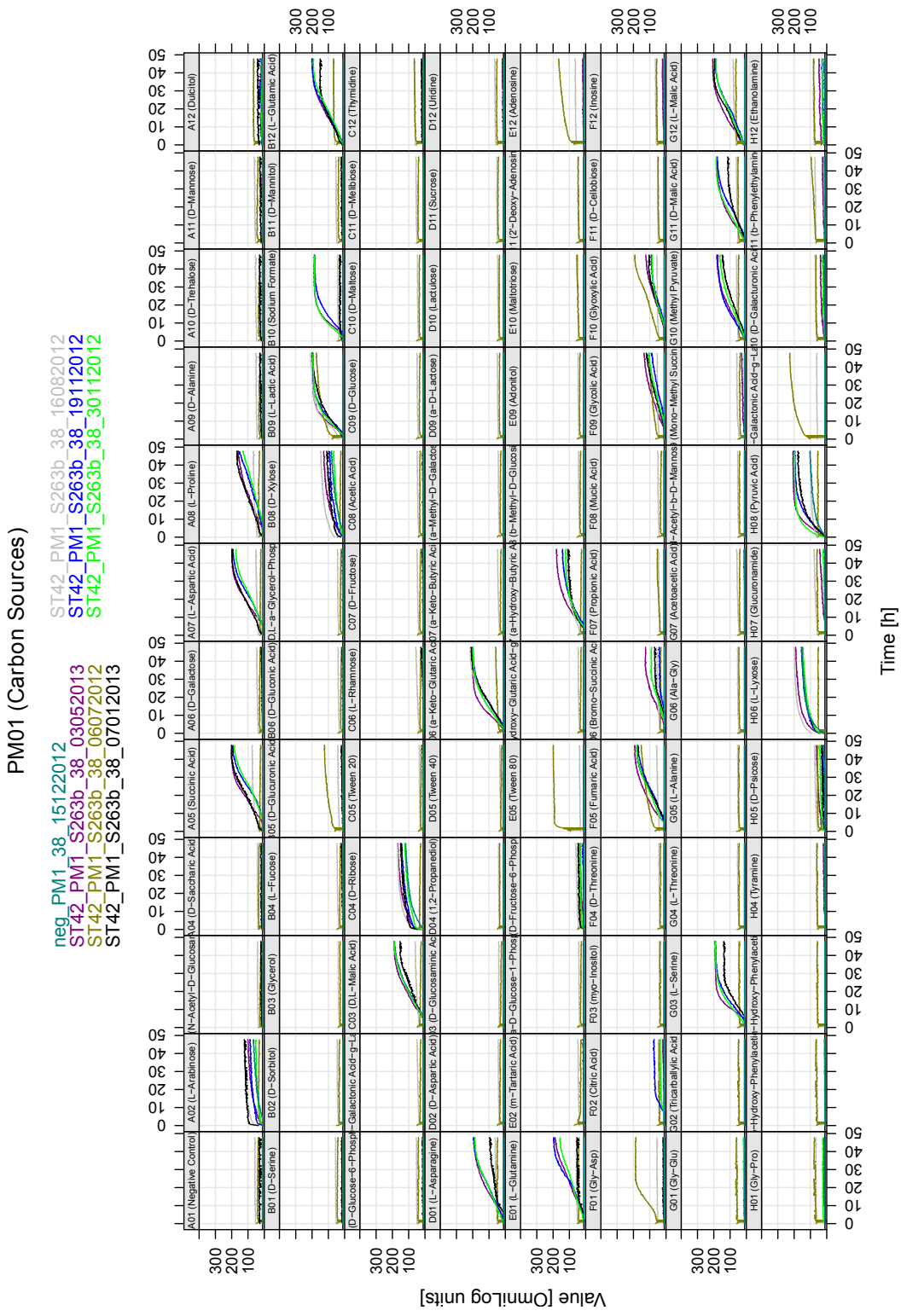


Figure A.53: xy plot for the *C. jejuni* isolate S263b in PM 1 at 38°C

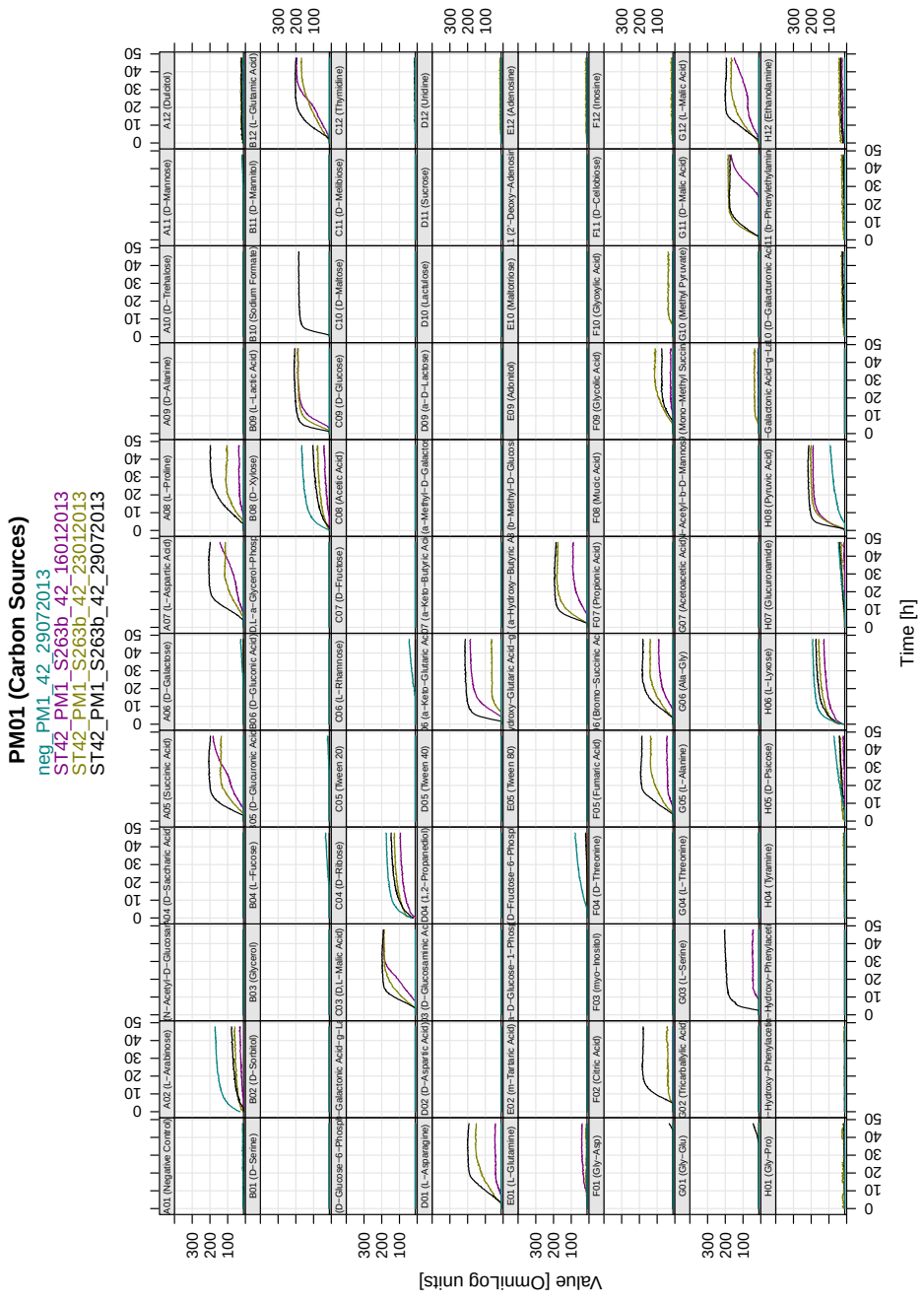


Figure A.54: xy plot for the *C. jejuni* isolate S263b in PM 1 at 42°C

A.2.28 Isolate S263b (ST-61) in PM 9 at 38°C and 42°C

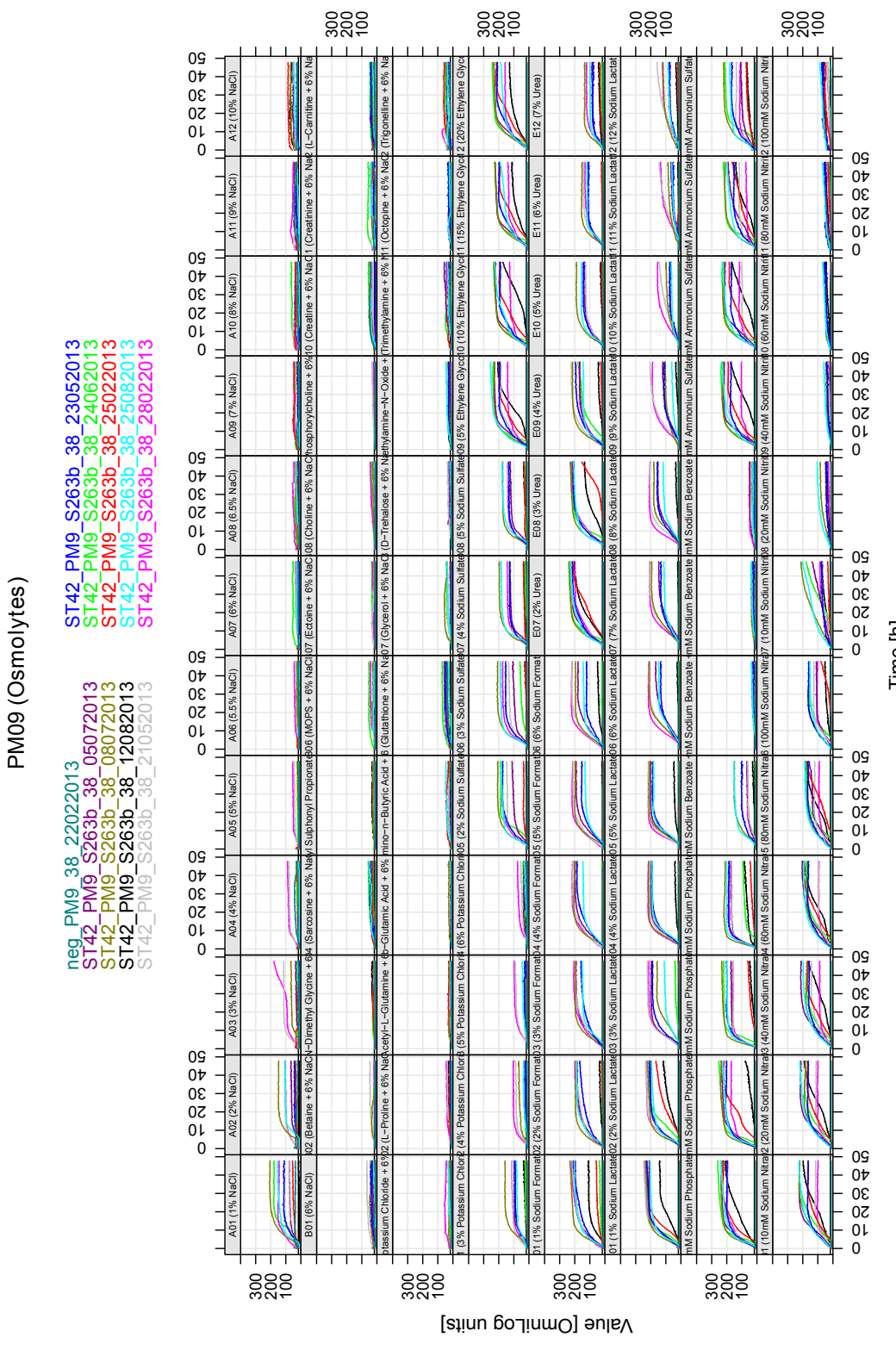


Figure A.55: xy plot for the *C. jejuni* isolate S263b in PM 9 at 38°C

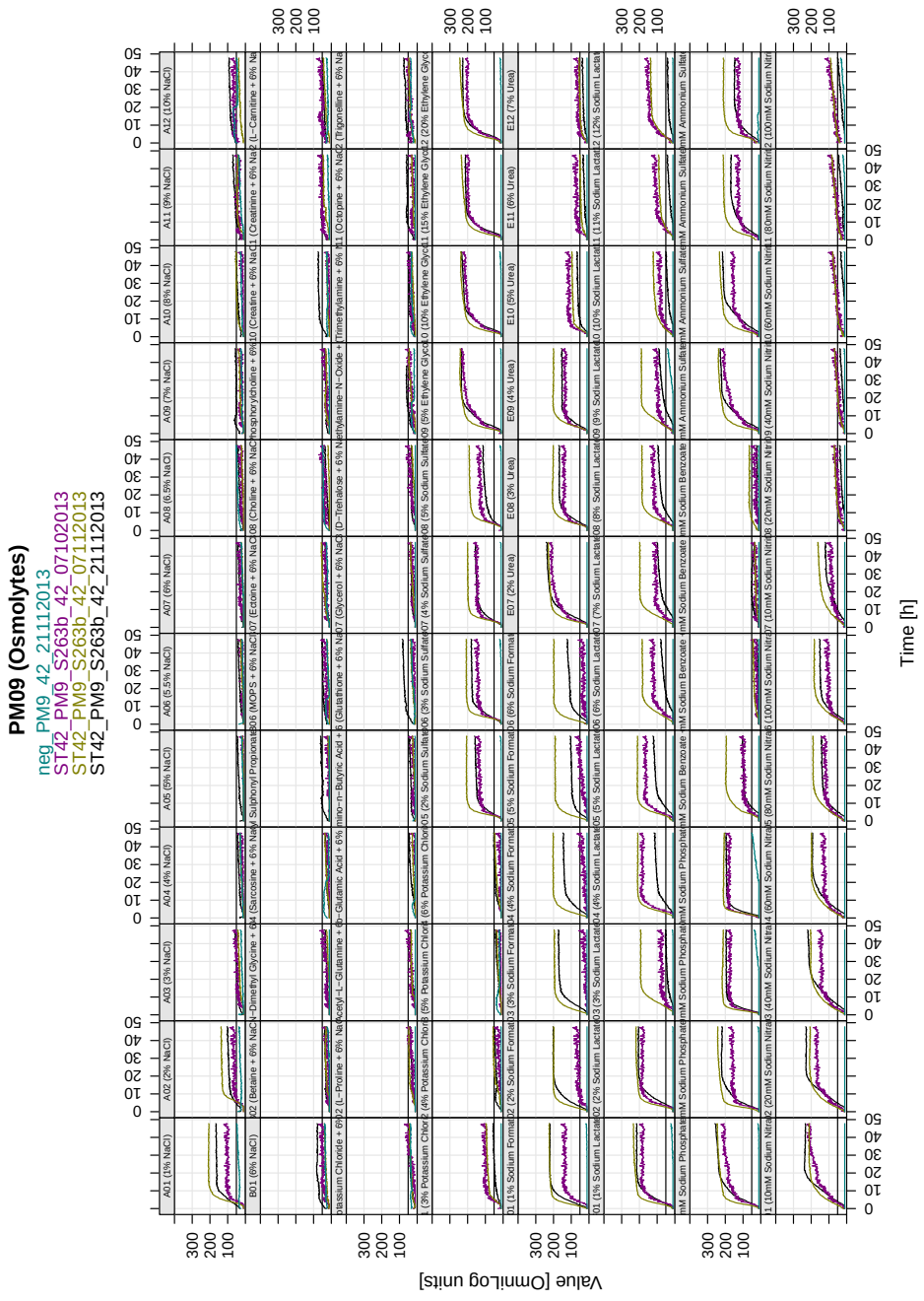


Figure A.56: xy plot for the *C. jejuni* isolate S263b in PM 9 at 42°C

A.2.29 Isolate H180 (ST-42) in PM 1 at 38°C and 42°C

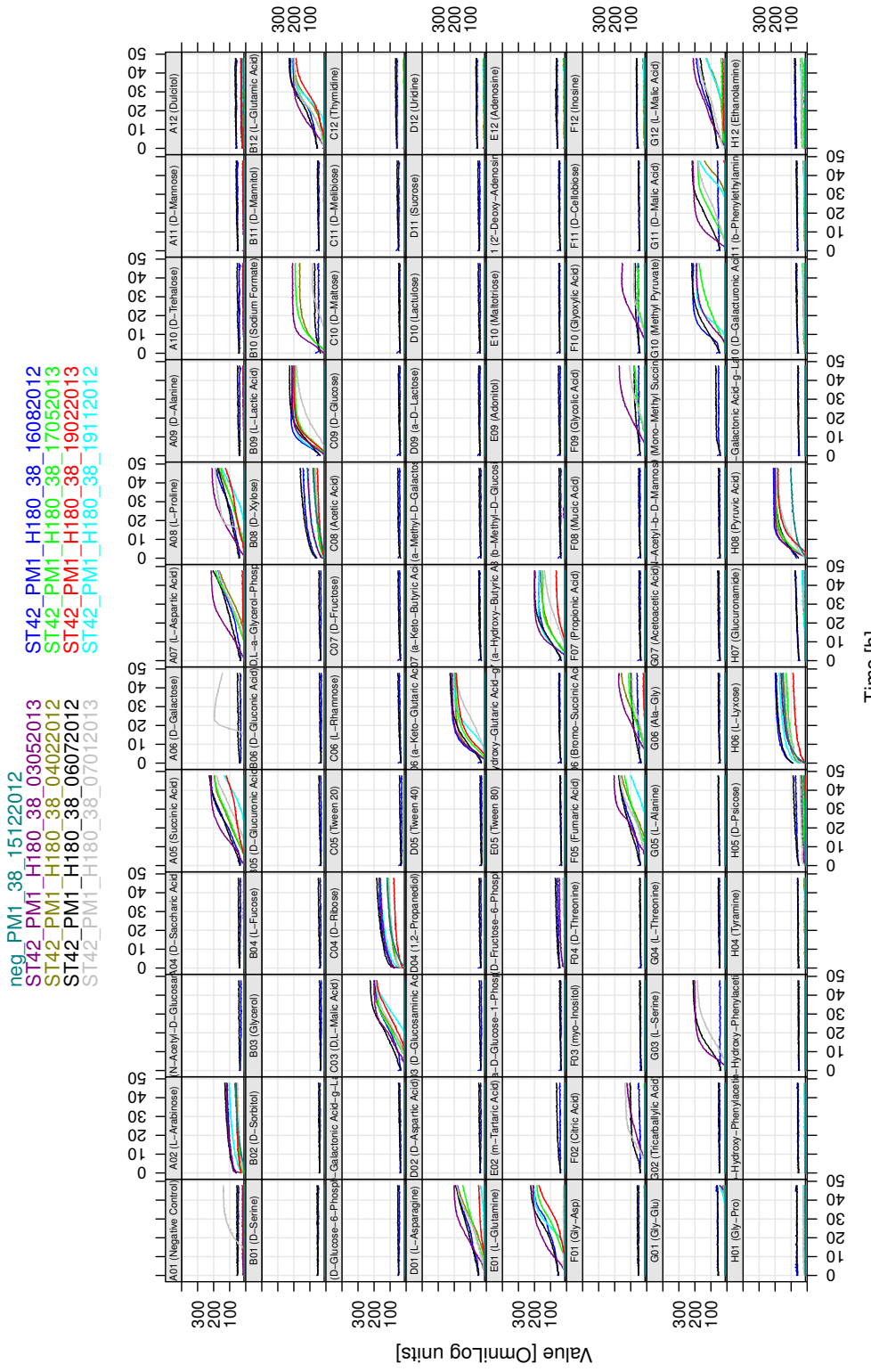


Figure A.57: xy plot for the *C. jejuni* isolate H180 in PM 1 at 38°C

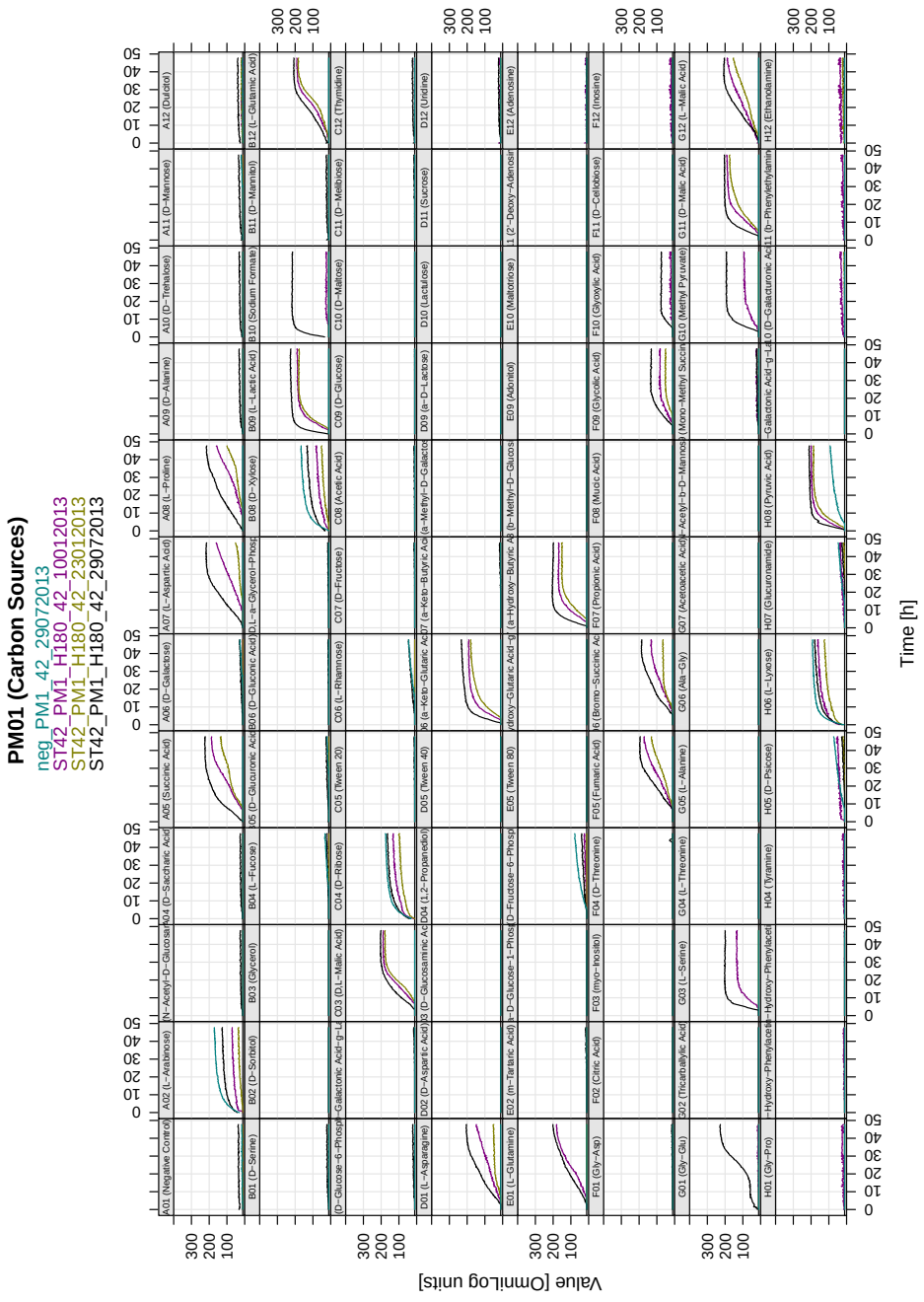


Figure A.58: xy plot for the *C. jejuni* isolate H180 in PM 1 at 42°C

A.2.30 Isolate H180 (ST-42) in PM 9 at 38°C and 42°C

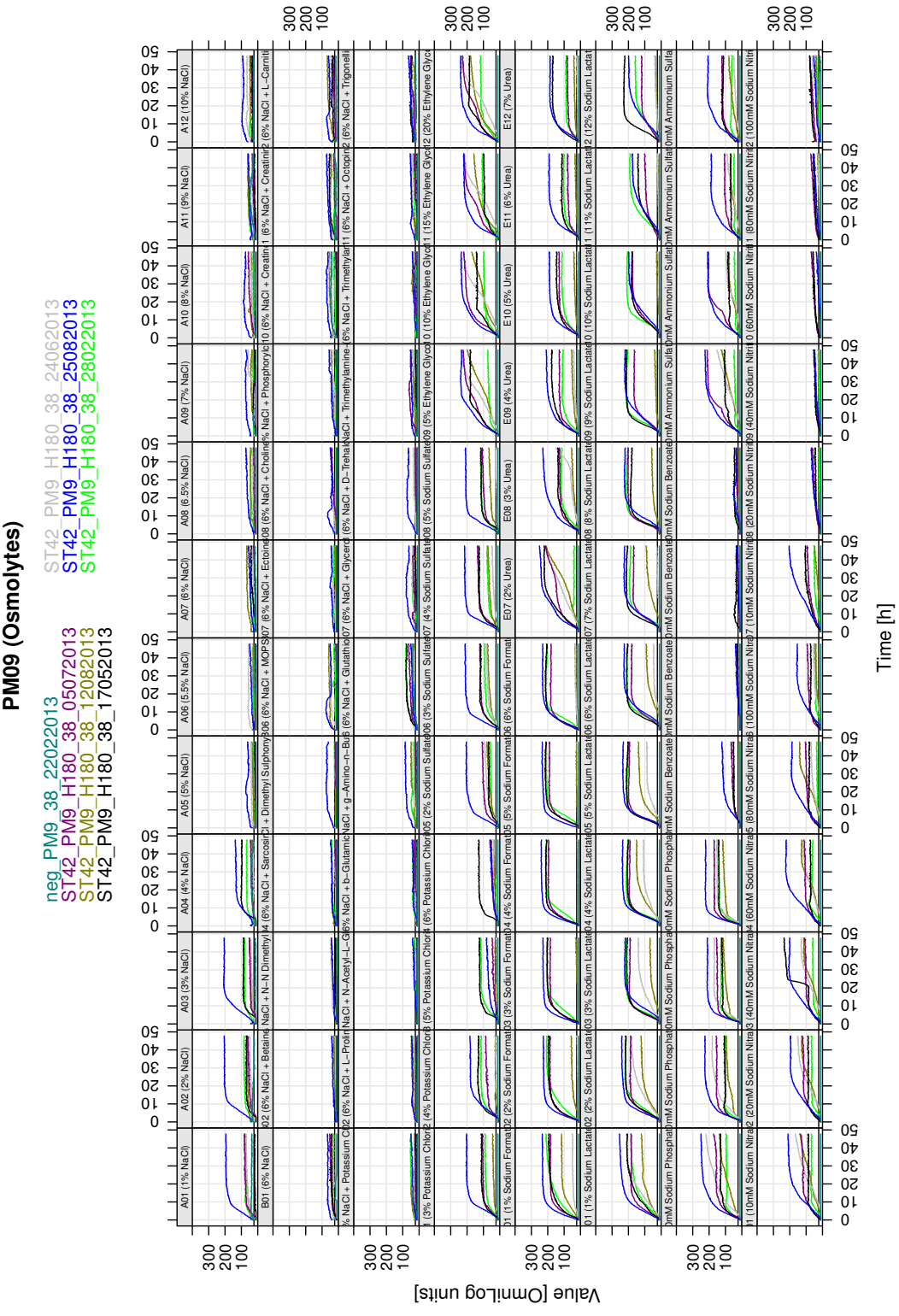


Figure A.59: xy plot for the *C. jejuni* isolate H180 in PM 9 at 38°C

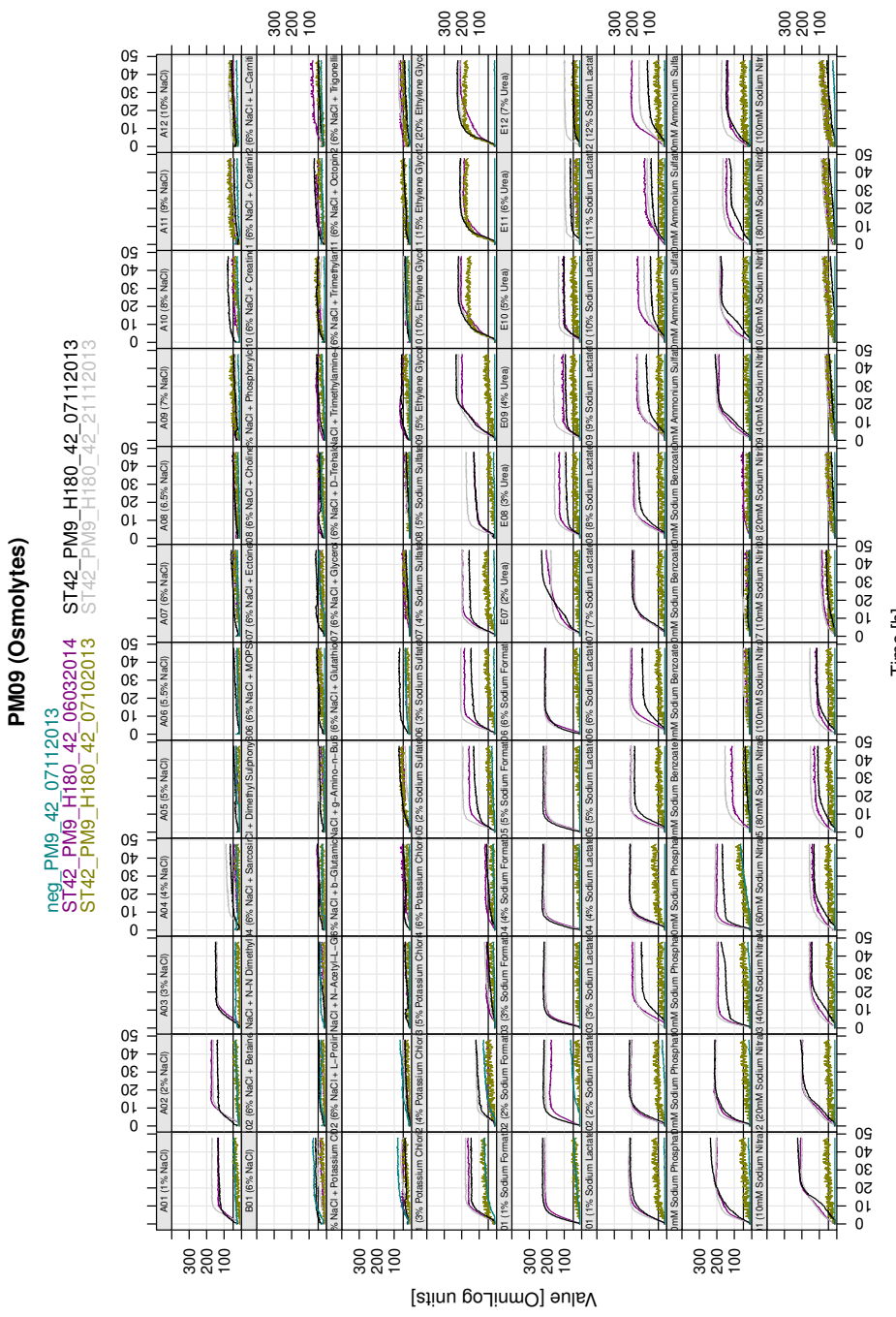


Figure A.60: xy plot for the *C. jejuni* isolate H180 in PM 9 at 42°C

A.3 heatmaps

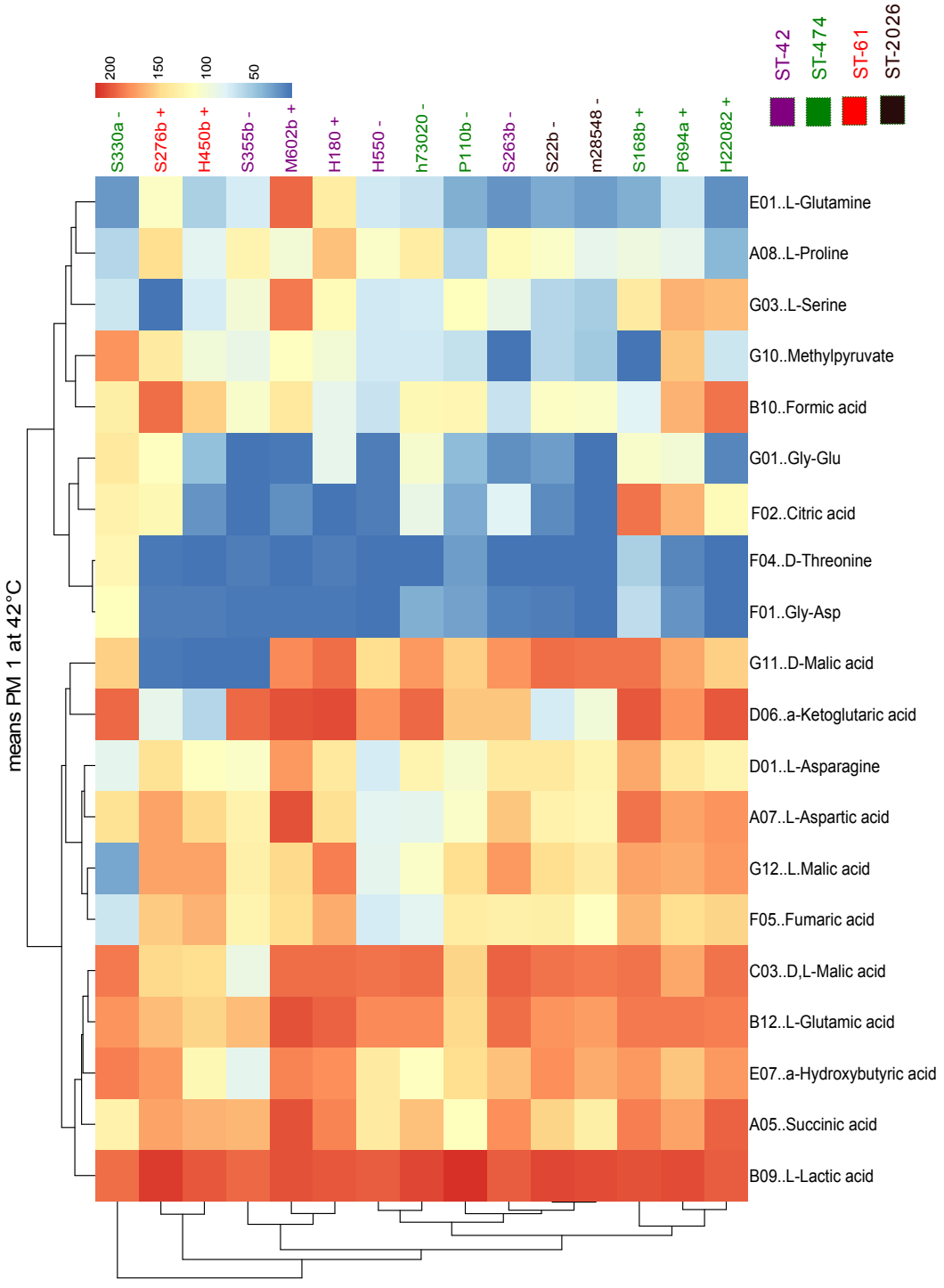


Figure A.61: Heatmap of the means of the 15 *C. jejuni* isolates in PM 1 at 42°C

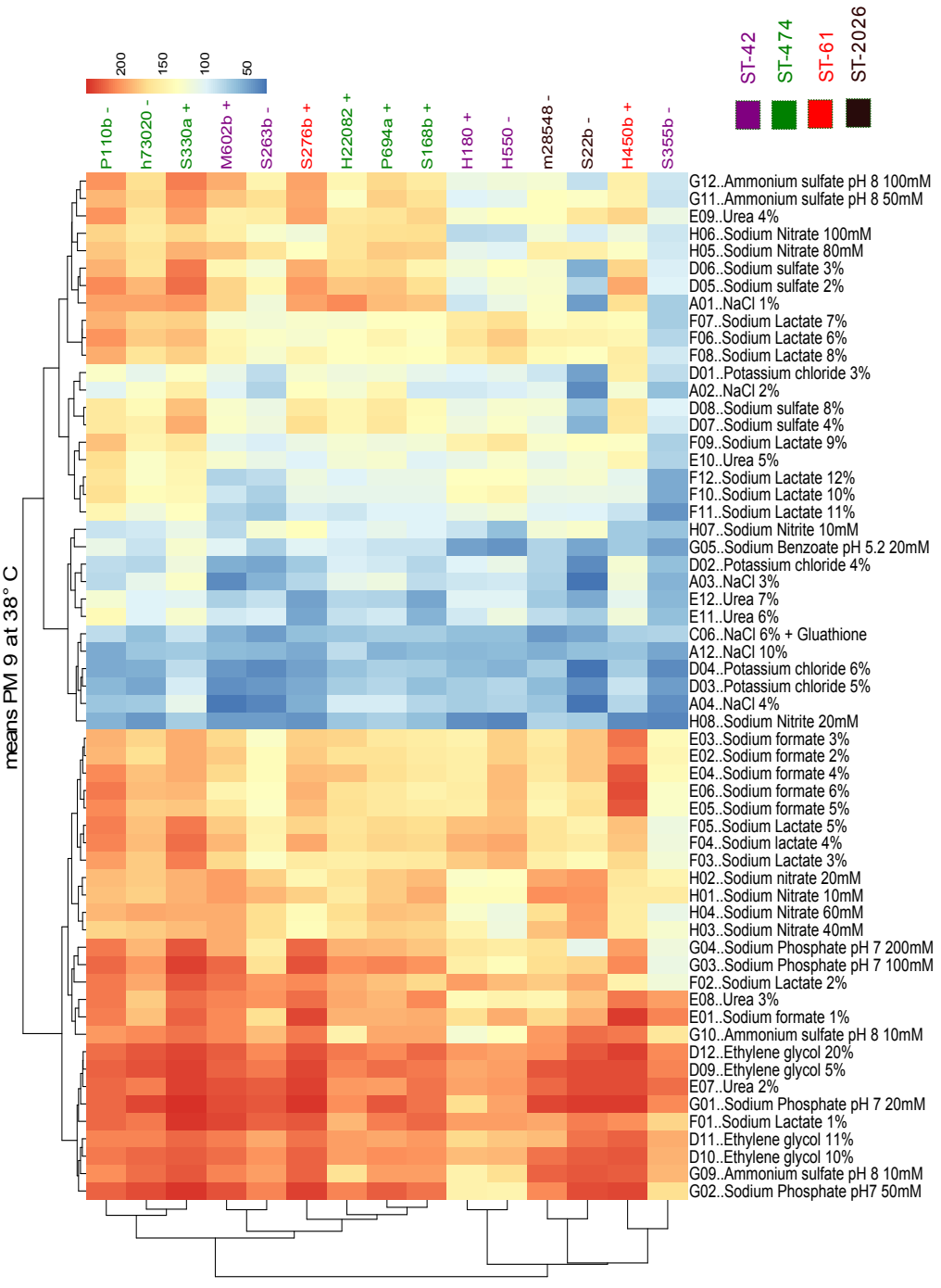


Figure A.62: Heatmap of the means of the 15 *C. jejuni* isolates in PM 9 at 38°C

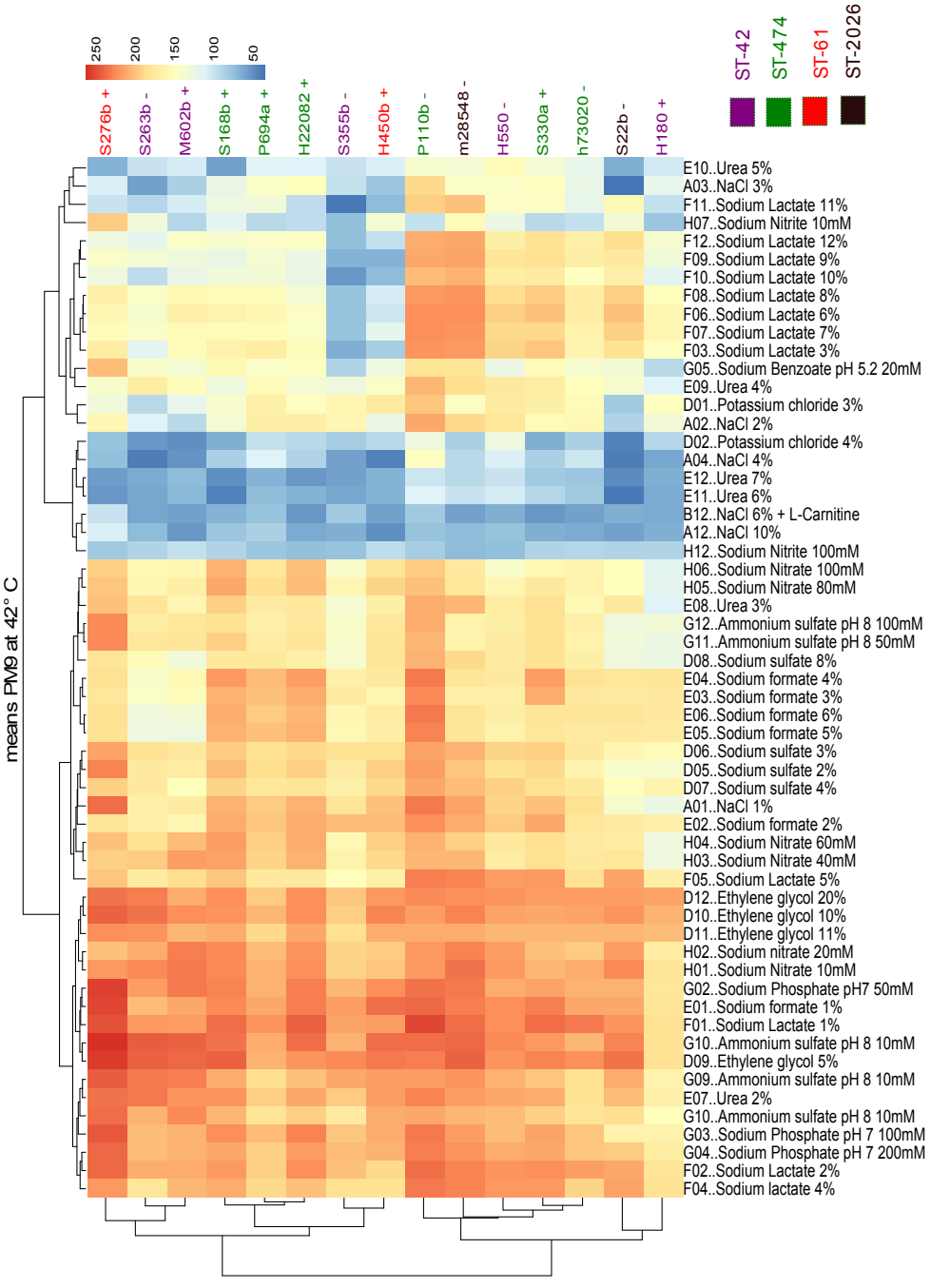


Figure A.63: Heatmap of the means of the 15 *C. jejuni* isolates in PM 9 at 42°C

Bibliography

Achtman, M.; Zurth, K.; Morelli, G.; Torrea, G.; Guiyoule, A., and Carniel, E. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proceedings of the National Academy of Sciences*, 96(24):14043, 1999.

Acke, E.; McGill, K.; Golden, O.; Jones, B.R.; Fanning, S., and Whyte, P. Prevalence of thermophilic *Campylobacter* species in household cats and dogs in Ireland. *Veterinary Record*, 164(2):44–47, 2009.

Adak, G.K.; Cowden, J.M.; Nicholas, S., and Evans, H.S. The public health laboratory service national case-control study of primary indigenous sporadic cases of *Campylobacter* infection. *Epidemiology and infection*, 115(01):15–22, 1995.

Adak, G.K.; Meakins, S.M.; Yip, H.; Lopman, B.A., and O'Brien, S.J. Disease risks from foods, England and Wales, 1996–2000. *Emerging Infectious Diseases*, 11(3):365–72, 2005.

Adékambi, T.; Berger, P.; Raoult, D., and Drancourt, M.l. *rpo*b gene sequence−based characterization of emerging non−tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 56(1):133–143, 2006.

Ailes, E.; Demma, L.; Hurd, S.; Hatch, J.; Jones, T. F; Vugia, D.; Cronquist, A.; Tobin-D'Angelo, M.; Larson, K., and Laine, E. Continued decline in the incidence of *Campylobacter* infections, FoodNet 1996−2006. *Foodborne Pathogens and Disease*, 5(3):329–337, 2008.

Alikhan, N.; Petty, N.K; Zakour, N.L.B., and Beatson, S.A. Blast ring image generator (brig): simple prokaryote genome comparisons. *BMC genomics*, 12(1):402, 2011.

Altizer, S.; Dobson, A.; Hosseini, P.; Hudson, P.; Pascual, M., and Rohani, P. Seasonality and the dynamics of infectious diseases. *Ecology Letters*, 9(4):467–484, 2006.

Arnold, C.; Westland, L.; Mowat, G.; Underwood, A.; Magee, J., and Gharbia, S.

Single−nucleotide polymorphism−based differentiation and drug resistance detection in *Mycobacterium tuberculosis* from isolates or directly from sputum. *Clinical Microbiology and Infection*, 11(2):122–130, 2005.

Arsenault, J.; Berke, O.; Michel, P.; Ravel, A., and Gosselin, P. Environmental and demographic risk factors for campylobacteriosis: do various geographical scales tell the same story? *BMC Infectious Diseases*, 12(1):318, 2012.

Ashgar, S.S.A.; Oldfield, N.J.; Wooldridge, K.G.; Jones, M.A.; Irving, G.J.; Turner, D.P.J., and Ala'Aldeen, D.A.A. CapA, an autotransporter protein of *Campylobacter jejuni*, mediates association with human epithelial cells and colonization of the chicken gut. *Journal of Bacteriology*, 189(5):1856–1865, 2007.

Ashkenasy, N.; Sánchez-Quesada, J.; Bayley, H., and Ghadiri, M.R. Recognizing a single base in an individual dna strand: a step toward dna sequencing in nanopores. *Angewandte Chemie*, 117(9):1425–1428, 2005.

Atack, J.M. and Kelly, D.J. Oxidative stress in *Campylobacter jejuni*: responses, resistance and regulation. *Future Microbiology*, 4(6):677–690, 2009.

Babakhani, F.K.; Bradley, G.A., and Joens, L.A. Newborn piglet model for campylobacteriosis. *Infection and Immunity*, 61(8):3466–3475, 1993.

Bacon, D.J.; Alm, R.A.; Burr, D.H.; Hu, L.; Kopecko, D.J.; Ewing, C.P., and Guerry, P. Involvement of a plasmid in virulence of *Campylobacter jejuni* 81-176. *Infection and Immunity*, 68(8):4384–4390, 2000.

bacterio.net, LPSN. List of prokaryotic names with standing in nomenclature. online, August 2014. URL `http://www.bacterio.net/-allnamesac.html`. last accessed on 5th of August 2014.

Baggesen, D.L.; Sorensen, G.; Nielsen, E.M., and Wegener, H.C. Phage typing of *Salmonella* typhimurium−is it still a useful tool for surveillance and outbreak investigation. *Eurosurveillance*, 15(4):19471, 2010.

Baker, M.G.; Wilson, N.A.; Ikram, R.; Chambers, S.; Shoemack, P., and Cook, G. Regulation of chicken contamination urgently needed to control New Zealand's serious campylobacteriosis epidemic. *New Zealand Medical Journal*, 119, 2006.

Barnes, If HA; Bagnall, Mary C; Browning, Darren D; Thompson, Stuart A; Manning, Georgina,

and Newell, Diane G. γ-glutamyl transpeptidase has a role in the persistent colonization of the avian gut by *Campylobacter jejuni*. *Microbial pathogenesis*, 43(5):198–207, 2007.

Bayjanov, J.R.; Molenaar, D.; Tzeneva, V.; Siezen, R.J., and van Hijum, S.A.F.T. PhenoLink−a web−tool for linking phenotype to˜ omics data for bacteria: application to gene-trait matching for *Lactobacillus plantarum* strains. *BMC Genomics*, 13(1):170, 2012.

Beaumont, H.J.E.; Gallie, J.; Kost, C.; Ferguson, G.C., and Rainey, P.B. Experimental evolution of bet hedging. *Nature*, 462(7269):90–93, 2009.

Beery, J.T.; Hugdahl, M.B., and Doyle, M.P. Colonization of gastrointestinal tracts of chicks by *Campylobacter jejuni*. *Applied and Environmental Microbiology*, 54(10):2365–2370, 1988.

Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27(2):573, 1999.

Berndtson, E.; Emanuelson, U.; Engvall, A., and Danielsson-Tham, M. A 1-year epidemiological study of campylobacters in 18 swedish chicken farms. *Preventive Veterinary Medicine*, 26(3): 167–185, 1996.

Best, E.L.; Fox, A.J.; Owen, R.J.; Cheesbrough, J., and Bolton, F.J. Specific detection of *Campylobacter jejuni* from faeces using single nucleotide polymorphisms. *Epidemiology and Infection*, 135(05):839–846, 2007.

Bi, P.; Cameron, A.S.; Zhang, Y., and Parton, K.A. Weather and notified *Campylobacter* infections in temperate and sub-tropical regions of Australia: An ecological study. *Journal of Infection*, 57(4):317–323, 2008.

Biggs, P.J.; Fearnhead, P.; Hotter, G.; Mohan, V.; Collins-Emerson, J.; Kwan, E.; Besser, T.E.; Cookson, A.; Carter, P.E., and French, N.P. Whole−genome comparison of two *Campylobacter jejuni* isolates of the same sequence type reveals multiple loci of different ancestral lineage. *PLoS One*, 6(11):e27121, 2011.

Bingham-Ramos, L.K. and Hendrixson, D.R. Characterization of two putative cytochrome c peroxidases of *Campylobacter jejuni* involved in promoting commensal colonization of poultry. *Infection and Immunity*, 76(3):1105–1114, 2008.

Biswas, D.; Itoh, K., and Sasakawa, C. Uptake pathways of clinical and healthy animal isolates of *Campylobacter jejuni* into int-407 cells. *FEMS Immunology & Medical Microbiology*, 29(3): 203–211, 2000.

Biswas, D.; Itoh, K., and Sasakawa, C. Role of microfilaments and microtubules in the invasion of int-407 cells by *Campylobacter jejuni*. *Microbiology and Immunology*, 47(6):469–473, 2003.

Blaser, M.J.; Cravens, J.; Powers, B.W.; Laforce, F.M., and Wang, W.L. *Campylobacter* enteritis associated with unpasteurized milk. *The American Journal of Medicine*, 67(4):715–718, 1979.

Bochner, B.R. New technologies to assess genotype–phenotype relationships. *Nature Reviews Genetics*, 4(4):309–314, 2003.

Bochner, B.R. Global phenotypic characterization of bacteria. *FEMS Microbiology Reviews*, 33 (1):191–205, 2009.

Bochner, B.R. Phenomics and phenotype microarrays: applications complementing metagenomics. *Handbook of Molecular Microbial Ecology I: Metagenomics and Complementary Approaches*, pages 533–540, 2011.

Bochner, B.R. and Savageau, M.A. Generalized indicator plate for genetic, metabolic, and taxonomic studies with microorganisms. *Applied and Environmental Microbiology*, 33(2):434–444, 1977.

Bochner, B.R.; Gadzinski, P., and Panomitros, E. Phenotype microarrays for high−throughput phenotypic testing and assay of gene function. *Genome Research*, 11(7):1246–1255, 2001.

Bolton, F.J.; Holt, A.V., and Hutchinson, D.N. *Campylobacter* biotyping scheme of epidemiological value. *Journal of Clinical Pathology*, 37(6):677, 1984.

Bourke, B.; Sherman, P.M.; Woodward, D.; Lior, H., and Chan, V.L. Pulsed−field gel electrophoresis indicates genotypic heterogeneity among *Campylobacter upsaliensis* strains. *FEMS Microbiology Letters*, 143(1):57–61, 1996.

Branton, D.; Deamer, D.W.; Marziali, A.; Bayley, H.; Benner, S.A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A., and Huang, X. The potential and challenges of nanopore sequencing. *Nature Biotechnology*, 26(10):1146–1153, 2008.

Brás, A.M.; Chatterjee, S.; Wren, B.W.; Newell, D.G., and Ketley, J.M. A novel *Campylobacter jejuni* two−component regulatory system important for temperature−dependent growth and colonization. *Journal of Bacteriology*, 181(10):3298–3302, 1999.

Bromham, L. and Penny, D. The modern molecular clock. *Nature Reviews Genetics*, 4(3): 216–224, 2003.

Brown, P.E.; Christensen, O.F.; Clough, H.E.; Diggle, P.J.; Hart, C.A.; Hazel, S.; Kemp, R.;

Leatherbarrow, A.J.H.; Moore, A., and Sutherst, J. Frequency and spatial distribution of environmental *Campylobacter* spp. *Applied and Environmental Microbiology*, 70(11):6501–6511, 2004.

Buchanan, A.V.; Sholtis, S.; Richtsmeier, J., and Weiss, K.M. What are genes for or where are traits from? What is the question? *Bioessays*, 31(2):198–208, 2009.

Buermans, H.P.J. and den Dunnen, J.T. Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1842: 1932–1941, 2014. doi: DOI:10.1016/j.bbadis.2014.06.015.

Bullman, S.; Corcoran, D.; O'Leary, J.; Lucey, B.; Byrne, D., and Sleator, R.D. *Campylobacter ureolyticus*: an emerging gastrointestinal pathogen? *FEMS Immunology & Medical Microbiology*, 61(2):228–230, 2011.

Cai, S.; Kabuki, D.Y.; Kuaye, A.Y.; Cargioli, T.G.; Chung, M.S.; Nielsen, R., and Wiedmann, M. Rational design of DNA sequence−based strategies for subtyping *Listeria monocytogenes*. *Journal of Clinical Microbiology*, 40(9):3319–3325, 2002.

Cameron, A.; Frirdich, E.; Huynh, S.; Parker, C.T., and Gaynor, E.C. Hyperosmotic stress response of *Campylobacter jejuni*. *Journal of Bacteriology*, 194(22):6116–6130, 2012.

Caro-Quintero, A.; Rodriguez-Castano, G.P., and Konstantinidis, K.T. Genomic insights into the convergence and pathogenicity factors of *Campylobacter jejuni* and *Campylobacter coli* species. *Journal of Bacteriology*, 191(18):5824–5831, 2009.

Carter, P.E.; McTavish, S.M.; Brooks, H.J.L.; Campbell, D.; Collins-Emerson, J.M.; Midwinter, A.C., and French, N.P. Novel clonal complexes with an unknown animal reservoir dominate *Campylobacter jejuni* isolates from river water in New Zealand. *Applied and Environmental Microbiology*, 75(19):6038–6046, 2009.

Cascales, E. and Christie, P.J. The versatile bacterial type IV secretion systems. *Nature Reviews Microbiology*, 1(2):137–149, 2003.

Castellano, I. and Merlino, A. γ-glutamyltranspeptidases: sequence, structure, biochemical properties, and biotechnological applications. *Cellular and Molecular Life Sciences*, 69(20): 3381–3394, 2012.

Chaban, B.; Musil, K.M.; Himsworth, C.G., and Hill, J.E. Development of cpn60-based real-time quantitative pcr assays for the detection of 14 *Campylobacter* species and application to

screening of canine fecal samples. *Applied and Environmental Microbiology*, 75(10):3055–3061, 2009.

Chaban, B.; Ngeleka, M., and Hill, J. Detection and quantification of 14 *Campylobacter* species in pet dogs reveals an increase in species richness in feces of diarrheic animals. *BMC Microbiology*, 10(1):73, 2010.

Chaloner, G.; Wigley, P.; Humphrey, S.; Kemmett, K.; Lacharme-Lora, L.; Humphrey, T., and Williams, N. Dynamics of dual infection with *Campylobacter jejuni* strains in chickens reveals distinct strain-to-strain variation in infection ecology. *Applied and Environmental Microbiology*, 80(20):6366–6372, 2014.

Champion, O.L.; Gaunt, M.W.; Gundogdu, O.; Elmi, A.; Witney, A.A.; Hinds, J.; Dorrell, N., and Wren, B.W. Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44):16043–16048, 2005.

Chang, C. and Miller, J.F. *Campylobacter jejuni* colonization of mice with limited enteric flora. *Infection and Immunity*, 74(9):5261–5271, 2006.

Chang, P.L.; Hsieh, W.S.; Chiang, C.L.; Tuohy, M.J.; Hall, G.S.; Procop, G.W.; Chang, H.T., and Ho, H.T. The *hsp65* gene patterns of less common *Mycobacterium* and *Nocardia* spp. by polymerase chain reaction-restriction fragment length polymorphism analysis with capillary electrophoresis. *Diagnostic Microbiology and Infectious Disease*, 58(3):315–323, 2007.

Chen, F.; Mackey, A.J.; Stoeckert, C.J., and Roos, D.S. Orthomcl-db: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Research*, 34(suppl 1):D363–D368, 2006a.

Chen, S.L.; Hung, C.S.; Xu, J.; Reigstad, C.S.; Magrini, V.i.; Sabo, A.; Blasiar, D.; Bieri, T.; Meyer, R.R., and Ozersky, P. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proceedings of the National Academy of Sciences*, 103(15):5977–5982, 2006b.

Chin, C.S.; Alexander, D.H.; Marks, P.; Klammer, A.A.; Drake, J.; Heiner, C.; Clum, A.; Copeland, A.; Huddleston, J., and Eichler, E.E. Nonhybrid, finished microbial genome assemblies from long-read smrt sequencing data. *Nature Methods*, 10(6):563–569, 2013.

Chipley, J.R. sodium benzoate and benzoic acid. *Antimicrobials in Food*, pages 11–48, 2010.

Clark, C.G.; Bryden, L.; Cuff, W.R.; Johnson, P.L.; Jamieson, F.; Ciebin, B., and Wang, G.

Use of the Oxford multilocus sequence typing protocol and sequencing of the flagellin short variable region to characterize isolates from a large outbreak of waterborne *Campylobacter* sp. strains in Walkerton, Ontario, Canada. *Journal of Clinical Microbiology*, 43(5):2080–2091, 2005.

Clarke, J.; Wu, H.; Jayasinghe, L.; Patel, A.; Reid, S., and Bayley, H. Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology*, 4(4):265–270, 2009.

Clarke, K. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, 18(1):117–143, 1993.

Cleaveland, S.; Laurenson, M.K., and Taylor, L.H. Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1411):991–999, 2001.

Close, M.; Dann, R.; Ball, A.; Savill, M.; Pirie, R., and Smith, Z. Microbial groundwater quality and its health implications for a border-strip irrigated dairy farm catchment, South Island, New Zealand. *Journal of Water and Health*, 6(1):83–98, 2008.

Cody, A.J.; Colles, F.M.; Sheppard, S.K., and Maiden, M.C.J. Where does *Campylobacter* come from? A molecular odyssey. *Hot Topics in Infection and Immunity in Children VI; Advances in Experimental Medicine and Biology*, 659:47–56, 2010.

Cody, A.J.; McCarthy, N.M.; Wimalarathna, H.L.; Colles, F.M.; Clark, L.; Bowler, I.C.J.W.; Maiden, M.C.J., and Dingle, K.E. A longitudinal 6-year study of the molecular epidemiology of clinical *Campylobacter* isolates in Oxfordshire, United Kingdom. *Journal of Clinical Microbiology*, 50(10):3193–3201, 2012.

Cody, A.J.; McCarthy, N.D.; van Rensburg, M.J.; Isinkaye, T.; Bentley, S.D.; Parkhill, J.; Dingle, K.E.; Bowler, I.C.J.; Jolley, K.A., and Maiden, M.C.J. Real-time genomic epidemiological evaluation of human *Campylobacter* isolates by use of whole-genome multilocus sequence typing. *Journal of Clinical Microbiology*, 51(8):2526–2534, 2013.

Cohan, F.M. and Koeppel, A.F. The origins of ecological diversity in prokaryotes. *Current Biology*, 18(21):R1024–R1034, 2008.

Coker, A.O.; Isokpehi, R.D.; Thomas, B.N.; Amisu, K.O., and Obi, C.L. Human campylobacteriosis in developing countries. *Emerging Infectious Diseases*, 8(3):237—244, 2002.

Colles, F.M. and Maiden, M.C.J. *Campylobacter* sequence typing databases: applications and future prospects. *Microbiology*, 158(Pt 11):2695–2709, 2012.

Colles, F.M.; Jones, K.; Harding, R.M., and Maiden, M.C.J. Genetic diversity of *Campylobacter jejuni* isolates from farm animals and the farm environment. *Applied and Environmental Microbiology*, 69(12):7409–7413, 2003.

Colles, F.M.; Dingle, K.E.; Cody, A.J., and Maiden, M.C.J. Comparison of *Campylobacter* populations in wild geese with those in starlings and free-range poultry on the same farm. *Applied and Environmental Microbiology*, 74(11):3583–3590, 2008.

Cookson, A.; Biggs, P.; Hotter, G.; Fearnhead, P.; Carter, P., and French, N. Recombination events associated with *Campylobacter jejuni* ST-474 drive clonal diversity, population structure and evolution. In *Campylobacter, Helicobacter and related organisms*, 2011.

Corander, J. and Marttinen, P. Bayesian identification of admixture events using multilocus molecular markers. *Molecular Ecology*, 15(10):2833–2843, 2006.

Corander, J. and Tang, J. Bayesian analysis of population structure based on linked molecular information. *Mathematical biosciences*, 205(1):19–31, 2007.

Corander, J.; Marttinen, P.; Sirén, J., and Tang, J. BAPS: Bayesian analysis of population structure, Manual v. 4.1. *Department of Mathematics and Statistics, University of Helsinki*, 2006.

Corporation, 2013 Life Technologies. The Ion PGM TM System, with 400-base read length chemistry, enables routine high-quality de novo assembly of small genomes. online, 2013. APPLICATION NOTE.

Coward, C.; van Diemen, P.M.; Conlan, A.J.K.; Gog, J.R.; Stevens, M.; Jones, M.A., and Maskell, D.J. Competing isogenic *Campylobacter* strains exhibit variable population structures *in vivo*. *Applied and Environmental Microbiology*, 74(12):3857–3867, 2008.

Cox, M.P.; Peterson, D.A., and Biggs, P.J. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*, 11(1):485–490, 2010.

Cox, N.A.; Richardson, L.J.; Buhr, R.J.; Northcutt, J.K.; Bailey, J.S.; Cray, P.F., and Hiett, K.L. Recovery of *Campylobacter* and *Salmonella* serovars from the spleen, liver and gallbladder, and ceca of six-and eight-week-old commercial broilers. *The Journal of Applied Poultry Research*, 16(4):477–480, 2007.

Craig, J.L. Habitat variation in the social organization of a communal gallinule, the pukeko,

Porphyrio porphyrio melanotus. *Behavioral Ecology and Sociobiology*, 5(4):331–358, 1979. ISSN 0340-5443. doi: 10.1007/BF00292523. URL `http://dx.doi.org/10.1007/BF00292523`.

Croucher, N.J.; Harris, S.R.; Grad, Y.H., and Hanage, W.P. Bacterial genomes in epidemiologypresent and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1614):20120202, 2013.

Crump, J.A.; Murdoch, D.R., and Baker, M.G. Emerging infectious diseases in an island ecosystem: the New Zealand perspective. *Emerging Infectious Diseases*, 7(5):767–772, 2001.

de Boer, P.; Wagenaar, J.A.; Achterberg, R.P.; van Putten, J.P.M.; Schouls, L.M., and Duim, B. Generation of *Campylobacter jejuni* genetic diversity *in vivo*. *Molecular Microbiology*, 44 (2):351–359, 2002.

de Haan, C.P.A.; Llarena, A.-K.; Revez, J., and Hänninen, M.-L. Association of *Campylobacter jejuni* metabolic traits with multilocus sequence types. *Applied and Environmental Microbiology*, 78(16):5550–5554, 2012.

Deamer, D.W. and Akeson, M. Nanopores and nucleic acids: prospects for ultrarapid sequencing. *Trends in Biotechnology*, 18(4):147–151, 2000.

Denes, A.S.; Lutze-Wallace, C.L.; Cormier, M.L., and Garcia, M.M. DNA fingerprinting of *Campylobacter fetus* using cloned constructs of ribosomal RNA and surface array protein genes. *Veterinary Microbiology*, 54(2):185–193, 1997.

Dethlefsen, L.; Huse, S.; Sogin, M.L., and Relman, D.A. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. *PLoS Biology*, 6 (11):e280, 2008.

Devane, M.L.; Nicol, C.; Ball, A.; Klena, J.D.; Scholes, P.; Hudson, J.A.; Baker, M.G.; Gilpin, B.J.; Garrett, N., and Savill, M.G. The occurrence of *Campylobacter* subtypes in environmental reservoirs and potential transmission routes. *Journal of Applied Microbiology*, 98(4): 980–990, 2005.

Didelot, X. and Falush, D. Inference of bacterial microevolution using multilocus sequence data. *Genetics*, 175(3):1251–1266, 2007.

Dingle, K.E.; Colles, F.M.; Wareing, D.R.A.; Ure, R.; Fox, A.J.; Bolton, F.E.; Bootsma, H.J.; Willems, R.J.L.; Urwin, R., and Maiden, M.C.J. Multilocus sequence typing system for *Campylobacter jejuni*. *Journal of Clinical Microbiology*, 39(1):14–23, 2001.

Dingle, K.E.; Colles, F.M.; Ure, R.; Wagenaar, J.A.; Duim, B.; Bolton, F.J.; Fox, A.J.; Wareing,

D.R., and Maiden, M.C. Molecular characterization of *Campylobacter jejuni* clones: a basis for epidemiologic investigation. *Emerging Infectious Diseases*, 8(9):949–55, 2002.

Dingle, K.E.; Colles, F.M.; Falush, D., and Maiden, M. Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*. *Journal of Clinical Microbiology*, 43(1):340–347, 2005.

Doi, M.; Wachi, M.; Ishino, F.; Tomioka, S.; Ito, M.; Sakagami, Y.; Suzuki, A., and Matsuhashi, M. Determinations of the DNA sequence of the textitmreB gene and of the gene products of the mre region that function in formation of the rod shape of *Escherichia coli* cells. *Journal of Bacteriology*, 170(10):4619–4624, 1988.

Domingues, A.R.; Pires, S.M.; Halasa, T., and Hald, T. Source attribution of human campylobacteriosis using a meta-analysis of case-control studies of sporadic infections. *Epidemiology and Infection*, 140(06):970–981, 2012.

Doolittle, W.F. Microbial evolution: stalking the wild bacterial species. *Current Biology*, 18 (13):R565–R567, 2008.

Dorrell, N.; Mangan, J.A.; Laing, K.G.; Hinds, J.; Linton, D.; Al-Ghusein, H.; Barrell, B.G.; Parkhill, J.; Stoker, N.G., and Karlyshev, A.V. Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Research*, 11(10):1706–1715, 2001.

Doyle, M.P. and Roman, D.J. Response of *Campylobacter jejuni* to sodium chloride. *Applied and Environmental Microbiology*, 43(3):561–565, 1982.

Drancourt, M.; Roux, V.; Tran-Hung, L.; Castex, D.; Chenal-Francisque, V.; Ogata, H.; Fournier, P.-E.; Crubézy, E., and Raoult, D. Genotyping, orientalis-like *Yersinia pestis*, and plague pandemics. *Emerging Infectious Diseases*, 10(9):1585–1592, 2004.

Duim, B.; Wassenaar, T.M.; Rigter, A., and Wagenaar, J. High-resolution genotyping of *Campylobacter* strains isolated from poultry and humans with amplified fragment length polymorphism fingerprinting. *Applied and Environmental Microbiology*, 65(6):2369–2375, 1999.

Eberhart-Phillips, J.; Walker, N.; Garrett, N.; Bell, D.; Sinclair, D.; Rainger, W., and Bates, M. Campylobacteriosis in New Zealand: results of a case-control study. *Journal of Epidemiology and Community Health*, 51(6):686–691, 1997.

Edgar, R.C. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797, 2004.

Egelman, H. Actins prokaryotic homologs. *Current Opinion in Structural Biology*, 13(2):244–248, 2003.

Eisenbach, M. Bacterial chemotaxis. *eLS*, 2011.

Ellsworth, D.L.; Rittenhouse, K.D., and Honeycutt, R.L. Artifactual variation in randomly amplified polymorphic dna banding patterns. *Biotechniques*, 14(2):214–217, 1993.

Enright, M.C. and Spratt, B.G. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: Identification of clones associated with serious invasive disease. *Microbiology*, 144(11): 3049–3060, 1998.

Espedido, B.A.; Steen, J.A.; Ziochos, H.; Grimmond, S.M.; Cooper, M.A.; Gosbell, I.B.; van Hal, S.J., and Jensen, S.O. Whole genome sequence analysis of the first australian oxa-48-producing outbreak-associated *Klebsiella pneumoniae* isolates: The resistome and *in vivo* evolution. *PloS one*, 8(3):e59920, 2013.

Evans, M.R.; Roberts, R.J.; Ribeiro, C.D.; Gardner, D., and Kembrey, D. A milk-borne *Campylobacter* outbreak following an educational farm visit. *Epidemiology and Infection*, 117(03): 457–462, 1996.

Falush, D.; Kraft, C.; Taylor, N.S.; Correa, P.; Fox, J.G.; Achtman, M., and Suerbaum, S. Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proceedings of the National Academy of Sciences*, 98(26):15056–15061, 2001.

Falush, D.; Stephens, M., and Pritchard, J.K. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4):1567, 2003.

Fang, Z.; Morrison, N.; Watt, B.; Doig, C., and Forbes, K.J. Is6110 transposition and evolutionary scenario of the direct repeat locus in a group of closely related *Mycobacterium tuberculosis* strains. *Journal of Bacteriology*, 180(8):2102–2109, 1998.

Fearnhead, P.; Smith, N.G.C.; Barrigas, M.; Fox, A., and French, N. Analysis of recombination in *Campylobacter jejuni* from MLST population data. *Journal of Molecular Evolution*, 61(3): 333–340, 2005.

Febriani, Y.; Levallois, P.; Gingras, S.; Gosselin, P.; Majowicz, S.E., and Fleury, M.D. The association between farming activities, precipitation, and the risk of acute gastrointestinal illness in rural municipalities of Quebec, Canada: a cross-sectional study. *BMC Public Health*, 10(1):48, 2010.

Feil, E.J. and Spratt, B.G. Recombination and the population structures of bacterial pathogens. *Annual Reviews in Microbiology*, 55(1):561–590, 2001.

Feil, E.J.; Maiden, M.C.; Achtman, M., and Spratt, B.G. The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Molecular Biology and Evolution*, 16(11):1496–1502, 1999.

Feil, E.J.; Enright, M.C., and Spratt, B.G. Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between *Neisseria meningitidis* and *Streptococcus pneumoniae*. *Research in Microbiology*, 151(6):465–469, 2000.

Fermer, C. and Engvall, E.O. Specific pcr identification and differentiation of the thermophilic campylobacters, *Campylobacter jejuni*, *C. coli*, *C. lari*, and *C. upsaliensis*. *Journal of Clinical Microbiology*, 37(10):3370–3373, 1999.

Fernando, U.; Biswas, D.; Allan, B.; Willson, P., and Potter, A.A. Influence of *Campylobacter jejuni fliA*, *rpoN* and *flgK* genes on colonization of the chicken gut. *International Journal of Food Microbiology*, 118(2):194–200, 2007.

Field, L.H.; Headley, V.L.; Payne, S.M., and Berry, L.J. Influence of iron on growth, morphology, outer membrane protein composition, and synthesis of siderophores in *Campylobacter jejuni*. *Infection and Immunity*, 54(1):126–132, 1986a.

Field, L.H.; Headley, V.L.; Underwood, J.L.; Payne, S.M., and Berry, L.J. The chicken embryo as a model for *Campylobacter* invasion: comparative virulence of human isolates of *Campylobacter jejuni* and *Campylobacter coli*. *Infection and Immunity*, 54(1):118–125, 1986b.

Filliol, I.; Driscoll, J.R.; van Soolingen, D.; Kreiswirth, B.N.; Kremer, K.; Valétudie, G.; Anh, D.D.; Barlow, R.; Banerjee, D., and Bifani, P.J. Snapshot of moving and expanding clones of *Mycobacterium tuberculosis* and their global distribution assessed by spoligotyping in an international study. *Journal of Clinical Microbiology*, 41(5):1963–1970, 2003.

Finn, R.D.; Bateman, A.; Clements, J.; Coggill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L., and Mistry, J. Pfam: the protein families database. *Nucleic Acids Research*, pages D222–D230, 2013.

Fitzgerald, C.; Helsel, L.O.; Nicholson, M.A.; Olsen, S.J.; Swerdlow, D.L.; Flahart, R.; Sexton, J., and Fields, P.I. Evaluation of methods for subtyping *Campylobacter jejuni* during an outbreak involving a food handler. *Journal of Clinical Microbiology*, 39(7):2386–2390, 2001.

Fitzgerald, C.; Sails, A.D., and Fields, P.I. *Campylobacter Molecular and Cellular Biology*,

chapter *Campylobacter jejuni* strain variation, pages 59–77. Horizon Bioscience, Norfolk, UK, 2005.

Flanagan, R.C.; Neal-McKinney, J.M.; Dhillon, A.S.; Miller, W.G., and Konkel, M.E. Examination of *Campylobacter jejuni* putative adhesins leads to the identification of a new protein, designated FlpA, required for chicken colonization. *Infection and Immunity*, 77(6):2399–2407, 2009.

Fologea, D.; Gershow, M.; Ledden, B.; McNabb, D.S.; Golovchenko, J.A., and Li, J. Detecting single stranded dna with a solid state nanopore. *Nano Letters*, 5(10):1905–1909, 2005.

Foster, P.G. and Hickey, D.A. Compositional bias may affect both dna-based and protein-based phylogenetic reconstructions. *Journal of Molecular Evolution*, 48(3):284–290, 1999.

Fournier, P.-E.; Zhu, Y.; Ogata, H., and Raoult, D. Use of highly variable intergenic spacer sequences for multispacer typing of *Rickettsia conorii* strains. *Journal of Clinical Microbiology*, 42(12):5757–5766, 2004.

Fournier, P.-E.; Zhu, Y.; Yu, X., and Raoult, D. Proposal to create subspecies of *Rickettsia sibirica* and an emended description of *Rickettsia sibirica*. *Annals of the New York Academy of Sciences*, 1078(1):597–606, 2006.

Fouts, D.E.; Mongodin, E.F.; Mandrell, R.E.; Miller, W.G.; Rasko, D.A.; Ravel, J.; Brinkac, L.M.; DeBoy, R.T.; Parker, C.T., and Daugherty, S.C. Major structural differences and novel potential virulence mechanisms from the genomes of multiple *Campylobacter* species. *PLoS Biology*, 3(1):e15, 2005.

Freese, E. The difference between spontaneous and base-analogue induced mutations of phage t4. *Proceedings of the National Academy of Sciences of the United States of America*, 45(4): 622, 1959a.

Freese, E. The specific mutagenic effect of base analogues on phage t4. *Journal of Molecular Biology*, 1(2):87–105, 1959b.

French, N.P. Enhancing surveillance of potentially foodborne enteric diseases in New Zealand: Human campylobacteriosis in the Manawatu. *Hopkirk Institute, Massey University, New Zealand*, 2008.

French, N.P.; Barrigas, M.; Brown, P.; Ribiero, P.; Williams, N.; Leatherbarrow, H.; Birtles, R.; Bolton, E.; Fearnhead, P., and Fox, A. Spatial epidemiology and natural population structure

of *Campylobacter jejuni* colonizing a farmland ecosystem. *Environmental Microbiology*, 7(8): 1116–1126, 2005.

French, N.P.; Midwinter, A.; Holland, B.; Collins-Emerson, J.; Pattison, R.; Colles, F., and Carter, P. Molecular epidemiology of *Campylobacter jejuni* isolates from wild-bird fecal material in children's playgrounds. *Applied and Environmental Microbiology*, 75(3):779–783, 2009.

French, N.P.; Muellner, P., and Pleydell, E. Attribution of potentially foodborne enteric diseases: human campylobacteriosis− ruminant attribution. Report, Massey University, 2010.

French, N.P.; Yu, S.; Biggs, P.; Holland, B.; Fearnhead, P.; Binney, B.; Fox, A.; Grove-White, D.; Leigh, J.W.; Miller, W.; Muellner, P., and Carter, P. *Campylobacter Ecology and Evolution*, chapter Evolution of *Campylobacter* species in New Zealand, pages 221–240. Caister Academic Press, 2014.

Friedman, C.R.; Neimann, J.; Wegener, H.C., and Tauxe, R.V. *Campylobacter*, chapter Epidemiology of *Campylobacter jejuni* infections in the United States and other industrialized nations, pages 121–138. ASM press, 2000.

Friis, C.; Wassenaar, T.M.; Javed, M.A.; Snipen, L.; Lagesen, K.; Hallin, P.F.; Newell, D.G.; Toszeghy, M.; Ridley, A., and Manning, G. Genomic characterization of *Campylobacter jejuni* strain M1. *PLoS One*, 5(8):e12253, 2010.

Gaasbeek, E.J.; van der Wal, F.J.; van Putten, J.P.M.; de Boer, P.; van der Graaf-van Bloois, L.; de Boer, A.G.; Vermaning, B.J., and Wagenaar, J.A. Functional characterization of excision repair and RecA-dependent recombinational DNA repair in *Campylobacter jejuni*. *Journal of Bacteriology*, 191(12):3785–3793, 2009.

Gardner, S.N. and Hall, B.G. When whole-genome alignments just won't work: ksnp v2 software for alignment-free snp discovery and phylogenetics of hundreds of microbial genomes. *PloS One*, 8(12):e81760, 2013.

Gardner, S.N. and Slezak, T. Scalable snp analyses of 100+ bacterial or viral genomes. *Journal of Forensic Research*, 1(2):107, 2010.

Garénaux, A.; Guillou, S.; Ermel, G.; Wren, B., and M.and RitzFederighi, M. Role of the cj1371 periplasmic protein and the cj0355c two-component regulator in the *Campylobacter jejuni* nctc 11168 response to oxidative stress caused by paraquat. *Research in Microbiology*, 159(9): 718–726, 2008a.

Garénaux, A.; Lucchetti-Miganeh, C.; Ermel, G.; Barloy-Hubler, F.; de Jonge, R.; Newell, D.;

Payot, S.; Federighi, M.; Tresse, O.; Giuillou, S., and Ritz, M. *Better understanding of the Campylobacter conundrum*. Nova Science Publishers, Inc., 2008b.

Ge, B.; Wang, F.; Sjölund-Karlsson, M., and McDermott, P.F. Antimicrobial resistance in *Campylobacter*: Susceptibility testing methods and resistance trends. *Journal of Microbiological Methods*, 95(1):57–67, 2013.

Gelman, A. and Rubin, D.B. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.

Gerner-Smidt, P.; Hise, K.; Kincaid, J.; Hunter, S.; Rolando, S.; Hyytiä-Trees, E.; Ribot, E.M., and Swaminathan, B. PulseNet USA: a five-year update. *Foodbourne Pathogens & Disease*, 3 (1):9–19, 2006.

Gibbons, C.L.; Mangen, M.-J..J; Plass, D.; Havelaar, A.H.; Brooke, R.J.; Kramarz, P.; Peterson, K.L.; Stuurman, A.L.; Cassini, A., and Fèvre, E.M. Measuring underreporting and under-ascertainment in infectious disease datasets: a comparison of methods. *BMC Public Health*, 14(1):147, 2014.

Gibson, J.R.; Sutherland, K., and Owen, R.J. Inhibition of DNAse activity in PFGE analysis of DNA from *Campylobacter jejuni*. *Letters in Applied Microbiology*, 19(5):357–358, 1994.

Gillespie, I.A.; OBrien, S.J.; Frost, J.A.; Adak, G.K.; Horby, P.; Swan, A.V.; Painter, M.J., and Neal, K.R. A case-case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: a tool for generating hypotheses. *Emerging Infectious Diseases*, 8(9):937–942, 2002.

Gilpin, B.J.; Thorrold, B.; Scholes, P.; Longhurst, R.D.; Devane, M.; Nicol, C.; Walker, S.; Robson, B., and Savill, M. Comparison of *Campylobacter jejuni* genotypes from dairy cattle and human sources from the Matamata-Piako District of New Zealand. *Journal of Applied Microbiology*, 105(5):1354–1360, 2008.

Glazunova, O.; Roux, V.; Freylikman, O.; Sekeyova, Z.; Fournous, G.; Tyczka, J.; Tokarevich, N.; Kovacava, E.; Marrie, T.J., and Raoult, D. *Coxiella burnetii* genotyping. *Emerging Infectious Diseases*, 11(8):1211–1217, 2005.

Gottesman, S. Bacterial regulation: global regulatory networks. *Annual Review of Genetics*, 18 (1):415–441, 1984.

Gow, A.G.; Gow, D.J.; Hall, E.J.; Langton, D.; Clarke, C., and Papasouliotis, K. Prevalence of potentially pathogenic enteric organisms in clinically healthy kittens in the UK. *Journal of Feline Medicine & Surgery*, 11(8):655–662, 2009.

Grajewski, B.A.; Kusek, J.W., and Gelfand, H.M. Development of a bacteriophage typing system for *Campylobacter jejuni* and *Campylobacter coli*. *Journal of Clinical Microbiology*, 22 (1):13–18, 1985.

Grant, C. C.; Konkel, M. E.; Cieplak, W., and Tompkins, L. S. Role of flagella in adherence, internalization, and translocation of *Campylobacter jejuni* in nonpolarized and polarized epithelial cell cultures. *Infection and immunity*, 61(5):1764–1771, 1993.

Gras, L.M.; Smid, J.H.; Wagenaar, J.A.; de Boer, A.G.; Havelaar, A.H.; Friesema, I.H.M.; French, N.P.; Busani, L., and van Pelt, W. Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: a combined case-control and source attribution analysis. *PloS One*, 7(8):e42599, 2012.

Gras, L.M.; Smid, J.H.; Wagenaar, J.A.; Koene, M.G.J.; Havelaar, A.H.; Friesema, I.H.M.; French, N.P.; Flemming, C.; Galson, J.D., and Graziani, C. Increased risk for *Campylobacter jejuni* and *C. coli* infection of pet origin in dog owners and evidence for genetic association between strains causing infection in humans and their pets. *Epidemiology and Infection*, 141 (12):2526–2535, 2013.

Gregory, E.; Barnhart, H.; Dreesen, D.W.; Stern, N.J., and Corn, J.L. Epidemiological study of *Campylobacter* spp. in broilers: source, time of colonization, and prevalence. *Avian Diseases*, pages 890–898, 1997.

Griekspoor, P.; Colles, F.M.; McCarthy, N.D.; Hansbro, P.M.; Ashhurst-Smith, C.; Olsen, B.; Hasselquist, D.; Maiden, M.C.J., and Waldenström, J. Marked host specificity and lack of phylogeographic population structure of *Campylobacter jejuni* in wild birds. *Molecular Ecology*, 22:1463–1472, 2013.

Gripp, E.; Hlahla, D.; Didelot, X.; Kops, F.; Maurischat, S.; Tedin, K.; Alter, T.; Ellerbroek, L.; Schreiber, K., and Schomburg, D. Closely related *Campylobacter jejuni* strains from different sources reveal a generalist rather than a specialist lifestyle. *BMC Genomics*, 12(1):584, 2011.

Groenen, P.; Bunschoten, A.E.; van Soolingen, D., and van Errtbden, J.D.A. Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Molecular Microbiology*, 10(5):1057–1065, 1993.

Gundogdu, O.; Bentley, S.; Holden, M.; Parkhill, J.; Dorrell, N., and Wren, B. Re-annotation and re-analysis of the *Campylobacter jejuni* NCTC11168 genome sequence. *BMC Genomics*, 8(1):162, 2007.

Gundry, C.N.; Vandersteen, J.G.; Reed, G.H.; Pryor, R.J.; Chen, J., and Wittwer, C.T. Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. *Clinical Chemistry*, 49(3):396–406, 2003.

Gygi, S.P.; Rochon, Y.; Franza, B.R., and Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology*, 19(3):1720–1730, 1999.

Habib, I.; Uyttendaele, M., and De Zutter, L. Survival of poultry-derived *Campylobacter jejuni* of multilocus sequence type clonal complexes 21 and 45 under freeze, chill, oxidative, acid and heat stresses. *Food Microbiology*, 27(6):829–834, 2010.

Hald, B. and Madsen, M. Healthy puppies and kittens as carriers of *Campylobacter* spp., with special reference to *Campylobacter upsaliensis*. *Journal of Clinical Microbiology*, 35(12):3351, 1997.

Hänel, I.; Borrmann, E.; Müller, J.; Müller, W.; Pauly, B.; Liebler-Tenorio, E.M., and Schulze, F. Genomic and phenotypic changes of *Campylobacter jejun* strains after passage of the chicken gut. *Veterinary Microbiology*, 136(1):121–129, 2009.

Harris, N.V.; Weiss, N.S., and Nolan, C.M. The role of poultry and meats in the etiology of *Campylobacter jejuni/coli* enteritis. *American Journal of Public Health*, 76(4):407–411, 1986.

Hartigan, J.A. and Wong, M.A. Algorithm as 136: A k-means clustering algorithm. *Applied Statistics*, 28:100–108, 1979.

Hasegawa, M.; Kishino, H., and Yano, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2):160–174, 1985.

Hazeleger, W.C.; Wouters, J.A.; Rombouts, F.M., and Abee, T. Physiological activity of *Campylobacter jejuni* far below the minimal growth temperature. *Applied and environmental microbiology*, 64(10):3917–3922, 1998.

Hearnden, M.; Skelly, C.; Eyles, R., and Weinstein, P. The regionality of campylobacteriosis seasonality in New Zealand. *International Journal of Environmental Health Research*, 13(4): 337–348, 2003.

Hendrixson, D. R. and DiRita, V. J. Identification of *Campylobacter jejuni* genes involved in commensal colonization of the chick gastrointestinal tract. *Molecular Microbiology*, 52(2): 471–484, 2004.

Henikoff, S. and Henikoff, J.G. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences*, 89(22):10915–10919, 1992.

Hepworth, P.J.; Ashelford, K.E.; Hinds, J.; Gould, K.A.; Witney, A.A.; Williams, N.J.; Leatherbarrow, H.; French, N.P.; Birtles, R.J., and Mendonca, C. Genomic variations define divergence of water/wildlife-associated *Campylobacter jejuni* niche specialists from common clonal complexes. *Environmental Microbiology*, 2011.

Hermans, D.; Pasmans, F.; Messens, W.; Martel, A.; Van Immerseel, F.; Rasschaert, G.; Heyndrickx, M.; Van Deun, K., and Haesebrouck, F. Poultry as a host for the zoonotic pathogen *Campylobacter jejuni*. *Vector-Borne and Zoonotic Diseases*, 12(2):89–98, 2012.

Herschleb, J.; Ananiev, G., and Schwartz, D.C. Pulsed-field gel electrophoresis. *Nature Protocols*, 2(3):677–684, 2007.

Herve, J.; Aissa, N.; Legrand, P.; Sorkine, M.; Calmette, MJ; Santin, A.; Roupie, E., and Renaud, B. *Campylobacter fetus* meningitis in a diabetic adult cured by imipenem. *European Journal of Clinical Microbiology & Infectious Diseases*, 23(9):722–724, 2004.

Heuvelink, A.E.; van Heerwaarden, C.; Zwartkruis-Nahuis, A.; Tilburg, J.J.H.C.; Bos, M.H; Heilmann, F.G.C.; Hofhuis, A.; Hoekstra, T., and de Boer, E. Two outbreaks of campylobacteriosis associated with the consumption of raw cows' milk. *International Journal of Food Microbiology*, 134(1):70–74, 2009.

Hiett, K.L.; Stern, N.J.; Fedorka-Cray, P.; Cox, N.A.; Musgrove, M.T., and Ladely, S. Molecular subtype analyses of *Campylobacter* spp. from Arkansas and California poultry operations. *Applied and Environmental Microbiology*, 68(12):6220–6236, 2002.

Hiett, K.L.; Stintzi, A.; Andacht, T.M.; Kuntz, R.L., and Seal, B.S. Genomic differences between *Campylobacter jejuni* isolates identify surface membrane and flagellar function gene products potentially important for colonizing the chicken intestine. *Functional & Integrative Genomics*, 8(4):407–420, 2008.

Hill, J.E.; Paccagnella, A.; Law, K.; Melito, P.L.; Woodward, D.L.; Price, L.; Leung, A.H.; Ng, L.-K.; Hemmingsen, S.M., and Goh, S.H. Identification of *Campylobacter* spp. and discrimination from *Helicobacter* and *Arcobacter* spp. by direct sequencing of pcr-amplified *cpn60* sequences and comparison to cpndb, a chaperonin reference sequence database. *Journal of Medical Microbiology*, 55(4):393–399, 2006.

Ho, H.-T.; Chang, P.-L.; Hung, C.-C., and Chang, H.-T. Capillary electrophoretic restriction fragment length polymorphism patterns for the mycobacterial *hsp65* gene. *Journal of Clinical Microbiology*, 42(8):3525–3531, 2004.

Hoe, N.; Nakashima, K.; Grigsby, D.; Pan, X.; Dou, S.J.; Naidich, S.; Garcia, M.; Kahn, E.;

Bergmire-Sweat, D., and Musser, J.M. Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains. *Emerging Infectious Diseases*, 5(2):254–263, 1999.

Hofreuter, D.; Tsai, J.; Watson, R.O.; Novik, V.; Altman, B.; Benitez, M.; Clark, C.; Perbost, C.; Jarvie, T., and Du, L. Unique features of a highly pathogenic *Campylobacter jejuni* strain. *Infection and Immunity*, 74(8):4694, 2006.

Hofreuter, D.; Novik, V., and Galán, J. Metabolic diversity in *Campylobacter jejuni* enhances specific tissue colonization. *Cell Host & Microbe*, 4(5):425–433, 2008.

Hogg, J.S.; Hu, F.Z.; Janto, B.; Boissy, R.; Hayes, J.; Keefe, R.; Post, J.C., and Ehrlich, G. Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of rd and 12 clinical nontypeable strains. *Genome Biology*, 8(6):R103, 2007.

Höök, H.; Fattah, M.A.; Ericsson, H.; Vågsholm, I., and Danielsson-Tham, M. Genotype dynamics of *Campylobacter jejuni* in a broiler flock. *Veterinary Microbiology*, 106(1):109–117, 2005.

Howard, S.L.; Jagannathan, A.; Soo, E.C.; Hui, J.P.M.; Aubry, A.J.; Ahmed, I.; Karlyshev, A.; Kelly, J.F.; Jones, M.A., and Stevens, M.P. *Campylobacter jejuni* glycosylation island important in cell charge, legionaminic acid biosynthesis, and colonization of chickens. *Infection and Immunity*, 77(6):2544–2556, 2009.

Hu, J. and Ng, P.C. Predicting the ffects of frameshifting indels. *Genome Biology*, 13(2):R9, 2012.

Hugdahl, M.B.; Beery, J.T., and Doyle, M.P. Chemotactic behavior of *Campylobacter jejuni*. *Infection and Immunity*, 56(6):1560–1566, 1988.

Hum, S.; Stephens, L. R., and Quinn, C. Diagnosis by ELISA of bovine abortion due to *Campylobacter fetus*. *Australian Veterinary Journal*, 68(8):272–275, 1991.

Humphrey, C.D.; Montag, D.M., and Pittman, F.E. Experimental infection of hamsters with *Campylobacter jejuni*. *Journal of Infectious Diseases*, 151(3):485–493, 1985.

Humphrey, S.; Chaloner, G.; Kemmett, K.; Davidson, N.; Williams, N.; Kipar, A.; Humphrey, T., and Wigley, P. *Campylobacter jejuni* is not merely a commensal in commercial broiler chickens and affects bird welfare. *mBio*, 5(4):e01364–14, 2014.

Humphrey, T.; O'Brien, S., and Madsen, M. Campylobacters as zoonotic pathogens: a food production perspective. *International Journal of Food Microbiology*, 117(3):237–257, 2007.

Huse, S.M.; Huber, J.A.; Morrison, H.G.; Sogin, M.L., and Welch, D.M. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biology*, 8(7):R143, 2007.

Huson, D.H. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*, 14(1): 68–73, 1998.

Huson, D.H. and Bryant, D. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2):254–267, 2006.

Hwang, M.N. and Ederer, G.M. Rapid hippurate hydrolysis method for presumptive identification of group B streptococci. *Journal of Clinical Microbiology*, 1(1):114–115, 1975.

Hyndman, R.J. and Khandakar, Y. Automatic time series for forecasting: the forecast package for R. Technical report, Monash University, Department of Econometrics and Business Statistics, 2007.

Hysom, D.A.; Naraghi-Arani, P.; Elsheikh, M.; Carrillo, A.C.; Williams, P.L., and Gardner, S.N. Skip the alignment: degenerate, multiplex primer and probe design using k-mer matching instead of alignments. *PloS One*, 7(4):e34560, 2012.

Imai, Y.; Kikuchi, M.; Matsuda, M.; Honda, M.; Fukuyama, M.; Tsukada, M., and Kaneuchi, C. Macro-fingerprinting analysis at the chromosomal genomic DNA level of isolates of thermophilic *Campylobacter coli* and *C. jejuni*, by pulsed-field gel electrophoresis. *Cytobios*, 78 (313):115–122, 1993.

Jackson, C.J.; Fox, A.J.; Wareing, D.R.A.; Hutchinson, D.N., and Jones, D.M. The application of genotyping techniques to the epidemiological analysis of *Campylobacter jejuni*. *Epidemiology and Infection*, 117(2):233–244, 1996.

Jackson, C.J.; Fox, A.J.; Jones, D.M.; Wareing, D.R.A., and Hutchinson, D.N. Associations between heat-stable (O) and heat-labile (HL) serogroup antigens of *Campylobacter jejuni*: evidence for interstrain relationships within three O/HL serovars. *Journal of Clinical Microbiology*, 36(8):2223–2228, 1998.

Jacobs-Reitsma, W.F.; Van de Giessen, A.W.; Bolder, N.M., and Mulder, R.W.A.W. Epidemiology of *Campylobacter* spp. at two Dutch broiler farms. *Epidemiology and Infection*, 114(03): 413–421, 1995.

Jerome, J.P.; Bell, J.A.; Plovanich-Jones, A.E.; Barrick, J.E.; Brown, C.T., and Mansfield, L.S. Standing genetic variation in contingency loci drives the rapid adaptation of *Campylobacter jejuni* to a novel host. *PLoS One*, 6(1):e16399, 2011.

Jin, S.; Joe, A.; Lynett, J.; Hani, E.K.; Sherman, P., and Chan, V.L. Jlpa, a novel surface-exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells. *Molecular Microbiology*, 39(5):1225–1236, 2001.

Johnson, W.M. and Lior, H. A new heat-labile cytolethal distending toxin (CLDT) produced by *Campylobacter* spp. *Microbial Pathogenesis*, 4(2):115–126, 1988.

Jolley, K.A.; Bliss, C.M.; Bennett, J.S.; Bratcher, H.B.; Brehony, C.; Colles, F.M.; Wimalarathna, H.; Harrison, O.B.; Sheppard, S.K., and Cody, A.J. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology*, 158(Pt 4):1005–1015, 2012.

Jonasson, J.; Olofsson, M., and Monstein, H.-J. Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16S rDNA fragments. *APMIS*, 110(3):263–272, 2002.

Jones, D.M.; Sutcliffe, E.M., and Curry, A. Recovery of viable but non-culturable *Campylobacter jejuni*. *Journal of General Microbiology*, 137(10):2477–2482, 1991.

Jones, K. *Campylobacters* in water, sewage and the environment. *Journal of Applied Microbiology*, 90(S6):68S–79S, 2001.

Jones, M.A.; Marston, K.L.; Woodall, C.A.; Maskell, D.J.; Linton, D.; Karlyshev, A.V.; Dorrell, N.; Wren, B.W., and Barrow, P.A. Adaptation of *Campylobacter jejuni* NCTC11168 to high-level colonization of the avian gastrointestinal tract. *Infection and Immunity*, 72(7):3769–3776, 2004.

Jukes, T H. and Cantor, C R. Evolution of protein molecules. *Mammalian protein metabolism*, 3:21–132, 1969.

Jureen, P.; Engstrand, L.; Eriksson, S.; Alderborn, A.; Krabbe, M., and Hoffner, S.E. Rapid detection of rifampin resistance in *Mycobacterium tuberculosis* by pyrosequencing technology. *Journal of Clinical Microbiology*, 44(6):1925–1929, 2006.

Kanungpean, D.; Kakuda, T., and Takai, S. Participation of CheR and CheB in the chemosensory response of *Campylobacter jejuni*. *Microbiology*, 157(5):1279–1289, 2011.

Kapperud, G.; Skjerve, E.; Bean, N.H.; Ostroff, S.M., and Lassen, J. Risk factors for sporadic *Campylobacter* infections: results of a case-control study in southeastern Norway. *Journal of Clinical Microbiology*, 30(12):3117–3121, 1992.

Karenlampi, R.; Rautelin, H., and Hanninen, M.L. Evaluation of genetic markers and molecular

typing methods for prediction of source of *Campylobacter jejuni* and *C. coli* infections. *Applied and Environmental Microbiology*, 73:1683–1685, 2007.

Kärenlampi, R.I.; Tolvanen, T.P., and Hänninen, M.-L. Phylogenetic analysis and PCR-restriction fragment length polymorphism identification of *Campylobacter* species based on partial *gro*EL gene sequences. *Journal of Clinical Microbiology*, 42(12):5731–5738, 2004.

Karlyshev, A.V.; Linton, D.; Gregson, N.A., and Wren, B.W. A novel paralogous gene family involved in phase-variable flagella-mediated motility in *Campylobacter jejuni*. *Microbiology*, 148(2):473–480, 2002.

Karlyshev, A.V.; Ketley, J.M., and Wren, B.W. The *Campylobacter jejuni* glycome. *FEMS Microbiology Reviews*, 29(2):377–390, 2005.

Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S., and Duran, C. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12):1647–1649, 2012.

Keim, P.; Price, L.B.; Klevytska, A.M.; Smith, K.L.; Schupp, J.M.; Okinaka, R.; Jackson, P.J., and Hugh-Jones, M.E. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *Journal of Bacteriology*, 182(10):2928–2936, 2000.

Keller, J.I. and Shriver, W.G. Prevalence of three *Campylobacter* species, *C. jejuni*, *C. coli*, and *C. lari*, using multilocus sequence typing in wild birds of the Mid-Atlantic region, USA. *Journal of Wildlife Diseases*, 50(1):31–41, 2014.

Kelly, A.F.; Park, S.F.; Bovill, R., and Mackey, B.M. Survival of *Campylobacter jejuni* during stationary phase: evidence for the absence of a phenotypic stationary-phase response. *Applied and Environmental Microbiology*, 67(5):2248–2254, 2001.

Khakhria, R. and Lior, H. Extended phage-typing scheme for *Campylobacter jejuni* and *Campylobacter coli*. *Epidemiology and Infection*, 108(03):403–414, 1992.

Kiehlbauch, J.A.; Plikaytis, B.D.; Swaminathan, B.; Cameron, D.N., and Wachsmuth, I.K. Restriction fragment length polymorphisms in the ribosomal genes for species identification and subtyping of aerotolerant *Campylobacter* species. *Journal of Clinical Microbiology*, 29(8): 1670–1676, 1991.

Killiny, N. and Almeida, R.P.P. Gene regulation mediates host specificity of a bacterial pathogen. *Environmental Microbiology Reports*, 3(6):791–797, 2011.

Kim, J.S.; Artymovich, K.A.; Hall, D.F.; Smith, E.J.; Fulton, R.; Bell, J.; Dybas, L.; Mansfield, L.S.; Tempelman, R., and Wilson, D.L. Passage of *Campylobacter jejuni* through the chicken reservoir or mice promotes phase variation in contingency genes Cj0045 and Cj0170 that strongly associates with colonization and disease in a mouse model. *Microbiology*, 158(Pt 5): 1304–1316, 2012.

Kimura, M. Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626, 1968.

Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16(2):111–120, 1980.

Kimura, M. *The neutral theory of molecular evolution*. Cambridge University Press, 1984.

King, E.O. Human infections with *Vibrio fetus* and a closely related vibrio. *Journal of Infectious Diseases*, 101(2):119–128, 1957.

Klena, J.D. and Konkel, M.E. *Campylobacter Molecular and Cellular Biology*, chapter Methods for epidemiological analysis of *Campylobacter jejuni*, pages 165–179. Horizon Scientific Press, 2005.

Kokotovic, B. and On, S.L.W. High-resolution genomic fingerprinting of *Campylobacter jejuni* and *Campylobacter coli* by analysis of amplified fragment length polymorphisms. *FEMS Microbiology Letters*, 173(1):77–84, 1999.

Kolde, R. pheatmap: Pretty heatmaps. r package version 0.7. 7, 2012. URL `http://cran.r-project.org/web/packages/pheatmap/index.html`.

Konkel, M.E.; Garvis, S.G.; Tipton, S.L.; Anderson, D.E., and Cieplak, W. Identification and molecular cloning of a gene encoding a fibronectin-binding protein (CadF) from *Campylobacter jejuni*. *Molecular Microbiology*, 24(5):953–963, 1997.

Konkel, M.E.; Christensen, J.E.; Dhillon, A.S.; Lane, A.B.; Hare-Sanford, R.; Schaberg, D.M., and Larson, C. *Campylobacter jejuni* strains compete for colonization in broiler chicks. *Applied and Environmental Microbiology*, 73(7):2297–2305, 2007.

Korczak, B.M.; Stieber, R.; Emler, S.; Burnens, A.P.; Frey, J., and Kuhnert, P. Genetic relatedness within the genus *Campylobacter* inferred from *rpoB* sequences. *International Journal of Systematic and Evolutionary Microbiology*, 56(5):937–945, 2006.

Korlach, J.; Marks, P.J.; Cicero, R.L.; Gray, J.J.; Murphy, D.L.; Roitman, D.B.; Pham, T.T.;

Otto, G.A.; Foquet, M., and Turner, S.W. Selective aluminum passivation for targeted immobilization of single DNA polymerase molecules in zero-mode waveguide nanostructures. *Proceedings of the National Academy of Sciences*, 105(4):1176–1181, 2008.

Korolik, V.; Alderton, M.R.; Smith, S.C.; Chang, J., and Coloe, P.J. Isolation and molecular analysis of colonising and non-colonising strains of *Campylobacter jejuni* and *Campylobacter coli* following experimental infection of young chickens. *Veterinary Microbiology*, 60(2):239–249, 1998.

Koskiniemi, S.; Gibbons, H.S.; Sandegren, L.; Anwar, N.; Ouellette, G.; Broomall, S.; Karavis, M.; McGregor, P.; Liem, A., and Fochler, E. Pathoadaptive mutations in *Salmonella enterica* isolated after serial passage in mice. *PloS One*, 8(7):e70147, 2013.

Kovanen, S.M.; Kivistö, R.I.; Rossi, M., and Hänninen, M.-L. A combination of MLST and CRISPR typing reveals dominant *Campylobacter jejuni* types in organically farmed laying hens. *Journal of Applied Microbiology*, 117:249–257, 2014.

Kovats, R.S.; Edwards, S.J.; Charron, D.; Cowden, J.; DSouza, R.M; Ebi, K.L.; Gauci, C.; Gerner-Smidt, P.; Hajat, S., and Hales, S. Climate variability and *Campylobacter* infection: an international study. *International Journal of Biometeorology*, 49(4):207–214, 2005.

Krause, R.; Ramschak-Schwarzer, S.; Gorkiewicz, G.; Schnedl, W.J.; Feierl, G.; Wenisch, C., and Reisinger, E.C. Recurrent septicemia due to *Campylobacter fetus* and *Campylobacter lari* in an immunocompetent patient. *Infection*, 30(3):171–174, 2002.

Krause-Gruszczynska, M.; Rohde, M.; Hartig, R.; Genth, H.; Schmidt, G.; Keo, T.; König, W.; Miller, W.G.; Konkel, M.E., and Backert, S. Role of the small Rho GTPases Rac1 and Cdc42 in host cell invasion of *Campylobacter jejuni*. *Cellular Microbiology*, 9(10):2431–2444, 2007.

Kühnert, D.; Wu, C.-H., and Drummond, A.J. Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. *Infection, Genetics and Evolution*, 11(8):1825–1841, 2011.

Kurtz, S.; Phillippy, A.; Delcher, A.L.; Smoot, M.; Shumway, M.; Antonescu, C., and Salzberg, S.L. Versatile and open software for comparing large genomes. *Genome Biology*, 5(2):R12.1–R12.9, 2004.

Lake, R. Transmission routes for campylobacteriosis in New Zealand. Technical report, Institute of Environmental Science & Research Limited, Christchurch Science Centre, 27 Creyke Road, Ilam, Christchurch, 2006.

Lal, A.; Ikeda, T.; French, N.; Baker, M.G., and Hales, S. Climate variability, weather and enteric disease incidence in New Zealand: Time series analysis. *PloS One*, 8(12):e83484, 2013.

Lan, R. and Reeves, P.R. When does a clone deserve a name? A perspective on bacterial species based on population genetics. *Trends in Microbiology*, 9(9):419–424, 2001.

Langmead, B. and Salzberg, S.L. Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9(4):357–359, 2012.

Lapierre, P. and Gogarten, J.P. Estimating the size of the bacterial pan-genome. *Trends in Genetics*, 25(3):107–110, 2009.

Lara-Tejero, M. and Galán, J. CdtA, CdtB, and CdtC form a tripartite complex that is required for cytolethal distending toxin activity. *Infection and Immunity*, 69(7):4358–4365, 2001.

Lara-Tejero, M. and Galán, J.E. A bacterial toxin that controls cell cycle progression as a deoxyribonuclease I-like protein. *Science*, 290(5490):354–357, 2000.

Lastovica, A.J. Emerging *Campylobacter* spp.: The tip of the iceberg. *Clinical Microbiology Newsletter*, 28(7):49–56, 2006.

Lee, M.D. and Newell, D.G. *Campylobacter* in poultry: filling an ecological niche. *Avian Diseases*, 50(1):1–9, 2006.

Lefébure, T. and Stanhope, M.J. Pervasive, genome-wide positive selection leading to functional divergence in the bacterial genus *Campylobacter*. *Genome Research*, 19(7):1224–1232, 2009.

Lefébure, T.; Bitar, P.D.P.; Suzuki, H., and Stanhope, M.J. Evolutionary dynamics of complete *Campylobacter* pan-genomes and the bacterial species concept. *Genome Biology and Evolution*, 2:646, 2010.

Lemey, P. and Posada, D. *The Phylogenetic Handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press, 2009.

Leonard II, E.E.; Takata, T.; Blaser, M.J.; Falkow, S.; Tompkins, L.S., and Gaynor, E.C. Use of an open-reading frame–specific *Campylobacter jejuni* DNA microarray as a new genotyping tool for studying epidemiologically related isolates. *Journal of Infectious Diseases*, 187(4): 691–694, 2003.

Levene, M.J.; Korlach, J.; Turner, S.W.; Foquet, M.; Craighead, H.G., and Webb, W.W. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*, 299(5607): 682–686, 2003.

Levesque, S.; Frost, E.; Arbeit, R.D., and Michaud, S. Multilocus sequence typing of *Campylobacter jejuni* isolates from humans, chickens, raw milk, and environmental water in Quebec, Canada. *Journal of Clinical Microbiology*, 46(10):3404, 2008.

Levinson, G. and Gutman, G.A. Slipped−strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution*, 4(3):203–221, 1987.

Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G., and Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16): 2078–2079, 2009a.

Li, J.; Gershow, M.; Stein, D.; Brandin, E., and Golovchenko, J.A. DNA molecules and configurations in a solid-state nanopore microscope. *Nature Materials*, 2(9):611–615, 2003a.

Li, L.; Stoeckert, C.J., and Roos, D.S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, 13(9):2178–2189, 2003b.

Li, W.; Chomel, B.B.; Maruyama, S.; Guptil, L.; Sander, A.; Raoult, D., and Fournier, P.-E. Multispacer typing to study the genotypic distribution of *Bartonella henselae* populations. *Journal of Clinical Microbiology*, 44(7):2499–2506, 2006.

Li, W.; Raoult, D., and Fournier, P.-E. Bacterial strain typing in the genomic era. *FEMS Microbiology Reviews*, 33(5):892–916, 2009b.

Lin, J.; Michel, L.O., and Zhang, Q. CmeABC functions as a multidrug efflux system in *Campylobacter jejuni*. *Antimicrobial Agents and Chemotherapy*, 46(7):2124–2131, 2002.

Lin, J.; Sahin, O.; Michel, L.O., and Zhang, Q. Critical role of multidrug efflux pump cmeabc in bile resistance and *in vivo* colonization of *Campylobacter jejuni*. *Infection and Immunity*, 71(8):4250–4259, 2003.

Lindstedt, B.-A. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis*, 26(13):2567–2582, 2005.

Lindstedt, B.-A.; Heir, E.; Vardund, T., and Kapperud, G. Fluorescent amplified-fragment length polymorphism genotyping of *Salmonella enterica* subsp. *enterica* serovars and comparison with pulsed-field gel electrophoresis typing. *Journal of Clinical Microbiology*, 38(4): 1623–1627, 2000.

Line, J.E.; Hiett, K.L.; Guard-Bouldin, J., and Seal, B.S. Differential carbon source utilization by *Campylobacter jejuni* 11168 in response to growth temperature variation. *Journal of Microbiological Methods*, 80(2):198–202, 2010.

Linton, D.; Owen, R.J., and Stanley, J. Rapid identification by PCR of the genus *Campylobacter* and of five *Campylobacter* species enteropathogenic for man and animals. *Research in Microbiology*, 147(9):707–718, 1996.

Lior, H.; Woodward, D.L.; Edgar, J.A.; Laroche, L.J., and Gill, P. Serotyping of *Campylobacter jejuni* by slide agglutination based on heat-labile antigenic factors. *Journal of Clinical Microbiology*, 15(5):761–768, 1982.

Liu, L.; Li, Y.; Li, S.; Hu, N.; He, Y.; Pong, R.; Lin, D.; Lu, L., and Law, M. Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, 2012, 2012. doi: doi:10.1155/2012/251364.

Lo, Y.M.D. and Chan, K.C.A. Introduction to the polymerase chain reaction. In *Clinical Applications of PCR*, pages 1–10. Springer, 2006.

Lombardo, M.; Aponyi, I., and Rosenberg, S.M. General stress response regulator RpoS in adaptive mutation and amplification in *Escherichia coli*. *Genetics*, 166(2):669–680, 2004.

López, C.M.; Giacoboni, G.; Agostini, A.; Cornero, F.J.; Tellechea, D.M., and Trinidad, J.J. Thermotolerant *Campylobacters* in domestic animals in a defined population in Buenos Aires, Argentina. *Preventive Veterinary Medicine*, 55(3):193–200, 2002.

Louis, V.R.; Gillespie, I.A.; O'Brien, S.J.; Russek-Cohen, E.; Pearson, A.D., and Colwell, R.R. Temperature−driven *Campylobacter* seasonality in England and Wales. *Applied and Environmental Microbiology*, 71(1):85–92, 2005.

Luber, P.; Bartelt, E.; Genschow, E.; Wagner, J., and Hahn, H. Comparison of broth microdilution, E Test, and agar dilution methods for antibiotic susceptibility testing of *Campylobacter jejuni* and *Campylobacter coli*. *Journal of Clinical Microbiology*, 41(3):1062–1068, 2003.

Lupski, J.R. and Weinstock, G.M. Short, interspersed repetitive DNA sequences in prokaryotic genomes. *Journal of Bacteriology*, 174(14):4525–4529, 1992.

Lushchak, V.I. Adaptive response to oxidative stress: Bacteria, fungi, plants and animals. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, 153(2):175–190, 2011.

MacPherson, J.M.; Eckstein, P.E.; Scoles, G.J., and Gajadhar, A.A. Variability of the random amplified polymorphic DNA assay among thermal cyclers, and effects of primer and DNA concentration. *Molecular and Cellular Probes*, 7(4):293–299, 1993.

Maiden, M.C.J.; Bygraves, J.A.; Feil, E.; Morelli, G.; Russell, J.E.; Urwin, R.; Zhang, Q.; Zhou,

J.; Zurth, K., and Caugant, D.A. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences*, 95(6):3140–3145, 1998.

Maiden, M.C.J.; van Rensburg, M.J.J.; Bray, J.E.; Earle, S.G.; Ford, S.A.; Jolley, K.A., and McCarthy, N.D. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature Reviews Microbiology*, 11:728–736, 2013.

Malloff, C.A.; Fernandez, R.C., and Lam, W.L. Bacterial comparative genomic hybridization: a method for directly identifying lateral gene transfer. *Journal of Molecular Biology*, 312(1): 1–5, 2001.

Manning, G.; Dowson, C.G.; Bagnall, M.C.; Ahmed, I.H.; West, M., and Newell, D.G. Multilocus sequence typing for comparison of veterinary and human isolates of *Campylobacter jejuni*. *Applied and Environmental Microbiology*, 69(11):6370–6379, 2003.

Marçais, G. and Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6):764–770, 2011.

Mardis, E.R. The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, 24(3):133–141, 2008.

Matsuda, M.; Tsukada, M.; Fukuyama, M.; Kato, Y.; Ishida, Y.; Honda, M., and Kaneuchi, C. Detection of genomic variability among isolates of *Campylobacter jejuni* from chickens by crossed-field gel electrophoresis. *Cytobios*, 82(329):73–79, 1994.

Mayr, E. Speciation and macroevolution. *Evolution*, 36(6):1119–1132, 1982.

McCarthy, N.D.; Colles, F.M.; Dingle, K.E.; Bagnall, M.C.; Manning, G.; Maiden, M.C.J., and Falush, D. Host-associated genetic import in *Campylobacter jejuni*. *Emerging Infectious Diseases*, 13(2):267–272, 2007.

McCarthy, N.D.; Gillespie, I.A.; Lawson, A.J.; Richardson, J.; Neal, K.R.; Hawtin, P.R.; Maiden, M.C.J., and O'Brien, S.J. Molecular epidemiology of human *Campylobacter jejuni* shows association between seasonal and international patterns of disease. *Epidemiology and Infection*, 1(1):1–9, 2012.

McCutcheon, J. P. and Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology*, 10(1):13–26, 2012.

McFarlane, D.; Salisbury, R.M.; Osborne, H.G., and Jebson, J.L. Investigation into sheep

abortion in New Zealand during the 1950 lambing season. *Australian Veterinary Journal*, 28 (9):221–226, 1952.

McTavish, S.M.; Pope, C.E.; Nicol, C.; Sexton, K.; French, N., and Carter, P.E. Wide geographical distribution of internationally rare *Campylobacter* clones within New Zealand. *Epidemiology and Infection*, 136(09):1244–1252, 2008.

Medini, D.; Donati, C.; Tettelin, H.; Masignani, V., and Rappuoli, R. The microbial pan-genome. *Current Opinion in Genetics & Development*, 15(6):589–594, 2005.

Melles, D.C.; van Leeuwen, W.B.; Snijders, S.V.; Horst-Kreft, D.; Peeters, J.K.; Verbrugh, H.A., and van Belkum, A. Comparison of multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and amplified fragment length polymorphism (AFLP) for genetic typing of *Staphylococcus aureus*. *Journal of Microbiological Methods*, 69(2):371–375, 2007.

Merriman, B.; Torrent, I.; Rothberg, J.M., and Team, R&D. Progress in Ion Torrent semiconductor chip based sequencing. *Electrophoresis*, 33(23):3397–3417, 2012.

Miller, W.G. and Mandrell, R.E. *Campylobacter Molecular and Cellular Biology*, chapter Prevalence of *Campylobacter* in the food and water supply: incidence, outbreaks, isolation and detection, pages 101–163. Horizon Bioscience, Norfolk, UK, 2005.

Miller, W.G.; On, S.L.W.; Wang, G.; Fontanoz, S.; Lastovica, A.J., and Mandrell, R.E. Extended multilocus sequence typing system for *Campylobacter coli*, *C. lari*, *C. upsaliensis*, and *C. helveticus*. *Journal of Clinical Microbiology*, 43(5):2315–2329, 2005.

Miller, W.G.; Chapman, M.H.; Yee, E.; On, S.L.W.; McNulty, D.K.; Lastovica, A.J.; Carroll, A.M.; McNamara, E.B.; Duffy, G., and Mandrell, R.E. Multilocus sequence typing methods for the emerging *Campylobacter* species *C. hyointestinalis*, *C. lanienae*, *C. sputorum*, *C. concisus*, and *C. curvus*. *Frontiers in Cellular and Infection Microbiology*, 2, 2012. doi: doi:10.3389/fcimb.2012.00045.

Mira, A.; Ochman, H., and Moran, N.A. Deletional bias and the evolution of bacterial genomes. *Trends in Genetics*, 17(10):589–596, 2001.

Mohan, V.; Stevenson, M.; Marshall, J.; Fearnhead, P.; Holland, B.R.; Hotter, G., and French, N.P. *Campylobacter jejuni* colonization and population structure in urban populations of ducks and starlings in New Zealand. *Microbiology Open*, 2(4):659–673, 2013.

Mollet, C.; Drancourt, M., and Raoult, D. *rpo*b sequence analysis as a novel basis for bacterial identification. *Molecular Microbiology*, 26(5):1005–1011, 1997.

Monteville, M.R.; Yoon, J.E., and Konkel, M.E. Maximal adherence and invasion of INT 407 cells by *Campylobacter jejuni* requires the CadF outer-membrane protein and microfilament reorganization. *Microbiology*, 149(1):153–165, 2003.

Moore, J.E.; Corcoran, D.; Dooley, J.S.G.; Fanning, S.; Lucey, B.; Matsuda, M.; McDowell, D.A.; Mégraud, F.; Cherie Millar, B., and OMahony, R. *Campylobacter. Veterinary Research*, 36(3):351–382, 2005.

Moran, N.A. Microbial minimalism: genome reduction in bacterial pathogens. *Cell*, 108(5): 583–586, 2002.

Moreno, G.S.; Griffiths, P.L.; Connerton, I.F., and Park, R.W.A. Occurrence of *Campylobacters* in small domestic and laboratory animals. *Journal of Applied Microbiology*, 75(1):49–54, 1993.

Morris, G.K.; El Sherbeeny, M.R.; Patton, C.M.; Kodaka, H.; Lombard, G.L.; Edmonds, P.; Hollis, D.G., and Brenner, D.J. Comparison of four hippurate hydrolysis methods for identification of thermophilic *Campylobacter* spp. *Journal of Clinical Microbiology*, 22(5):714, 1985.

Mortimer, P. and Arnold, C. FAFLP: last word in microbial genotyping? *Journal of Medical Microbiology*, 50(5):393, 2001.

Mount, D.W. Comparison of the PAM and BLOSUM amino acid substitution matrices. *Cold Spring Harbor Protocols*, 2008(6), 2008.

Moxon, R.; Bayliss, C., and Hood, D. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annual Review of Genetics*, 40:307–333, 2006.

Muellner, P. *Estimating the contribution of different sources to the burden of human campylobacteriosis and salmonellosis.* PhD thesis, Massey University, 2009.

Muellner, P.; Marshall, J.C.; Spencer, S.E.F.; Noble, A.D.; Shadbolt, T.; Collins-Emerson, J.M.; Midwinter, A.C.; Carter, P.E.; Pirie, R., and Wilson, D.J. Utilizing a combination of molecular and spatial tools to assess the effect of a public health intervention. *Preventive Veterinary Medicine*, 102(3):242–253, 2011.

Muellner, P.; Pleydell, E.; Pirie, R.; Baker, M.G.; Campbell, D.; Carter, P.E., and French, N.P. Molecular-based surveillance of campylobacteriosis in New Zealand–from source attribution to genomic epidemiology. *Eurosurveillance*, 18(3), 2013.

Mullner, P.; Jones, G.; Noble, A.; Spencer, S.E.F.; Hathaway, S., and French, N.P. Source

attribution of food-borne zoonoses in New Zealand: A modified Hald model. *Risk Analysis*, 29(7):970–984, 2009.

Müllner, P.; Spencer, S.E.F.; Wilson, D.J.; Jones, G.; Noble, A.D.; Midwinter, A.C.; Collins-Emerson, J.M.; Carter, P.; Hathaway, S., and French, N.P. Assigning the source of human campylobacteriosis in New Zealand: a comparative genetic and epidemiological approach. *Infection, Genetics and Evolution*, 9(6):1311–1319, 2009.

Müllner, P.; Collins-Emerson, J.M.; Midwinter, A.C.; Carter, P.; Spencer, S.E.F.; van der Logt, P.; Hathaway, S., and French, N.P. Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry. *Applied and Environmental Microbiology*, 76(7):2145–2154, 2010a.

Müllner, P.; Shadbolt, T; Collins-Emerson, J.M.; Midwinter, A.C.; Spencer, S.E.F.; Marshall, J.; Carter, P.E.; Campbell, D.M.; Wilson, D.J., and Hathaway, S. Molecular and spatial epidemiology of human campylobacteriosis: source association and genotype-related risk factors. *Epidemiology and Infection*, 138(10):1372–1383, 2010b.

Muraoka, W.T. and Zhang, Q. Phenotypic and genotypic evidence for l-fucose utilization by *Campylobacter jejuni*. *Journal of Bacteriology*, 193(5):1065–1075, 2011.

Murphy, C.; Carroll, C., and Jordan, K. N. Induction of an adaptive tolerance response in the foodborne pathogen, *Campylobacter jejuni*. *FEMS Microbiology Letters*, 223(1):89–93, 2003.

Nachamkin, I.; Blaser, M.J., and Tompkins, L.S. *Campylobacter jejuni: current status and future trends*. American Society for Microbiology, 1992.

Nachamkin, I.; Yang, X.-H., and Stern, N.J. Role of *Campylobacter jejuni* flagella as colonization factors for three-day-old chicks: analysis with flagellar mutants. *Applied and Environmental Microbiology*, 59(5):1269–1273, 1993.

Neal, K.R. and Slack, R.C.B. Diabetes mellitus, anti-secretory drugs and other risk factors for *Campylobacter* gastro-enteritis in adults: a case-control study. *Epidemiology and Infection*, 119(03):307–311, 1997.

Nesbakken, T.; Eckner, K.; Hřidal, H.K., and Rřtterud, O.J. Occurrence of *Yersinia enterocolitica* and *Campylobacter* spp. in slaughter pigs and consequences for meat inspection, slaughtering, and dressing procedures. *International Journal of Food Microbiology*, 80(3): 231–240, 2003.

New Zealand Food Safety Authority, 2008. *Campylobacter* Risk Management Strategy 2008

- 2011. Technical report, New Zealand Food Safety Authority, 2008. URL `http://www.foodsafety.govt.nz/elibrary/industry/Campylobacter_Risk-Aims_Acheive.pdf`.

Newell, D.G. Animal models of *Campylobacter jejuni* colonization and disease and the lessons to be learned from similar *Helicobacter pylori* models. *Journal of Applied Microbiology*, 90 (S6):57S–67S, 2001.

Newell, D.G. The ecology of *Campylobacter jejuni* in avian and human hosts and in the environment. *International Journal of Infectious Diseases*, 6:S16–S21, 2002.

Newell, D.G. and Fearnley, C. Sources of *Campylobacter* colonization in broiler chickens. *Applied and Environmental Microbiology*, 69(8):4343–4351, 2003.

Neysens, P.; Messens, W.; Gevers, D.; Swings, J., and De Vuyst, L. Biphasic kinetics of growth and bacteriocin production with *Lactobacillus amylovorus* DCE 471 occur under stress conditions. *Microbiology*, 149(4):1073–1082, 2003.

Nylen, G.; Dunstan, F.; Palmer, S.R.; Andersson, Y.; Bager, F.; Cowden, J.; Feierl, G.; Galloway, Y.; Kapperud, G., and Megraud, F. The seasonal distribution of *Campylobacter* infection in nine European countries and New Zealand. *Epidemiology and Infection*, 128(03):383–390, 2002.

Oelschlaeger, T.A.; Guerry, P., and Kopecko, D.J. Unusual microtubule-dependent endocytosis mechanisms triggered by *Campylobacter jejuni* and *Citrobacter freundii*. *Proceedings of the National Academy of Sciences*, 90(14):6884–6888, 1993.

Ogden, I.D.; Dallas, J.F.; MacRae, M.; Rotariu, O.; Reay, K.W.; Leitch, M.; Thomson, A.P.; Sheppard, S.K.; Maiden, M., and Forbes, K.J. *Campylobacter* excreted into the environment by animal sources: prevalence, concentration shed, and host association. *Foodborne Pathogens and Disease*, 6(10):1161–1170, 2009.

Oliver, J.D. The viable but nonculturable state in bacteria. *The Journal of Microbiology*, 43(1): 93–100, 2005.

Oliver, J.D. Recent findings on the viable but nonculturable state in pathogenic bacteria. *FEMS Microbiology Reviews*, 34(4):415–425, 2010.

On, S.L. Identification methods for *Campylobacters*, *Helicobacters*, and related organisms. *Clinical Microbiology Reviews*, 9(3):405, 1996.

On, S.L. and Holmes, B. Effect of inoculum size on the phenotypic characterization of *Campylobacter* species. *Journal of Clinical Microbiology*, 29(5):923–926, 1991.

On, S.L.W. Taxonomy of *Campylobacter*, *Arcobacter*, *Helicobacter* and related bacteria: current status, future prospects and immediate concerns. *Journal of Applied Microbiology*, 90(S6): 1S–15S, 2001.

On, S.L.W. Isolation, identification and subtyping of *Campylobacter*: Where to from here? *Journal of Microbiological Methods*, 95(1):3–7, 2013.

On, S.L.W. and Harrington, C.S. Identification of taxonomic and epidemiological relationships among *Campylobacter* species by numerical analysis of AFLP profiles. *FEMS Microbiology Letters*, 193(1):161–169, 2000.

On, S.L.W. and Harrington, C.S. Evaluation of numerical analysis of PFGE-DNA profiles for differentiating *Campylobacter fetus* subspecies by comparison with phenotypic, PCR and 16S rDNA sequencing methods. *Journal of Applied Microbiology*, 90(2):285–293, 2001.

On, S.L.W.; Dorrell, N.; Petersen, L.; Bang, D.D.; Morris, S.; Forsythe, S.J., and Wren, B.W. Numerical analysis of DNA microarray data of *Campylobacter jejuni* strains correlated with survival, cytolethal distending toxin and haemolysin analyses. *International Journal of Medical Microbiology*, 296(6):353–363, 2006.

Pacific Biosciences of California, Inc., . Pacific Biosciences introduces new chemistry with longer read lengths to detect novel features in DNA sequence and advance genome studies of large organisms. online, October 2013. URL `http://investor.pacificbiosciences.com/releasedetail.cfm?ReleaseID=794692`.

Page, R.D.M. and Holmes, E.C. *Molecular evolution: a phylogenetic approach*. Blackwell Science Ltd, Oxford, 1998.

Pajaniappan, M.; Hall, J.E.; Cawthraw, S.A.; Newell, D.G.; Gaynor, E.C.; Fields, J.A.; Rathbun, K.M.; Agee, W.A.; Burns, C.M., and Hall, S.J. A temperature-regulated *Campylobacter jejuni* gluconate dehydrogenase is involved in respiration-dependent energy conservation and chicken colonization. *Molecular Microbiology*, 68(2):474–491, 2008.

Pallen, M.J. and Wren, B.W. Bacterial pathogenomics. *Nature*, 449(7164):835–842, 2007.

Palyada, K.; Sun, Y.-Q.; Flint, A.; Butcher, J.; Naikare, H., and Stintzi, A. Characterization of the oxidative stress stimulon and PerR regulon of *Campylobacter jejuni*. *BMC Genomics*, 10 (1):481, 2009.

Papin, J.A.; Reed, J.L., and Palsson, B.O. Hierarchical thinking in network biology: the unbiased

modularization of biochemical networks. *Trends in Biochemical Sciences*, 29(12):641–647, 2004.

Park, S.F. The physiology of *Campylobacter* species and its relevance to their role as foodborne pathogens. *International Journal of Food Microbiology*, 74(3):177–188, 2002.

Parker, C.T.; Quinones, B.; Miller, W.G.; Horn, S.T., and Mandrell, R.E. Comparative genomic analysis of *Campylobacter jejuni* strains reveals diversity due to genomic elements similar to those present in *C. jejuni* strain RM1221. *Journal of Clinical Microbiology*, 44(11):4125, 2006.

Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, T.; Davies, R.M.; Feltwell, T., and Holroyd, S. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature*, 403(6770):665–668, 2000.

Parsons, B.N.; Cody, A.J.; Porter, C.J.; Stavisky, J.H.; Smith, J.L.; Williams, N.J.; Leatherbarrow, A.J.H.; Hart, C.A.; Gaskell, R.M., and Dingle, K.E. Typing of *Campylobacter jejuni* isolates from dogs by use of multilocus sequence typing and pulsed-field gel electrophoresis. *Journal of Clinical Microbiology*, 47(11):3466, 2009.

Parsons, B.N.; Porter, C.J.; Stavisky, J.H.; Williams, N.J.; Birtles, R.J.; Miller, W.G.; Hart, C.A.; Gaskell, R.M., and Dawson, S. Multilocus sequence typing of human and canine *C. upsaliensis* isolates. *Veterinary Microbiology*, 157(3):391–397, 2012.

Patrick, M.E.; Christiansen, L.E.; Wainø, M.; Ethelberg, S.; Madsen, H., and Wegener, H.C. Effects of climate on incidence of *Campylobacter* spp. in humans and prevalence in broiler flocks in Denmark. *Applied and Environmental Microbiology*, 70(12):7474–7480, 2004.

Patton, C.M.; Barrett, T.J., and Morris, G.K. Comparison of the Penner and Lior methods for serotyping *Campylobacter* spp. *Journal of Clinical Microbiology*, 22(4):558–565, 1985.

Patton, C.M.; Nicholson, M.A.; Ostroff, S.M.; Ries, A.A.; Wachsmuth, I.K., and Tauxe, R.V. Common somatic O and heat-labile serotypes among *Campylobacter* strains from sporadic infections in the United States. *Journal of Clinical Microbiology*, 31(6):1525–1530, 1993.

Pearson, B.M.; Pin, C.; Wright, J.; I'Anson, K.; Humphrey, T., and Wells, J.M. Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. *FEBS Letters*, 554(1-2):224–230, 2003.

Pei, Z. and Blaser, M.J. PEB1, the major cell-binding factor of *Campylobacter jejuni*, is a

homolog of the binding component in gram-negative nutrient transport systems. *Journal of Biological Chemistry*, 268(25):18717–18725, 1993.

Penner, G.A.; Bush, A.; Wise, R.; Kim, W.; Domier, L.; Kasha, K.; Laroche, A.l.; Scoles, G.; Molnar, S.J., and Fedak, G. Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. *Genome Research*, 2(4):341–345, 1993.

Penner, J.L. The genus *Campylobacter*: a decade of progress. *Clinical Microbiology Reviews*, 1 (2):157–172, 1988.

Penner, J.L. and Hennessy, J.N. Passive hemagglutination technique for serotyping *Campylobacter fetus* subsp. *jejuni* on the basis of soluble heat-stable antigens. *Journal of Clinical Microbiology*, 12(6):732, 1980.

Pesci, E.C.; Cottle, D.L., and Pickett, C.L. Genetic, enzymatic, and pathogenic studies of the iron superoxide dismutase of *Campylobacter jejuni*. *Infection and Immunity*, 62(7):2687–2694, 1994.

Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C., and Ferrin, T.E. UCSF Chimeraa visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

Pickett, C.L. and Whitehouse, C.A. The cytolethal distending toxin family. *Trends in Microbiology*, 7(7):292–297, 1999.

Pickett, C.L.; Auffenberg, T.; Pesci, E.C.; Sheen, V.L., and Jusuf, S.S. Iron acquisition and hemolysin production by *Campylobacter jejuni*. *Infection and Immunity*, 60(9):3872–3877, 1992.

Pinheiro, J.; Bates, D.; DebRoy, S., and Sarkar, D. nlme: an R package for fitting and comparing Gaussian linear and nonlinear mixed-effects models. *See http://www. stats.bris.ac.uk/R*, 2006.

Pinheiro, José; Bates, Douglas, and Maintainer, R. Package nlme, 2013.

Pires, S.M.; Evers, E.G.; van Pelt, W.; Ayers, T.; Scallan, E.; Angulo, F.J.; Havelaar, A., and Hald, T. Attributing the human disease burden of foodborne infections to specific sources. *Foodborne Pathogens and Disease*, 6(4):417–424, 2009.

Pourcel, C.; Salvignol, G., and Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151(3):653–663, 2005.

Power, E.G.M. RAPD typing in microbiologya technical review. *Journal of Hospital Infection*, 34(4):247–265, 1996.

Price, E.P.; Huygens, F., and Giffard, P.M. Fingerprinting of *Campylobacter jejuni* by using resolution-optimized binary gene targets derived from comparative genome hybridization studies. *Applied and Environmental Microbiology*, 72(12):7793–7803, 2006.

Price, E.P.; Smith, H.; Huygens, F., and Giffard, P.M. High-resolution DNA melt curve analysis of the clustered, regularly interspaced short-palindromic-repeat locus of *Campylobacter jejuni*. *Applied and Environmental Microbiology*, 73(10):3431–3436, 2007.

Price, M.N.; Dehal, P.S., and Arkin, A.P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS One*, 5(3):e9490, 2010.

Purdy, D. and Park, S.F. Cloning, nucleotide sequence and characterization of a gene encoding superoxide dismutase from *Campylobacter jejuni* and *Campylobacter coli*. *Microbiology*, 140 (5):1203–1208, 1994.

Purdy, D.; Buswell, C.M.; Hodgson, A.E.; McAlpine, K.; Henderson, I., and Leach, S.A. Characterisation of cytolethal distending toxin (CDT) mutants of *Campylobacter jejuni*. *Journal of Medical Microbiology*, 49(5):473–479, 2000.

Quintana-Hayashi, M.P. and Thakur, S. Phylogenetic analysis reveals common antimicrobial resistant *Campylobacter coli* population in antimicrobial-free (ABF) and commercial swine systems. *PloS One*, 7(9):e44662, 2012.

R Development Core Team, . *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

Ragimbeau, C.; Salvat, G.; Colin, P., and Ermel, G. Development of a multiplex pcr gene fingerprinting method using *gyrA* and *pflA* polymorphisms to identify genotypic relatedness within *Campylobacter jejuni* species. *Journal of Applied Microbiology*, 85(5):829–838, 1998.

Rainey, P.B.; Beaumont, H.J.; Ferguson, G.C.; Gallie, J.; Kost, C.; Libby, E., and Zhang, X.-X. The evolutionary emergence of stochastic phenotype switching in bacteria. *Microbial Cell Factories*, 10(Suppl 1):S14, 2011.

Ramos, J.L.; Gallegos, M.; Marqués, S.; Ramos-González, M.; Espinosa-Urgel, M., and Segura, A. Responses of gram-negative bacteria to certain environmental stressors. *Current Opinion in Microbiology*, 4(2):166–171, 2001.

Reid, A. and Buckley, M. Microbial evolution. *American Academy of Microbiology*, page 17, 2011.

Richardson, P.T. and Park, S.F. Enterochelin acquisition in *Campylobacter coli*: characterization of components of a binding-protein-dependent transport system. *Microbiology*, 141(12):3181–3191, 1995.

Ridley, K.A.; Rock, J.D.; Li, Y., and Ketley, J.M. Heme utilization in *Campylobacter jejuni*. *Journal of Bacteriology*, 188(22):7862–7875, 2006.

Ring, M.; Zychowska, M.A., and Stephan, R. Dynamics of *Campylobacter* spp. spread investigated in 14 broiler flocks in Switzerland. *Avian Diseases*, 49(3):390–396, 2005.

Ringoir, D.D. and Korolik, V. Colonisation phenotype and colonisation potential differences in *Campylobacter jejuni* strains in chickens before and after passage *in vivo*. *Veterinary Microbiology*, 92(3):225–235, 2003.

Rollins, D.M. and Colwell, R.R. Viable but nonculturable stage of *Campylobacter jejuni* and its role in survival in the natural aquatic environment. *Applied and Environmental Microbiology*, 52(3):531–538, 1986.

Rosef, O.; Rettedal, G., and Lågeide, L. Thermophilic cfampylobacters in surface water: a potential risk of campylobacteriosis. *International Journal of Environmental Health Research*, 11(4):321–327, 2001.

Rozen, Y. and Belkin, S. Survival of enteric bacteria in seawater. *FEMS Microbiology Reviews*, 25(5):513–529, 2001.

Sahin, O.; Fitzgerald, C.; Stroika, S.; Zhao, S.; Sippy, R. J.; Kwan, P.; Plummer, P. J.; Han, J.; Yaeger, M. J., and Zhang, Q. Molecular evidence for zoonotic transmission of an emergent, highly pathogenic *Campylobacter jejuni* clone in the United States. *Journal of Clinical Microbiology*, 50(3):680–687, 2012.

Sails, A.D.; Swaminathan, B., and Fields, P.I. Clonal complexes of *Campylobacter jejuni* identified by multilocus sequence typing correlate with strain associations identified by multilocus enzyme electrophoresis. *Journal of Clinical Microbiology*, 41(9):4058, 2003.

Salama, N.; Guillemin, K.; McDaniel, T.K.; Sherlock, G.; Tompkins, L., and Falkow, S. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proceedings of the national Academy of Sciences*, 97(26):14668–14673, 2000.

Salama, S.M.; Bolton, F.J., and Hutchinson, D.N. Application of a new phagetyping scheme

to *Campylobacters* isolated during outbreaks. *Epidemiology and Infection*, 104(03):405–411, 1990.

Samie, A.; Obi, C.L.; Barrett, L.J.; Powell, S.M., and Guerrant, R.L. Prevalence of *Campylobacter* species, *Helicobacter pylori* and *Arcobacter* species in stool samples from the Venda region, Limpopo, South Africa: Studies using molecular diagnostic methods. *Journal of Infection*, 54 (6):558–566, 2007.

Samuel, M.C.; Vugia, D.J.; Shallow, S.; Marcus, R.; Segler, S.; McGivern, T.; Kassenborg, H.; Reilly, K.; Kennedy, M., and Angulo, F. Epidemiology of sporadic *Campylobacter* infection in the United States and declining trend in incidence, FoodNet 1996–1999. *Clinical Infectious Diseases*, 38(Supplement 3):S165, 2004.

Sánchez-Gracia, A. and Castresana, J. Impact of deep coalescence on the reliability of species tree inference from different types of DNA markers in mammals. *PloS One*, 7(1):e30239, 2012.

Sanger, F.; Nicklen, S., and Coulson, A.R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12):5463–5467, 1977.

Sarkar, Deepayan. *Lattice: Multivariate data visualization with R*. Springer, New York, 2008. URL `http://lmdvr.r-forge.r-project.org`. ISBN 978-0-387-75968-5.

Savill, M.G.; Hudson, J.A.; Ball, A.; Klena, J.D.; Scholes, P.; Whyte, R.J.; McCormick, R.E., and Jankovic, D. Enumeration of *Campylobacter* in New Zealand recreational and drinking waters. *Journal of Applied Microbiology*, 91(1):38–46, 2001.

Schmees, C.; Prinz, C.; Treptau, T.; Rad, R.; Hengst, L.; Voland, P.; Bauer, S.; Brenner, L.; Schmid, R.M., and Gerhard, M. Inhibition of t-cell proliferation by *Helicobacter pylori* γ-glutamyl transpeptidase. *Gastroenterology*, 132(5):1820–1833, 2007.

Schouls, L.M.; Reulen, S.; Duim, B.; Wagenaar, J.A.; Willems, R.J.L.; Dingle, K.E.; Colles, F.M., and Van Embden, J.D.A. Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination. *Journal of Clinical Microbiology*, 41(1):15–26, 2003.

Schwartz, D.C. and Cantor, C.R. Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*, 37(1):67–75, 1984.

Sears, A.; Baker, M.G.; Wilson, N.; Marshall, J.; Muellner, P.; Campbell, D.M.; Lake, R.J., and

French, N.P. Marked campylobacteriosis decline after interventions aimed at poultry, New Zealand. *Emerging Infectious Diseases*, 17(6):1007–1015, 2011.

Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, page btu153, 2014.

Sela, R.J. and Simonoff, J.S. Reemtree: Regression trees with random effects, 2010. URL `http://cran.r-project.org/web/packages/REEMtree/index`.

Sela, R.J. and Simonoff, J.S. Re-em trees: a data mining approach for longitudinal and clustered data. *Machine Learning*, 86(2):169–207, 2012.

Selander, R.K.; Caugant, D.A.; Ochman, H.; Musser, J.M.; Gilmour, M.N., and Whittam, T.S. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Applied and Environmental Microbiology*, 51(5):873, 1986.

Shendure, J.; Porreca, G.J.; Reppas, N.B.; Lin, X.; McCutcheon, J.P.; Rosenbaum, A.M.; Wang, M.D.; Zhang, K.; Mitra, R.D., and Church, G.M. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309(5741):1728–1732, 2005.

Sheppard, S.K.; McCarthy, N.D.; Falush, D., and Maiden, M.C.J. Convergence of *Campylobacter* species: Implications for bacterial evolution. *Science*, 320(5873):237, 2008.

Sheppard, S.K.; Dallas, J.F.; Strachan, N.J.C.; MacRae, M.; McCarthy, N.D.; Wilson, D.J.; Gormley, F.J.; Falush, D.; Ogden, I.D., and Maiden, M.C.J. *Campylobacter* genotyping to determine the source of human infection. *Clinical Infectious Diseases*, 48(8):1072, 2009.

Sheppard, S.K; Colles, F.M.; McCarthy, N.D.; Strachan, N.J.C.; Ogden, I.D.; Forbes, K.J.; Dallas, J.F., and Maiden, M.C.J. Niche segregation and genetic structure of *Campylobacter jejuni* populations from wild and agricultural host species. *Molecular Ecology*, 20(16):3484–3490, 2011a.

Sheppard, S.K.; McCarthy, N.D.; Jolley, K.A., and Maiden, M.C.J. Introgression in the genus *Campylobacter*: generation and spread of mosaic alleles. *Microbiology*, 157(4):1066, 2011b.

Sheppard, S.K.; Didelot, X.; Jolley, K.A.; Darling, A.E.; Pascoe, B.; Meric, G.; Kelly, D.J.; Cody, A.; Colles, F.M., and Strachan, N.J.C. Progressive genome−wide introgression in agricultural *Campylobacter coli*. *Molecular Ecology*, 22(4):1051–1064, 2013a.

Sheppard, S.K.; Didelot, X.; Meric, G.; Torralbo, A.; Jolley, K.A.; Kelly, D.J.; Bentley, S.D.; Maiden, M.C.J.; Parkhill, J., and Falush, D. Genome-wide association study identifies vitamin b5 biosynthesis as a host specificity factor in *Campylobacter*. *Proceedings of the National Academy of Sciences*, 110(29):11923–11927, 2013b.

Sheppard, S.K.; Cheng, L.; Méric, G.; Haan, C.; Llarena, A.-K.; Marttinen, P.; Vidal, A.; Ridley, A.; Clifton-Hadley, F., and Connor, T.R. Cryptic ecology among host generalist *Campylobacter jejuni* in domestic animals. *Molecular Ecology*, 23:2442–2451, 2014.

Shibayama, K.; Wachino, J.; Arakawa, Y.; Saidijam, M.; Rutherford, N.G., and Henderson, P.J.F. Metabolism of glutamine and glutathione via γ-glutamyltranspeptidase and glutamate transport in *Helicobacter pylori*: possible significance in the pathophysiology of the organism. *Molecular Microbiology*, 64(2):396–406, 2007.

Siemer, B.L.; Harrington, C.S.; Nielsen, E.M.; Borck, B.; Nielsen, N.L.; Engberg, J., and On, S.L.W. Genetic relatedness among *Campylobacter jejuni* serotyped isolates of diverse origin as determined by numerical analysis of amplified fragment length polymorphism (AFLP) profiles. *Journal of Applied Microbiology*, 96(4):795–802, 2004.

Sikorski, J. and Nevo, E. Patterns of thermal adaptation of *Bacillus simplex* to the microclimatically contrasting slopes of Evolution Canyons I and II, Israel. *Environmental microbiology*, 9 (3):716–726, 2007.

Sims, G.E.; Jun, S.-R.; Wu, G.A., and Kim, S.-H. Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions. *Proceedings of the National Academy of Sciences*, 106(8):2677–2682, 2009.

Smith, T. Spirilla associated with disease of the fetal membranes in cattle (infectious abortion). *The Journal of Experimental Medicine*, 28(6):701–719, 1918.

Smith, T. The etiological relation of spirilla (*Vibrio fetus*) to bovine abortion. *The Journal of Experimental Medicine*, 30(4):313–323, 1919.

Snyder, L.A.; Loman, N.J.; Faraj, L.A.; Levi, K.; Weinstock, G.; Boswell, T.C.; Pallen, M.J., and AlaAldeen, D.A. Epidemiological investigation of *Pseudomonas aeruginosa* isolates from a six-year-long hospital outbreak using high-throughput whole genome sequencing. *Eurosurveillance*, 18(42), 2013.

Song, Y.C.; Jin, S.; Louie, H.; Ng, D.; Lau, R.; Zhang, Y.; Weerasekera, R.; Al Rashid, S.; Ward, L.A., and Der, S.D. FlaC, a protein of *Campylobacter jejuni* TGH9011 (ATCC43431) secreted through the flagellar apparatus, binds epithelial cells and influences cell invasion. *Molecular Microbiology*, 53(2):541–553, 2004.

Sopwith, W.; Birtles, A.; Matthews, M.; Fox, A.; Gee, S.; Painter, M.; Regan, M.; Syed, Q., and Bolton, E. Identification of potential environmentally adapted *Campylobacter jejuni* strain, United Kingdom. *Emerging Infectious Diseases*, 14(11):1769, 2008.

Southern, E.M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, 98(3):503–517, 1975.

Spencer, S.E.F.; Marshall, J.; Pirie, R.; Campbell, D.; Baker, M.G., and French, N.P. The spatial and temporal determinants of campylobacteriosis notifications in New Zealand, 2001–2007. *Epidemiology and Infection*, 140(09):1663–1677, 2012.

Stackebrandt, E.B.M.G. and Goebel, B.M. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology*, 44(4):846–849, 1994.

Stahl, M.; Friis, L.M.; Nothaft, H.; Liu, X.; Li, J.; Szymanski, C.M., and Stintzi, A. L-fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proceedings of the National Academy of Sciences*, 108(17):7194–7199, 2011.

Stamatakis, A. Raxml-vi-hpc: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21):2688–2690, 2006.

Stanley, K. and Jones, K. Cattle and sheep farms as reservoirs of *Campylobacter*. *Journal of Applied Microbiology*, 94(s1):104–113, 2003.

Stephens, B.B.; Loar, S.N., and Alexandre, G. Role of CheB and CheR in the complex chemotactic and aerotactic pathway of *Azospirillum brasilense*. *Journal of Bacteriology*, 188(13): 4759–4768, 2006.

Stintzi, A. Gene expression profile of *Campylobacter jejuni* in response to growth temperature variation. *Journal of Bacteriology*, 185(6):2009–2016, 2003.

Stokes, E.J. Antibiotic sensitivity testing. *British Medical Journal*, 2(5763):707, 1971.

Storz, G. and Imlayt, J.A. Oxidative stress. *Current Opinion in Microbiology*, 2(2):188–194, 1999.

Strachan, N.J.C.; Rotariu, O.; Smith-Palmer, A.; Cowden, J.; Sheppard, S.K.; O'Brien, S.J.; Maiden, M.C.J.; Macrae, M.; Bessell, P.R., and Matthews, L. Identifying the seasonal origins of human campylobacteriosis. *Epidemiology and Infection*, 141(06):1267–1275, 2013.

Stucki, U.R.S.; Frey, J.; Nicolet, J., and Burnens, A.P. Identification of *Campylobacter jejuni* on the basis of a species-specific gene that encodes a membrane protein. *Journal of Clinical Microbiology*, 33(4):855–859, 1995.

Suerbaum, S.; Smith, J.M.; Bapumia, K.; Morelli, G.; Smith, N.H.; Kunstmann, E.; Dyrek, I.,

and Achtman, M. Free recombination within *Helicobacter pylori*. *Proceedings of the National Academy of Sciences*, 95(21):12619, 1998.

Suerbaum, S.; Lohrengel, M.; Sonnevend, A.; Ruberg, F., and Kist, M. Allelic diversity and recombination in *Campylobacter jejuni*. *Journal of Bacteriology*, 183(8):2553–2559, 2001.

Taboada, E.; MacKinnon, J.; Luebbert, C.; Gannon, V.; Nash, J., and Rahn, K. Comparative genomic assessment of multi-locus sequence typing: rapid accumulation of genomic heterogeneity among clonal isolates of *Campylobacter jejuni*. *BMC Evolutionary Biology*, 8(1):229, 2008.

Tang, J.Y.H.; Carlson, J.; Mohamad Ghazali, F.; Abdul Aziz, S.; Nishibuchi, M.; Nakaguchi, Y., and Radu, S. Phenotypic MicroArray (PM) profiles (carbon sources and sensitivity to osmolytes and pH) of *Campylobacter jejuni* ATCC 33560 in response to temperature. *International Food Research Journal*, 17(4):837–844, 2010.

Tenover, F.C.; Arbeit, R.D.; Goering, R.V.; Mickelsen, P.A.; Murray, B.E.; Persing, D.H., and Swaminathan, B. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *Journal of Clinical Microbiology*, 33(9):2233, 1995.

Tenover, F.C.; Vaughn, R.R.; McDougal, L.K.; Fosheim, G.E., and McGowan, J.E. Multiple-locus variable-number tandem-repeat assay analysis of methicillin-resistant *Staphylococcus aureus* strains. *Journal of Clinical Microbiology*, 45(7):2215–2219, 2007.

Tettelin, H.; Masignani, V.; Cieslewicz, M.J.; Donati, C.; Medini, D.; Ward, N.L.; Angiuoli, S.V.; Crabtree, J.; Jones, A.L., and Durkin, A.S. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial pan-genome. *Proceedings of the National Academy of Sciences of the United States of America*, 102(39):13950–13955, 2005.

Tettelin, H.; Riley, D.; Cattuto, C., and Medini, D. Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, 11(5):472–477, 2008.

The Institute of Environmental Science and Research Ltd., . Notifiable and other diseases in new zealand: Annual report 2006. porirua (nz): Institute of environmental science and research limited; 2007. online, May 2007. URL `https://surv.esr.cri.nz/PDF_surveillance/AnnualRpt/AnnualSurv/`. cited 2014 July 28.

The Institute of Environmental Science and Research Ltd., . Notifiable and other diseases in

new zealand: Annual report 2012. online, April 2013. URL `https://surv.esr.cri.nz/PDF_surveillance/AnnualRpt/AnnualSurv/2012/2012AnnualSurvRpt.pdf`.

Thomas, D.K.; Lone, A.G.; Selinger, L.B.; Taboada, E.N.; Uwiera, R.R.E.; Abbott, D.W., and Inglis, G.D. Comparative variation within the genome of *Campylobacter jejuni* NCTC 11168 in human and murine hosts. *PloS One*, 9(2):e88229, 2014.

Thomas, L.M.; Long, K.A.; Good, R.T.; Panaccio, M., and Widders, P.R. Genotypic diversity among *Campylobacter jejuni* isolates in a commercial broiler flock. *Applied and Environmental Microbiology*, 63(5):1874–1877, 1997.

Timp, W.; Mirsaidov, U.M.; Wang, D.; Comer, J.; Aksimentiev, A., and Timp, G. Nanopore sequencing: electrical measurements of the code of life. *Nanotechnology, IEEE Transactions on*, 9(3):281–294, 2010. doi: 10.1109/TNANO.2010.2044418.

Toft, C. and Andersson, S.G.E. Evolutionary microbial genomics: insights into bacterial host adaptation. *Nature Reviews Genetics*, 11(7):465–475, 2010.

Tu, Q.V.; McGuckin, M.A., and Mendz, G.L. *Campylobacter jejuni* response to human mucin MUC2: modulation of colonization and pathogenicity determinants. *Journal of Medical Microbiology*, 57(7):795–802, 2008.

Tyler, K.D.; Wang, G.; Tyler, S.D., and Johnson, W.M. Factors affecting reliability and reproducibility of amplification-based DNA fingerprinting of representative bacterial pathogens. *Journal of Clinical Microbiology*, 35(2):339, 1997.

Vaas, L.A.; Sikorski, J.; Michael, V.; Göker, M., and Klenk, H.-P. Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics. *PLoS One*, 7(4):e34846, 2012.

Vaas, L.A.; Sikorski, J.; Hofner, B.; Fiebig, A.; Buddruhs, N.; Klenk, H.-P., and Göker, M. opm: an R package for analysing OmniLog® phenotype microarray data. *Bioinformatics*, 29(14): 1823–1824, 2013.

Van Bergen, M.A.P.; Dingle, K.E.; Maiden, M.C.J.; Newell, D.G.; van der Graaf-Van Bloois, L.; van Putten, J.P.M., and Wagenaar, J.A. Clonal nature of *Campylobacter fetus* as defined by multilocus sequence typing. *Journal of Clinical Microbiology*, 43(12):5888–5898, 2005.

Van de Giessen, A. W.; Heuvelman, C. J.; Abee, T., and Hazeleger, W.C. Experimental studies on the infectivity of non-culturable forms of *Campylobacter* spp. in chicks and mice. *Epidemiology and Infection*, 117(03):463–470, 1996.

van den Berg, R.J.; Schaap, I.; Templeton, K.E.; Klaassen, C.H.W., and Kuijper, E.J. Typing and subtyping of *Clostridium difficile* isolates by using multiple−locus variable−number tandem−repeat analysis. *Journal of Clinical Microbiology*, 45(3):1024–1028, 2007.

van den Ent, F.; Amos, L.A., and LoÈwe, J. Prokaryotic origin of the actin cytoskeleton. *Nature*, 413(6851):39–44, 2001.

Van Der Beek, M.T.; Claas, E.C.J.; Mevius, D.J.; Van Pelt, W.; Wagenaar, J.A., and Kuijper, E.J. Inaccuracy of routine susceptibility tests for detection of erythromycin resistance of *Campylobacter jejuni* and *Campylobacter coli*. *Clinical Microbiology and Infection*, 16(1): 51–56, 2010.

Van Pelt, W.; Van De Giessen, A .W.; Van Leeuwen, W. J.; Wannet, W.; Henken, A. M., and G., Evers E. Oorsprong, omvang en kosten van humane salmonellose. Deel 1. Oorsprong van human salmonellose met betrekking tot varken, rund, kip, ei en overige bronnen. *Infectiezikten Bulletin*, pages 240–243, 1999.

Vandamme, P. *Campylobacter*, chapter Taxonomy of the family Campylobacteraceae, pages 3–26. Campylobacter. ASM Press, 2000. ISBN 1-55581-165-5.

Vandamme, P. and De Ley, J. Proposal for a new family, *Campylobacteraceae*. *International Journal of Systematic Bacteriology*, 41(3):451–455, 1991.

Vandamme, P. and Goossens, H. Taxonomy of *Campylobacter*, *Arcobacter*, and *Helicobacter*: a review. *Zentralblatt fur Bakteriologie: International Journal of Medical Microbiology*, 276(4): 447, 1992.

Vandamme, P. and On, S.L. Recommendations of the subcommittee on the taxonomy of *Campylobacter* and related bacteria. *International Journal of Systematic and Evolutionary Microbiology*, 51(2):719–721, 2001.

Vandamme, P.; Falsen, E.; Rossau, R.; Hoste, B.; Segers, P.; Tytgat, R., and De Ley, J. Revision of *Campylobacter*, *Helicobacter*, and *Wolinella* taxonomy: emendation of generic descriptions and proposal of *Arcobacter* gen. nov. *International Journal of Systematic Bacteriology*, 41(1): 88–103, 1991.

Vegge, C.S.; Brøndsted, L.; Li, Y.-P.; Bang, D.D., and Ingmer, H. Energy taxis drives *Campylobacter jejuni* toward the most favorable conditions for growth. *Applied and Environmental Microbiology*, 75(16):5308–5314, 2009.

Vegge, C.S.; Brøndsted, L.; Ligowska-Marzketa, M., and Ingmer, H. Natural transformation of *Campylobacter jejuni* occurs beyond limits of growth. *PloS One*, 7(9):e45467, 2012.

Vergnaud, G. and Denoeud, F. Minisatellites: mutability and genome architecture. *Genome Research*, 10(7):899–907, 2000.

Véron, M. and Chatelain, R. Taxonomic study of the genus *Campylobacter* Sebald and Véron and designation of the neotype strain for the type species, *Campylobacter fetus* (Smith and Taylor) Sebald and Véron. *International Journal of Systematic Bacteriology*, 23(2):122–134, 1973.

Vilchez-Vargas, R.; Junca, H., and Pieper, D.H. Metabolic networks, microbial ecology and omics technologies: towards understanding in situ biodegradation processes. *Environmental Microbiology*, 12(12):3089–3104, 2010.

Vos, P.; Hogers, R.; Bleeker, M.; Reijans, M.; van De Lee, T.; Hornes, M.; Friters, A.; Pot, J.; Paleman, J., and Kuiper, M. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, 23(21):4407–4414, 1995.

Waldenström, J.; Broman, T.; Carlsson, I.; Hasselquist, D.; Achterberg, R.P.; Wagenaar, J.A., and Olsen, B. Prevalence of *Campylobacter jejuni*, *Campylobacter lari*, and *Campylobacter coli* in different ecological guilds and taxa of migrating birds. *Applied and Environmental Microbiology*, 68(12):5911–5917, 2002.

Walder, M.; Sandstedt, K., and Ursing, J. Phenotypic characteristics of thermotolerant *Campylobacter* from human and animal sources. *Current Microbiology*, 9(5):291–295, 1983.

Wang, Y. and Taylor, D.E. Natural transformation in *Campylobacter* species. *Journal of Bacteriology*, 172(2):949–955, 1990.

Wanscher, J.H. The history of Wilhelm Johannsen's genetical terms and concepts from the period 1903 to 1926. *Centaurus*, 19(2):125–147, 1975.

Wassenaar, T.M. and Newell, D.G. Genotyping of *Campylobacter* spp. *Applied and Environmental Microbiology*, 66(1):1, 2000.

Wassenaar, T.M.; van der Zeijst, B.A.M.; Ayling, R., and Newell, D.G. Colonization of chicks by motility mutants of *Campylobacter jejuni* demonstrates the importance of flagellin a expression. *Journal of General Microbiology*, 139(6):1171–1175, 1993.

Weisent, J.; Seaver, W.; Odoi, A., and Rohrbach, B. The importance of climatic factors and

outliers in predicting regional monthly campylobacteriosis risk in Georgia, USA. *International Journal of Biometeorology*, 58:1865–1878, 2014.

Welkos, S.L. Experimental gastroenteritis in newly-hatched chicks infected with *Campylobacter jejuni*. *Journal of Medical Microbiology*, 18(2):233–248, 1984.

Wetterstrand, K.A. Dna sequencing costs: Data from the nhgri genome sequencing program (gsp). online, October 2014. URL `www.genome.gov/sequencingcosts`.

Whiley, H.; van den Akker, B.; Giglio, S., and Bentham, R. The role of environmental reservoirs in human campylobacteriosis. *International Journal of Environmental Research and Public Health*, 10(11):5886–5907, 2013.

Whitehouse, C.A.; Balbo, P.B.; Pesci, E.C.; Cottle, D.L.; Mirabito, P.M., and Pickett, C.L. *Campylobacter jejuni* cytolethal distending toxin causes a G2-phase cell cycle block. *Infection and Immunity*, 66(5):1934–1940, 1998.

Wichelhaus, T.A.; Hunfeld, K.-P.; Böddinghaus, B.; Kraiczy, P.; Schäfer, V., and Brade, V. Rapid molecular typing of methicillin resistant *Staphylococcus aureus* by PCR-RFLP. *Infection Control and Hospital Epidemiology*, 22(5):294–298, 2001.

Wiesner, R.S.; Hendrixson, D.R., and DiRita, V.J. Natural transformation of *Campylobacter jejuni* requires components of a type II secretion system. *Journal of Bacteriology*, 185(18): 5408–5418, 2003.

Willenbrock, H.; Petersen, A.; Sekse, C.; Kiil, K.; Wasteson, Y., and Ussery, D.W. Design of a seven-genome *Escherichia coli* microarray for comparative genomic profiling. *Journal of Bacteriology*, 188(22):7713–7721, 2006.

Williams, M.; Gummer, H.; Powlesland, R; Robertson, H., and Taylor, G. *Migrations and movements of birds to New Zealand and surrounding seas*. Science & Technical Publishing, Department of Conservation, 2006.

Williams, N.J.; Jones, T.R.; Leatherbarrow, H.J.; Birtles, R.J.; Lahuerta-Marin, A.; Bennett, M., and Winstanley, C. Isolation of a novel *Campylobacter jejuni* clone associated with the bank vole, *Myodes glareolus*. *Applied and Environmental Microbiology*, 76(21):7318–7321, 2010.

Wilson, D.J.; Gabriel, E.; Leatherbarrow, A.J.H.; Cheesbrough, J.; Gee, S.; Bolton, E.; Fox, A.; Fearnhead, P.; Hart, C.A., and Diggle, P.J. Tracing the source of campylobacteriosis. *PLoS Genetics*, 4(9):e1000203, 2008.

Wilson, D.J.; Gabriel, E.; Leatherbarrow, A.J.H.; Cheesbrough, J.; Gee, S.; Bolton, E.; Fox, A.; Hart, C.A.; Diggle, P.J., and Fearnhead, P. Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. *Molecular Biology and Evolution*, 26(2):385–397, 2009.

Winn, W.; Allen, S.; Janda, W.; Koneman, E.; Procop, G.; Schreckenberger, P., and Woods, G. *Koneman's color atlas and textbook of diagnostic microbiology*. Lippincott Williams & Wilkins, 6th edition, 2006.

Woese, C.R. Bacterial evolution. *Microbiological Reviews*, 51(2):221, 1987.

Woodall, C.A.; Jones, M.A.; Barrow, P.A.; Hinds, J.; Marsden, G.L.; Kelly, D.J.; Dorrell, N.; Wren, B.W., and Maskell, D.J. *Campylobacter jejuni* gene expression in the chick cecum: evidence for adaptation to a low-oxygen environment. *Infection and Immunity*, 73(8):5278–5285, 2005.

Woodward, D.L. and Rodgers, F.G. Identification of *Campylobacter* heat-stable and heat-labile antigens by combining the Penner and Lior serotyping schemes. *Journal of Clinical Microbiology*, 40(3):741–745, 2002.

Wösten, M.M.S.M.; Wagenaar, J.A., and van Putten, J.P.M. The FlgS/FlgR two-component signal transduction system regulates the fla regulon in *Campylobacter jejuni*. *Journal of Biological Chemistry*, 279(16):16214–16222, 2004.

Wu, Z.; Sippy, R.; Sahin, O.; Plummer, P.; Vidal, A.; Newell, D., and Zhang, Q. Genetic diversity and antimicrobial susceptibility of *Campylobacter jejuni* isolates associated with sheep abortion in the United States and Great Britain. *Journal of Clinical Microbiology*, 52 (6):1853–1861, 2014.

Xu, F.; Zeng, X.; Haigh, R.D.; Ketley, J.M., and Lin, J. Identification and characterization of a new ferric enterobactin receptor, CfrB, in *Campylobacter*. *Journal of Bacteriology*, 192(17): 4425–4435, 2010.

Yang, J.; Roy, A., and Zhang, Y. Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, 29 (20):2588–2595, 2013a.

Yang, J.; Roy, A., and Zhang, Y. BioLiP: a semi-manually curated database for biologically relevant ligand–protein interactions. *Nucleic Acids Research*, 41(D1):D1096–D1103, 2013b.

Yao, R.; Burr, D.H., and Guerry, P. CheY-mediated modulation of *Campylobacter jejuni* virulence. *Molecular Microbiology*, 23(5):1021–1031, 1997.

Young, K.T.; Davis, L.M., and DiRita, V.J. *Campylobacter jejuni*: molecular biology and pathogenesis. *Nature Reviews Microbiology*, 5(9):665–679, 2007.

Yu, S.; Fearnhead, P.; Holland, B.R.; Biggs, P.; Maiden, M., and French, N. Estimating the relative roles of recombination and point mutation in the generation of single locus variants in *Campylobacter jejuni* and *Campylobacter coli*. *Journal of Molecular Evolution*, 74(5-6): 273–280, 2012.

Zappe Pasturel, B.; Cruz-Cano, R.; Rosenberg Goldstein, R.E.; Palmer, A.; Blythe, D.; Ryan, P.; Hogan, B.; Jung, C.; Joseph, S.W., and Wang, M.Q. Impact of rurality, broiler operations, and community socioeconomic factors on the risk of campylobacteriosis in Maryland. *American Journal of Public Health*, 103(12):2267–2275, 2013.

Zautner, A.E.; Herrmann, S.; Corso, J.; Tareen, A.M.; Alter, T., and Groß, U. Epidemiological association of different *Campylobacter jejuni* groups with metabolism-associated genetic markers. *Applied and Environmental Microbiology*, 77(7):2359–2365, 2011.

Zerbino, D.R. and Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18(5):821–829, 2008.

Zhang, W.; Li, F., and Nie, L. Integrating multiple omics analysis for microbial biology: application and methodologies. *Microbiology*, 156(2):287–301, 2010.

Zhao, S.; Mitchell, S.E.; Meng, J.; Kresovich, S.; Doyle, M.P.; Dean, R.E.; Casa, A.M., and Weller, J.W. Genomic typing of *Escherichia coli* O157: H7 by semi-automated fluorescent AFLP analysis. *Microbes and Infection*, 2(2):107–113, 2000.

Zilbauer, M.; Dorrell, N.; Wren, B.W., and Bajaj-Elliott, M. *Campylobacter jejuni* mediated disease pathogenesis: an update. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 102(2):123–129, 2008.

Zuckerkandl, E. and Pauling, L. Evolutionary divergence and convergence in proteins. *Evolving Genes and Proteins*, 97:97–166, 1965.

Zuckerkandl, E. and Pauling, L.B. *Molecular disease, evolution, and genetic heterogeneity*. Academic Press, New York, 1962. URL `http://www.evolocus.com/Publications/Zuckerkandl1962.pdf`.