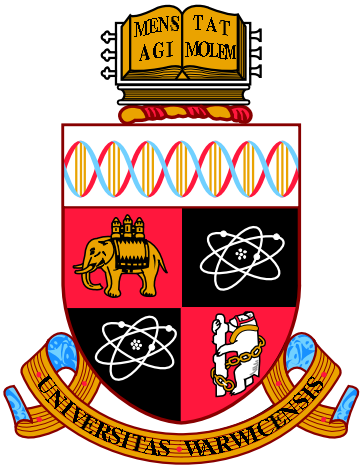

Transcriptomic analysis predicts the identities and biological roles of *Campylobacter jejuni* small RNAs

by

Stephen Li

Thesis

Submitted to the University of Warwick

for the degree of Interdisciplinery Biomedical Science

Doctor of Philosophy

Supervisors: Dr Chrystala Constantinidou, Dr Tauqeer Alam

Warwick Medical School

April 2022

Contents

List of Tables

List of Figures

Acknowledgments

First I would like express my gratitude to my primary supervisor Dr. Chrystala Constantinidou, who have been supportive throughout both for technical and personal challenges. Thank you for the patience and guidance that have navigated through countless obstacles during the experiments and analysis through the past 4 years, and also for showing understanding to my personal difficulties. Also a big thanks to my second supervisor Dr. Mohammad Tauqeer Alam, especially for the mentorship that allows me improve my bioinformatics skills despite having limited bioinformatics experience prior to the PhD project. Also thank you for understanding the challenge I faced as a foreigner during the COVID lockdown. Moreover, I would also like to thank Dr. Emma Denham, who has been welcoming when I first applied to Warwick and for the later collaboration that have subsequently led to a publication.

I also appreciate the support and advice from the advisory panel Dr. Meera Unnkrishnan and Dr. Sascha Ott. In particular, those advice from our latest PhD review has reminded me the importance of finding a clear research direction and to translate my computational analysis into a biological answer. Also thank for Sascha for linking me up with Dr. Nigel Dyer, who has kindly taken his time to advise on the RNA crosslinking analysis.

Also thank you for the support from MRC-DTP, especially for Sally and Jonathan. I am especially thankful that you have taken my unexpected circumstances into consideration and providing the extension and support required. I also want to express my appreciation to the rest of the MRC cohort, for providing both academic

and social support throughout the year. I will remember those time where we play squash, football and judo together.

To former and present members of M116 (Shath, Jeff, Lucy, Leo, Grace, Joe, Alexia, Kate, Jack, Swati, Arnaud, Sandra, Nina, Giri, Amy, Elena and those who joined after I left UK due to COVID), thank you for making M116 a friendly environment to work in. Also thank you Leo for the help with the submitted manuscript. Among all labs I have worked in so far, M116 is one of the friendliest and most supportive lab I have worked in so far. To those who have finished before me, all the best to your future career and hopefully we will cross path again in the future either as colleagues or collaborators. To those who started after me, all the best with your PhD and may you have the perseverance to push through the rest of your PhD.

Special thanks to the Constantinidou lab. Thank you to Jenna Lam for the day to day supervision, and for conversations on all sorts of topics ranging from Hong Kong, anime and Warwick ABACUS. Also a huge thanks for helping with proofreading our paper and my thesis despite having a full time job. Thank you to Dagnija Tupina for sharing the same love for growing *C. jejuni*. Also a huge respect for recently voicing out for Ukraine. Also thanks Emily Stoakes for generating those RNA-seq data, which was used in the sRNA prediction. And to Tom and Andrea, all the best to the rest of the PhD and for growing *C. jejuni*.

I would also like to express my gratitude to friends outside work. First to my housemates Danny and Dan, thank you for sharing countless anime banter and intellectual discussion on quotes such as "Food is just food". Hope you will both do well at handling your thesis defense. Also thank you to other friends from the Christian Union and Christian Postgraduate and Staff Network Warwick (e.g. Judith, Joanna, Andy, Manus, Martin, James, Harry, Vivian, Clytie, E-Wen, Theo, Venessa ....), for showing support especially when Hong Kong was having a difficult time politically, and for reminding me to stand firm in faith and keep seeking God's justice in the midst of hardship. Also thank you to friends in Hong Kong (Kingston,

Cheris, Seth, Jeffrey, Mark, Caila, Ellie, Kan, Wing, Jason, Kenneath, Steven ...) , especially after I returned there due to the COVID situation. Being in this city hasn't been easy, given what is happening here. Your support has enabled me to keep going with PhD and personal life here in HK.

Special mention to my family, for the tremendous support for my education in all these years. Thank you for sending me to the UK for education, where I realised my passion in biochemistry and decided to pursue a research career. And special thanks to Agnes Wong, for being there to go through all those struggles. May we keep moving forward together as we faces more challenges in the future.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

Some parts of Chapters 3 and 4 have been included in a submitted to a manuscript, which is under revision at Microbiology Spectrum.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- The transcriptomic data were generated by Dr. Jenna Lam and Dr. Emily Stoakes.
- Parts of the scripts for RNA crosslinking analysis was written by Adam Callan-Sidat.

Abstract

Campylobacter jejuni (*C. jejuni*) is a zoonotic pathogen responsible for millions of bacterial gastroenteritis cases each year worldwide. The World Health Organisation (WHO) has also highlighted *Campylobacter* as one of the pathogens with high priority for research due to its emerging antibiotic resistance. One remarkable feature of *C. jejuni* is its persistence under various environmental stresses throughout transmission. However, very little is known about its genetic regulatory network. Annotation of the *C. jejuni* genome has only identified three sigma factors, suggesting additional regulatory mechanisms.

One potential regulatory element that requires further investigation is small RNA (sRNA). Studies in other pathogens have shown sRNA mediates post-transcriptional regulation by affecting translation initiation and mRNA stability, regulating processes involved in stress survival, virulence and antibiotic resistance. However, experimental discovery of *C. jejuni* sRNA is hindered by its fastidious nature and difficulties of experimental investigation. Several published transcriptomic studies on *C. jejuni* have discovered relatively few sRNAs, with limited insights into their biological activities. Moreover, most published RNA-seq data were obtained from standard laboratory conditions, limiting the exploration of stress adaptation by sRNAs.

This study aims to elucidate the condition-specific expression and activities of *C. jejuni* sRNAs. sRNA prediction was carried out using multiple tools that detect sRNA positions from genomic and transcriptomic data. The prediction produced a list of putative sRNAs. Further analysis of putative sRNA binding targets was carried

out using in-house transcriptomic data, including 21 experimental conditions and aids in understanding sRNA activities under non-standard conditions. The bioinformatics analysis has generated a global sRNA-target regulatory network, followed by a detailed analysis of sRNAs related to food storage conditions and iron stress. In order to complement the limitations of bioinformatics prediction, sRNA-target interactions were identified experimentally using RNA crosslinking. The results have identified interactions that are absent from the computational prediction. The difference between RNA crosslinking and computational prediction may indicate different sRNA-regulatory mechanisms. Overall, this study has highlighted RNA-RNA interactions for future detailed analysis.

Chapter 1

Introduction

1.1 *C. jejuni* biology

Campylobacter jejuni (*C. jejuni*) is a leading cause of foodborne diarrhoea worldwide. Infection with *Campylobacter* bacterium (campylobacteriosis) causes about 400 to 500 million cases of diarrhoea annually around the globe (Young et al., 2007). Even well-resourced countries suffer from millions of cases of Campylobacteriosis each year. For example, the Centers for Disease Control and Prevention records 1.3 million annual cases of Campylobacteriosis in the United States, while an estimated half a million *Campylobacter* infections are reported in England and Wales annually (EFSA, 2019). In addition, *Campylobacter* is the most common cause of bacterial gastroenteritis in the European Union (EU), with nearly 250,000 confirmed cases in the European Union in 2018 (EFSA, 2019). The number of confirmed cases may not fully highlight the medical impact of Campylobacteriosis due to the potential under-reporting of clinical Campylobacteriosis (O'Mahony et al., 2011; EFSA, 2010).

While most *C. jejuni* infection induces self-limited diarrheal symptoms, Campylobacteriosis can lead to chronic complications such as reactive arthritis (RA) or Guillain-Barré syndrome (GBS) (Mishu and Blaser, 1993) and other long-term physical and cognitive impairments for children (Nataro and Guerrant, 2017). According to a report in 1995, Campylobacter-associated GBS cost around 136 million to 1.3 billion US dollars (Buzby et al., 1997). Moreover, *Campylobacter* can cause fatal dehydrating diarrhoea and dysentery, especially in poorer regions. Data from the Global Burden of Disease Cause of Death (CoD) database showed *Campylobacter* infection caused over 30,000 deaths in young children in 2015, mostly from less

developed countries (COD, 2017).

C. jejuni infects the gastrointestinal (GI) tract by penetrating the intestinal epithelial layer by reorganising microtubules and microfilaments (Monteville et al., 2003; Oelschlaeger et al., 1993). The internalisation of *C. jejuni* enables immune evasion and the induction of the MAP kinase signal pathway that leads to the production of interleukin-8 (IL-8), a cytokine that activates the host immune response that results in inflammatory diarrhoea (Hickey et al., 1999; Watson and Galan, 2005). IL-8 induction is associated with a cytolethal distending toxin (CDT), for which the exact molecular mechanism remains elusive (Hickey et al., 2000). The contribution of CDT to Campylobacteriosis is questionable as some CDT-negative clinical isolates showed no difference in clinical severity (Abuoun et al., 2005; Mortensen et al., 2011). The induced signal transduction also activates the production of interleukin-10 (IL-10), which facilitates intracellular survival during infection (Li et al., 2011).

While most cases of Campylobacteriosis are self-limiting, the rare GBS complications leads to an autoimmune response that damages peripheral nerves. The autoimmune response is due to the false recognition of lipopolysaccharide (LPS) by host macrophages and T cells. The structural similarity between *C. jejuni* LPS and the host ganglioside leads to the production of antibodies that falsely target the host ganglioside as foreign agents, subsequently causing the damage of the peripheral nervous system (Rees et al., 1995). Moreover, the activated macrophages also release cytokines and free radicals that contribute to demyelination which blocks nerve conduction and triggers axon degeneration (Nyati et al., 2010). Demyelination is partly a result of the cytokines released from activated T cells (reviewed in Nyati and Nyati, 2013).

Campylobacter also demonstrates differential pathogenicity in different hosts. Diarrhoea among avians as a result of *C. jejuni* infection only occurs sporadically (Sanyal et al., 1984). Meanwhile, colonisation in pregnant sheep, goats and cattle leads to abortion (Sahin et al., 2017), thus causing an economic loss for the farming industry. In contrast, *C. jejuni* that has colonised the human gastrointestinal tract induces inflammation, leading to bloody diarrhoea, fever and abdominal pain (Young et al., 2007). On some less common occasions, the inflammatory response induced by *C. jejuni* can also induce GBS (Rhodes and Tattershield, 1982).

1.1.1 *C. jejuni* reservoirs

Campylobacter excels at surviving in various environments along the transmission cycle (Figure 1.1). *Campylobacter* species reside in broiler chickens as a primary

reservoir that accounts for up to 80 % of infections (Sheppard et al., 2009; O'Mahony et al., 2011; Gormley et al., 2014). In chickens, *C. jejuni* resides typically as a commensal through colonisation of the mucosal layer of the gastrointestinal tract without penetrating the intestinal epithelial layer (Beery et al., 1988), with the possibility of spreading to extra-intestinal tissues through acquiring virulence factors (Laconi et al., 2021). Chicken colonisation by *Campylobacter* is associated with the administration of antibiotics for poultry farming, as antibiotics disrupt the gut microbiota and thus remove the protective barrier against *Campylobacter* colonisation (Han et al., 2019).

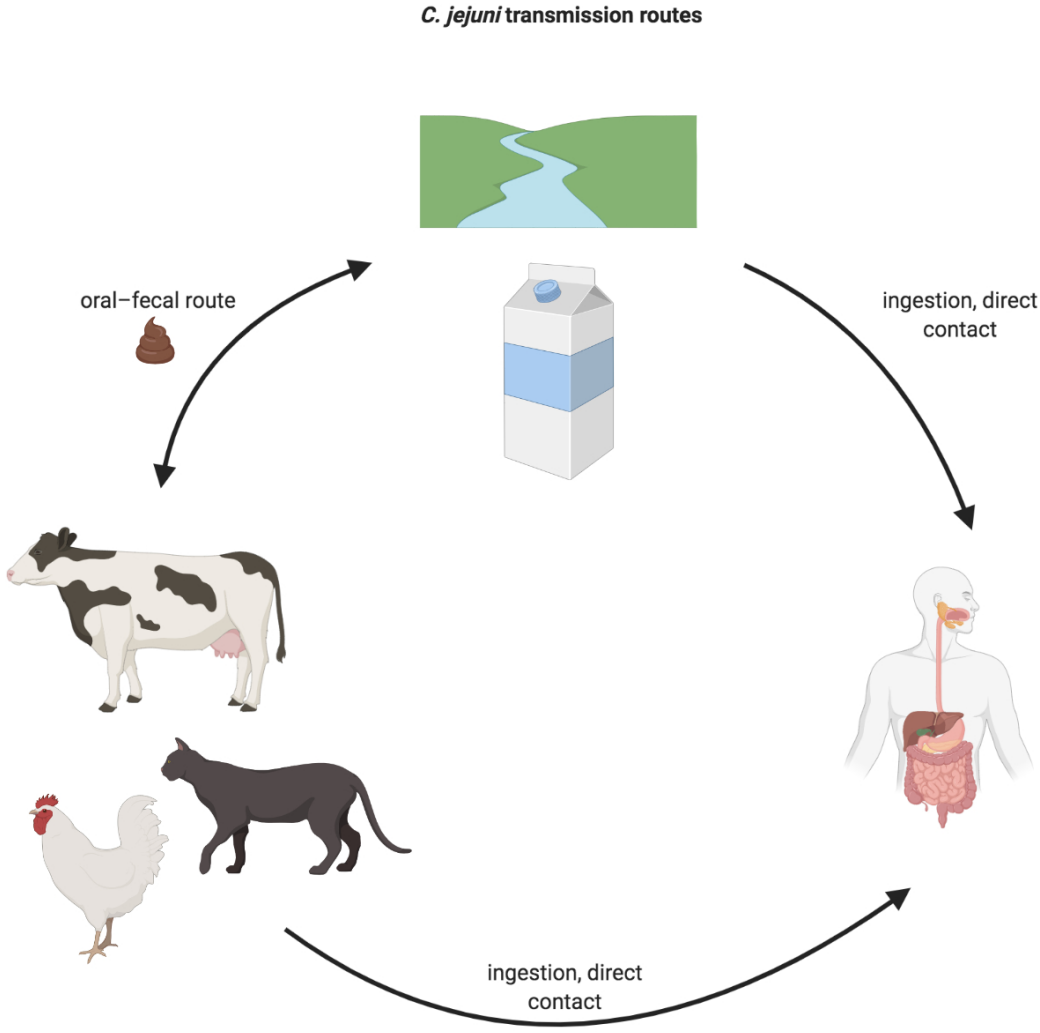

Figure 1.1: **Summary of *C. jejuni* transmission route.** Possible transmission routes include direct transmission through contract or ingestion of animal hosts (primarily chicken), or indirectly via external sources such as water or milk. Created in BioRender.com.

Other animal sources of *Campylobacter* include cattle. A study of *Campylobacter* diversity has identified strains that specialise in cattle adaptation. The cattle adaptation phenotype emerges as modern intensive livestock farming grows, allowing faster horizontal gene transfer (Hull et al., 2021). For instance, several flagellin O-linked glycosylation system genes are missing in the cattle specialised strains. The absence of flagellin O-linked glycosylation genes coincides with reduced cell hydrophobicity, autoagglutination and biofilm formation (Mourkas et al., 2020). Such changes help *Campylobacter* attach to different intestinal surfaces. Moreover, cattle strains acquired thiamine biosynthesis genes by recombination, which may help maintain a sufficient amount of Vitamin B1, which is less abundant in cattle intestines compared to poultry (Mourkas et al., 2020).

Apart from livestock, a recent genotyping study also identified *Campylobacter* in nearly 40 % of tested domestic pets such as cats and dogs. Pets might have contracted *Campylobacter* through direct contact with livestock or the environment (Santaniello et al., 2021). *Campylobacter* in pets needs serious attention as pets are responsible for 10–25 % of clinical cases (Thépault et al., 2020) while whole-genome sequencing of *C. jejuni* isolates from diseased animals (cats, dogs and cattles) identified 41 and 61.1 % of fluoroquinolone-resistance and tetracyclines-resistance isolates, respectively (Moser et al., 2020; Ma et al., 2021).

Campylobacter species transmit from animal reservoirs to human hosts mainly through ingesting contaminated food, such as uncooked chicken meat. *Campylobacter* can also migrate between animals and humans through the environmental route. For example, bacterial cells from the oral-faecal route leave one animal host through faecal particles before being ingested by another host. Such a route requires survival outside the hosts. For example, faecal chicken particles sometimes carry *Campylobacter* into drinking water. While drinking water is low in nutrient content, *C. jejuni* can survive in water for days at temperatures ranging from 4 to 25 °C. (González and Hänninen, 2012). Based on samples from Czech Republic, Strakova et al. (2021) has identified higher occurrence of *C. jejuni* and *C. coli* in wastewater compared to surface water, especially during autumn and in the presence of ammonium (above 0.2 mg/L) and chloride ion (above 60 mg/L) (Strakova et al., 2022). Moreover, different strains prefer different habitats. For example, *C. jejuni* strain 81116 and 81-176 are more adapted to water and milk, respectively (Trigui et al., 2017). *C. jejuni* strains in milk have caused several Campylobacteriosis outbreaks via raw milk consumption (Jaakkonen et al., 2020).

1.1.2 Metabolic pathways

The whole *C. jejuni* genome sequence reveals some peculiar features. It is a compact genome of about 1.6 Mb, in contrast to the genomes of *Salmonella enterica* and *Shigella flexneri*, which are longer than 4.6 Mb. Moreover, the *C. jejuni* genome distinguishes itself from model organism genomes (e.g. *E. coli*) with an AT content of nearly 70 %. Genome annotation of *C. jejuni* has only identified three sigma factors, namely sigma-70, sigma-54 and sigma-28 (Parkhill et al., 2000). That is relatively few compared to the seven known sigma factors in the *E. coli* genome. Comparison between *C. jejuni* and *E. coli* sigma-70 promoters showed similar -10 region but completely different -35 sequences (Wösten et al., 1998).

Several classical gene systems exhibit different genetic structures, which increases the difficulty of gene function annotation by sequence homology. For example, *C. jejuni* lacks the classical glycolysis pathway and other pathways for carbohydrate catabolism. For instance, most *C. jejuni* genome does not encode glucokinase that phosphorylates glucose to glucose-6-phosphate, nor the 6-phosphofructokinase that irreversibly phosphorylates fructose 6-phosphate to fructose-1,6-biphosphate (Velayudhan and Kelly, 2002). Other glucose metabolic pathways are also lacking or scarcely populated. For instance, a recent systematic search revealed that only 1.7 % of over 6000 *C. jejuni* and *C. coli* genomes on the PubMLST database comprise of a complete Entner-Doudoroff (ED) pathway (Vegge et al., 2016). All these are evidence that *C. jejuni* metabolises alternative nutrients other than glucose.

Recent metabolomics data of *C. jejuni* demonstrated complete depletion of aspartic acid, glutamine, methionine, proline, and serine from the growth medium within four hours, while degradation products of fucose were also differentially enriched upon L-fucose supplementation (van der Hooft et al., 2018). Amino acid utilisation occurs sequentially, with aspartate and serine used preferentially before glutamate (van der Hooft et al., 2018).

Another study also showed that *C. jejuni* utilised dipeptides and tripeptides by extracellular uptake using Cj0917 (homologous to carbon starvation protein A (CstA) from *Escherichia coli*) (Rasmussen et al., 2013). These results confirm that fucose and amino acids are the primary energy source for *C. jejuni* and fits previous characterisation of the *C. jejuni* fucose and amino acid metabolic pathways.

The fucose permease FucP (Cj0486) imports free fucose from intestinal mucins and host glycans produced by commensal bacteria. The impairment of fucose import by Cj0486 mutation led to a competitive disadvantage when colonising birds and

when used in piglet models (Stahl et al., 2011; Muraoka and Zhang, 2011). When co-culturing with *Bacteroides vulgatus*, *C. jejuni* exhibited enhanced growth and increased fucose uptake under the reduced concentration of amino acids, suggesting *C. jejuni* may scavenge free fucose released by the fucosidase reaction from other microbiota members (Garber et al., 2020). The crystal structure of fucose dehydrogenase FucX also suggests it reduces L-fucose and D-arabinose into pyruvate and other TCA cycle intermediates, indicating L-fucose and D-arabinose as alternate carbon sources under amino acid shortage (Garber et al., 2020). Catabolism of other sugars is rare among *Campylobacter* species.

Glutamate may also act as an entry point to the central metabolism of proline and glutamine. After entering the cytosol, proline converts into glutamate through the action of PutA before being further metabolised to aspartate. Both PutA and PutP, which are responsible for proline utilisation and import, exhibited upregulation during stationary phase entry (Wright et al., 2009). The result suggested that proline metabolism persists throughout the stationary phase, contrary to aspartate and serine consumption that ends after the mid-log phase.

In contrast, glutamate is metabolised throughout growth and more rapidly during the stationary phase. That suggests proline and glutamate serve as substrates that maintain stationary phase survival instead of growth (Wright et al., 2009). *C. jejuni* 81-176, but not strain NCTC11168, can also utilise glutamine for growth, as it encodes a secreted glutamyltranspeptidase (GGT) which metabolises periplasmic glutamine to glutamate. The glutamate enters the cells by Peb1a, an aspartate/glutamate-binding protein (Hofreuter et al., 2008). In addition to the central metabolism pathway, the leucine, isoleucine, and valine (LIV) branched-chain amino acid transport system also contributes to acquiring leucine, isoleucine, and valine and promoting chick colonisation (Ribardo and Hendrixson, 2011). In addition, *Campylobacter* peptide transporter A (CptA) mediates sulphur metabolism by metabolising host-derived cysteine-containing peptides (Vorwerk et al., 2014).

Besides amino acids and fucose, *C. jejuni* can also consume short-chain fatty acids, including acetate, which is abundant in the lower chicken intestinal tracts. Acetates are products of pyruvate acetogenesis during the exponential growth phase by the AckA–Pta pathway and Acs gene. As serine and aspartate are depleted during entry to the stationary phase, the reversed AckA–Pta pathway converts acetate to acetyl-CoA to enter the TCA cycle (Wright et al., 2009). The increased acetate level correlates to the upregulation of amino acid transporter genes such as *peb1c* and *ggt*, suggesting microbiota-derived acetate may signal chicken lower intestine

colonisation (Luethy et al., 2017). *C. jejuni* can also metabolise lactate. Microbiota-derived lactate enters the cells via the lactate transporter LctP before conversion into pyruvate by lactate dehydrogenases (*lldEFG*/*lutABC*) (Thomas et al., 2011).

1.1.3 Survival under external stress

Despite displaying metabolic pathways and genetic structure distinctive from other bacterial species, *C. jejuni* remains a successful pathogen. Transmission from natural hosts to humans presents *C. jejuni* with environmental challenges. For instance, the body temperature of humans and chickens are 37 and 42 °C, respectively. *C. jejuni* may also encounter lower temperatures (4 to 10 °C) outside hosts such as in natural water or refrigerated milk/food. *C. jejuni* showed growth at both 37 and 42 °C, while also able to persist at 4 to 10 °C for months (Haddad et al., 2009).

Apart from the temperature difference, *C. jejuni* is exposed to reactive oxygen species (ROS) when external to host bodies or when subjected to ROS produced by host cells during host colonisation. The exposure to ROS presents a challenge for *C. jejuni*, which is a microaerophile prefering growth in low oxygen concentration (Handley et al., 2015). Oxidative stress is associated with iron acquisition, as iron regulates the level of ROS by the Fenton and Haber-Weiss reaction. Iron acquisition inside the human intestine is not an easy process due to the competition from the gut microbiota (Raines et al., 2016). Additionally, host iron usually forms complexes with haemoglobin or myoglobin and becomes insoluble, thus increasing the difficulty of bacterial iron uptake (Stintzi et al., 2008).

However, *C. jejuni* lacks genetic homology with specific and global regulators such as SoxRS, OxyR or RpoS, while only carrying a single copy of superoxide dismutase (SodB). Nevertheless, *C. jejuni* is resistant to oxidative stress due to the presence of ROS regulators such as KatA catalase, alkyl hydroperoxide reductase (AphC) and superoxide dismutase (SodB) (Flint et al., 2012, 2014). These stress regulators are under the regulation of PerR and Fur. PerR and Fur regulate the expression of KatA and the alkyl hydroperoxide reductase (AhpC) in response to the changes of iron and ROS concentrations. (Christiansen et al., 2006; Sittka et al., 2008).

1.1.4 Flagella, secretion systems and antibiotic resistance

C. jejuni only has a few identified secretion systems. One characterised pathway of virulence protein secretion is the type VI secretion system (T6SS). The T6SS facilitates *C. jejuni* invasion, colonisation and survival during infection, including tolerance to oxidative stress and bile salts (Lertpiriyapong et al., 2012; Liaw et al.,

2019). Flagella also play an active role in *C. jejuni* virulence. A flagellar export apparatus analogous to the type III secretion system is essential for Campylobacter invasion antigens (Cia) excretion (Konkel et al., 2004). In particular, internalised CiaD activates the host Erk 1/2 signalling pathway to promote host actin reorganisation and bacterial internalisation (Negretti et al., 2021).

The regulation of *C. jejuni* flagellar biosynthesis is distinctive from model organisms. Unlike *E. coli* and *Salmonella* where flagellum-associated genes are clustered within the same operon, *C. jejuni* flagellar genes showed different hierarchies as they are scattered throughout the genome in various operons (Wösten et al., 2004). In addition, transposon mutagenesis showed that flagellar protein biogenesis is mainly dependent on sigma-54 and sigma-28. The activation of both sigma factors depends on the two-component system FlgR/FlgS, instead of the typical master regulator like the *E. coli flhDC* (Hendrixson et al., 2001; Wösten et al., 2004, 2010a).

Other putative regulators of *C. jejuni* flagellar biosynthesis include FliW and CsrA. Deleting FliW, a binding target to CsrA, resulted in the synthesis of shorter flagellar, which indicates repressed FlaA synthesis (Radomska et al., 2016; Li et al., 2018). The results suggested a negative feedback mechanism where free CsrA binds to *flaA* mRNA and represses *flaA* translation when a high concentration of FlaA sequesters FliW. When FlaA becomes less abundant, the released FliW sequesters CsrA to remove the translational repression on FlaA (Radomska et al., 2016; Li et al., 2018).

Moreover, the selection pressure from antibiotic administration in poultry farming leads to increasing resistance against fluoroquinolone and tetracyclin. Such a trend has prompted the World Health Organisation (WHO) to highlight *Campylobacter* on the High Priority list for researching and developing novel antibiotics (WHO, 2017). Genotyping of 283 stool specimens from hospitalised children in Iran suggested the emerging antibiotic-resistance correlates with increased expression of the CmeABC Efflux pump, the presence of the *tetO* gene and a point mutation of *gyrA* (Sharifi et al., 2021).

1.1.5 Annotated ncRNAs

C. jejuni's ability to persist under environmental stress and demonstrate diverse virulence potential suggests a complicated genetic regulatory network. However, as mentioned earlier, the *C. jejuni* genome is relatively small (1.5 Mb) and encodes only three annotated sigma factors while also lacking standard stress response genes such as *rpoS* (Parkhill et al., 2000). Such discrepancy suggests the presence of additional

genetic and transcription regulatory elements.

Eukaryotic microRNA (miRNA) usually targets 3'-untranslated regions (3'-UTRs) of mRNA targets and silences translation by promoting mRNA degradation (Lee et al., 1993). Some reports also suggest miRNA binds to coding regions and 5'-untranslated regions (5'-UTRs) (Forman et al., 2008; Zhang et al., 2018). Another eukaryotic ncRNA family, namely long ncRNA (lncRNA), mediates gene expression through recruitment of epigenetic effectors (Yap et al., 2010). In contrast to the well-characterised eukaryotic ncRNAs, most insights into bacterial ncRNAs have only emerged in the last two decades. For example, the riboswitch is a cis-acting mRNA that undergoes secondary structural rearrangements upon ligand binding to adjust translational efficiency and transcript stability. Translational control by ligand binding enables feedback regulation of metabolic pathways according to the availability of metabolites (Patterson-Fortin et al., 2013). Another example is the small non-coding CRISPR RNAs (crRNAs), known for their involvement in the genome editing CRISPR-Cas system. crRNAs mediate bacterial immunity against viral infection through direct hybridisation and subsequent cleavage of foreign viral nucleic acids (Barrangou et al., 2007) as well as regulating antibiotic resistance (Sampson et al., 2014).

Rfam, a database containing collections of ncRNA families (Kalvari et al., 2018), has identified *C. jejuni* non-coding RNAs including the thiamin pyrophosphate (TPP) riboswitch, the bacterial signal recognition particle (SRP), tmRNA and the RNA component of RNase P (Parkhill et al., 2000; Gundogdu et al., 2007). The TPP riboswitch is an RNA aptamer that is the most widely distributed riboswitch among bacterial species (Pavlova et al., 2019). The TPP riboswitch binds to thiamine pyrophosphate and attenuates ThiC production, creating a negative feedback regulation for thiamine pyrophosphate biosynthesis (Rodionov et al., 2002). Also, SRP RNA is the RNA component of ribonucleoprotein that mediates co-translational membrane-targeting by recognising signal peptide sequence emerging from the ribosome (Regalia et al., 2002; Rosenblad et al., 2003). Furthermore, transfer-messenger RNA (tmRNA) forms complexes with small protein B (SmpB) to rescue ribosomes stalled by truncated mRNA. The tmRNA attaches to the 3' end of the nascent peptides, enabling translation elongation to continue. The final translated peptides undergo degradation to free up the stalled ribosomes (Weis et al., 2010). Moreover, the RNA component of RNase P is responsible for catalysing tRNA maturation (Kazantsev and Pace, 2006).

A recent study has identified CzcD at the 5'-UTR of the *czcD* gene, which encodes

an ABC transporter for Zn(II) export. CzcD is an RNA thermometer (RNAT) that senses temperature variation and undergoes secondary structure changes to regulate translation. Under elevated host temperature, secondary structure rearrangement of CzvD improves ribosome accessibility for *czcD* mRNA synthesis to confer Zn(II) resistance (Barnawi et al., 2020).

C. jejuni also carries four copies of CRISPR RNAs (crRNA 1-4) and a copy of trans-activating CRISPR RNA (tracrRNA). Similar to species such as *Streptococcus pyogenes*, *C jejuni* also contains a Cas9-encoding operon (CjCas9 operon) which has shown potential for *in vivo* genome editing for treating age-related muscular degeneration (Kim et al., 2017). However, the CjCas9 operon only encodes small or degenerate CRISPR arrays, thus having a relatively minor role in viral defence. Unlike the commonly used *Streptococcus pyogenes* Cas9 (SpCas9), CjCas9 for genome editing is less common as it leads to a higher rate of host DNA damage and more apoptotic induction (Saha et al., 2020b). Transcriptomic analysis showed that CjCas9 activates DNA damage pathways, including p53, ATM (Ataxia Telangiectasia Mutated Protein) and pro-inflammatory (Saha et al., 2020a). CjCas9 may also mediate endogenous mRNA stability. Co-immunoprecipitation followed by RNA-Seq identified about 100 *C. jejuni* endogenous mRNAs bound to CjCas9, potentially through base-pairing with crRNAs and tracrRNA. The mRNA binding precedes mRNA cleavage mediated by the CjCas9 HNH domain (Dugar et al., 2018). In a recent preprint paper, rationally engineered CjCas9 possessed increased cleavage activity *in vitro* and *in vivo*. The engineered CjCas9 may serve as an alternative genome engineering tool, with the advantage of being smaller than spCas9 that allows it to fit easily into an adeno-associated virus vector (Nakagawa et al., 2021).

1.2 sRNA biology of *C. jejuni*

1.2.1 Post-transcriptional regulation by sRNA

In addition to previously mentioned ncRNAs, computational and biochemical approaches have increasingly identified small non-coding RNAs (sRNAs) in recent years (Mraheil et al., 2011). As the name suggests, sRNAs are short RNA molecules between 50 to 500 nucleotides long and possess no protein-coding activities. Experimental evidence suggests that these small transcripts can impact the global transcriptomic landscape. For example, deletion of ArcZ sRNA affects about 15 % of total mRNA expression in *Photorhabdus* and *Xenorhabdus* (Neubacher et al.,

2020). Other earlier studies have shown that sRNAs are responsible for the post-transcriptional regulation of at least half of *E. coli* and *Salmonella* total mRNA expression (Melamed et al., 2016; Waters et al., 2017; Chao et al., 2012; Hör et al., 2018).

External stimuli such as nutrient availability and cell density mediate differential expression of sRNAs (Gurung et al., 2015). sRNAs fine-tune gene expression of biological pathways ranging from TCA cycle, amino acid uptake, heat shock, oxidative traits, small toxic protein, DNA replication initiator to the heme uptake system (Miyakoshi et al., 2015, 2018; Fritsch et al., 2018; Sharma et al., 2011; Andresen et al., 2020; Li et al., 2020a; Wilson et al., 2021). Since *C. jejuni* lacks canonical stress response regulators, sRNA becomes an alternative regulatory factor. Studies in other bacterial species have demonstrated the potential of sRNAs to mediate the global stress response. For example, *Legionella pneumophila* (Lp) sRNA Lpr10 acts as a negative feedback regulator of the sigma factor RpoS, which is essential for survival in water. A high level of RpoS increases Lpr10 expression, which then represses the translation of RpoS mRNA, thereby optimising the level of RpoS protein (Saoud et al., 2020).

Some sRNAs directly regulate bacterial virulence genes. In *Pseudomonas aeruginosa*, sRNA 179 and Hfq inhibit virulence factors such as the type III secretion system (T3SS) and the cAMP-Vfr regulons (Janssen et al., 2020). Moreover, deletion of MicC sRNA relieves repression of *ompA* and *ompC*, which encode an outer membrane protein in *Salmonella* Enteritidis (AU - Meng et al., 2021). *H. pylori*, a close relative of *C. jejuni*, also regulates virulence with sRNAs. *H. pylori* NikS sRNA expression is under the control of nickel availability. When nickel availability is low, NikS represses multiple virulence gene mRNAs and regulates host cell internalisation and epithelial barrier disruption (Eisenbart et al., 2020). That agrees with a later publication, which showed that NikS is regulated by T-repeats upstream of the NikS coding region. T-repeat elongation during *in vivo* infection leads to decreased NikS expression, which relieves the repression on outer membrane proteins (OMPs) and the oncogene CagA that promotes host colonisation and may contribute to oncogenesis. The possible link of NikS to oncogenesis matches the lower NikS expression of clinical isolates from gastric cancer patients (Kinoshita-daitoku et al., 2021).

Notably, it is becoming increasingly clear that sRNAs also modulate bacterial-host interactions, including survival within host cells (Gao et al., 2019), intestinal colonisation (Melson and Kendall, 2019) and biofilm production (Sonnleitner et al., 2011; Sonnleitner and Haas, 2011). In particular, modulation of the host immune response

by sRNA facilitates symbiosis. A recent study showed that *Vibrio fischeri* (*V. fischeri*) SsrA sRNA is one of the most highly expressed sRNA in the outer membrane vesicle (OMV) during the infection of squid light organs, suggesting OMV delivers SsrA sRNA to the host. An *ssrA* deletion mutant colonised the light organ initially before declining in abundance (Moriano-Gutierrez et al., 2020). Also, deletion of *ssrA* leads to lower host retinoic-acid inducible gene-I (RIG-I) expression. The reduced RIG-I expression activates the host immune response, which may explain the decreased symbiotic colonisation (Moriano-Gutierrez et al., 2020). In addition to promoting symbiosis, there are also examples of sRNA-mediated virulence through modulating host gene expression. For example, dual RNA-Seq of infection assays of *Salmonella* and HeLa cells showed *Salmonella* PinT sRNA regulates the expression of *Salmonella* virulence effectors that perturb the host immune response (Westermann et al., 2016).

Apart from targeting bacterial virulence mRNAs, bacterial sRNAs also regulate the expression of host virulence mRNAs. For example, *P. aeruginosa* delivers sRNAs into host epithelial cells to attenuate host immune response (de Bruijn and Verhoeven, 2018; Koeppen et al., 2016). Such a mechanism also occurs in *H. pylori*. Two sRNAs from *H. pylori*, sR-2509025 and sR-989262, are enriched in OMVs and AGS cells exposed to OMV. Deletion of these two sRNAs leads to increased IL-8 secretion. Conversely, transfection of AGS cells with sR-2509025 and sR-989262 resulted in reduced IL-8 secretion by AGS cells. These observations suggested that sR-2509025 and sR-989262 are delivered into host cells to modulate the host immune response (Zhang et al., 2020). However, it remains unclear whether *C. jejuni* also delivers sRNAs to the host via OMVs.

In addition to infectious pathways, some sRNAs are responsible for emerging antibiotic resistance (Sahni et al., 2019) or transient persistence against antibiotics. For instance, transient upregulation of RyhB sRNAs during iron starvation conditions lead to reduced biogenesis of respiratory complex responsible for gentamicin uptake (Chareyre et al., 2019). Similarly, *V. cholerae* VadR sRNA post-transcriptionally inhibits *crvA* mRNA and improves resistance against penicillin G through cell shape maintenance (Peschek et al., 2020).

The impact of sRNAs on virulence and antibiotic resistance has led to clinical and industrial applications. For instance, sRNAs can become the therapeutic targets of RNA-based therapies against multidrug-resistant pathogens, with several studies conducted on sRNAs from *E. coli* and *Salmonella* strains (reviewed in Parmeciano et al., 2019). Secreted extracellular sRNAs that mediate host-pathogen communi-

cations can serve as biomarkers for infectious diseases circulating in human tissue. For industrial applications, metabolic engineering by sRNAs is more convenient and less time-consuming than the conventional synthetic biology approach as it does not involve building or rearranging large genetic elements such as promoters or gene orders. Regulating multiple mRNAs by sRNAs is more effective than random mutagenesis and recombination due to their low mutation rates. Metabolic engineering by sRNAs is also easier to implement than CRISPR-Cas9 as sRNA engineering only requires one genetic element, thus imposing a lower metabolic burden (Na et al., 2013).

Moreover, metabolic engineering by sRNA does not require the fusion of scaffold sequences as in synthetic riboswitches. Several studies have illustrated the potential of fine-tuning gene expression with sRNAs. A protocol published in 2013 demonstrated a constructed synthetic sRNA library that fine-tunes metabolic circuits in *E. coli* (Yoo and Na, 2013). This protocol has facilitated tyrosine and cadaverine production by *E. coli* (Na et al., 2013), using a small sRNA library with different promoters to adjust expression levels. The sRNA library targeted six enzymes in the beta-carotene production pathways, with different levels of promoter strength for each sRNA. This sRNA-tuned system showed a higher beta-carotene production rate than the T7-promoter tuned system. However, all designed sRNAs are repressive. Thus, future work may include activating sRNAs into the pool (Ghodasara and Voigt, 2017; Noh et al., 2017).

Plasmid expressed sRNAs, which carry complementary sequences that recognise mRNA targets and a scaffold sequence that recruits Hfq, are also applicable to other organisms. One example is improving glutamate production in *Corynebacterium glutamicum* through gene knockout, which is otherwise tricky due to the genome stability of *Corynebacterium glutamicum* (Sun et al., 2019). Also, a recent study modulated *Pseudomonas putida* gene expression for industrial purposes using plasmid expression of sRNAs that regulate expression levels of endogenous and heterologous mRNAs via antisense interaction. The result showed both upregulation and downregulation of target genes (depending on where it targets), which allows the re-designing of *P. putida* metabolism (Apura et al., 2020).

Interestingly, sRNAs have demonstrated a potential application for novel antibiotic development. Most antibiotics are broad-spectrum compounds that unspecifically eliminate multiple microbial species within the gut microbiota. Short antisense oligonucleotides (ASO) provide an alternative to broad-spectrum antibiotics by mimicking sRNAs to target specific mRNAs. ASO couples with a protein that

carries the ASO into the cell and targets the mRNA of essential genes. The sRNA-regulatory mechanisms can serve as a reference for ASO design (reviewed by Vogel, 2020). However, replacing broad-spectrum antibiotics with ASO faces technical hurdles, such as optimising the ASO dosage and improving ASO stability.

1.2.2 sRNA mode of action

The genes encoding sRNAs are located in intergenic regions, 3'- and 5'-UTRs, antisense of a protein-coding gene or even inside an open reading frame (Tsai et al., 2015; Dar and Sorek, 2018). For example, in contrast to riboswitches and thermosensors, some 5'-UTR derived sRNAs are transcripted as separate transcripts (Luo et al., 2021). Moreover, a *Pseudomonas aeruginosa* 5'-UTR transcript RhlS is an sRNA derived from the 5'-UTR of the *rhll* locus, which encodes an enzyme responsible for synthesising AHL N-butanoyl-homoserine lactone (C4-HSL) for quorum sensing. Disruption of the RhlS terminator reduced C4-HSL, and RhlS mutants exhibited lower levels of Rhll (Thomason et al., 2019). The result suggests RhlS acts as a positive regulator for Rhll, possibly by sequestering its antisense counterpart.

Similarly, 3'-UTRs of some mRNAs can be excised into separate short transcripts by RNase digestion (Chao et al., 2017). For example, *Salmonella* Typhimurium sRNAs DapZ sRNA is derived from the 3'-UTR of *dapB* and acts as a trans-acting repressor of ABC transporters such as DppA and OppA (Chao et al., 2012). Another example is CpxQ sRNA, an sRNA synthesised by RNase E cleavage of cpxP mRNA. CpxQ sRNA is upregulated in response to inner membrane protein misfolding and represses the synthesis of inner membrane proteins, which protect bacteria against inner membrane damage (Chao and Vogel, 2016). 3'-UTR derived sRNAs can also act in a cis-encoded manner. A recent TIER-seq analysis (transiently- inactivating-an-endoribonuclease-followed-by-RNA-seq) featured *Vibrio cholerae* OppZ and CarZ sRNAs from 3'-UTR of the OppABCDF and *carAB* operons. Both sRNAs act as negative feedback regulators of their corresponding operons (Papenfort et al., 2020). Meanwhile, a *cis*-acting sRNA in *Clostridium tetani* near the 3'-UTR regions of *tent* gene (responsible for tetanus toxin (TeNT) production) carries a 14-nucleotide region that is complementary to the 5' end of the *tent* gene. Knockdown of that sRNA resulted in a threefold increase of TeNT production (Brüggemann et al., 2021).

The interactions between an sRNA and an mRNA depend on the sequence complementarity between the interacting base pairs. *Cis*-acting sRNAs are usually antisense to their mRNA targets, allowing stable sRNA-mRNA duplex formation through complete complementation to the neighbouring opposite stranded mRNA

target (Zhang and Liu, 2019). In contrast to the sequence similarity and physical proximity of the *cis*-acting sRNAs, *trans*-acting sRNAs target mRNAs from distant genomic loci, which are only partially complementary and thus lead to less stable base-pairing interactions (Vogel and Luisi, 2015). To compensate for the incomplete complementarity, RNA chaperones such as Hfq or ProQ act as scaffolds for RNA binding and provide extra stabilisation energy (Morita et al., 2005). The flexibility of sRNAs to bind to both fully and partially complementary targets enables fine adjustment of a wide range of molecular pathways under different conditions (Mai et al., 2019).

While miRNAs can only inhibit translation by stimulating mRNA degradation, post-transcriptional regulation by sRNAs both inhibit and activate translation initiation, with or without regulation of mRNA degradation. sRNA can prevent translation initiation by physically occupying the ribosomal binding site (RBS) (Nielsen et al., 2009; Sievers et al., 2015) (Figure 1.2a) or through duplexing with target mRNAs and promoting RNase E degradation (Figure 1.2b). An example of such a mechanism includes RyhB sRNA, which regulates iron homeostasis by promoting degradation of transcripts encoding iron-binding proteins such as *sodB* (Massé et al., 2003; Desnoyers et al., 2009; Ikeda et al., 2011; Bossi et al., 2012). Apart from acting as translation inhibitors, sRNA-mRNA interactions can activate translation initiation by exposing the RBS by disrupting secondary structures (Figure 1.2c) (Balbontín et al., 2016; Quereda et al., 2014; Melson and Kendall, 2019). sRNAs can also act as translation activators by blocking RNase cleavage sites and thus inhibiting RNase digestion (Fröhlich et al., 2013; Richards and Belasco, 2019) (Figure 1.2d).

Furthermore, recent studies have revealed non-canonical mechanisms of sRNA-mediated post-transcriptional regulation. Instead of targeting the RBS, *E.coli* GCV-B sRNA targets the AC rich region upstream of the RBS to inhibit *gltI* mRNA translation (Sharma et al., 2007). Similarly, Spot42 inhibits *sdhC* mRNA translation without causing degradation. However, its binding site is also too far upstream from the RBS. Instead of blocking the RBS, Spot42 acts as a recruiter of Hfq, which directly binds to the Hfq binding site near the RBS and prevents ribosome attachment (Desnoyers and Massé, 2012). A more recent example is SgrS sRNA. Under glucose-phosphate stress conditions, SgrS represses the translation of *manY* mRNAs by adhering to the translation enhancer upstream of the RBS and Hfq to interfere with S1-dependent translation. The example of SgrS sRNA suggests an non-canonical mechanism of translation initiation inhibition that targets the enhancer sites instead of the RBS (Azam and Vanderpool, 2020).

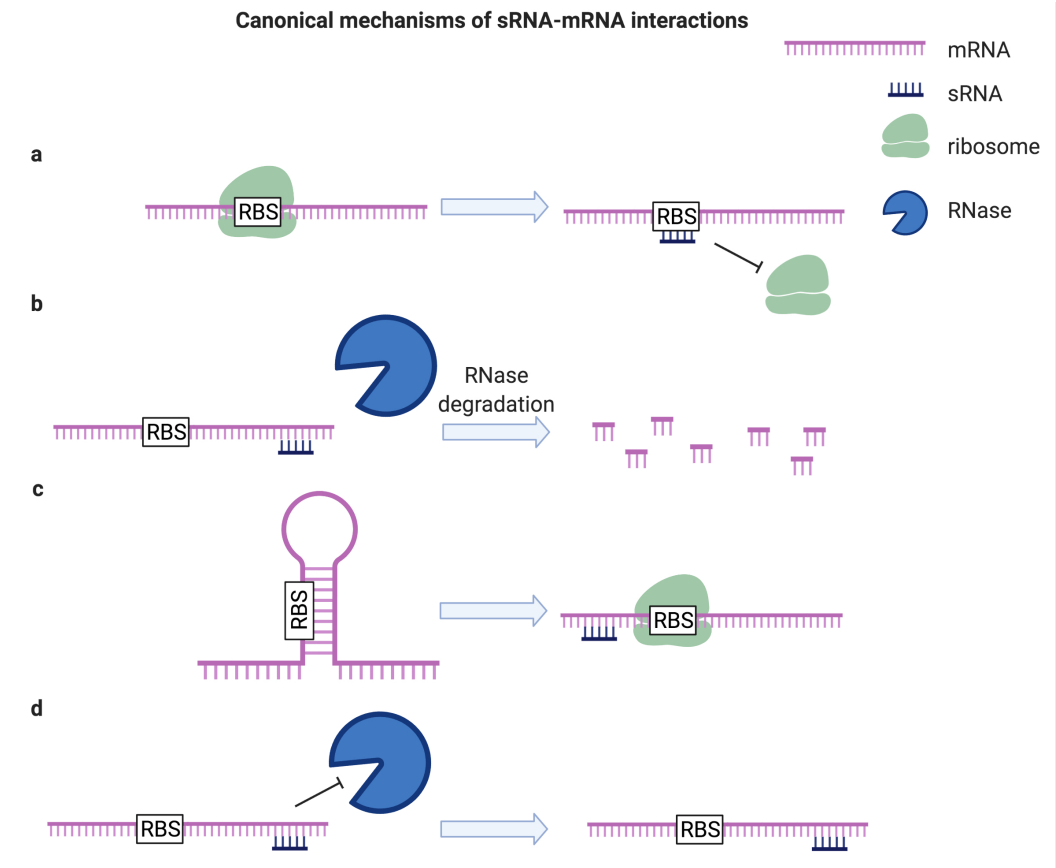

Figure 1.2: **Canonical mechanisms of post-transcriptional regulation by sRNAs.** (a) Translation inhibition by RBS occupation. (b) Translation inhibition by promoting RNase degradation. (c) Translation activation by structural rearrangement. (d) Translation activation by mRNA stabilisation. Created in BioRender.com.

Moreover, some sRNAs target protein-coding regions, such as *Salmonella* MicC sRNA which silences *ompC* mRNA within the +67 to +78 position of the coding regions to promote mRNA degradation (Pfeiffer et al., 2009). Similarly, *E. coli* OmrA/B sRNA binds to the +34 to +44 position of the *fepA* mRNA. This sRNA-mRNA interaction destabilises a stem-loop structure that otherwise activates translation initiation (Jagodnik et al., 2017).

Besides regulating translation initiation and mRNA degradation, sRNA can also regulate transcriptional termination. In *E. coli*, sRNAs DsrA, ArcZ, and RprA are upregulated under stress and act as anti-termination factors by binding the *rpoS* 5'-UTR to suppress premature Rho-dependent transcription (Sedlyarova et al., 2016). sRNAs can also provide a negative feedback mechanism for transcription termination. MS2-affinity purification coupled with RNA sequencing (MAPS), an approach that aims to identify all binding interactions of a specific sRNA, revealed that *Salmonella* SraL sRNAs showed significant transcript enrichment transcriptional termination factor Rho. Surprisingly, SraL acts as a positive regulator for *rho* mRNA levels despite not having any significant impact on RNA stability. Instead, SraL sRNA prevents Rho-dependent termination of the *rho* transcript by blocking the interaction between the nascent *rho* mRNA transcripts and Rho (Silva et al., 2019).

A deeper layer of post-transcriptional regulation of sRNA involves titrating sRNAs by sponge sRNAs. Sponge sRNAs can either sequester or degrade their sRNA partners, or both. The first identified sponge RNA is the *Salmonella chb* transcript. In the absence of chitooligosaccharide, the constitutively expressed ChiX sRNA binds to the *chiP* mRNA 5' end and represses the translation of *chiP* mRNA. The presence of chitooligosaccharides induces the expression of *chb* transcript, which binds to and promotes ChiX sRNA degradation, thus allowing ChiP synthesis for more chitooligosaccharide uptake (Figueroa-Bossi et al., 2009).

Similarly in *E. coli*, while GcvB sRNA represses the Glu/Asp transporter expression by promoting the degradation of *gltl* mRNA, the *gltl* 3'-UTR derived sponge sRNA SroC sequesters and controls the level of GcvB (Miyakoshi et al., 2015). Besides, AgvB mimics dipeptide transporter (*dppA*) to displace GcvB sRNA, which frees up *dppA* transcript for translation *E. coli* (Tree et al., 2014). A more recent study has identified the ArcZ sponge sRNA that inhibits CyaR sRNA expression, thus controlling the level of CyaR during the transition phase and may contribute to stationary phase/biofilm development (Iosub et al., 2020). Premature termination is also a mechanism for sRNA sponges. 3'-end mapping has detected an *E. coli* sRNA

sponge derived from 5'-UTR and internal ORF, suggesting premature termination as a mechanism for deriving sRNA sponge. In particular, ChiX sRNA inhibits *chiP* mRNA translation by blocking ribosome binding and inducing Rho-dependent termination. The activity of ChiX sRNA generates ChiZ that act as sponges to inhibit ChiX activity (Adams et al., 2021).

In addition to titrating other sRNA species, sRNAs can also act as proteins sponges. For example, RsmY and RsmZ act as sponge sRNAs that sequester RNA binding protein RsmA from its targeted mRNAs. The sequestration reduces the production of the type III secretion system in *Pseudomonas aeruginosa* (Janssen et al., 2018). Moreover, carbon catabolite induction of *Pseudomonas aeruginosa* leads to the overexpression of the CrcZ sRNA. The overexpressed CrcZ sRNA associates with Hfq protein and relieves the repression of catabolic genes like *amiE* (Sonnleitner and Bla, 2014).

Another novel aspect of sRNA mechanisms is their spatial regulation. sRNA and Hfq accumulate in the cell pole when under stress, which may increase post-transcriptional regulation of polar-localised mRNAs. Upon overexpression, sRNAs such as SgrS, GlmZ, OxyS, RyhB shows increased cytoplasmic localisation (Fei et al., 2015; Sheng et al., 2017). Similarly, a polar transcriptome analysis of non-stressed *E. coli* demonstrated the polar localisation of over 19 sRNAs, with two-thirds of them showing enriched polar localisation under osmotic stress (Kannaiah et al., 2019).

1.2.3 *C. jejuni* sRNAs

Despite the biological significance of sRNAs, understanding of *C. jejuni* sRNAs is insufficient compared to other pathogens due to technical and biological factors. Firstly, *C. jejuni* is a fastidious organism to culture. Upon exposure to suboptimal growth conditions, it enters the viable but nonculturable (VBNC) state, a physiological state where cells remain alive yet metabolically inactive (Rollins and Colwell, 1986). Efficient growth requires specialised culture media, modified low oxygen atmospheric conditions, a narrow temperature range, and long doubling times (Davis and DiRita, 2008). Another technical difficulty is the lack of effective *in vitro* and *in vivo* models. Compared to other enteric bacterial pathogens, *C. jejuni* demonstrates a lower *in vitro* human intestinal cell infection rate and lacks ideal animal models for *in vivo* pathogenesis analysis (reviewed in Burnham and Hendrixson, 2018). Moreover, *C. jejuni* strains are genetically heterogeneous and non-clonal due to their fast genome rearrangement and natural transformation. Such a genetic diversity contributes to heterogeneous experimental data (Boer et al., 2002). In addition,

limited genetic tools such as reporter systems for genetic manipulation are available for *C. jejuni* (Davis et al., 2008; Buchanan et al., 2017). These factors have held back experimental investigation of *C. jejuni* sRNA.

Global RNA-binding proteins such as Hfq and ProQ act as baits for high-throughput sRNA discovery by co-immunoprecipitation in *E. coli* and *Salmonella* (Sittka et al., 2008; Vogel et al., 2003; Zhang et al., 2003). However, this approach is currently impossible for *C. jejuni* due to the lack of identified global RNA-binding proteins. Due to the lack of protein baits for co-immunoprecipitation, current information on *C. jejuni* sRNAs is mainly obtained from RNA-Seq datasets (Dugar et al., 2013; Porcelli et al., 2013; Butcher and Stintzi, 2013; Taveirne et al., 2013). However, northern blot has only confirmed relatively few sRNAs, with limited insights into their biological implications. In addition, those RNA-Seq data convey limited information on condition-specific sRNA expression as these studies cultured *C. jejuni* under a narrow range of conditions.

1.3 Published and in-house transcriptomic data

1.3.1 Published transcriptomic data

The primary transcriptome of four *C. jejuni* strains (NCTC11168, 81-176, 81116, RM1221) in the standard laboratory condition (37 °C in the exponential phase) was first identified by differential RNA-Seq, with transcription start sites (TSS) annotated across all four genomes (Dugar et al., 2013). 3377 TSSs were detected, with 1035 and 1067 conserved among all four strains and 2 or 3 strains, respectively. The list of TSSs consisted of 1905 from the standard laboratory strain NCTC11168. Subsequent northern blot analysis has validated about 20 intergenic ncRNAs, including some previously annotated ncRNAs such as SRP RNA, RnpB, tmRNA, crRNA2, crRNA4 and TracrRNA. However, crRNA1 and crRNA3 of the minimal CRISPR loci were only found in dRNA-Seq without northern blot validation (Dugar et al., 2013). It is also noteworthy that BLASTn (McGinnis and Madden, 2004) analysis suggested sequence conservation among ncRNAs is primarily limited in *C. jejuni*, but not among other *Campylobacter* species. The lack of sequence conservation increases the difficulty of predicting sRNA homologs.

Another dRNA-Seq has also characterised the *C. jejuni* transcriptomic landscape (Porcelli et al., 2013) where a periodic signal for the -35 sequence of the sigma-70 promoter was identified and many mRNAs were designated as leaderless as they lack recognisable RBS. Moreover, a comparison against the *H. pylori* transcriptome

showed no conservation between these two closely related species. The species-specificity is likely due to *C. jejuni* genome reorganisations and gene reshuffling. Nonetheless, the study has identified 992 TSSs, and northern blot analysis has confirmed the expression of eight sRNAs (Porcelli et al., 2013). All identified sRNAs except CjNC9 are also found in Dugar et al., 2013.

Both dRNA-seq studies obtained expression data under standard laboratory conditions. Another RNA-Seq study under iron-stress also discovered novel *C. jejuni* sRNAs transcribed in intergenic regions, suggesting that iron availability also modulates *C. jejuni* sRNA expression. (Butcher and Stintzi, 2013). Another RNA-Seq paper obtained transcriptomic data from chicken colonisation and *in vitro* culturing of strain 81-176. The study has discovered several novel sRNAs, possibly due to differences in experimental conditions (Taveirne et al., 2013). Similarly, RNA-Seq on C. jejuni (NCTC11168) and its *rpoN* mutant identified five intergenic sRNA candidates using evolutionary conservation and structural prediction of large intergenic regions (Chaudhuri et al., 2011). However, all sRNAs mentioned above lack northern blot confirmation.

Some RNA-Seq datasets illustrate other *C. jejuni* stress adaptation mechanisms. One study investigated the regulatory and phenotypic roles of PerR. PerR inactivation facilitated survival under hydrogen peroxides and aerobic conditions (Handley et al., 2015) and increased the expression of peroxidase proteins, including KatA, TrxB, Rrc and AhpC. However, contrary to previous microarray analysis (Palyada et al., 2009), PerR inactivation did not affect motility and ability to kill wax moth larvae. It is noteworthy that this study has identified eight novels TSSs, which might represent transcription uniquely suppressed by PerR. In addition, some transcriptomic datasets obtained microarray analysis (Xu et al., 2015; Hao et al., 2013). For example, one microarray analysis under osmotic stress showed a crosstalk between oxidative and osmotic stress response (Cameron et al., 2012).

1.3.2 In-house RNA-Seq data

Our laboratory previously generated a dataset that represented both the transcriptional profiles (RNAtag-seq) and the primary transcriptome (Cappable-seq) of *C. jejuni* in 21 growth and stress conditions (Lam, 2019). Those conditions were relevant to colonisation in the chicken host as a commensal, human host infection, or transmission between the hosts.

The in-house RNAtag-seq dataset covers the experimental conditions that model the following environmental factors:

Growth phase

Cells undergo rapid division in the exponential phase when nutrients are abundant. The cells transition into the stationary phase as nutrients deplete. In the standard laboratory setting, prolonged growth until nutrient shortage models the early and late stationary phases.

Entry into the stationary phase involves physiological changes such as halted cell replication. Transition into the stationary phase allows the cell to persist under adverse environments, such as limited amino acid availability and accumulated toxic metabolites (Jaishankar and Srivastava, 2017). Different growth phases also lead to the rewiring of *C. jejuni* metabolic pathways. For instance, aspartate and serine are metabolised throughout the exponential phase and depleted upon the entry of the stationary phase, while proline metabolism occurs throughout the stationary phase. Meanwhile, glutamate consumption takes place at both the exponential and stationary phases. Glutamate utilisation is more rapid during the stationary phase than exponential growth (Wright et al., 2009).

Also, upon stationary phase entry, *C. jejuni* showed increased cell membrane integrity under physical pressure, which coincides with the change in fatty lipid composition. That includes the increase in cyclopropane composition, which offers protection against acid stress, freezing damage and ethanol exposure in *E. coli* (Martı and Mackey, 2005). However, increased resistance against acid or heat was absent *C. jejuni*.

C. jejuni lacks the RpoS sigma factor responsible for stationary phase entry in other Enterobacteriaceae species (Kelly et al., 2001). Instead, several other regulators appear to promote the growth phase transition. That includes CosR, with a 4-fold increase in protein expression but reduced transcript abundance in the stationary phase compared to the exponential phase. That suggests negative autoregulation through promoter regulation. CosR is also a DNA binding protein that regulates protein expressions related to ROS detoxification, antibiotic efflux pump and biofilm formation (Turonova et al., 2017). Other mechanisms involved in the stationary phase entry include chemotaxis proteins that promote migration towards nutrient-rich environments (Turonova et al., 2017). In addition, SpoT is another stringent response regulator that may regulate stationary phase survival (Gaynor et al., 2005). Moreover, the late stationary phase also coincided with reduced transcript levels for six of seven genes directly influenced by the two-component system DccRS, suggesting DccRS may also play a role in stationary phase survival (Wösten et al.,

2010b).

Growth temperature

37 °C to 42 °C are the body temperatures of human and avian species respectively (Haddad et al., 2009). Notably, *C. jejuni* demonstrated distinctive transcriptomic landscapes under these two temperatures. 20 % of annotated genes experience differential expression due to temperature changes, most notably for energy metabolism and macromolecules biosynthesis (Stintzi, 2003).

As mentioned earlier, *C. jejuni* is more virulent in humans than in avians. At human body temperature, *C. jejuni* showed elevated chemotaxis and quorum-sensing compared to avian body temperature (Khanna et al., 2006). Meanwhile, 59 out of 181 proteins in the outer membrane vesicle (OMVs) proteome showed differential expression. Among all differentially regulated proteins inside OMVs, virulence-associated proteins showed more enrichment at 37 °C (Taheri et al., 2019).

Iron limitation and oxidative stress

Iron is necessary for bacterial growth as it acts as a cofactor for housekeeping metabolic enzymes. However, host free iron availability is limited as free iron tends to form an insoluble complex with host proteins. Iron-binding host proteins, including haemoglobin and transferrin, leave limited free iron available to the pathogens. Many bacterial species capture host iron by secreting siderophore or recognising heme/haemoglobin by cell surface receptors. The high iron-binding affinity of those iron-chelators allows siderophores to scavenge insoluble iron from host protein and transport it to outer membrane receptors that translocate the iron ions into the pathogen cytoplasm. The cells import captured iron via ABC transporters or the siderophore receptors. Gram-negative pathogens also express transferrin (TF) or lactoferrin (LF) binding proteins, which scavenge iron from host TF and LF. Heme oxygenase degrades imported heme to release free iron (Cassat and Skaar, 2013). However, *C. jejuni* do not produce siderophores. Instead, they compete for siderophores from other species through multiple uptake systems for different siderophores. For example, *C. jejuni* CfrA and CfrB complex with *E .coli* enterobactin-Fe^{3+} to translocate the enterobactin-Fe^{3+} complex into the periplasm. The periplasmic ABC transporter CeuE further translocates the iron complex into the cytoplasm by binding to the hydrolysis product of enterobactin (Raines et al., 2016)).

The insolubility of Fe^{3+} increases the difficulty of iron acquisition. The redox states

of iron also promote the formation of reactive oxygen species (ROS) such as hydrogen peroxide (H_2O_2) and superoxide radicals which damage proteins and DNA (Ezraty et al., 2017). *C. jejuni* also experiences oxidative stress as the host immune system counters infection by ROS production. ROS damages DNA, proteins and metal cofactors. ROS damages protein structure by oxidising Cys and Met, causing non-native disulphide bonds and conformational change (Ezraty et al., 2017). Hence excessive intracellular iron tends to associate with oxidative stress.

A close relationship between iron availability and ROS generation leads to crosstalk between iron and oxidative stress responses. Superoxide dismutase (SOD) inhibits ROS accumulation by converting hydroxyl radical and hydrogen peroxide into oxygen molecules. Other oxidative stress defence proteins such as catalase (Kat) and Alkyl hydroperoxide reductase C (AhpC) also prevent ROS accumulation (Gare´naux et al., 2008; Handley et al., 2015). In addition, the acquisition of Fe^{2+} also reduces ROS. The abundance of Fe^{2+} acquisition and ROS inhibitors are under the control of regulators responding to both iron availability and ROS accumulation. Unlike well-studied aerobes, *C. jejuni* lacks specific global regulators such as SoxRS, OxyR or RpoS and only a single copy of superoxide dismutase (SodB). Alternative genes include the iron homeostasis pathway and PerR and Fur that regulate the expression of KatA and the alkyl hydroperoxide reductase (AhpC) (Gare´naux et al., 2008; Handley et al., 2015). However, global differential gene expression remains unexplored. A proteomic study under oxidative stress led to overexpression of proteins related to iron homeostasis (Cft), virulence (CadF, FlaA) and general stress response, as well as periplasmic protein Cj1371 (Gare´naux et al., 2008; Handley et al., 2015).

Prolonged aerobic exposure increases the promoter activities of antioxidant genes *ahpC*, *katA*, and *sodB*. PerR, which is downregulated upon exposure to oxidative stress, promotes the expression of *ahpC*, *katA*, and *sodB* (Kim et al., 2015). PerR activity is also regulated by Fe^{2+}, which induces conformational change by binding to three histidine residues and two aspartates which are highly conserved across bacteria (Kim et al., 2011). Antioxidant genes are also regulated by Fur-Fe^{2+}, which inhibits the expression of iron uptake membrane proteins and regulates the expression of KatA and PerR. In addition, co-immunoprecipitation of Fur leads to enrichment of genes including iron transporters such as *ceuB*, *chuA* and *cfbpA*, and transporters for other substrates such as *kpsT*, *zupT* and *modB* (Butcher et al., 2012). That may suggest crosstalk between iron uptake and the acquisition of other ions. Meanwhile, CosR, which is repressed by superoxide radicals except H_2O_2, reg-

ulates the expression of both oxidative stress response genes (*sodB*, *ahpC*, *katA*) and iron uptake regulators such as Cj1658 and Cj1659, *dps* (Hwang et al., 2011a) and PerR (Park et al., 2021). CosR-knockdown also leads to a 1.7-fold downregulation of iron *katA* and iron transporters such as Cj1658 and Cj1659 transcript (Hwang et al., 2012). Another regulator, known as DNA binding protein from starved cells (Dps), is similarly co-regulated by iron availability and ROS. Dps acts as DNA binding protein under the presence of Fe^{2+} or H_2O_2, while also exhibiting protection for DNA against oxidative damage and enzymatic degradation (Huergo et al., 2013). Moreover, T6SS effector *tssD* promotes oxidative stress survival, putatively by upregulating expression of *katA*, *sodB* and *ahpC*, and facilitates survival and invasion in chicken (Liaw et al., 2019). Other antioxidant proteins include thioredoxin (Trx) or glutaredoxin (Grx). Both proteins reduce non-native disulphide bonds created by ROS (Zheng et al., 1998; Arts et al., 2016).

Hyper-aerotolerant *C. jejuni* isolates also exhibit enhanced biofilm formation and cross-protection against heat, osmotic and cold stress (Mouftah et al., 2021). Increased iron concentration leads to more biofilm and ROS formation, but not antioxidants. That suggests iron stimulates biofilm formation by inducing oxidative stress (Oh et al., 2018). Most iron acquisition genes also show increased expression during biofilm growth (Tram et al., 2020).

Acid shock

Acid shock models the exposure to gastric pH, where food is retained for up to an hour before reaching the infection site in the small intestine. Relatively little information is available on how *C. jejuni* cope with such prolonged exposure to an acidic environment. Current knowledge suggests several mechanisms are involved.

A study exposed *C. jejuni* to pH 3.5 and showed induced expression of sigma-54 regulated motility genes, including the flagellar basal body and hook (Le et al., 2012; Askoura et al., 2016). Such increased motility may facilitate adhesion and internalisation in *Acanthamoeba polyphaga*, standard transmission vehicles of gut pathogens. *Acanthamoeba polyphaga* provides a protective environment for *C. jejuni* against low pH (Axelsson-Olsson et al., 2010).

The acid stress response is also associated with iron homeostasis and ROS detoxification. Fur deletion mutants are more vulnerable to acid stress (Askoura et al., 2016). Moreover, a microarray analysis on acidic and aerobic conditions showed upregulation of iron transporter genes including *p19*, Cj1660, Cj1662, Cj1663, *cfbpA*, *aphC*, *katA* and *dps*, with P19 showing the largest upregulation Varsaki et al., 2015).

Similarly, proteomic and qRT-PCR analysis with radioactively labelled methionine was conducted on *C. jejuni* strains exposed to HCl. The acid stress causes increased proteomic expression of proteins involved in iron homeostasis and oxidative stress, such as P19, Dps, TrxB, SodB, AphC (Birk et al., 2012).

In addition, *C. jejuni* also copes with low pH by pumping out H^+. Acidic and aerobic conditions lead to upregulation of a putative Na^+/H^+ antiporter (Cj0832c) which pumps out intracellular protons (Varsaki et al., 2015). Acid stress also appears to affect other cellular stress responses, as shown by increased survival rates in other stress conditions, including nutrient starvation and osmotic stress (Kumar-Phillips et al., 2013).

Bile stress

Bile compounds influence the composition of the intestinal microbiome. They can benefit the microbiota by acting as signalling molecules, nutrients and electron acceptors. However, they also destroy biomolecules, including membranes, proteins and DNA, and chelate iron and calcium. They also affect expression of host genes that mediate defence and immunity. Examples of bile compounds include cholic acid and chenodoxycholic acid, primary bile acids synthesised in the liver. After entering the intestine, they transform into secondary bile acids such as deoxycholic acid and lithocholic acid (Urdaneta and Casadesús, 2017). Sodium deoxycholate also affects other biological processes, such as chemotaxis, flagellin production and multidrug efflux system production (Urdaneta and Casadesús, 2017).

Moreover, bile stress increases ROS, possibly because membrane disruption inhibits electron transport chain activity (Negretti et al., 2017). It is also likely that bile salt enhances *C. jejuni* virulence. Exposure to a low concentration of bile affects the contents of outer membrane vesicles (OMVs), which increases adhesion to intestinal epithelial cells. These OMVs further enhance intestinal epithelial cells adhesion and invasion of other *C. jejuni* cells that have not experienced bile stress (Taheri et al., 2018).

Pathogens have evolved defence mechanisms to survive under high bile concentration, including DNA repair, the downregulation of outer membrane porins and the upregulation of efflux pump. Expression of the genes encoding the CmeABC multidrug efflux pump is repressed by a TetR family repressor CmeR. CmeR repression is inhibited by bile salts, thus leading to the upregulation of CmeABC. In addition to pumping out bile salts, CmeABC can also confer antibiotic resistance and other antimicrobial substances, leading to enhanced survival and chicken colonisa-

tion (Lin et al., 2003, 2005). In addition, CmeDEF acts as a secondary efflux pump that maintains bile resistance without CmeABC (Akiba et al., 2006). Bile exposure also stimulates virulence mechanisms. Sodium deoxycholate induces *C. jejuni* virulence by increasing Invasion antigen B (*ciaB*) gene expression without altering adherence or motility (Malik-Kale et al., 2008). Deleting the two-component system response regulator CbrR also leads to susceptibility to bile salt. However, the mechanism by which CbrR confers bile resistance has yet to be elucidated (Raphael et al., 2005).

Heat stress

Heat stress at 48 °C, and 55 °C, which models the defeathering and scalding steps in poultry processing, decreases culturability and viability (Klančnik et al., 2014). HspR and HrcA regulate the gene expression of heat shock proteins, including chaperonins, through their cooperative DNA binding activities (Palombo et al., 2020). For instance, Expressions of GroES and GroEL, which are chaperonins that assist protein folding, are repressed by heat shock regulator HspR (Holmes et al., 2010). Under 37 °C, heat shock promoters are co-repressed by HrcA and HspR interactions. Such interactions are removed when the temperature rises to 42 °C, where HrcA, but not HspR, showed reduced DNA binding activity. Temperature rise may alter HrcA conformation, which reduces its DNA binding capacity, thus allowing the transcription of heat shock genes (Versace et al., 2021).

Anaerobic stress

C. jejuni experiences anaerobic conditions in the host intestine, especially in the chicken caecum, which carries a high population of obligate anaerobes. Despite its difficulty to grow under atmospheric oxygen levels, insufficient oxygen affects cell electron transport as oxygen is a terminal electron acceptor (Kendall et al., 2014). Among all electron acceptors, oxygen is the most preferred choice and is essential for DNA synthesis (Sellars et al., 2002). The lack of oxygen as an electron acceptor significantly hampered cell growth, ATP generation and motility (van der Stel et al., 2017). The requirement for oxygen prompts *C. jejuni* to migrate to regions with higher oxygen levels, including the mucous layer and intestinal crypt.

Nonetheless, anaerobic respiration in chicken caecum is possible in the presence of alternative electron acceptors. For example, disruption of periplasmic electron transport enzymes nitrate reductase *napA* and nitrite reductase *nrfA* leads to a significant decrease in anaerobic growth (Weingarten et al., 2008). NapA receives

electrons from the electron transport chain, which reduces nitrate to nitrite and subsequently ammonium ion by NrfA (Pittman et al., 2007). Similarly, the addition of nitrate and nitrite alongside other electron acceptors such as fumarate, trimethylamine-N-oxide (TMAO) and dimethyl sulfoxide (DMSO) increases both the growth rate and final cell density (Sellars et al., 2002).

Starvation stress

Starvation stress involves growing cells to early-stationary phase followed by resuspension in Ringer's solution (2.25 g/L NaCl, 0.08 g/L $CaCl_2.2H_2O$, 0.05 g/L $NaHCO_3$) for 5 hours. Such treatment models nutrient deprivation during the *C. jejuni* transmission cycle. When compared against heat shock and oxygen exposure, 15 hours of starvation stress lead to significantly more growth defects, lower viability and reduced adhesion and invasion of Caco-2 cells (Mihaljevic et al., 2007). Starvation increased the number of cells in the VBNC state and induced heat stress resistance (Klancnik et al., 2009). Starvation stress also leads to upregulation of Cj0917, which encodes a homologue of *E. coli* carbon starvation protein A (CstA) (Rasmussen et al., 2013).

Cold stress and chicken exudate

C. jejuni must survive refrigeration during poultry processing before cooking (Bhaduri and Cottrell, 2004). Oxygen solubility increases during low temperatures, thereby increasing oxidative stress. Thus cold exposure is also related to the reduced oxidative metabolism TCA cycle, pentose phosphate pathway and electron transport, or maybe by using secondary metabolites to bypass ROS intermediate (reviewed in Tribelli and López, 2018)

Many bacteria species cope with cold stress by reducing fatty acid saturation to maintain membrane fluidity while reducing stiffness during cold exposure. However, C. jejuni exhibits no fatty acid alteration. When exposed to heat treatment after treatment at 6 °C for 24 hours, *C. jejuni* showed better heat tolerance. That might be due to less fatty acid alteration during cold exposure, preventing membrane leakage during heating afterwards (Hughes et al., 2009). Such cold exposure also leads to reduced electron transport activities (Hughes et al., 2010). In addition, *pebC* was differentially downregulated when subjected to 46 °C heat stress after exposing to 4 °C. That suggests a decrease in glutamate metabolism during cold stress (Duqué et al., 2021).

The "chicken exudate" model mimics the cold stress and the food packaging environ-

ment. The food packaging environment comprises organic materials (e.g. chicken carcass and other foodstuffs) that provide nutrients for bacterial growth. The exudate was extracted and filtered from commercially frozen chickens after defrosting at 4 °C, which is the equivalent temperature of refrigeration. Compared to *C. jejuni* incubated in standard laboratory media brain–heart infusion (BHI), growth in chicken juice is linked to increased biofilm formation and enhanced growth (Brown et al., 2014)

Hyperosmotic stress

Food pathogens experience high salt concentrations during food processing. Food preservatives such as NaCl increase the extracellular osmolarity, leading to water efflux and cell dehydration (Burgess et al., 2016). The biological impact of hyperosmotic stress varies according to NaCl concentration. *C. jejuni* strain 81-176 dies at 2.0 % NaCl, and displayed filamentation at 1.0 and 1.5 % NaCl. Examples of *C. jejuni* osmotic stress regulator includes sigma-54 RpoN, as *rpoN* mutant showed increased growth defect with abnormal elongated cell morphology and lower viability under 0.8% NaCl (Hwang et al., 2011b). In addition, deletion of polyphosphate kinase (*ppk1*), which is responsible for poly-phosphate accumulation, also reduced survival at 0.25 M NaCl (Candon et al., 2007).

Nitrosative stress

C. jejuni encounters nitric oxide upon infection as the host macrophage uses nitric oxide to eliminate pathogens. Nitric oxide reacts with metal centres such as haem centres Fe-S clusters. Such reactivity leads to inhibition of enzymes such as terminal oxidases and transcription regulators. *C. jejuni* counters nitrosative stress (induced by nitrosating agents S-nitrosoglutathione (GSNO)) by expression of Campylobacter globin (Cgb) and NssR (Cj0466) (Elvers et al., 2004, 2005).

1.3.3 Results of Cappable-seq and RNAtag-seq

Previous work in the host laboratory has generated Cappable-seq and RNAtag-seq data for the conditions mentioned above (Lam, 2019). Cappable-seq labels the 5'-triphosphates with biotinylated GTP. Streptavidin then captures labelled 5'-triphosphates for primary transcript enrichment (Ettwiller et al., 2016). Pooling RNA from 21 conditions together allows identification of TSSs expressed in non-standard conditions, which complement the published dataset obtained from the standard laboratory condition (Dugar et al., 2013; Porcelli et al., 2013). The

Cappable-seq data analysis identified 5197 TSSs after filtering with the control sample (the sample that lacks streptavidin enrichment) and clustering TSSs within ten nucleotides away. Among them, 3613 TSSs were previously unknown (Figure 1.3a). Those promoter motifs for sigma-70 and sigma-54 were similar to the motif found in earlier reports (Dugar et al., 2013; Porcelli et al., 2013). However, sigma-28 displayed no clear consensus motifs. Most identified TSS are internal to protein-coding genes. Other than that, antisense TSSs occur the most frequently, making up about a quarter of sigma-70 and sigma-28 promoter motifs (Figure 1.3b).

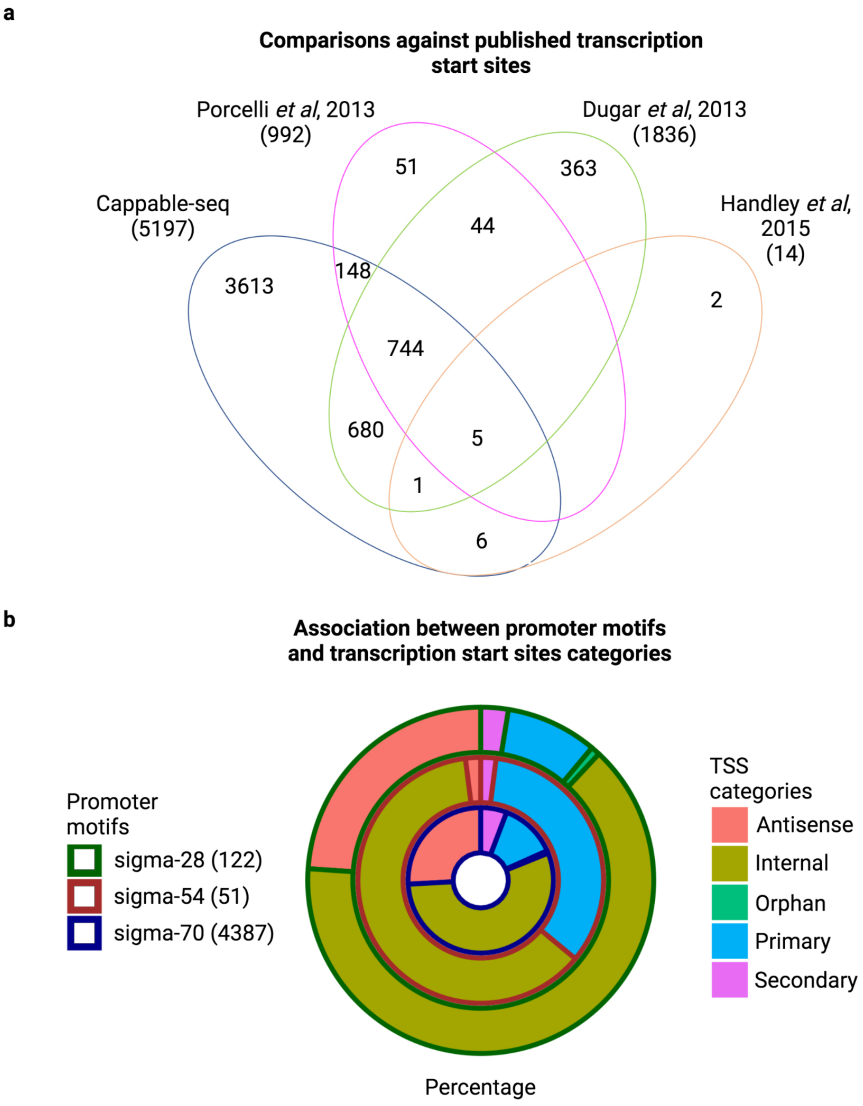

Figure 1.3: **Summary of Cappable-seq analysis.** (a) Number of identified TSSs from published data and in-house Cappable-seq. (b) Distribution of TSS types among promoter motifs. Adapted from previous Cappable-seq analysis from the host laboratory (Lam, 2019).

Analysis of the host research group predicted novel sRNAs from the Cappable-seq output using toRNAdo, which searches for an expression peak 5-fold above the background level (Hermansen et al., 2018). toRNAdo identified 371 putative sRNAs, including 208 absent in prior publications (Dugar et al., 2013; Porcelli et al., 2013; Lam, 2019). The predicted sRNAs were included in subsequent differential expression analysis, using transcriptomic data generated from RNAtag-seq. Instead of pooling all samples together like Cappable-seq, the RNAtag-seq protocol involves barcoding rRNA depleted RNA samples, enabling downstream demultiplexing to distinguish transcriptional signals from individual replicates (Shishkin et al., 2015). Replicates of the same biological conditions showed clustering of expression patterns, with the nutrient deprived conditions demonstrating expression patterns distrinctive from other conditions (Figure 1.4). Pairwise comparisons between 21 conditions have highlighted significantly enriched pathways. For instance, ABC transporters for iron(III) transporters showed significant differential expression across most conditions, suggesting iron import plays a crucial role in stress adaptation.

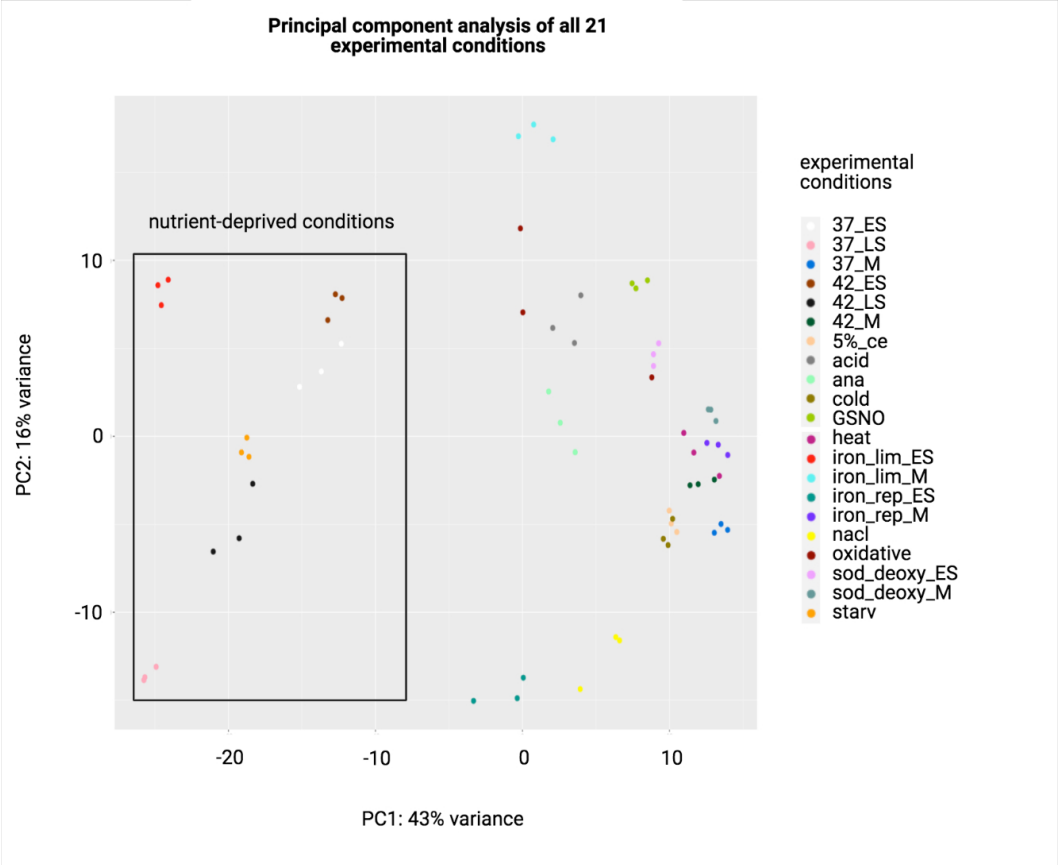

Figure 1.4: **PCA analysis of RNAtag-seq results.** Expression of annotated genes and predicted sRNAs were included. Among them, nutrient-deprived conditions showed distinctive expression patterns from the rest of the RNAtag-seq conditions. Adapted from RNAtag-seq analysis from previous in-house research (Lam, 2019).

1.4 Project outline

With the availability of published and in-house transcriptomic data, it becomes possible to predict novel sRNAs using gene expression coverage. Un-annotated regions with high expression coverage may indicate the presence of an sRNA. Notably, the in-house RNAtag-seq data covers a wide range of non-standard conditions, which may help identifying sRNA expressed and sRNA-mRNA interactions in response to specific biological conditions.

1.4.1 sRNA prediction based on genomic and transcriptomic data

In some of the above examples, sRNA prediction has been made using transcriptomic data. Unannotated regions with high expression coverage are more likely to be predicted as sRNAs. However, RNA-Seq coverage can be affected by spurious transcription, which may lead to false sRNA prediction. Such kind of false prediction is particularly likely from *C. jejuni's* AT-rich genome sequence. Another difficulty is to define the sRNA boundary accurately. toRNAdo determines the start and end of the sRNA using an arbitrarily defined threshold, which is easily affected by spurious transcription and artefacts such as PCR duplication arising during sequencing library preparation. Some publications define the sRNA boundary by manual visualisation. Such an approach is not feasible when handling a large number of datasets. Alternately, several published tools are available to predict sRNA from the whole genome sequence. This study will attempt to identify sRNA using both genomic and transcriptomic data and evaluate the result from different tools and generate a finalised list of novel sRNAs.

1.4.2 Characterisation of predicted sRNAs by *in vivo* crosslinking

The sRNA-mRNA interactome is another poorly characterised aspect of the *C. jejuni* post-transcriptional network. In order bacterial species, intermolecular RNA interactions have been identified using protein baits like Hfq (Iosub et al., 2020; Melamed et al., 2016) and RNase R (Waters et al., 2017). However, due to the lack of a characterised global RNA chaperon in *C. jejuni*, sRNA-mRNA interactions need to be identified without using protein baits. To avoid disrupting the native *in vivo* sRNA-mRNA interactome during experimental procedures, psoralen and the psoralen derivative 4-aminomethyltrioxsalen (AMT) can be utilised to stabilise the intracellular RNA duplex. Psoralen and AMT are crosslinking reagents that promote the formation of a covalent linkage between uridine base pairs of two

neighbouring RNA molecules (Wilms et al., 1997; Calvet and Pederson, 1981). The crosslinking creates RNA duplexes that can be extracted using standard RNA extraction, ligation and finally linearisation to create a linear chimeric RNA molecule. The linearised RNA can serve as a template for RNA-Seq. Hence, this study will also aim to perform and optimise the RNA crosslinking protocol in *C. jejuni*.

1.4.3 sRNA function prediction using in-house transcriptomic data

RNA crosslinking is unlikely to identify all condition-specific sRNA-mRNA interactions. The expensive reagents cost and the time required for optimising individual experimental steps would limit the number of experimental conditions tested throughout this project. Therefore, RNA crosslinking alone cannot identify the sRNA-mRNA interactions in other conditions. Hence, analysis of transcriptomic expression patterns and sequence complementarity can explore sRNA-mRNA networks under a broader range of conditions. The in-house RNAtag-seq results enable an analysis of overall gene expression patterns across 21 conditions and associate the constructed sRNA-mRNA network to particular stress conditions. Moreover, *in silico* target prediction based on sequence complementarity and secondary structure prediction can elucidate the binding mechanism. Moreover, exploring the transcriptomic landscape under such a vast number of conditions allows the identification of pathway crosstalks responsible for adapting to different biological conditions.

1.5 Aims and Objectives

This study will predict novel sRNAs from *C. jejuni* genome sequences and transcriptomic data, followed by functional annotation of predicted sRNA using experimental validation and computational prediction. *In silico* prediction of novel sRNAs will be carried out using selected tools with various prediction parameters. The prediction performance will be evaluated using validated *C. jejuni* sRNAs. Moreover, available RNA-Seq data will be analysed to construct an sRNA-target regulatory network. An *in vivo* RNA crosslinking protocol will further explore the sRNA-target network. This information will be integrated to build a global sRNA-mRNA regulatory network. Accomplishing these aims will answer questions regarding the regulatory role of sRNAs in *C. jejuni*, an area that is currently poorly understood.

Chapter 2

Materials and Methods

2.1 RNA crosslinking protocol

2.1.1 Culturing of *C. jejuni*

Motile clinical isolates of strain NCTC 11168 (donated by Professor Charles Penn) were stored in Microbank bacterial preservation tubes (Fischer Scientific) at -80 °C. A loop of strain NCTC11168 was streaked and transferred onto a plate with blood-free selective agar (CCD agar) containing charcoal ferrous sulfate, sodium pyruvate, casein hydrolysates, cefazolin, and sodium deoxycholate (Bolton et al., 1984). After growing on CCD agar plates overnight, the isolated cells were and transferred to cation-adjusted Mueller-Hinton 2 (MH2) (Sigma-Aldrich) 1.5 % agar (BactoTMAgar) plates for 24-48 hours. Then the cells were transferred to fresh MH2 agar plates and cultured for another 24-48 hours. Afterwards, a loop of cells was incubated into 5 mL of MH2 broth and cultured overnight for 12-13 hours at a shaking speed of 250 rpm. The overnight culture was diluted into 30 ml of MH2 broth at an (optical density) OD_{600} value of around 0.2. The culture was then grown by shaking at 250 rpm for 13 and 24 hours to the exponential and early stationary phase, respectively. All incubations took place in a variable atmosphere incubator (VAIN) (Whitley VA500 workstation cabinet) with 90 % N_2, 6 % CO_2, and 4 % O_2.

2.1.2 RNA crosslinking

C. jejuni cells were harvested from the liquid cultures by centrifugation for 20 minutes, 8000 x g at 4 °C. Cell pellets were resuspended in phosphate buffered saline

(PBS) and 4'-aminomethyltrioxsalen hydrochloride (AMT) to a final concentration of 1 mM of AMT and transferred to 6-well plates (Greiner).

RNA residues were covalently crosslinked inside an UV crosslinker (BIO-LINK BLX-254), where cells were irradiated by 365 nm UV light for 5 minutes at an intensity of 0.120 Jcm^{-2}. During irradiation the 6-well plate was kept surrounded by ice to prevent RNA degradation. After the crosslinking step, 1 ml of cold killing buffer (20 mM Tris-HCl (pH 7.5), 5 mM MgCl2, 20 mM NaN3) stored at 4°C was immediately added to stabilize the cells. Bacterial cells were collected by centrifuging at 5000-10000 x g for 5 minutes and snap frozen with liquid nitrogen prior to storage at -20°C.

2.1.3 RNA extraction by a modified bead-beating protocol

The pellet was re-suspended in 600 µL of LETS buffer (0.1 M LiCl, 0.01 M Na_2EDTA, 0.01 M Tris-Cl (pH7.4), 0.2 % SDS). The suspended solution was transferred to Lysing Matrix B, 0.1 mm tubes (MP Biomedicals), which were then placed into a FastPrep^{TM} Homogenizer (MP Biomedicals) to perform three rounds of bead-beating at a speed of 6.0 m/s for 40 sec, with 5 minutes of pause time between each round. Afterwards, proteins and nucleic acids were separated from the cell debris by centrifugation at 13000 rpm at 4 °C for 10 minutes. The RNA was phase-separated by adding 1 volume phenol/chloroform/isoamyl alcohol (PCI) at the ratio of 125:24:1 for two times, followed by another round of phase separation using chloroform/isoamyl alcohol at the ratio of 24:1. The upper phase was transferred to a new tube and mixed with 0.1 volume of 3M sodium acetate (pH 5.2) and 1 volume of isopropanol. After mixing by inverting several times, the samples were precipitated at -20 °C overnight.

Precipitated RNA was washed with 1 volume (600 µL) of 70 % ethanol after the overnight precipitation. The ethanol was then removed by centrifugation at 13000 rpm at 4 °C for 15 minutes. After air-drying for 15 minutes, the pellets were dissolved in 50 µL of nuclease-free water.

2.1.4 DNase treatment

The extracted RNA was treated with DNase following the protocol from the Precision DNase kit (by Primer Design, catalogue #DNASE-50). To minimise RNA degradation, 1 µL of SUPERase In^{TM} RNase Inhibitor (Invitrogen) was added to each sample. The DNase-treated samples were purified using RNA Clean and Concentrator^{TM} -5 kit (Zymo).

To ensure no DNA was left after the DNase treatment, all samples were PCR amplified using forward primer Bac27F (`AGAGTTTGGATCMTGGCTCAG`) and reverse primer Univ1492R (`CGGTTACCTTGTTACGACTT`) that targets 16S rRNA DNA sequences. The following 25ul PCR reaction was set up: 12.5 µL of 2X GoTaq Green master mix (Promega), 1 µL of 10 pM Bac27F, 1 µL of 10 pM Univ1492R, 1 µL of RNA sample and 9.5 µL of nuclease-free water. The PCR mixture was amplified with the heat cycles summarised in Table 2.1. The depletion of DNA in the PCR product was confirmed by gel electrophoresis under 1% Tris-acetate-EDTA (TAE) gel.

Table 2.1: **The heat cycle for PCR amplification of 16S rRNA sequences.**

steps	**temperature**	**duration**	**number of** cycles
initial denaturation	95 °C	3 minutes	1
denaturation annealing extension	95 °C 55 °C 72 °C	30s 30s 1 min	30
final extension	72 °C	5 min	1
storage	10 °C	infinite	-

After ensuring complete DNA removal, RNA quality was examined by running the Agilent 2100 Bioanalyzer RNA 6000 Pico kit (Agilent, UK). RNA samples with RNA integrity number (RIN) equal to or above 8.0 were carried forward for further processing.

2.1.5 Ribosomal RNA depletion with RiboMinus

Ribosomal RNA was removed from the DNase-treated RNA samples by following the protocol from RiboMinus Transcriptome Isolation Kit, for bacteria (by Life Technologies, Catalog # K155004). The efficiency of rRNA depletion was assessed using the Agilent 2100 Bioanalyzer RNA 6000 Pico kit (Agilent, UK).

2.1.6 S1 digestion of linear RNA

400 ng of rRNA depleted RNA samples were denatured at 85 °C for 2 minutes before mixing with 2 µL of S1 reaction buffer (Promega), 2 µL of 1/100 S1 nuclease (Promega) and RNase-free water to make up 20 µL of reaction. The mixture was incubated at 37 °C for 30 minutes and then RNA was purified with RNA Clean and Concentrator^{TM} -5 (by Zymo Research, Catalog # R1013).

2.1.7 Proximity ligation

50 ng of S1 digested RNA was mixed with 2 µL of CircLigase ssDNA ligase buffer and RNase-free water to make up 18 µL of reaction mixture. After incubating at 85 °C for 2 minutes and on ice for 1 minute, 1 µL of 1 mM ATP and 1 µL of CircLigase ssDNA ligase (by Lucigen,Catalog # CL4115K) were added and reaction was incubated for 1 hour at 60 °C.

2.1.8 RNase R treatment

Linear RNA was digested using RNase R by mixing the following: 0.5 µL of 1/33 RNase R (by Lucigen, Catalog # RNR07250), 2 µL of RNase-free water and 2.5 µL of RNase R buffer. The reaction mixture was incubated at 37 °C for 10 minutes.

2.1.9 Reverse crosslinking

Crosslinked RNA was linearised by irradiation with 254 nm UV for 5 minutes in an ice-cooled 96-well plate.

2.1.10 RNA concentration measurement

Throughout the steps mentioned above, RNA concentration was determined with Qubit RNA broad-range assay or Qubit RNA high-sensitivity assay, depending on the expected RNA concentration and following manufacturer's instructions.

2.1.11 Library preparation for MiSeq and NextSeq

cDNA libraries were prepared using the Illumina TruSeq stranded mRNA library preparation kit (Illumina, UK) and according to the manufacturer's guidelines,.

2.2 rRNA depletion by Depletion of Abundant Sequences by Hybridisation (DASH)

2.2.1 Designing sgRNA templates

As depicted in Figure 2.1, the single guide RNA (sgRNA) template consists of the following components from the 5' end to 3' end: T7 promoter sequence (`TTCTAATA CGACTCACTATA`), the guide RNA (gRNA) sequence, a scaffold sequence (`GTTTTAGAG CTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTT TTTT`) and a short fill-in sequence (`ACGATGTCGCAGAGTATGCC`) for primer binding in

the later *in-vitro* transcription. All rRNA sequences of the *C. jejuni* NCTC11168 genome were aligned by Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar, 2004) to identify conserved rRNA regions. The alignment was visualised in JalView (Clamp et al., 2004). Based on conserved sequences among rRNA copies, a library of target sequences was obtained using the webserver of CRISPOR (Concordet and Haeussler, 2018). 524 guide RNA (gRNA) sequences remained after removing candidates with off-target effects. To ensure efficient transcription, the first 2 bp after the T7 -10 promoter sequence must be "G". If the conserved gRNA did not start with 2 "G" s, "G" (s) were added to the beginning of the template. The sgRNA template library was ordered from Integrated DNA Technologies (IDT) at 10 pmol/oligo concentration.

2.2.2 Synthesising sgRNA pools by PCR and *in-vitro* transcription

The fill-in sequence (Figure 2.1) of the sgRNA templates was targeted by the following primer sequence: `GGCATACTCTGCGACATCGT`. Second strand synthesis of the single-strand sgRNA template was carried out in the following PCR reaction:

- 2X KAPA HiFi HotStart ReadyMix (KAPA Biosystems): 12.5 µL
- 10 ng/ul fill-in primer :1 µL
- 10 ng/ul DNA template (oligo Pool) :1 µL
- Add nuclease-free water to make up 25 µL

The PCR was carried out for 1 cycle: denaturation at 95 °C for 3 minutes, annealing at 60 °C for 20 seconds, and extension at 72 °C for 1 minute. The PCR product was purified by following the QIAquick PCR Purification Kit protocol (Qiagen). 300 ng of the double-stranded template was transcribed with the MEGAshortscript T7 Transcription Kit (Thermo Fisher Scientific) following manufacturer's recommended protocol and then purified by Zymo Research Oligo Clean and Concentrator^{TM} -25 kit (Figure 2.1).

2.2.3 rRNA removal by DASH

DASH was used to remove remaining rRNA from the prepared cDNA Illumina TruSeq libraries by setting up a reaction mixture consisting of a 35000:3500:1 molar ratio of sgRNA:spCas9 (NEB Catalog # M0386T):cDNA. Cas9 and the sgRNA pool were pre-incubated at 37 °C for 15 minutes. The purified cDNA library was then

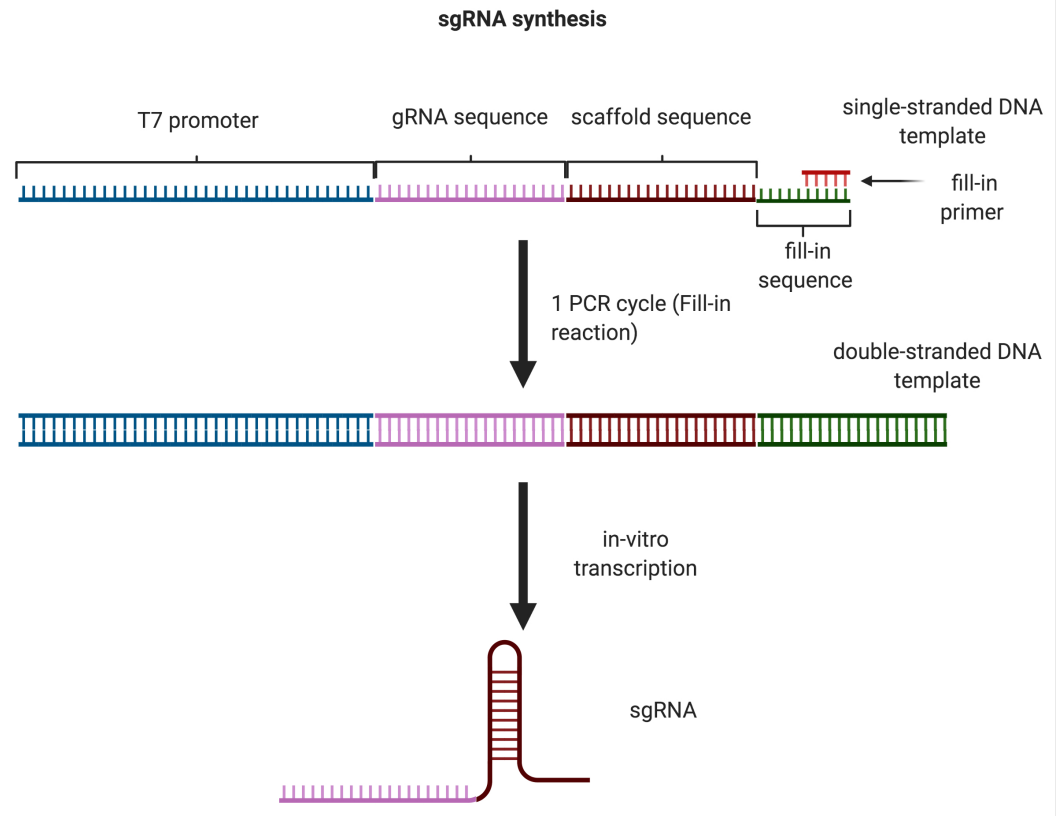

Figure 2.1: **sgRNA template design and *in-vitro* transcription.** Single-stranded DNA templates were PCR amplifed by fill-in primers that targets the fill-in sequence to form double-stranded DNA templates, which were then converted into sgRNAs by *in-vitro* transcription reaction. Created in BioRender.com

added to the pre-incubated Cas9-sgRNA complex and incubated for a further 2 hours at 37 °C. After the rRNA digestion, spCas9 was removed by adding 1 µL of proteinase K (NEB Catalog # P8107S) for 15 minutes at 37 °C followed by heat-inactivation for 15 minutes at 95 °C (Figure 2.2). The resulting DASH-treated samples were purified using the mRNA TruSeq protocol with AMPure XP (Agencourt) beads.

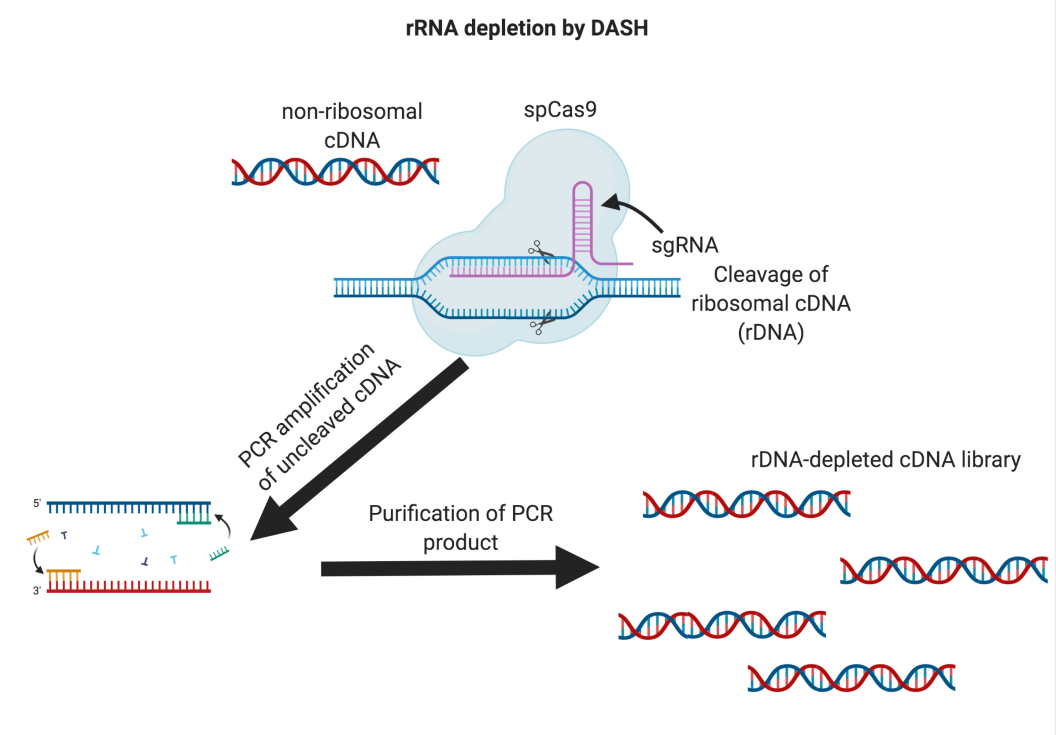

Figure 2.2: **Schematic illustration of rRNA removal by DASH.** sgRNAs bind DNA sequences encoding target of interest (rRNA in this case) and trigger Cas9-mediated digestion. That enabled the non-target sequences (non-rRNA sequences in this case) to be PCR amplified. Created in BioRender.com

2.2.4 Estimating rRNA depletion efficiency

rRNA depletion efficiency by DASH was assessed through high-throughput sequencing using Illumina MiSeq Reagent Kit v2 Nano (2 x 151 cycles). This is described in more detail in Chapter 5.

2.3 RNA crosslinking data analysis

2.3.1 RNA-Sequencing

Final cDNA libraries from the RNA-crosslinked samples were diluted to 4 nm (see the "Denature and dilute libraries guide" of Illumina, UK). The libraries were sequenced using NextSeq 550 System using a high output cartridge (2 x 76 cycles) (Illumina, UK) .

2.3.2 Read trimming

The NextSeq data was demultiplexed from BCL files to fastq files with bcl2fastq v2.20.0.422 (Illumina) using the following command: bcl2fastq –no-lane-splitting. The option "–no-lane-splitting" was included to prevent generating fastq files that were split according to the flowcell lanes.

For the MiSeq data generated to assess DASH efficiency, the fastq files were automatically demultiplexed in BaseSpace during the sequencing run. The Illumina adapter sequences were trimmed with Cutadapt (version 1.3) (Martin, 2011).

2.3.3 STAR alignment

STAR was selected as the alignment tool for analysing NextSeq data generated by RNA crosslinking data (Version STAR_2.6.0c_08-11) (Dobin et al., 2013). This tool was originally designed to analyse eukaryotic splicing events and the resulting chimeric reads created from two distant RNA fragments. The forward and reverse reads obtained from the paired-end sequencing, were mapped separately using the single-end alignment mode. Any chimeric reads with over five nucleotides from the shorter end should be detected by setting the "–chimSegmentMin" parameter as 5. The "outFilterMismatchNoverLmax" parameter was adjusted to 0.05 to make the mapping more stringent to ensure accurate mapping of the shorter ends of chimeric reads..

2.3.4 Annotating and filtering sequence alignments

The output from STAR was then annotated using featureCounts (Liao et al., 2014). The annotation for chimeric reads would lead to multiple annotations of the same read, while the annotation for non-chimeric reads should only result in one feature, except for those reads with multiple mapping or those that map to two adjacent features.

2.3.5 Counting chimeric reads and statistical analysis

After the annotation, annotated features mapping to the same read would be considered interacting pairs. These features would be added to the interaction count matrix, which summarises the total number of interactions between each pair of features. A read map to both geneA and geneB will be counted as a geneA-geneB duplex.

Read counts from all three replicates were pooled together to increase total read counts for sRNAs and mRNAs. After the merging, the interaction count matrix allowed for statistical analysis of each interacting pair using hypergeometric or binomial tests. Moreover, reads were also filtered out if they mapped to two adjacent features or if the same RNA-RNA pairs also appeared in the negative controls (Figure 2.3).

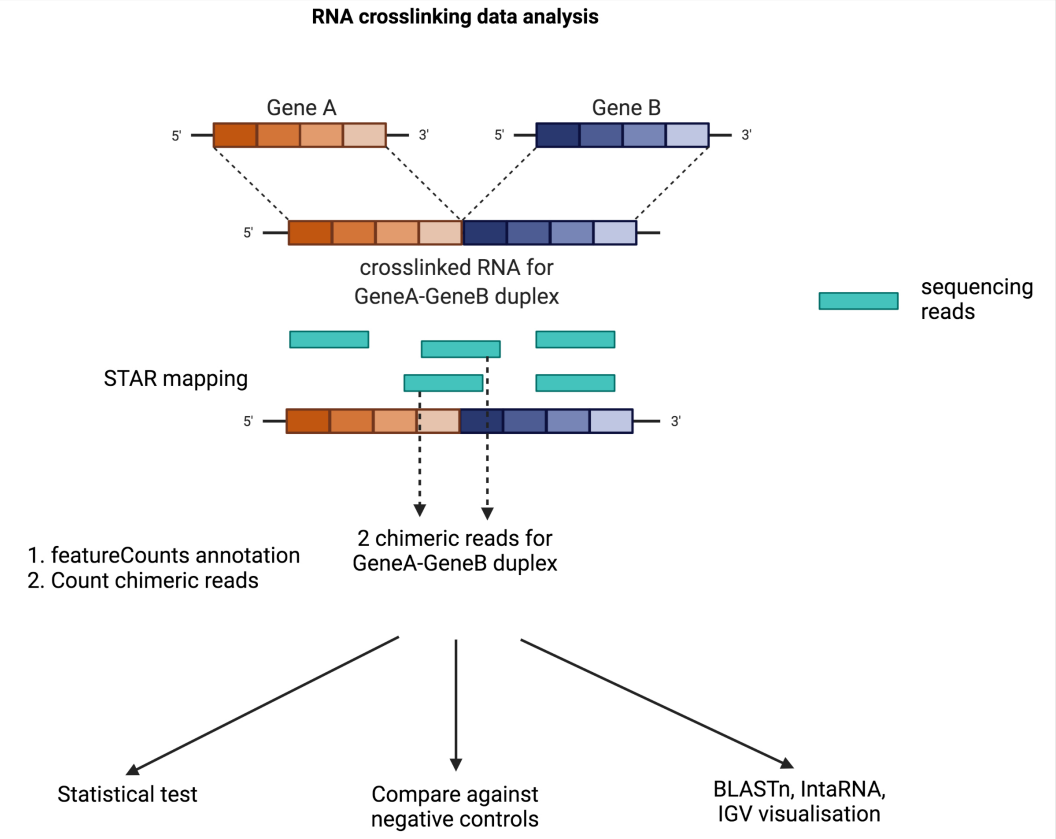

Figure 2.3: **Summary of chimeric read alignment, annotation and downstream analysis.** Sequencing reads were aligned by STAR followed by feature annotation by FeatureCounts. Reads that were annotated by multiple features at distinctive genomics coordinates were identified as chimeric reads, which were subjected to further downstream validation by statistical tests, negative controls and further bioinformatics tools such as BLASTn, IntaRNA and IGV. Created in BioRender.com

2.3.6 Downstream chimeric reads analysis

The chimeric junctions were visualised using Integrative Genomics Viewer (IGV) (Thorvaldsdottir et al., 2012), with BAM files and the GFF file as the input. Moreover, the conservation of chimeric reads was analysed using BLASTn (McGinnis and Madden, 2004) and RNACentral web servers (Consortium, 2017).

2.4 sRNA prediction

2.4.1 Bioinformatics tools for novel sRNA prediction

Novel sRNAs were predicted from the *C. jejuni* genome sequences using RNAdetect (Chen et al., 2018), RNAz 2.0 (Gruber et al., 2010), RNAz (Washietl et al., 2005), a SVM (support vector machine) tool based on trinucleotide composition (designated as SVM-Tri in this study) (Barman et al., 2017) and StructRNAfinder (Arias-Carrasco et al., 2018). For prediction that requires multiple sequence alignment (RNAz 2.0, RNAdetect, RNAz), the genome sequence of strain NCTC11168 (NCBI accession: AL111168.1) was aligned with sequences from strain 81-176 (NCBI accession: CP000538.1), strain 81116 (NCBI accession: CP000814.1), strain CJ017CCUA (NCBI accession: CP012212.1) and strain RM1221 (NCBI accession: NC_003912.7). The alignment was performed by either Clustalw2 (Larkin et al., 2007) (for RNAz 2.0) or with progressiveMauve (Darling et al., 2010) (for RNAdetect and RNAz). SVM-Tri and StructRNAfinder only made use of the NCTC11168 genome sequence for sRNA detection.

The sensitivity of each window size was compared by their ability to detect benchmark *C. jejuni* sRNAs. A window detected the sRNAs if the starts site and ends site were within the window's boundary.

2.4.2 Data source

The in-house Cappable-seq and RNAtag-seq data are available in the ArrayExpress database (Parkinson et al., 2007), under the accession number E-MTAB-11308 and E-MTAB-11309, respectively.

2.4.3 sRNA prediction by ANNOgesic and toRNAdo

ANNOgesic was executed using author's guidelines at: `https://github.com/Sung-Huan/ANNOgesic` (Yu et al., 2018). ANNOgesic predictions were made using wig files created from the sorted bam files of each sample, using bam2wig.py from

RSeQC (Wang et al., 2012). The command used was: bam2wig.py -i [sorted BAM files] -s [text file with the chromosome length and name] -o [output wig files] -t 100000000 -d "++,–" (for running single-end data). All sequencing coverage were normalised with the -t value, which represents the specified wigsum. For instance, -t 100,000,000 means coverage achieved by 1,000,000 reads that are 100 bp long. This value needs to be the same for all RNA-Seq datasets to normalise sequencing libraries with different depths.

The transcript boundaries were predicted using both the wig files (ie, expression level) and the TSS positions, as derived by the in-house Cappable-Seq data (Lam, 2019) and data from Dugar et al., 2013, Porcelli et al., 2013 and Handley et al., 2015, using the ANNOgesic transcript prediction function. The command used for the ANNOgesic transcript boundary prediction was as follows:

annogesic transcript –project_path annogesic/ –annotation_files [GFF file] –modify_transcript merge_overlap –tss_files [File with all TSS positions] –frag_libs [all input wig files] –replicate_frag all_1

The output file of transcript predictions was then included in the ANNOgesic sRNA prediction function. The predictions were further filtered based on the normalised coverage value computed by ANNOgesic. An example of ANNOgesic sRNA prediction command is as following :

annogesic srna –project_path annogesic/ –transcript_files [ANNOgesic transcript prediction output] –annotation_files [GFF file] –fasta_files [fasta file] –filter_info tss –tss_files [File with all TSS positions] –frag_libs [all input wig files] –replicate_frag all_1 –tss_source –detect_srna_in_cds –utr_derived_srna –tss_intergenic_antisense_tolerance 500 –tss_5utr_tolerance n_500 –tss_3utr_tolerance n_500 –tss_intercds_tolerance n_500

Details of the commands are documented in `https://annogesic.readthedocs.io/en/latest/`.

The toRNAdo source code and instructions were obtained from (Hermansen et al., 2018). The sorted BAM files were first processed by running the provided script Prepare_for_toRNAdo_v2.py in the same directory, which converted BAM files into Bedtools 2.24.0. Bedtools output was processed by running the provided The_toRNAdo_script.py to predict a list of ncRNAs.

2.4.4 Sequencing read visualisation

Sequencing reads were visualised with Integrative Genomics Viewer (Thorvaldsdottir et al., 2012), using the wig and GFF files as input.

2.4.5 Conservation analysis of predicted sRNAs

Using NCBI's BLASTn tool, homologous sequences were searched for conservation in predicted sRNAs. As part of the conservation analysis, the query sRNA sequences were compared to those from the Epsilonproteobacteria strains used by Dugar et al. (2013).

2.5 Building a post-transcriptional regulatory network

2.5.1 Co-expression analysis

The construction of co-expression networks was conducted using the Weighted correlation network analysis (WGCNA) package (Langfelder and Horvath, 2008). The co-expression network was then visualised using Cytoscape (Shannon et al., 2003). All Cytoscape images were generated automatically using the Bioconductor package RCy3 (Gustavsen et al., 2019).

2.5.2 Differential gene analysis and enrichment analysis

The differential gene analysis was conducted using the DESeq2 package (Love et al., 2014). Functional enrichment analysis with Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) (including enrichment of KEGG PATHWAY, GO and Pfam) was carried out for protein-coding genes from each co-expression module using the STRINGdb Bioconductor package (Szklarczyk et al., 2019).

2.5.3 Genome-wide sRNA target prediction IntaRNA prediction

IntaRNA prediction was carried out using the input parameter "–personality= IntaRNAsTar". The parameter has been optimised using a set of benchmark sRNA-target data, which optimise the value of seed, accessibility and interaction constraints for sRNA-target prediction Raden et al., 2020.

p-values of the predicted binding energies were estimated using IntaRNA_CSV_p-value.R. The predicted binding sites were visualised using IntaRNA_plotRegions.R. Both scripts were provided in `https://github.com/BackofenLab/IntaRNA/tree/master/R`.

The 5'-UTR regions were estimated by linking TSS to protein-coding genes to include the 5'-UTR positions in the extracted mRNA sequences. The estimated 5'-UTR areas were between the furthest primary or secondary TSS and the 5' end of the protein-coding region (Figure 2.4).

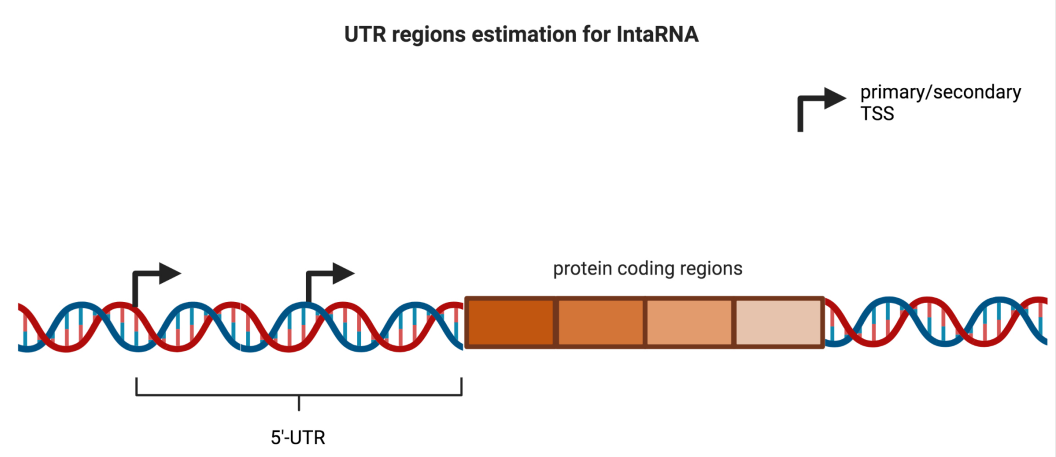

Figure 2.4: **5'-UTR estimation for mRNA genes for IntaRNA analysis.** The 5'-UTRs were defined using the regions between the protein-coding genes and most distant TSSs. Created in BioRender.com

2.5.4 sRNA structure prediction

RNA structures were predicted using the RNAfold algorithm (Zuker and Stiegler, 1981; McCaskill, 1990) in the Vienna RNA Websuite (Gruber et al., 2008). The input sequences were the same as the RNA sequences extracted for genome-wide sRNA target prediction by IntaRNA.

2.5.5 Source code

The source code for transcriptomic analysis can be found in `https://github.com/StephenLi55/c.-jejuni-integrative-analysis` if not specified otherwise.

Chapter 3

Novel sRNA prediction from genomic and transciptomic data

3.1 Introduction

The number of identified sRNAs in *C. jejuni* is relatively low compared to organisms such as *E. coli*. Due to the lack of recognised protein baits, limited availability of molecular tools and the difficulty of culturing *C. jejuni*, the high-throughput experimental discovery of *C. jejuni* sRNAs is difficult. In *C. jejuni* sRNAs also lack primary sequence conservation, making sRNA discovery by sequence homology impossible (Dugar et al., 2013). The publicly available RNA-seq data by and large have been collected from samples grown in standard laboratory conditions and therefore may have not detected sRNAs expressed explicitly under particular conditions. The in-house Cappable-seq and RNAtag-seq data from 63 independent *C. jejuni* samples exposed to 21 different conditions (186) has provided (Lam, 2019) has provided us with a very large and invaluable dataset. But using only RNA-seq datasets for defining sRNAs, can fail to detect some sRNA with poor expression under one or more conditions. Aiming to maximise the number of *C. jejuni* sRNAs captured by this study, we decided to utilise *in silico* tools that predict sRNAs systematically from genome features (e.g. secondary structures, nucleotide compositions) in combination with the expression coverage information from our RNA-seq dateset.

sRNAs and other ncRNAs show structural stability and conservation (Bussotti et al.,

2013) despite sequence divergence. Support vector machine (SVM) is a supervised machine learning approach that separates the training data into positive and negative groups based on features of interest. The groups were classified by mapping annotated training dataset (e.g. ncRNA VS random sequence) to high-dimension space and separating two groups with a hyperplane using predictive features (e.g structural stability). Then, the input testing dataset is mapped into the high-dimension space and classified based on the position of the data points relative to the hyperplane. For instance, data points that are closer to the side of the positive traning dataset will be annotated as the positive group (e.g. ncRNAs).

In the case of ncRNA prediction, SVM tools attempt to classify the testing DNA sequence as either ncRNA or not ncRNA. Minimal free energy (MFE) and structural conservation index (SCI) are therefore widely utilised as predictive features among existing ncRNA detection tools such as QRNA, MultiFind, RNAz and RNAz2.0 (Rivas and Eddy, 2001; Washietl et al., 2005; Gruber et al., 2010). In this study, SVM-based models such as RNAz (Washietl et al., 2005), RNAz 2.0 (Gruber et al., 2010) and RNAdetect (Chen et al., 2018) were selected. RNAz 2.0 further incorporates the dinucleotide content as another predictive feature. RNAdetect was also included as its predictive features also include generalised ensemble defect (GED) and the n-gram model (NGM), which enhance prediction performance by considering suboptimal secondary folding and nucleotide sequence similarities. This study has also included another SVM model (designated as "SVM-Tri") that discovers novel sRNA sequences based on trinucleotide compositions (Barman et al., 2017). This study will also compare the performance of supervised machine learning tools against a knowledge-based tool, namely StructRNAfinder. StructRNAfinder is an automatic pipeline that integrates Infernal and RNAfold. The pipeline searches for consensus sequence and structural alignments for RNA families in the Rfam database (Arias-Carrasco et al., 2018).

The majority of tools described above rely on databases populated by well characterised sRNAs with known structures. On the other hand, transcriptomic data enables sRNA detection from expression coverage and transcription start site (TSS) positions, regardless of structural conservation and stability allowing for the identification of novel sRNAs with distinctive secondary structures. Initial analysis of the in-house RNA-Seq and Cappable-Seq data was carried out with toRNAdo (Lam, 2019; Hermansen et al., 2018), which utilises sRNA expression by searching for intergenic regions where normalised expression levels are above a user-defined threshold. The defined threshold and the genome annotation that defines intergenic regions

help classify predicted sRNAs, with predicted sRNAs flanking protein-coding genes assigned as UTR-derived sRNAs. Any predicted sRNAs within the intergenic region that are not near any annotated genes are classified as intergenic sRNAs. Furthermore, toRNAdo also considers putative sRNAs in the reverse strand of annotated genes as antisense sRNAs.

Another potential tool for sRNA prediction from transcriptomic data is ANNOgesic (Yu et al., 2018). ANNOgesic is a versatile tool that can predict sRNA based on sequence homology from ncRNA databases. It can also recognise non conserved sRNA sequences using the coverage profile and TSS position. By default, all predictions must be within 30 - 500 nt long or other user-selected values. Like toRNAdo, ANNOgesic also looks for expression coverage above a threshold defined by the user. ANNOgesic also integrates TSS locations to define the 5' end of its prediction (Yu et al., 2018) (Figure 3.1). The in-house produced Cappable-seq and RNAtag-seq data (Lam, 2019), in-house data from Emily Stoakes (Stoakes, 2017) as well as publicly available dRNA-seq data (Dugar et al., 2013; Porcelli et al., 2013; Negretti et al., 2017; Dugar et al., 2016, 2018) will therefore be used in this study with ANNOgesic to more accurately predict the 5' ends of predicted sRNA.

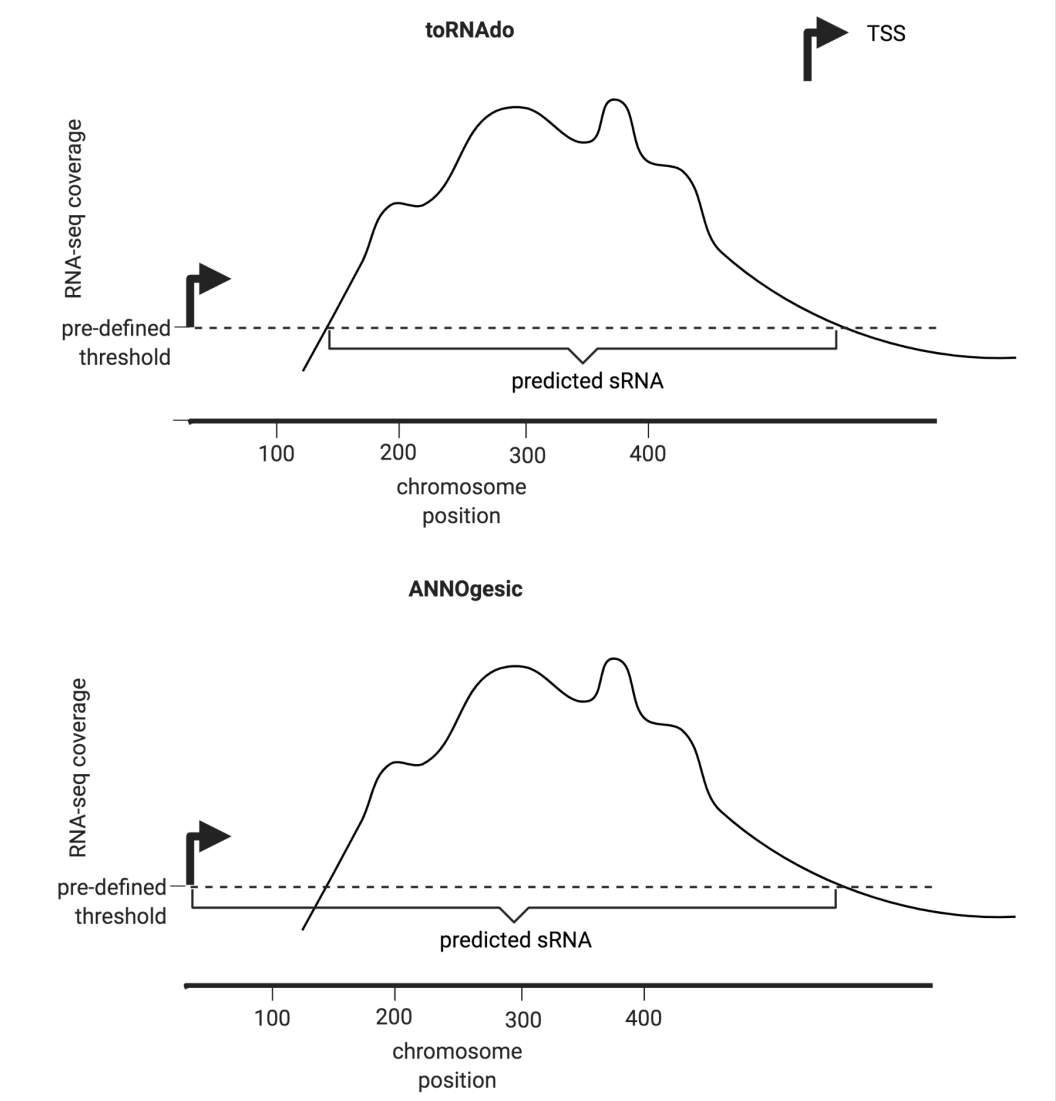

Figure 3.1: **The difference between toRNAdo and ANNOgesic.** Both toRNAdo and ANNOgesic discover sRNA by searching for RNA-seq coverage regions above the user-defined threshold. In addition, ANNOgesic defines the 5' end of the predicted sRNA using the TSS in proximity. Created in BioRender.com.

Chapter Aims:

- Predict sRNA using genome sequences and transcriptomic data
- Compare prediction of different tools and select the best prediction using benchmark sRNAs
- Evaluate putative transcript boundaries by manual inspection
- Assess the conservation of predicted sRNAs.

3.2 sRNA prediction from the genome sequences

3.2.1 Novel sRNA prediction using tools with various prediction parameters and models

So far no systematic evaluation of available tools in predicting *C. jejuni* sRNA has been carried out. As there is not a gold standard sRNA prediction tool for *C. jejuni*, multiple prediction approaches were used in this study to enable direct comparisons. The overall workflow of sRNA prediction using both transcriptomic and genomic datasets is summarised in Figure 3.2. In order to discover sRNA sequences from the *C. jejuni* genome sequence, predictions were made using five tools designed for sRNA discovery from just the genome sequence (Table 3.1).

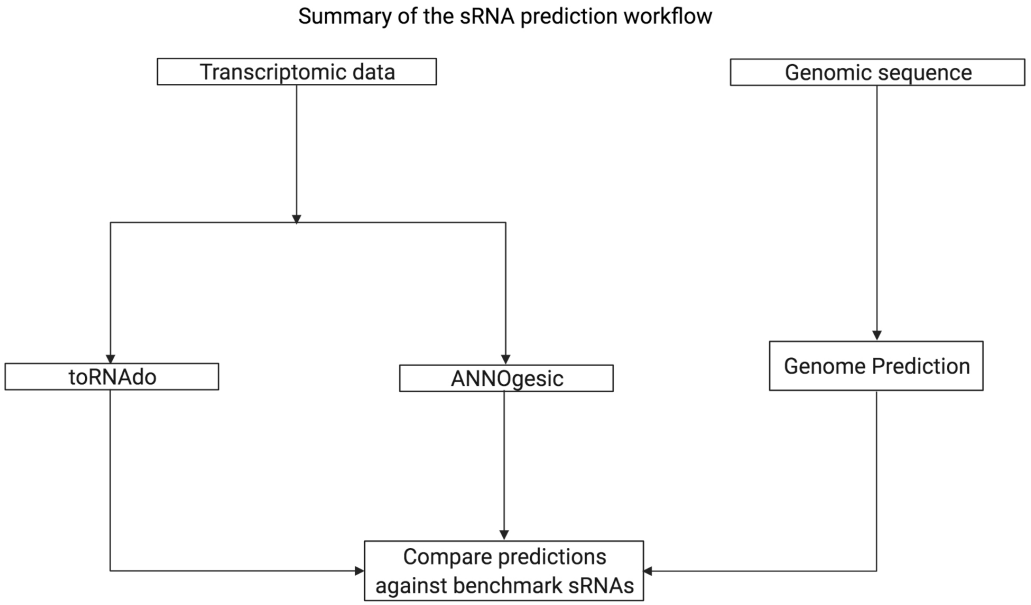

Figure 3.2: **sRNA prediction using genomic and transcriptomic data.** ANNOgesic defines sRNA boundaries using both expression coverage and TSS positions, while toRNAdo prediction defines sRNA boundaries using expression coverage alone. The genome sequence prediction (Genome Prediction) consists of five tools that use different models or parameters. Created in BioRender.com.

Table 3.1: **All selected Genome Prediction tools.**

Tool name	Prediction model	Prediction parameter	Reference
RNAz	SVM	MFE, SCI	(Washietl et al., 2005)
RNAz 2.0	SVM	MFE, SCI, dinucleotide content	(Gruber et al., 2010)
RNAdetect	SVM	MFE, SCI, GED, NGM	(Chen et al., 2018)
SVM-Tri	SVM	trinucleotide composition	(Barman et al., 2017)
StructRNAfinder	knowledge-based	sequence and structure homology	(Arias-Carrasco et al., 2018)

An accurate prediction set would demonstrate high detection sensitivity for real sRNAs and high specificity with minimal false positives. This study estimated the sensitivity of each tool by comparing the number of benchmark sRNAs detected. Twenty-one sRNAs were selected from previous publications as the benchmark to determine which set of predictions was the most accurate (Dugar et al., 2013; Porcelli et al., 2013). All benchmark sRNAs have been detected by both *C. jejuni* NCTC11168 dRNA-seq and northern blot, with defined 5' and 3' ends. The criteria of a detected benchmark sRNA require that the prediction covers the entire region of the benchmark sRNA. That means if the putative sRNA covered only part of the benchmark sRNA, the benchmark sRNA is still considered as "not detected". The result showed that StructRNAfinder, RNAz 2.0 and RNAz identified less than half of the benchmark sRNAs. StructRNAfinder showed particularly low sensitivity as it found only two benchmark sRNAs. In contrast, RNAdetect and SVM-Tri displayed better sensitivity as both of them covered more than half of the benchmark sRNAs (Figure 3.3). A closer look at benchmark sRNA detected by each tool further revealed that none of the prediction tools found sRNAs such as *rnpB*, Cjnc30 or CJnc21 (Figure 3.4).

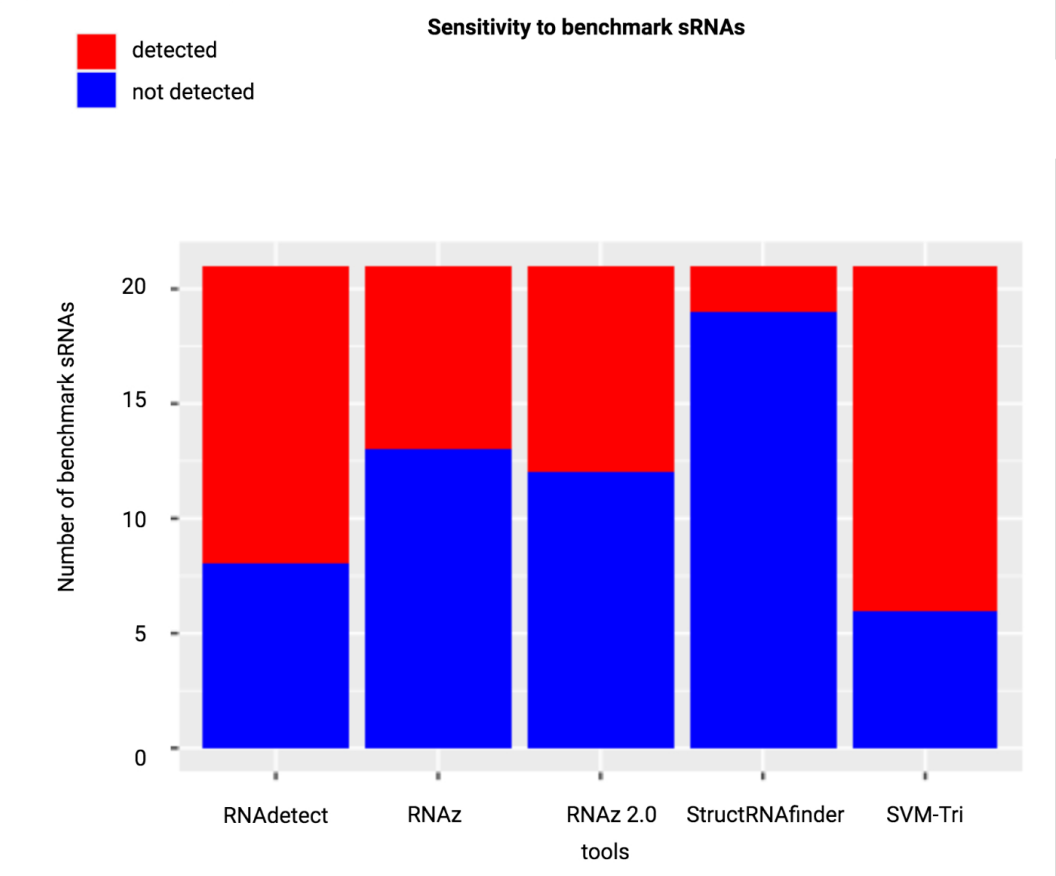

Figure 3.3: **The number of benchmark sRNAs detected by each set of predictions**. The benchmark dataset consist of 21 sRNAs validated by northern blots in previous publications (Dugar et al., 2013; Porcelli et al., 2013).

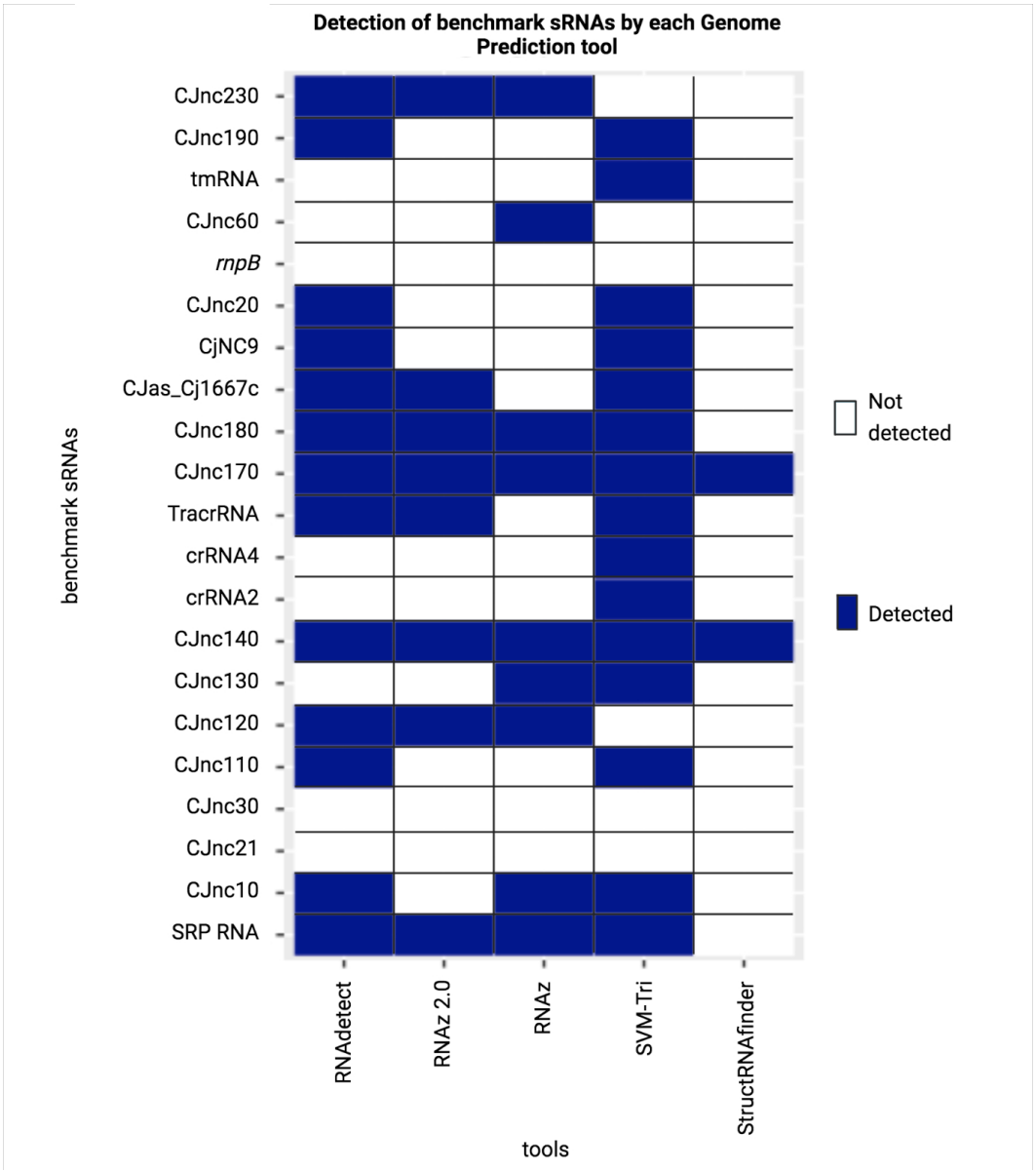

Figure 3.4: **The detection of benchmark sRNAs by each genome prediction tool.** A benchmark sRNA was considered as detected if its full sequence was covered by predicted sRNAs. Only benchmark sRNAs fully covered by the prediction were considered as detected. The stringent criteria was due to the lengths of the genome prediction sRNAs, which can reach above 1000 nucleotides after merging overlapping predictions.

While RNAdetect and SVM-Tri detected more than half of the benchmark sRNAs, such sensitivity may come with a more significant proportion of false positives as predictions from both tools covered about 60 % of the whole genome sequence (Figure 3.5). Such percentages suggested excessive regions were predicted as sRNAs, with most regions being false positives. Based on the estimation so far, it appeared that neither of the selected tools showed sufficient sensitivity or specificity for accurate sRNA prediction. Some benchmark sRNA was undetected, possibly because the putative sRNA only covers part of the benchmark sRNA sequence. Meanwhile, the predictions from RNAdetect and SVM-Tri may consist of lots of false positives. Hence, results from all five tools have been merged to improve the detection sensitivity of benchmark sRNAs (3.6).

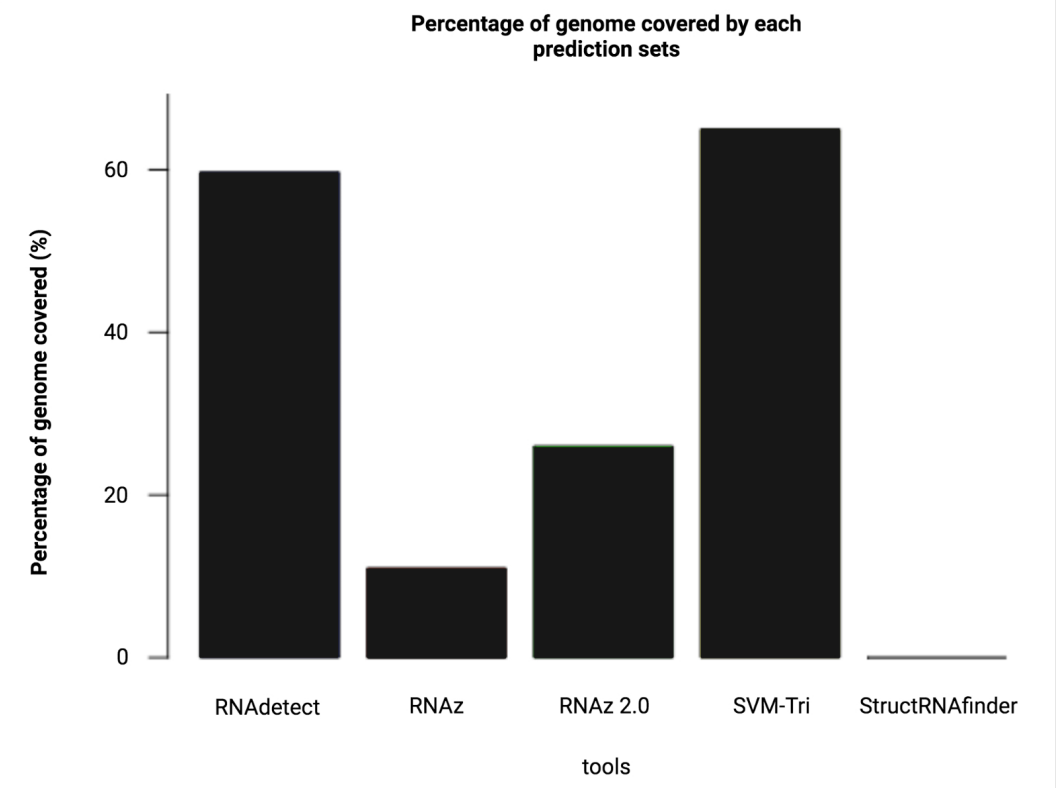

Figure 3.5: **The percentage of genome covered by each set of predictions.** The numbers were calculated by dividing the number of nucleotides covers by all predicted sRNAs with the total number of nucleotides in the *C. jejuni* NCTC11168 genome sequence.

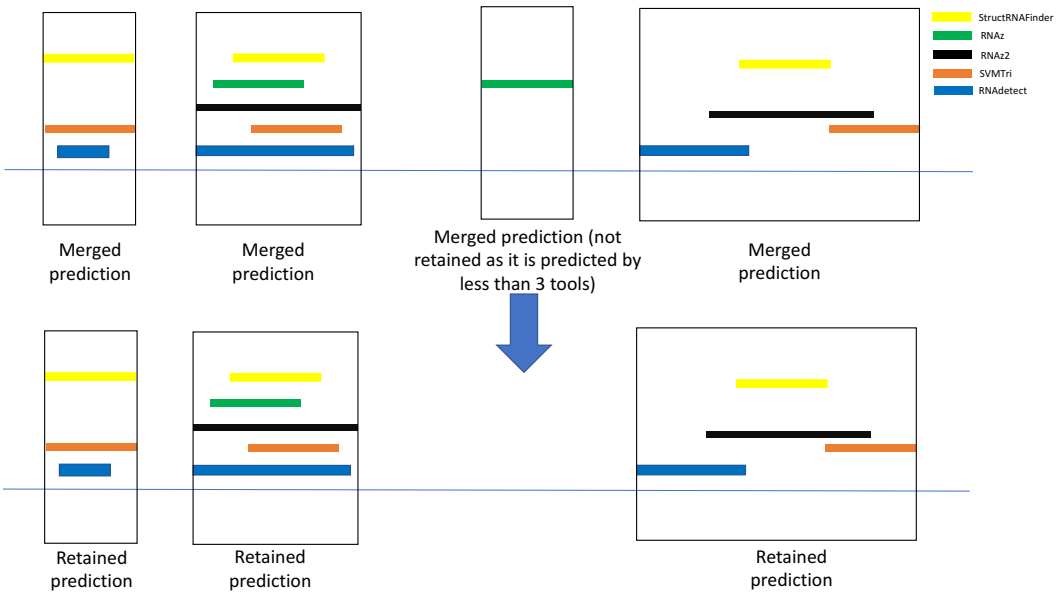

Figure 3.6: **The merging and filtering of consensus results from least 3 different tools.** Only predicted sRNAs predicted by at least three tools were retained for downstream evaluation.

The merged results detected 19 out of 21 sRNAs in total, with each benchmark sRNA covered by results originating from at least three different tools. Notably, *rnpB* was only detected after the merging. That means individual algorithms only covered part of the *rnpB* sequence. The merged predictions from RNAdetect, RNAz2.0, RNAz and StructRNAfinder mapped the entire *rnpB* sequence. However, CJnc30 and CJnc21 remain undetected even after the merging (Figure 3.7).

Only predictions from at least three different tools were retained to improve sensitivity while filtering out false positives (Figure 3.7). The merging retained 1703 predictions, which is quite a lot compared to the Rfam database. For example, Rfam showed 227 and 197 experimentally validated or homologously aligned sRNAs for *Salmonella enterica* subsp. enterica serovar Typhimurium str. LT2 and *Escherichia coli* str. K-12 substr. MG1655, respectively (Kalvari et al., 2018). Moreover, some merged predictions are over 1000 nucleotides long, casting doubts on the accuracy of the predicted transcript boundaries. Further comparison against transcriptomic prediction is required to decide if prediction using the genome sequence provides the most accurate prediction set for *C. jejuni*.

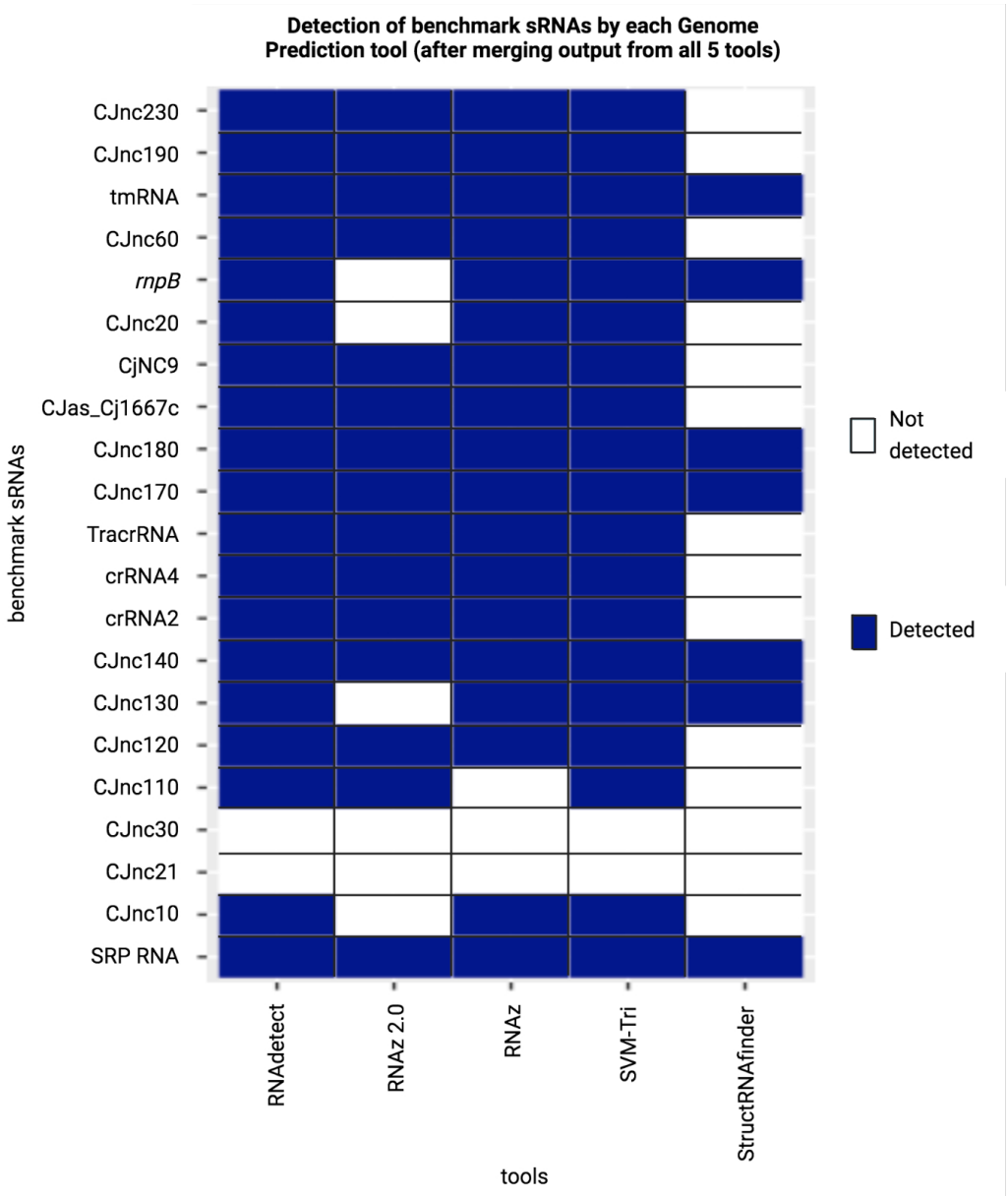

Figure 3.7: **Benchmark sRNA detected after merging tools output.** Predicted sRNAs with overlapping regions were merged into a longer sRNA. Note that *rnpB* was detected by the merged prediction from RNAdetect, RNAz2.0, RNAz and StructRNAfinder

.

3.3 sRNA prediction from transcriptomic data

Despite detecting more benchmark sRNAs, merging results of at least three Genome Prediction tools still led to over 1700 predictions. Moreover, SVM-Tri, RNAdetect, RNAz and RNAZ 2.0 predicted sRNA sequencs from screening through the genome sequence using a sliding window, with the window size and window length being 100 and 50, respectively. However, merging positive predictions from multiple overlapping sliding windows created predicted sRNAs longer than 1000 bp, which means less accurate transcript boundaries. Hence, it would be interesting to see if prediction from transcriptomic data can provide a more accurate set of predictions with similar or better sensitivity, less predicted sRNAs and more accurate transcript boundaries. RNA-seq and TSS data were selected from published studies to identify sRNA expression from transcriptomic data (Dugar et al., 2013; Porcelli et al., 2013; Dugar et al., 2018, 2016; Negretti et al., 2017). These datasets were used alongside our in-house transcriptomic RNAtag-seq and Cappable-seq data which unlike the majority of the publicly available data include mostly non-standard laboratory conditions that can enable the detection of condition-specific sRNA expression.

Using published and in-house transcriptomic data, with both the ANNOgesic and toRNAdo tools we detected 116 and 997 putative sRNAs, respectively. The accuracy of the transcript boundary for all three approaches were assessed by TSS positions. Predictions with TSS near the 5' end of a putative sRNA (500 bp upstream or 50 bp downstream) indicated accurate starting position. Of the three approaches used to predict sRNA, Genome Prediction and toRNAdo only have less than two-thirds of their sRNA predictions associated with a TSS. In contrast, 103 out of 116 ANNOgesic sRNAs had 5'-terminals defined by TSS. That suggested that ANNOgesic predicted transcript boundaries were more accurate than Genome Prediction and toRNAdo (Figure 3.8).

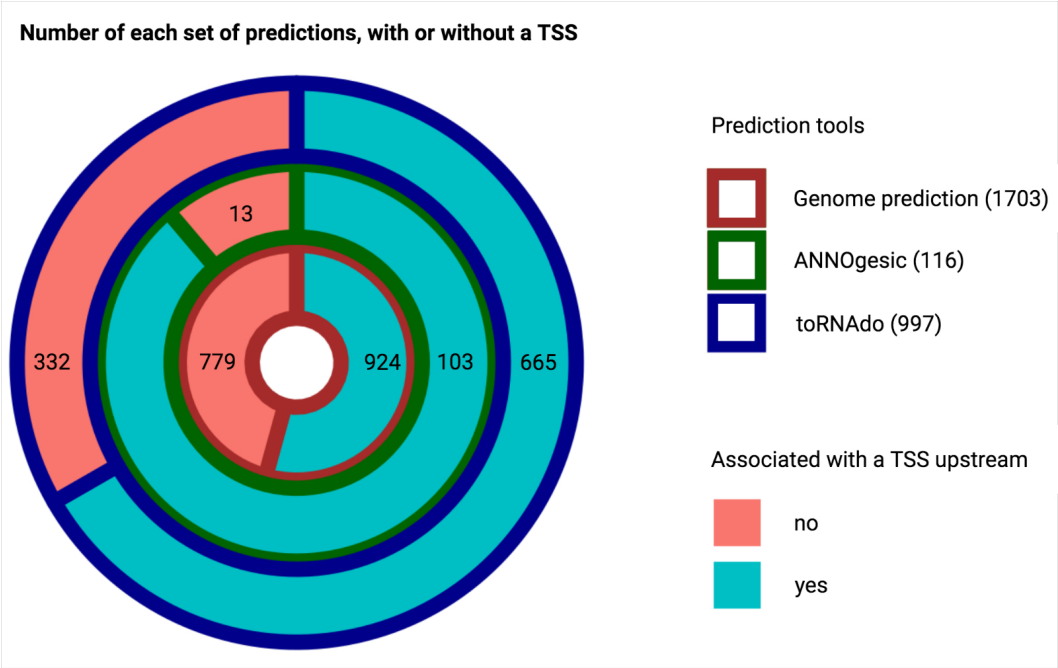

Figure 3.8: **The number of predictions with a TSS upstream.** The numbers following the prediction tools indicate the total number of predictions used for comparison.

To further evaluate the performance of each prediction approach, the three prediction tools ANNOgesic, toRNAdo and Genome Prediction were further compared by their detection sensitivity of the benchmark sRNAs used in Section 3.2.1. ANNOgesic detected 19 out of the 21 benchmark sRNAs, with all of them coupled to at least one TSS within 500 bp upstream of the 5' end (Table 3.2). While Genome Prediction also detected 19 benchmark sRNAs, 6 of them did not have a TSS near the 5' end. Meanwhile, toRNAdo only predicted 12 benchmark sRNAs (Figure 3.9).

Table 3.2: **All northern blot validated sRNAs from previous publications and their corresponding predicted sRNAs.**

validated sRNA	**predicted sRNA**
CjNC9	CjSA4
SRP RNA	CjSA9
CJnc10	CjSA14
CJnc20	CjSA22
rnpB	CjSA51
CJnc60	CjSA64
CJnc110	CjSA88
CJnc120	CjSA90
6S RNA (CJnc130)	CjSA91
CJnc140	CjSA93
tmRNA	CjSA97
crRNA2	CjSA103
crRNA4	CjSA103
TracrRNA	CjSA103
CJnc170	CjSA108
CJnc180	CjSA109
CJnc190	CjSA110
CJnc230	CjSA116
CJas_Cj1667c	CjSA112

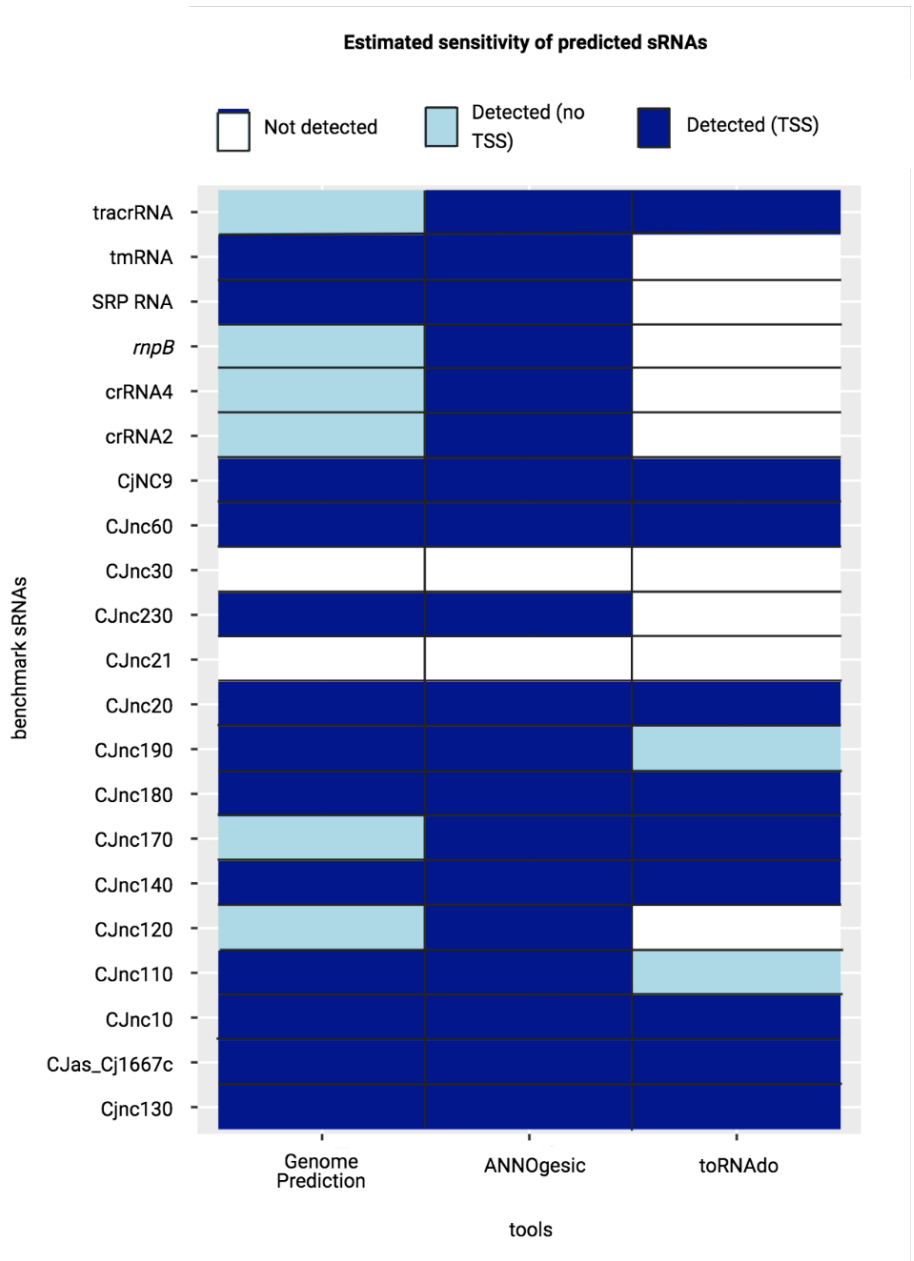

Figure 3.9: **The estimated sensitivity of ANNOgesic, toRNAdo and Genome Prediction.** Among all three prediction approaches, ANNOgesic showed better sensitvity than toRNAdo, while also showed better association with Cappabble-seq TSSs.

Apart from displaying the highest estimated detection sensitivity and accurate 5' ends position, ANNOgesic output also demonstrated the best specificity. The prediction specificity was calculated by dividing the total number of TSS-associated predictions by the number of detected benchmark sRNA. As mentioned earlier, the number of TSS-associated predictions from ANNOgesic, toRNAdo and Genome Prediction were 103, 665 and 924, respectively (Figure 3.10). By dividing these numbers by the number of detected benchmark sRNAs, the estimated specificity for ANNOgesic was 5.4. ANNOgesic's output was more specific than Genome Prediction and toRNAdo, which had an estimated specificity of 71.1 and 66.5, respectively (Figure 3.10). Taking all these together, 116 ANNOgesic-predicted sRNA were assigned the CjSAX designation, where X is an integer between 1 and 116 (with 1 being closest to the point of origin) and were all carried forward for subsequent analysis.

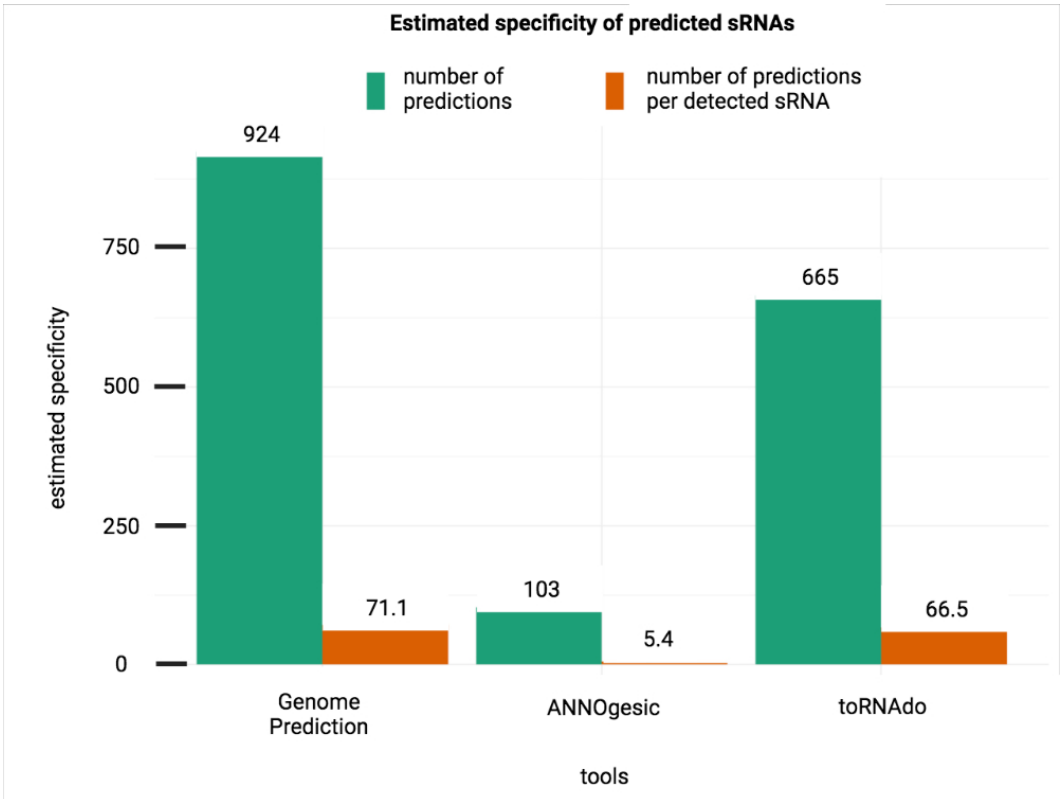

Figure 3.10: **The estimated specificity of ANNOgesic, toRNAdo and Genome Prediction.** The total number of TSS-associated predicted sRNAs and the number of predictions per detected benchmark sRNAs for each tool.

3.4 Improving ANNOgesic prediction by IGV visualisation

3.4.1 Improving sRNA 3' ends of ANNOgesic prediction

While most ANNOgesic predictions were associated with a TSS, it remained unclear if their 3' ends were defined correctly. Moreover, the 5' end positions of many sRNA were associated with multiple TSSs that were close to the starting point of observed gene expression. Looking closer at the benchmark RNAs, most of the ANNOgesic-predicted boundaries using the TSS position defined by our Cappable-seq and dRNA-seq, shared identical or similar 5' coordinates. Nonetheless, some 5' ends were defined with less accuracy especially when several TSS were nearby. Furthermore, the 3' ends that were determined based on expression coverage, appeared to be more affected by RNA-seq background noise. Hence, manual visualisation of the genome and transcriptional coverage and profile at the 5' and 3' positions for all 103 ANNOgesic-predicted sRNA was carried out to allow us to further refine manually the quality of the predicted transcript boundaries.

The problem with defining transcript boundaries was exemplified by CjSA51 and *rnpB*. CjSA51 was the ANNOgesic prediction responsible for detecting *rnpB* (Figure 3.11). Hence, it was reasonable to expect CjSA51 and *rnpB* start and end in similar positions. The predicted 3' end of CjSA51 was over 200 bp longer than the *rnpB* 3' end. The normalised coverage of the region after the *rnpB* 3' end was slightly above the pre-defined threshold for ANNOgesic transcript boundary prediction (normalised coverage > 10). The normalised coverage after the *rnpB* 3' end was likely to be background noise as the coverage was over 1000-fold lower than the expression peak for *rnpB*. Another uncertainty was that the CjSA51 5' end was slightly upstream of the *rnpB* 5' end. The difference was possibly caused by an extra TSS (TSS1 in Figure 3.11) upstream of the TSS responsible for *rnpB* expression (TSS2 in Figure 3.11). As a result, ANNOgesic assigned TSS2 as the 5' end of CjSA51 (Figure 3.11). While such a difference may represent a mistake in ANNOgesic prediction, it is also possible that TSS1 regulates *rnpB* expression in some conditions and TSS2 in others.

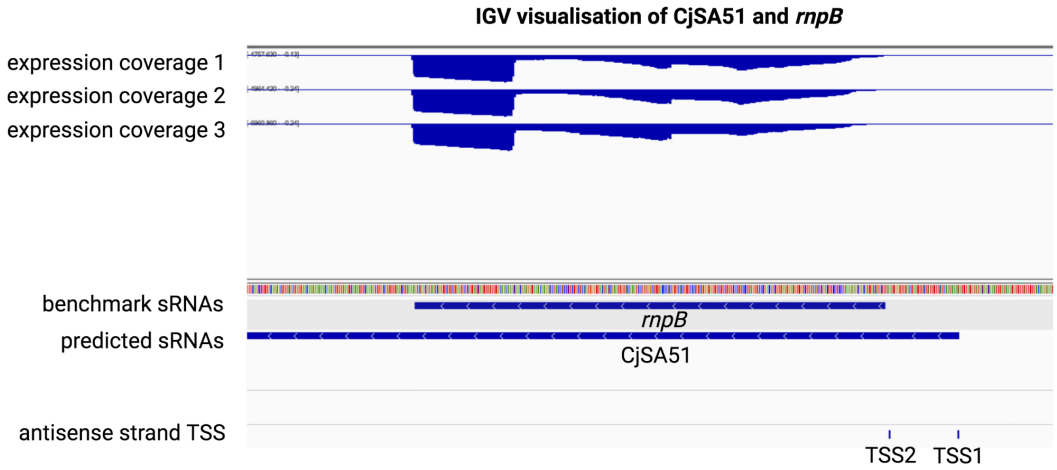

Figure 3.11: **IGV visualisation the expression coverage and transcript boundary of CjSA51 and *rnpB*.** Only the expression coverage of the antisense strand is shown here.

A similar approach used in the filtering step of toRNAdo for sRNA detection (`https://github.com/pavsaz/toRNAdo`) was applied to enable the removal of 3' end sequences originating from background noise. In the previous analysis, the default setting of toRNAdo defined the background noise level to be 5-fold below the maximum expression level of the expression peak of a putative sRNA (Hermansen et al., 2018). Hence, for each ANNOgesic prediction, the 3' end was similarly trimmed using the expression cut-off at one-fifth of the maximum expression value of the prediction. Using this approach, the 3' end of CjSA51 became shorter matching more closely (within 10bp) the benchmark sRNA *rnpB* predicted sequence (Figure 3.12).

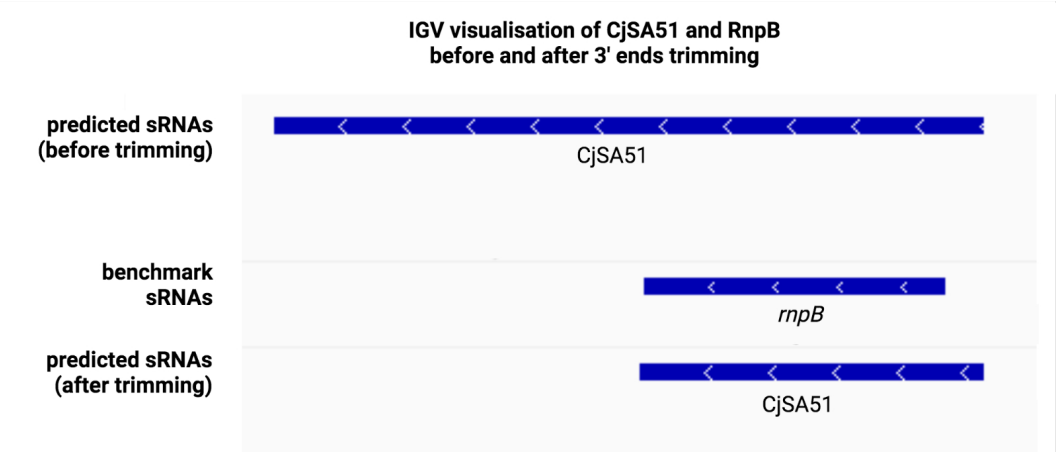

Figure 3.12: **Transcript boundaries of *rnpB* and CjSA51 before and after manual trimming.** The predicted 3' ends were trimmed by removing regions with expression coverage below one-fifth of the maximum coverage of their corresponding predicted sRNAs.

Similarly, the 3' end trimming approach also further refined the transcript boundaries of CjSA9 and CJnc60, making them more similar to the transcript boundaries of SRP RNA and CJnc60, respectively (Figures 3.13 and 3.14).

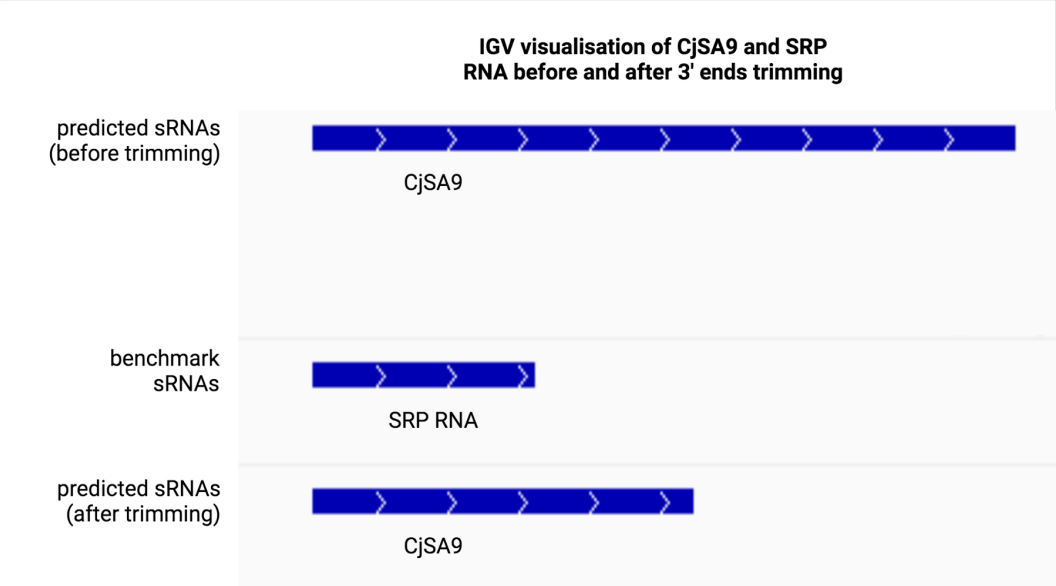

Figure 3.13: **Transcript boundaries of SRP RNA and CjSA9 before and after trimming.** The predicted 3' ends were trimmed by removing regions with expression coverage below one-fifth of the maximum coverage of their corresponding predicted sRNAs.

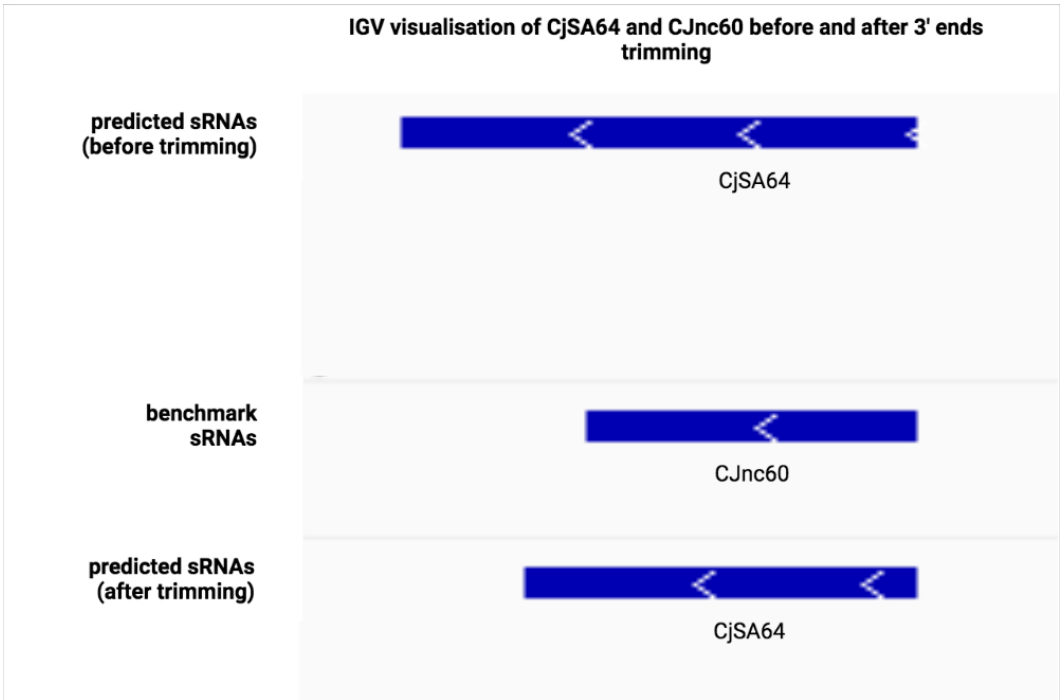

Figure 3.14: **Transcript boundaries of CJnc60 and CjSA64 before and after trimming.** The predicted 3' ends were trimmed by removing regions with expression coverage below one-fifth of the maximum coverage of their corresponding predicted sRNAs. The trimming step reduced the distance between the 3' ends of CjSA64 and the 3' ends of CJnc60.

The 3' end trimming has also improved the accuracy of other predictions corresponding to benchmark sRNAs. For example, CjSA109 and CjSA110 shared similar genomic coordinates as the antisense pair CJnc180 and CJnc190. The predicted lengths of CjSA109 and CjSA110 were both longer than CJnc180 and CJnc190 but the trimming reduced the length of their 3' ends (Figure 3.15).

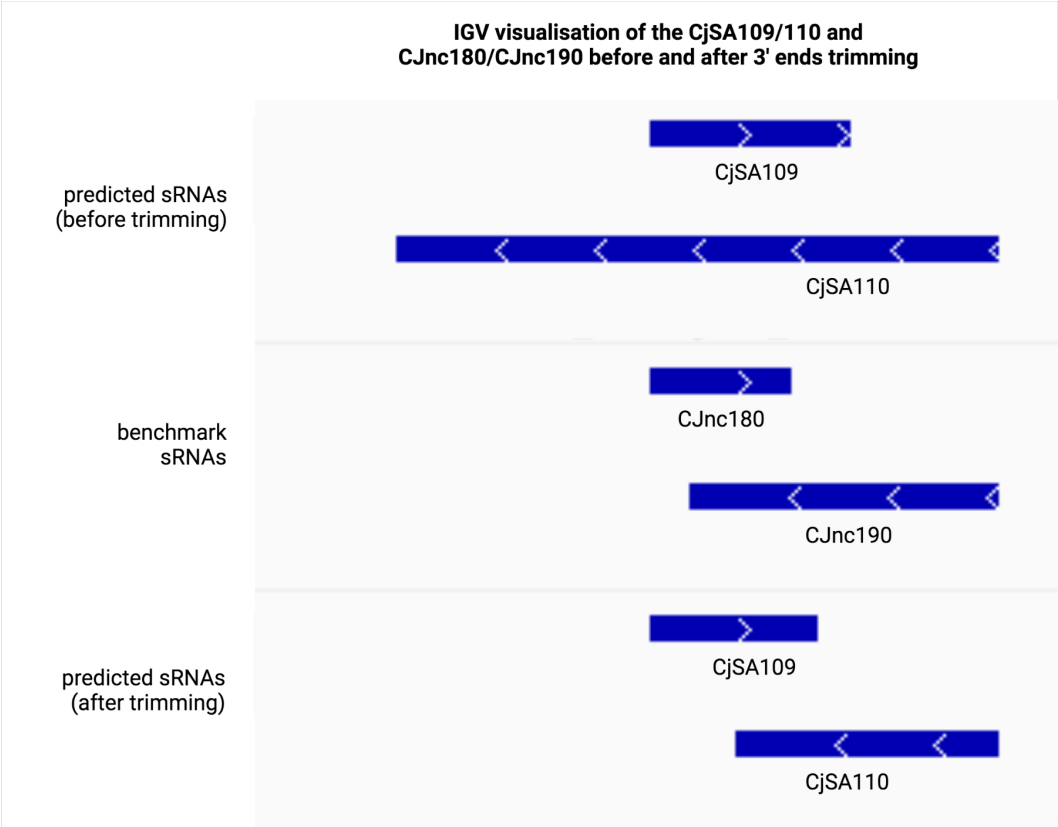

Figure 3.15: **Transcript boundaries of CJnc180/CJnc190 and CjSA109/CjSA110 before and after trimming.** The trimming step reduced the distance between the 3' ends of CjSA109 and CjSA110 and the 3' ends of CJnc180 and CJnc190.

The finalised details of CjSA1-116 can be downloaded from the following link: https://github.com/StephenLi55/c.-jejuni-integrative-analysis/blob/main/ANNOgesic_summary_concatanated.xlsx. The table includes details of CjSA1-116, including the finalised genome corrdinates and strands, samples which showed expression that were responsible for the predictions The table has also highlighted whether predictions retained for downstream analysis and annotations from previous publications. A simplified version of the table is shown as Table 3.3.

Table 3.3: **Details of CjSA1 - 116.** That includes the genome coordinates, strand and whether it was retained for donwstream analysis. More details can be found in https://github.com/StephenLi55/c.-jejuni-integrative-analysis/blob/main/ANNOgesic_summary_concatanated.xlsx.

predicted sRNAs	start	end	strand	Excluded/retained
CjSA1	4604	4926	+	Retained
CjSA2	8103	8214	+	Retained
CjSA3	13572	13687	-	Retained
CjSA4	38708	38814	-	Retained
CjSA5	38883	39255	+	Retained
CjSA6	40741	41056	+	Excluded (Annotated features)
CjSA7	41200	41407	+	Retained
CjSA8	44733	44881	-	Excluded (lack of TSS)
CjSA9	66641	66825	+	Retained
CjSA10	66653	66744	-	Excluded (lack of TSS)
CjSA11	66840	67215	-	Retained
CjSA12	81949	82219	+	Retained
CjSA13	82804	83220	+	Retained
CjSA14	94249	94352	+	Retained
CjSA15	104350	104731	+	Excluded (Annotated features)
CjSA16	132040	132347	-	Retained
CjSA17	151672	152097	-	Excluded (Annotated features)
CjSA18	165860	166263	+	Retained
CjSA19	167067	167249	+	Excluded (Annotated features)
CjSA20	213375	213470	-	Retained
CjSA21	226840	227186	+	Retained
CjSA22	245248	245380	-	Retained
CjSA23	294365	294462	+	Retained

CjSA24	302735	303015	+	Retained
CjSA25	303066	303437	+	Retained
CjSA26	303794	304123	+	Retained
CjSA27	308396	308657	+	Retained
CjSA28	320722	320959	+	Retained
CjSA29	339036	339285	-	Excluded (lack of TSS)
CjSA30	376027	376101	+	Retained
CjSA31	383248	383425	+	Retained
CjSA32	385166	385420	+	Retained
CjSA33	385545	385804	+	Retained
CjSA34	385986	386298	+	Retained
CjSA35	389864	390283	+	Retained
CjSA36	390809	391031	+	Retained
CjSA37	393758	394136	+	Retained
CjSA38	395622	395936	+	Excluded (Annotated features)
CjSA39	396081	396525	+	Retained
CjSA40	399614	399764	-	Excluded (lack of TSS)
CjSA41	399826	400227	+	Retained
CjSA42	408877	409027	+	Retained
CjSA43	418415	418534	+	Retained
CjSA44	426501	426587	+	Retained
CjSA45	430270	430366	+	Retained
CjSA46	433752	434199	+	Excluded (Annotated features)
CjSA47	437723	437881	+	Retained
CjSA48	438698	438934	+	Retained
CjSA49	460270	460371	+	Retained
CjSA50	465478	465837	+	Retained
CjSA51	513132	513500	-	Retained
CjSA52	513143	513441	+	Excluded (lack of TSS)
CjSA53	514272	514368	+	Retained
CjSA54	528585	528700	-	Retained
CjSA55	553074	553193	-	Retained
CjSA56	588000	588357	+	Retained
CjSA57	588920	589223	+	Excluded (Annotated features)
CjSA58	639972	640124	+	Retained
CjSA59	641848	641967	+	Retained
CjSA60	644204	644648	+	Retained

CjSA61	649010	649160	+	Retained
CjSA62	656881	656993	-	Retained
CjSA63	661698	661794	-	Retained
CjSA64	671463	671797	-	Retained
CjSA65	688882	688979	+	Retained
CjSA66	696054	696430	+	Retained
CjSA67	697916	698117	+	Retained
CjSA68	698375	698775	+	Retained
CjSA69	701908	702058	-	Excluded (lack of TSS)
CjSA70	722609	722739	-	Retained
CjSA71	724144	724257	+	Retained
CjSA72	761267	761368	-	Retained
CjSA73	789955	790083	+	Retained
CjSA74	826215	826569	+	Retained
CjSA75	846940	847208	+	Retained
CjSA76	849472	849893	-	Retained
CjSA77	872433	872884	+	Excluded (Annotated features)
CjSA78	911559	911866	+	Retained
CjSA79	911995	912397	+	Retained
CjSA80	917608	917811	-	Retained
CjSA81	943278	943636	+	Excluded (Annotated features)
CjSA82	943755	943943	+	Retained
CjSA83	948661	948761	+	Retained
CjSA84	1034188	1034336	-	Retained
CjSA85	1088294	1088432	+	Retained
CjSA86	1110862	1111049	-	Retained
CjSA87	1119877	1120016	+	Retained
CjSA88	1127993	1128134	+	Retained
CjSA89	1149153	1149261	-	Retained
CjSA90	1150936	1151405	+	Retained
CjSA91	1179498	1179874	+	Retained
CjSA92	1179602	1179763	-	Excluded (lack of TSS)
CjSA93	1188925	1189104	+	Retained
CjSA94	1193164	1193311	-	Retained
CjSA95	1287685	1287980	-	Excluded (Annotated features)
CjSA96	1293323	1293664	+	Excluded (lack of TSS)
CjSA97	1293350	1293675	-	Retained

CjSA98	1323508	1323980	+	Retained
CjSA99	1330420	1330500	-	Retained
CjSA100	1353323	1353396	-	Retained
CjSA101	1368654	1368947	+	Retained
CjSA102	1450025	1450118	-	Retained
CjSA103	1455167	1455608	+	Retained
CjSA104	1493935	1494182	+	Retained
CjSA105	1538468	1538684	+	Retained
CjSA106	1543251	1543594	+	Retained
CjSA107	1547379	1547889	+	Retained
CjSA108	1559373	1559998	+	Retained
CjSA109	1575015	1575130	+	Retained
CjSA110	1575074	1575257	-	Retained
CjSA111	1583112	1583492	+	Excluded (Annotated features)
CjSA112	1589585	1589667	+	Retained
CjSA113	1603941	1604093	-	Retained
CjSA114	1626481	1626725	+	Excluded (Annotated features)
CjSA115	1632050	1632391	-	Retained
CjSA116	1637798	1638412	-	Retained

3.4.2 Manual correction

After using the toRNAdo setting to trim the 3' end of the ANNOgesic-predicted sRNA by estimating background expression, a manual inspection and correction where appropriate, was carried out. Some sRNAs were removed as they were likely to be false predictions due to transcription signals from annotated genes. For example, CjSA114 was derived based on three separate expression peaks, which shared similar locations as tRNA-Ser, tRNA-Ala and tRNA-Val. Hence CjSA114 was removed as it was likely to be a false prediction made from transcription signals from the three tRNA copies (Figure 3.16). Similarly, CjSA95 was eliminated as it was predicted based on the transcription signal from a tRNA-Ser copy (Figure 3.17).

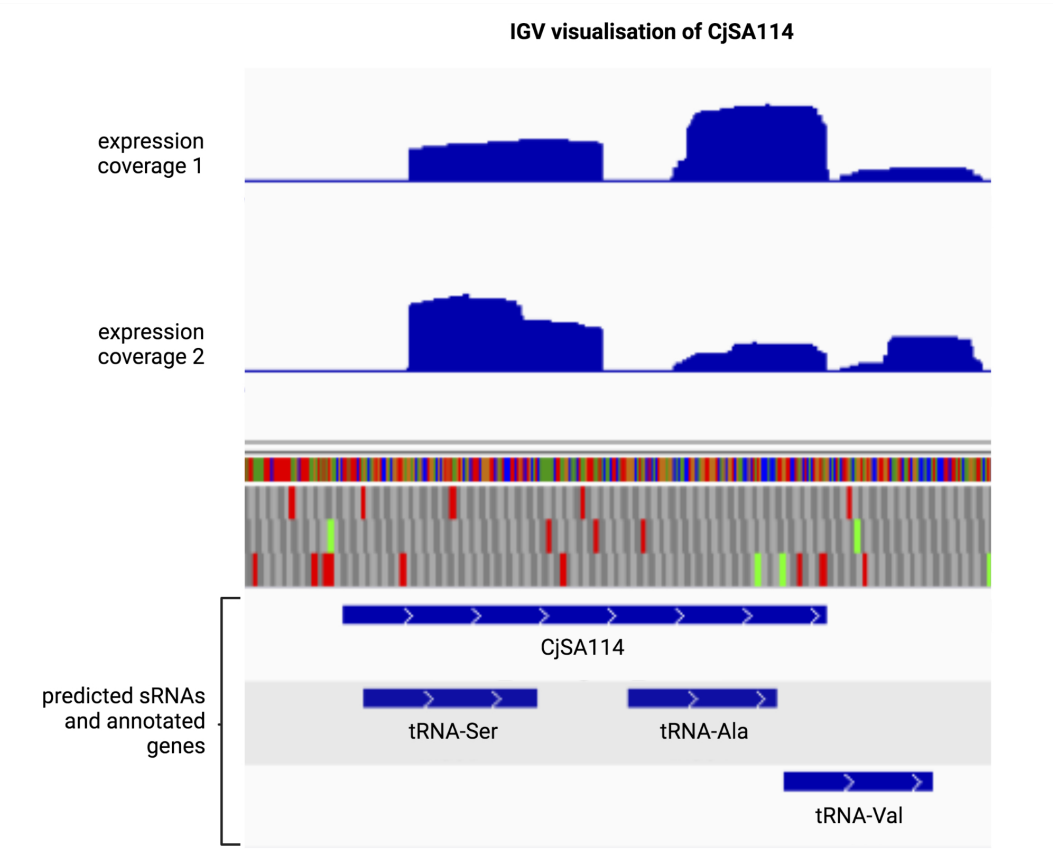

Figure 3.16: **The visualisation of the CjSA114 transcript boundary.** The predicted sRNA was removed as its expression coverage corresponds to other annotated features.

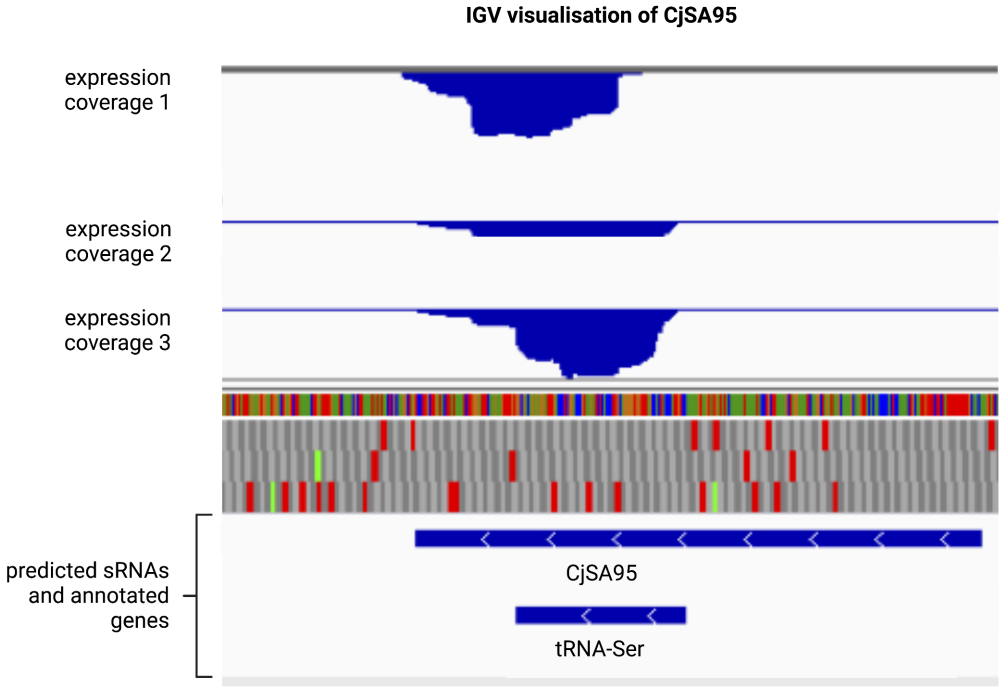

Figure 3.17: **The visualisation of the CjSA95 transcript boundary.** The predicted sRNA was removed as its expression coverage corresponds to other annotated features.

Annotated boundaries of benchmark sRNAs were also used to correct some ANNOgesic outputs. For example, the anticipated length of CjSA90 was too long when compared against CJnc120 in a similar position. The extra length was likely to be a result of the transcription signals from the upstream *groES-groEL* operon and the background noise downstream. While trimming regions with low expression coverage corrected some 3'-terminals, the 5' end was corrected manually using the 5' end of CJnc120. The corrected CjSA90 shared the same transcript boundary as CJnc120 (Figure 3.18).

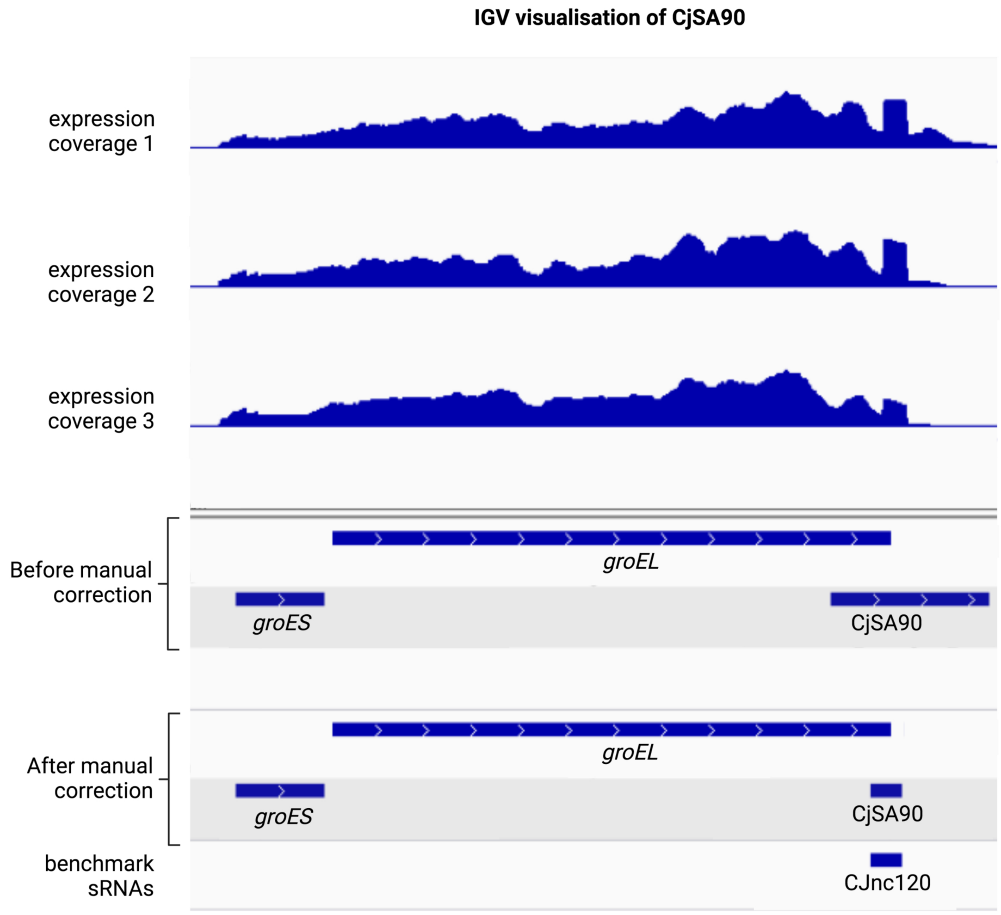

Figure 3.18: **The visualisation of the CjSA90 transcript boundary.** The transcript boundary of the predicted sRNA was corrected using the 5' and 3' ends of the nearest benchmark sRNAs.

3.5 Understanding the putative biological activities of predicted sRNAs

toRNAdo boundary refinement and manual inspection of the original 103 ANNOgesic-predicted sRNA, led to the rejection of seven sRNA and generated a final list of 96 high-confidence sRNAs. The biological significance of the majority of these is unknown. Their activities can be further explored by identifying their mRNA targets and co-expressed genes, as reported in chapters 4 and 5. Other than that, sRNA activities can also be explored based on several factors, including their genomic context, condition-specific expression, and sequence conservation.

3.5.1 Categories and data source of predicted sRNA

The genomic context of sRNA can be significant in their biological role. For instance, antisense sRNA may regulate the activities of its antisense mRNA. Similarly, UTR-derived sRNAs may regulate the same biological processes as their corresponding coding genes. All predicted sRNAs were therefore further evaluated by manually categorising them based on their genome position using IGV visualisation. Notably, more than one-third of all predictions were antisense sRNAs, while 24 out of 96 predictions were inside an operon (a cluster of genes regulated by the same TSS). Another 28 predictions were either 5'-UTR or 3'-UTR with only 9 predictions being intergenic, while the remaining 2 were intragenic (Figure 3.19). Such an observations are probably the result of the small *C. jejuni* compact genome which is only 1.64Mb in length with a total of 1762 annotated genes, pseudogenes and RNA genes.

Figure 3.19: **The proportion of different sRNA categories for the 96 finalised predicted sRNAs.** All sRNAs were categorised by IGV visualisation, using their genomic positions relative to the genome annotation.

Another way to understand the biological role of each sRNA is their data source. Published dRNA-seq data (Dugar et al., 2013; Porcelli et al., 2013) was obtained from standard laboratory conditions. In contrast, our in-house RNAtag-seq data represented a more comprehensive range of non-standard laboratory conditions. Hence, it would be interesting to investigate how many predictions were generated from non-standard laboratory conditions, as those sRNAs might strongly influence stress adaptation. The results indicated that 27 predicted sRNAs were made from RNAtag-seq data alone, while 55 predicted sRNAs were obtained from ANNOgesic output using published data from standard laboratory conditions. The remaining 14 predictions came from data from both standard and non-standard laboratory conditions (Figure 3.20).

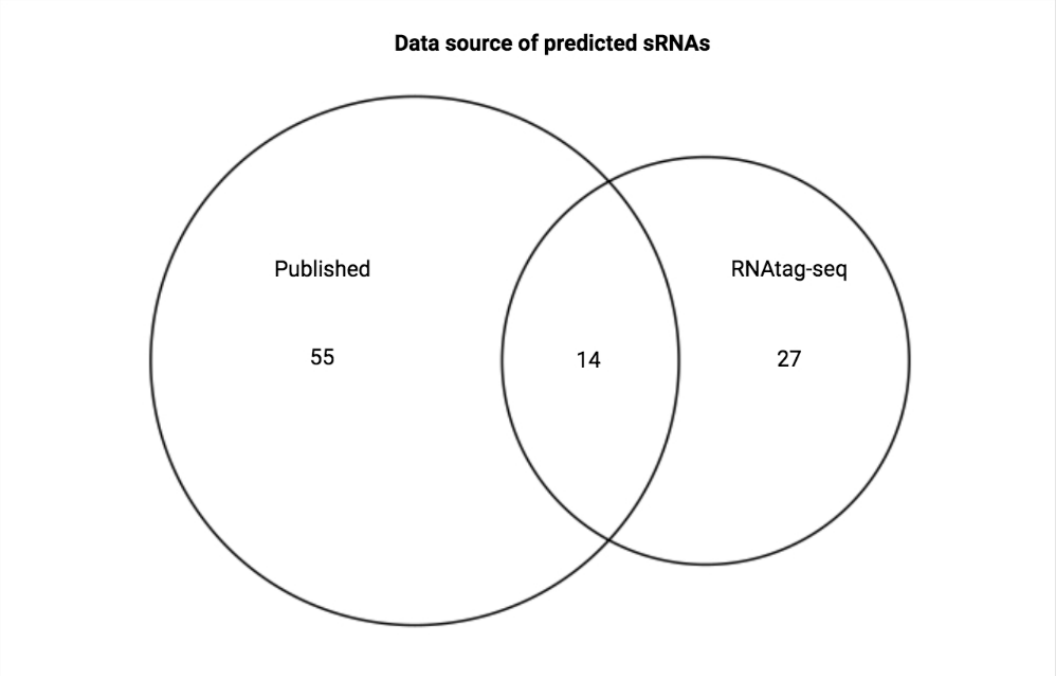

Figure 3.20: **The data source of the 96 finalised predicted sRNAs.** The numbers indicate whether an ANNOgesic prediction was made using published data, RNAtag-seq data or both.

While ANNOgesic made 69 predictions from published datasets, not all of those predictions were reported in those publications, possibly due to different prediction approaches used. That suggested ANNOgesic may produce a novel set of predicted sRNAs.

In order to confirm the novelty of our predicted sRNAs, it is worth checking if they have already been postulated from the published dRNA-seq analysis. Besides validated benchmark sRNAs, published dRNA-seq data have also suggested other sRNAs without northern blot verification. Hence, any novel ANNOgesic predicted sRNAs would neither be detected by northern blot nor dRNA-seq analysis. Among all published and predicted sRNAs, 15 predicted sRNAs were found in all 3 studies. Interestingly, 65 out of 96 ANNOgesic predictions were never detected in published dRNA-seq studies. That means nearly two-thirds of the predicted sRNAs in this study were novel (Figure 3.21).

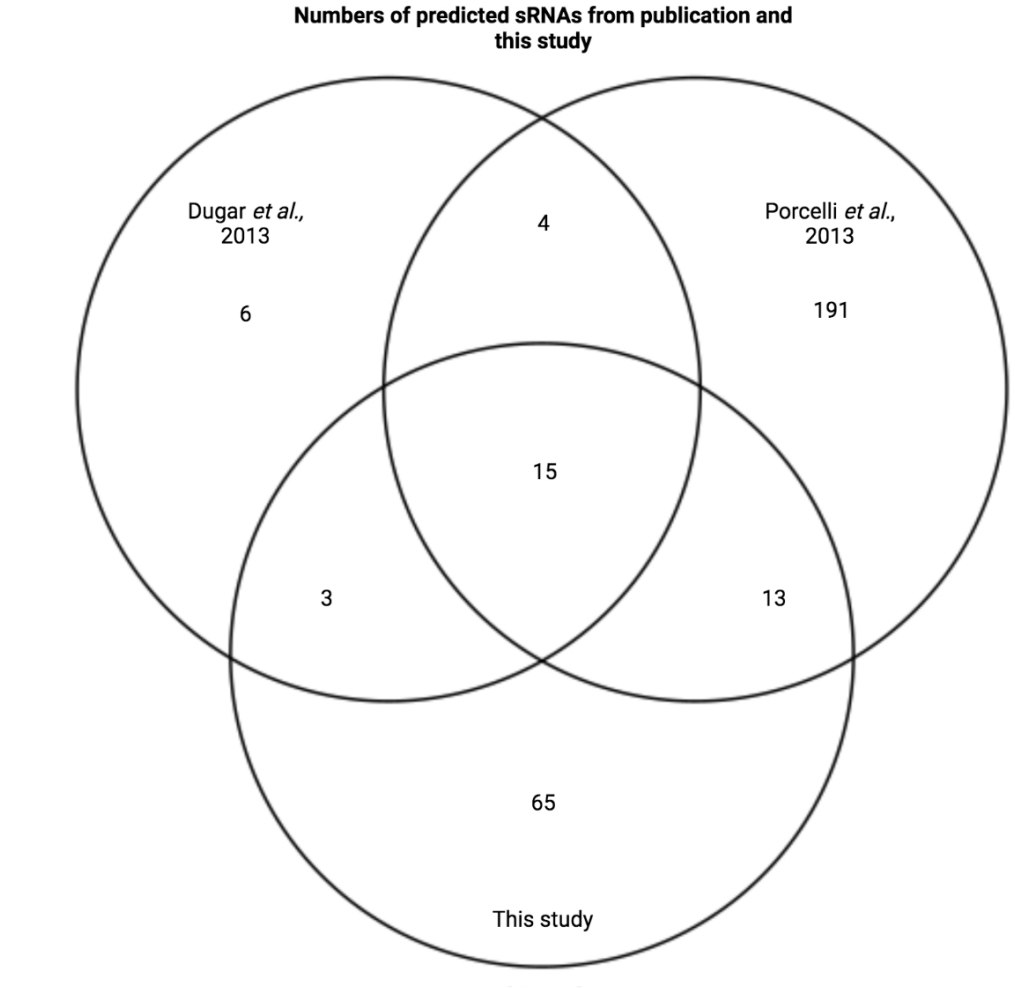

Figure 3.21: **Comparisons of ANNOgesic prediction against sRNAs detected from published dRNA-seq studies (Dugar et al., 2013; Porcelli et al., 2013**. The comparison was made against published sRNAs, both the experimentally validated ones and those that have yet to be confirmed by northern blot.

3.5.2 Conservation of most sRNAs among *C. jejuni* strains

The biological roles of predicted sRNAs can also be anticipated based on their sequence conservation. Some sRNAs play conserved regulatory roles across *Campylobacter* and *Helicobacter* and thus may regulate conserved or housekeeping pathways (Dugar et al., 2013). Other sRNAs may affect processes in specific *C. jejuni* strains only. Conservation analysis of all ANNOgesic predictions was conducted using BLASTn search against genomes from *Campylobacter* and *Helicobacter* strains. The *Campylobacter* and *Helicobacter* strains were the same as those chosen for the conservation analysis in Dugar et al. 2013.

The result of the BLASTn search (available at `https://github.com/StephenLi55/c.-jejuni-integrative-analysis/tree/main/blastn_output`) showed that most of the 96 ANNOgesic output were conserved only among *C. jejuni* strains and were poorly conserved or unconserved in other *Campylobacter* or *Helicobacter* strains (Figure 3.22).

There were exceptions where some sRNAs were poorly conserved even among *C. jejuni* strains other than NCTC11168. Examples of those included CjSA101 and CjSA50. These sRNAs may play regulatory roles in a specific strain only. On the other hand, several sRNAs were highly conserved among all selected *Campylobacter* and *Helicobacter* strains. Examples of such included CjSA21 and CjSA90 (Figure 3.23). These candidates may be involved in regulatory pathways conserved among Epsilonproteobacteria strains.

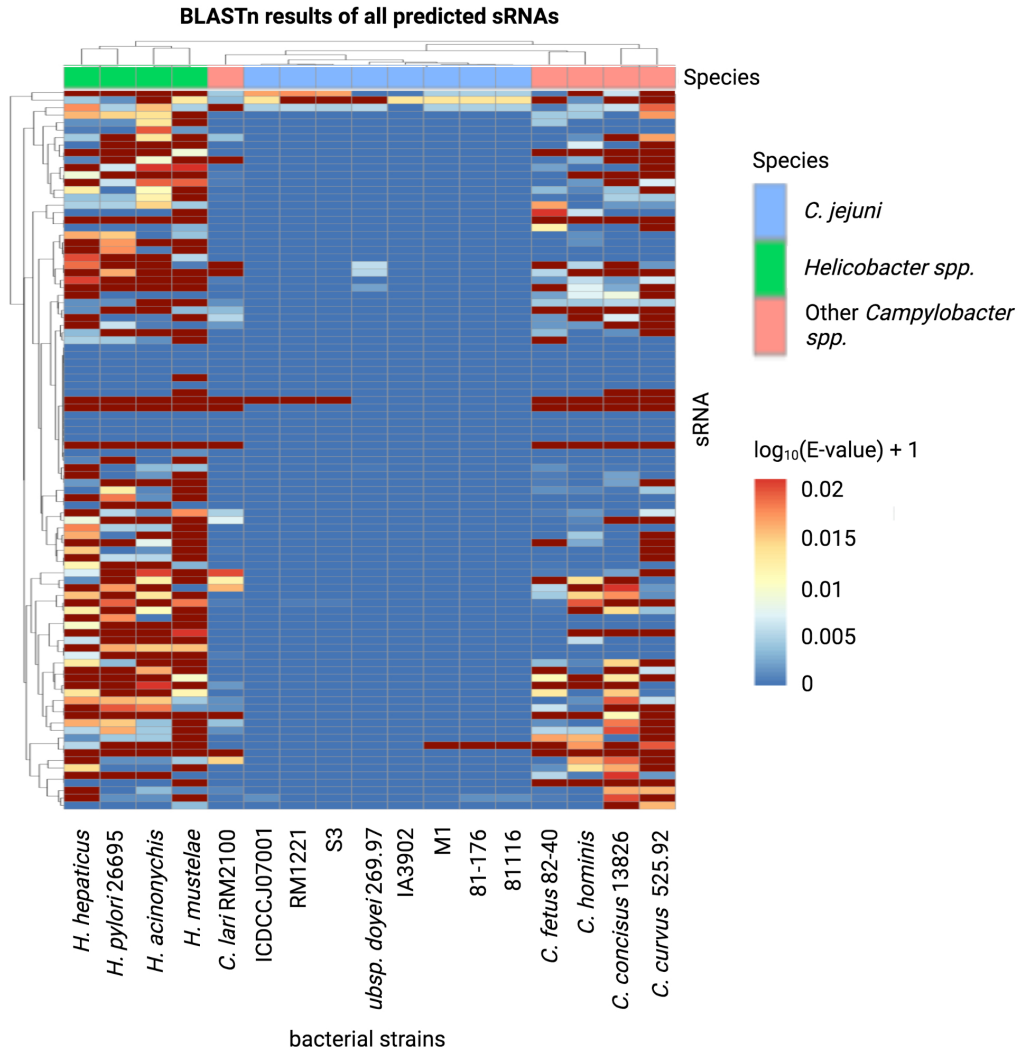

Figure 3.22: **Conservation of predicted sRNAs among that Epsilonproteobacteria strains.** All 96 sRNAs from ANNOgesic were included to compare against genomes from *Campylobacter* and *Helicobacter* strains selected in Dugar et al. 2013. The dark red boxes indicates that the absence of conserved sequences.

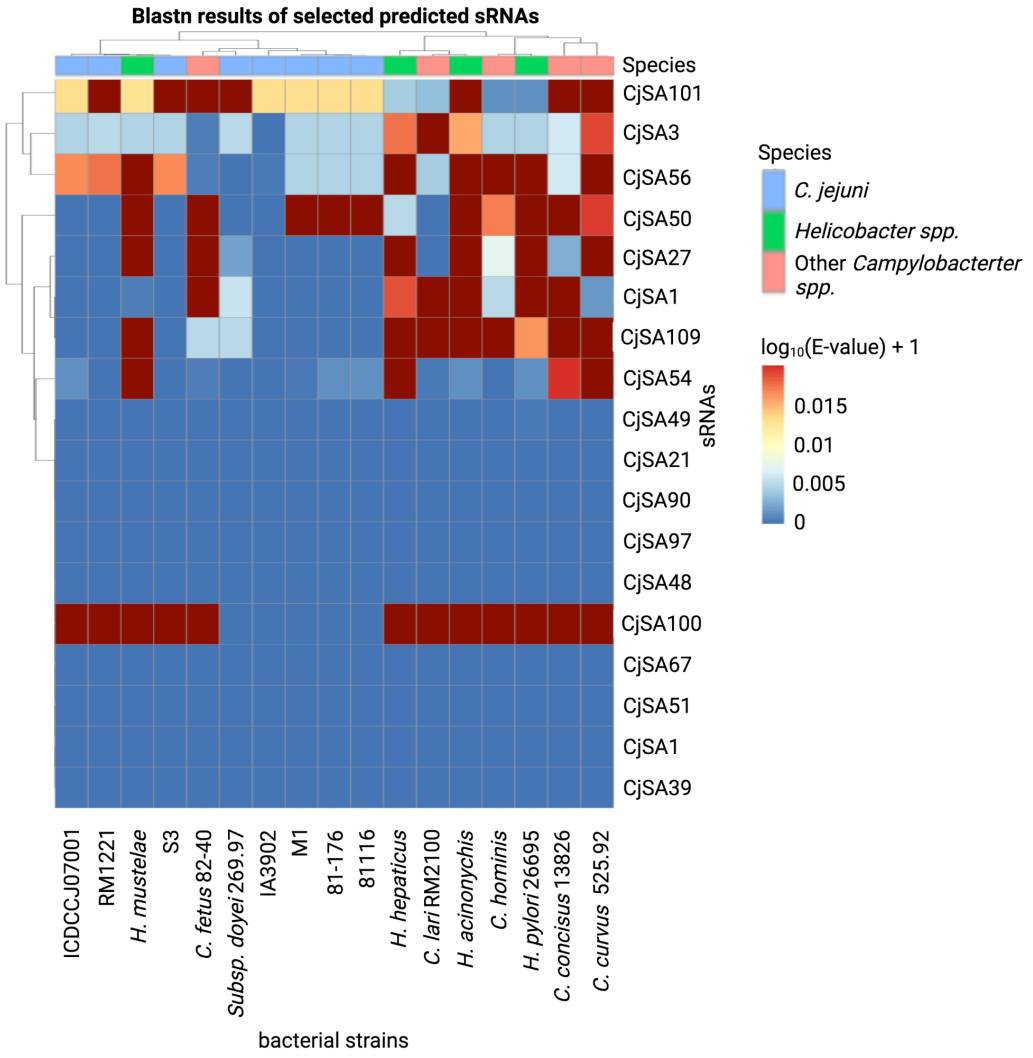

Figure 3.23: **Conservation of selected predicted sRNAs among Epsilonproteobacteria strains.** Only a subset of sRNAs that highly conserved among Epsilonproteobacteria strains or poorly conserved among *C. jejuni* strains were included. The Epsilonproteobacteria strains were same as those selected for conservation analysis in Dugar et al. 2013. The dark red boxes indicates that the absence of conserved sequences.

3.6 Discussion

3.6.1 Comparison between prediction tools

For species like *C. jejuni*, computational discovery of sRNAs is a challenging task as *C. jejuni* is genetically distant from model organisms with the most abundant discovered sRNAs such as *E. coli*. Moreover, *C. jejuni* sRNA sequences and structures are poorly conserved even among closely related species, including *H .pylori* (Porcelli et al., 2013).

Several published tools used supervised machine learning classifiers (SVM) to predict sRNAs from genome sequences. The SVM models were trained with training data sets of validated ncRNA from model organisms such as *E. coli* and *Salmonella*. However, SVM tools selected for this study displayed low detection sensitivity for benchmark sRNAs. Moreover, predictions from tools such as RNAdetect covered nearly 60 % of the genome sequence, which is likely to yield excessive false positives. The poor performance of SVM tools was likely the result of over-fitting, as most of the training data come from genetically distant species from *C. jejuni*. Such a problem can theoretically be resolved by building a novel SVM model trained by species that are more genetically related. However, such an approach is not feasible at this stage. The number of experimentally confirmed sRNAs from *C. jejuni* and closely related species is insufficient to train a reliable supervised machine learning model.

In contrast, StructRNAfinder looked for conserved RNA families from the Rfam database. While such an approach excels at picking out highly conserved RNA sequences, results in this chapter showed that StructRNAfinder was less sensitive to *C. jejuni* benchmark sRNAs, possibly due to the poor sequence and structural conservation of *C. jejuni* sRNAs. Such an outcome indicated that a knowledge-based approach is less capable of discovering new sRNAs from the *C. jejuni* genome than machine learning algorithms.

This chapter also assessed the consensus results from at least three genome prediction tools, alongside ANNOgesic and toRNAdo that include transcriptomic data in their analysis. ANNOgesic showed the best performance among the three approaches in terms of estimated sensitivity, estimated specificity, and transcript boundary accuracy. Such an outstanding performance might be because ANNOgesic predictions have considered TSS positions while also tolerating a few nucleotides below the threshold expression level to prevent actual sRNA signals from being falsely omitted due to transcriptional noise.

Results in this chapter further confirmed ANNOgesic as an accurate tool for sRNA detection among bacteria. Apart from *C. jejuni* NCTC11168, ANNOgesic has also been used to predict sRNAs from *H. pylori* 26695 and *C. jejuni* 81116 (Sharma et al., 2010; Bischler et al., 2015). ANNOgesic prediction followed by manual inspection is becoming more widely used in sRNA prediction for other species, such as in *Fusobacterium nucleatum*, which makes up part of the oral microflora. One of the conserved predictions, FoxI sRNA, was shown to be induced by molecular oxygen and it repressed the major outer membrane porin FomA (Ponath et al., 2021). Likewise, a transcriptomic study in 2020 on *Bacteroides thetaiotaomicron* used ANNOgesic to map TSSs, promoter motifs, terminators, sORFs and regulatory RNAs (riboswitches, RNA thermometers, sRNA). Notably, this sRNA prediction agreed with the subsequent northern blot analysis (Ryan et al., 2020). Similarly, ANNOgesic has recently been used to predicted a list of sRNAs from *B. pertussis* transcriptomic data, with ten selected predictions confirmed by northern blot (Moon et al., 2021).

ANNOgesic's sRNA prediction has been improved by visualisation of genomic features and their genomic context, expression coverage and integrating the filtering step from toRNAdo. While other sRNA prediction tools are available, ANNOgesic appears to be the optimal prediction tool for this study. A recent publication comparing a number of sRNA prediction tools showed that APERO, ANNOgesic, and TLA (from RNA-eXpress) performed better than other sRNA prediction tools. ANNOgesic is the only tool that is capable of working with single-end sequencing data (APERO can only work with paired-end sequencing data) and is designed explicitly for bacteria, whereas TLA was not tailored for predicting bacterial sRNA (Leonard et al., 2019). Hence, to our current knowledge, ANNOgesic is still the most effective tool available for predicting sRNAs from our transcriptomic dataset.

3.6.2 The challenge of accurately defining the transcript boundaries

The biggest challenge for identifying sRNAs in *C. jejuni* was to define the transcript boundaries accurately. This is inherently difficult due to the very compact *C. jejuni* genome, with very few intergenic spaces, and only 14 annotated loop structures which could potentially act as transcriptional termination signals. Such difficulties were partially addressed by trimming the predicted 3' ends by estimating the background expression level. While there were still boundary differences after trimming such differences might be the result of transcription readthrough due to the small

number of a transcription terminator sequences in *C. jejuni*, especially downsream of highly expressed genes.

The presence of multiple TSS near the 5' end of predicted sRNA occasionally perturbed the sRNA 5' boundary.The high TSS density and background noise are partly a result of the AT-rich *C. jejuni* genome, which is prone to spurious transcription (Wade and Grainger, 2017). Moreover, Cappable-seq data analysis from Lam, 2019 on transcripts derived from 21 different growth conditions has identified over 3600 novel TSS, suggesting there might be condition-specific TSS activities under non-standard laboratory conditions.

While refining and confirming predicted sRNA solely by visualising the expression profile mapped to the genome as shown in this study, is possible, such an approach will be highly time-consuming applied to all RNAtag-seq and published data. Hence this study has narrowed down the choice of sRNA sequences with ANNOgesic before refining the transcript boundaries by manual inspection.

3.6.3 Comparison between in-house datasets and published datasets

The ANNOgesic prediction in this study highlighted the importance of our in-house RNAtag-seq generated from conditions absent in other published RNA-Seq data. The in-house data set was responsible for 27 out of 96 ANNOgesic predictions that published data have failed to identify. Notably, ANNOgesic has predicted 65 sRNAs that were not found by either Dugar et al. (2013) nor Porcelli et al. (2013) which only used *C. jejuni* growing under standard laboratory conditions to different growth phases. That suggested that this study has discovered a repertoire of novel sRNAs that could only be detected because of the increased sensitivity of the TSS detection and the diverse growth conditions we used as well as the application for the first time in *C. jejuni* NCTC11168 research of ANNOgesic, a relatively new tool published in 2018.

The final 96 sRNA ANNOgesic predictions were comprised predominantly of antisense and intra-operon sRNAs. Inferences can be made with regards to sRNA biological functions based on their genomic context. However, such organisation could merely be the result of the small compact *C. jejuni* genome or spurious transcription from the AT-rich sequence. It has been suggested that spurious transcription plays a role in the function of bacterial cells (Wade and Grainger, 2017) therefore more research is required to understand the activity of these intragenic sRNA and their associated TSS and promoters. Future work is also necessary to confirm expression and investigate the condition-specific expression of the individual sRNAs in order

to elucidate their precise biological activities.

3.6.4 Conservation of predicted sRNA

BLASTn analysis in this study showed that most predicted sRNAs are only conserved among *C. jejuni* strains, similar to the previous conservation analysis in Dugar et al. 2013. Such a pattern suggests that *C. jejuni* genomes may carry a unique repertoire of sRNAs absent in closely related species or other *Campylobacter* species such as *Campylobacter coli*. This also means *C. jejuni* is an example that demonstrates the diversity of bacterial sRNAs.

19 out of 21 annotated sRNAs were represented by the finalised list of ANNOgesic predictions, for which there is limited information about their biological functions. Other novel sRNAs such as CjSA21 and CjSA90 have primary sequences conserved in *H. pylori*. Hence they may represent post-transcriptional regulatory pathways also found in *H. pylori* sRNAs. Interestingly, most of these conserved sRNAs are UTR-derived instead of antisense. That may illustrate the importance of the regulatory roles of UTR-derived sRNAs across species.

3.7 Summary

In this chapter, sRNAs were predicted from both genomic and transcriptomic data. Results from different approaches were evaluated, which demonstrated the superiority of ANNOgesic for detecting novel sRNAs for *C. jejuni*. The transcript boundaries of ANNOgesic predictions were then refined by systematically trimming 3' end sequences caused by background noise such as transcriptional readthrough. Further refinements in sRNA prediction involved manually distinguishing genuine sRNA expression from transcription readthrough from annotated features. The final list of predicted *C. jejuni* sRNAs included candidates corresponding to previously annotated sRNAs, predictions made from RNA-seq data from our data-sets derived non-standard laboratory conditions, novel sRNAs, and highly conserved sequences *among Campylobacter* and *Helicobacter*. These sRNAs will be used for the further exploration of our dataset in the chapters that follow: chapter 4 explores the condition-specific expression patterns of these sRNA and other co-expressing genes and chapter 5 aims to identify sRNA-mRNA interactions through in vivo RNA crosslinking.

Chapter 4

Integrating datasets from complementary conditions to predict sRNA-target interactions

4.1 Introduction

The biological activities of sRNAs depend on their mRNA binding targets, which are difficult to identify as *C. jejuni* does not contain any identified global RNA binding proteins as scaffolds for co-immunoprecipitation. Moreover, sRNA-mRNA interactions may vary upon biological conditions. Studying these transient binding activities require data from multiple conditions. However, transcriptomic study of *C. jejuni* are conducted under limited sets of conditions.

This study has predicted 96 sRNAs in *C. jejuni* strain NCTC11168, including 65 that have not been described in previous publications (refer to Chapter 3). However, their mRNA binding activities in adapting to non-standard laboratory conditions remain largely unclear. To address this, the host laboratory has previously generated an RNAtag-seq dataset for exploring stress adaptation mechanisms (refer to Chapter 1). The RNAtag-seq experiment covers 21 experimental conditions that model various growth phases, growth temperatures, and environmental stresses throughout the transmission and food processing cycle. This dataset allows simultaneous analysis of multiple environments with minimal technical batch effects, as all RNA samples

processing happened in the same laboratory setup. Hence, this chapter aims to analyse the RNAtag-seq data to elucidate condition-specific sRNA-target interaction. A previous study on the *Staphylococcus aureus* RNA-seq dataset, which constructed an interaction network by combining co-expression analysis, differential expression analysis and *in silico* genome-wide target prediction, has provided an example for predicting the sRNA-target interactome (Subramanian et al., 2019).

Co-expression analysis predicts biological functions of novel genes by partitioning genes with shared biological roles into the same co-expression module, which shows a correlated expression pattern across all RNA samples. Here the co-expression analysis was conducted using the widely used Bioconductor package WGCNA (Langfelder and Horvath, 2008; Bordini et al., 2021; Lu et al., 2021). WGCNA is a standard workflow that assumes a biological network follows a scale-free topology, where only a few central nodes have many interactions (Barabási and Oltvai, 2004). Several examples constructed ncRNA-mRNA network by co-expression analysis. Some most recent examples include the mRNA-lncRNA co-expression network during Severe SARS-CoV-2 Infection (Mukherjee et al., 2021), the lncRNA-mRNA co-expression networks to elucidate the potential role of lncRNA in myocardial infarction (Zhang et al., 2021), and the lncRNA-mRNA co-expression in pre-eclampsia (He et al., 2021). Here, WGCNA will cluster all *C. jejuni* genes into networks. Extracting network edges connecting an sRNA and another gene will form an sRNA-target network. WGCNA clustering can also help predict novel/uncharacterised gene function by the Guilt-By-Association (GBA) principle, which assumes genes with similar biological functions would show expression correlation (Oliver, 2000; Ballouz et al., 2015). The GBA principle suggests that novel sRNAs may share similar physiological functions as other genes in the same cluster.

While WGCNA addresses the overall correlation between gene expression across all samples, it provides less information on the functional significance of individual genes under specific conditions. Hence this study will look for experimental conditions where both a predicted sRNA and sRNA targets are differentially regulated and whether the differential expression pattern matches the gene-gene correlation. Differential expression analysis and co-expression analysis were integrated to identify critical genes related to atrial fibrillation (Li et al., 2020b), mammalian lactation processes (Farhadian et al., 2021) and idiopathic pulmonary fibrosis (Xia et al., 2020a). For this study, an sRNA and its target might display negative co-expression. Conditions of interest are those with one gene upregulated and the other downregulated. For sRNA-target interactions showing positive co-expression, the desired differen-

tial expression pattern involves both genes exhibiting differential upregulation or downregulation together. This study looks for interactions with desired differential expression patterns in at least two biological conditions to ensure the predicted interaction is biologically relevant.

Gene pairs may share similar or opposite co-expression and differential expression patterns either because of direct binding interactions or being transcriptionally co-regulated by other regulations. Several tools compute RNA-RNA base-pairing interactions based on sequence complementarity and structural stability. A study compared different RNA-RNA interaction prediction tools using benchmark RNA-RNA interactions datasets from Archaea, Bacteria and Eukaryotes. The results showed that RNAup, IntaRNA and RNAplex are the tools with the leading Matthews correlation coefficient (MCC), the geometric mean of sensitivity and precision (Umu and Gardner, 2017). IntaRNA showed a slightly higher area under ROC-like curve (AUC) than RNAplex and RNAup (Pain et al., 2015) in another comparative study. IntaRNA can also fit the binding energy values into a generalised extreme value (GEV) distribution to compute the p-value and false discovery rate (FDR) based on the distribution of binding energy values to identify the outstanding actual base pairing from all background binding energy values. Another reason for selecting IntaRNA is the availability of optimised parameters for sRNA-target prediction (Raden et al., 2020). All these suggest IntaRNA is the optimal choice for computing sRNA-target from RNAtag-seq data.

To generate a list of putative sRNA-target interactions, the RNAtag-seq dataset and predicted sRNAs from the previous chapter were analysed using WGCNA, DESeq2 and IntaRNA. To understand the significance of the predicted interactions, we further analysed some of the interactions of interest in-depth.

Chapter Aims:

- Identify pathways and genes that facilitate stress adaptations.
- Construct a global sRNA-target interaction network by integrating WGCNA, DESeq2 and IntaRNA.
- Overview analysis of the sRNA-target interaction network to identify biological conditions, biological pathways and co-expression modules of interest.
- Detailed analysis of some sRNA-target interactions as examples to understand their biological significance.

4.2 Overview of condition-specific transcriptional landscapes and predicted sRNA-target interactions

Previous work in the host laboratory investigated the *C. jejuni* transcriptional landscape under both standard and non-standard conditions. sRNA expression under these conditions facilitates the transient sRNA-mRNA interactions under condition of interests. Table 4.1 summarises all experimental conditions used in the RNAtag-seq dataset and their corresponding abbreviation. All pairwise comparisons selected from the RNAtag-seq dataset are shown in Table 4.2, covering comparisons related to growth phases, growth temperatures, iron availability, bile salt availability and various environmental stress.

Table 4.1: Experimental conditions for Cappable-seq and RNAtag-seq. All conidtions involved growing cells in MH2 broth unless specified otherwise.

Sample name	**Initial growth**	**Treatment**
37_M	Grow to exponential phase at 37 °C	NA
37_ES	Grow to early stationary phase at 37 °C	NA
37_LS	Grow to late stationary phase at 37 °C	NA
42_M	Grow to exponential phase at 42 °C	NA
42_ES	Grow to early stationary phase at 42 °C	NA
42_LS	Grow to late stationary phase at 42 °C	NA
cold	Grow to exponential phase at 37 °C	then resuspended in MH2 broth at 4 °C for 24 hours. (Brown et al., 2014).
5%_ce	Grow to exponential phase at 37 °C	incubated in MH2 broth supplemented with 5% chicken exudate at 4 °C for 24 hours (Birk et al., 2004).
acid	Grow to exponential phase at 37 °C	resuspended in MH2 broth at pH 3.5 for 10 minutes

		(Le et al., 2012).
ana	Grow to exponential phase at 37 °C	incubate in anerobic chamber for 1 hour (Klančnik et al., 2014).
heat	Grow to exponential phase at 37 °C	incubated in 55 °C for 3 minutes (Klančnik et al., 2014).
iron_lim_M	Grow to exponential phase at 37 °C	the growth media was MEMα supplemented with 10 M pyruvate (Butcher and Stintzi, 2013).
iron_lim_ES	Grow to early statioanry phase at 37 °C	the growth media was MEMα supplemented with 10 M pyruvate (Butcher and Stintzi, 2013).
iron_rep_M	Grow to exponential phase at 37 °C	the growth media was MEMα supplemented with 10 M pyruvate and 40 M $FeSO_4$ (Butcher and Stintzi, 2013).
iron_rep_ES	Grow to early stationary phase at 37 °C	the growth media was MEMα supplemented with 10 M pyruvate and 40 M $FeSO_4$ (Butcher and Stintzi, 2013).
nacl	Grow to exponential phase at 37 °C	incubated in 1.5% NaCl for 2 hours (Cameron et al., 2012).
oxidative	Grow to exponential phase at 37 °C	added 3mM H_2O_2 for 10 minutes (Klancnik et al., 2006).
starv	Grow to early stationary phase at 37 °C	resuspend in Ringer's solution for 5 hours (Klancnik et al., 2009).
GSNO	Grow to exponential phase at 37 °C	then incubate in 1.5 mM GSNO for 2 hours (Elvers et al., 2005).
sod_deoxy_M	Grow to exponential phase at 37 °C	the growth media was supplemented with 0.1 % sodium deoxycholate (Malik-Kale et al., 2008).
sod_deoxy_ES	Grow to early stationary	the growth media was supplemented

	phase at 37 °C	with 0.1 % sodium deoxycholate (Malik-Kale et al., 2008).

Table 4.2: All pairwise comparisons selected from the RNAtag-seq dataset.

sample	**control**	**stress conditions**
37_ES	37_M	growth phase
37_LS	37_M	growth phase
37_LS	37_ES	growth phase
5%_ce	37_M	temperature
acid	37_M	acid
ana	37_M	anaerobic
cold	37_M	temperature
GSNO	37_M	nitrosative
heat	37_M	temperature
nacl	37_M	hyperosmotic
oxidative	37_M	oxidative
starv	37_ES	starvation
42_ES	42_M	growth phase
42_LS	42_M	growth phase
42_LS	42_ES	growth phase
42_M	37_M	temperature
42_ES	37_ES	temperature
42_LS	37_LS	temperature
iron_lim_ES	iron_rep_ES	iron limitation
iron_lim_M	iron_rep_M	iron limitation
iron_lim_ES	iron_lim_M	growth phase
iron_rep_ES	iron_rep_M	growth phase
sod_deoxy_ES	37_ES	bile salt
sod_deoxy_M	37_M	bile salt
sod_deoxy_ES	sod_deoxy_M	growth phase

4.2.1 The main driving force behind stress adaptation

sRNA-mRNA binding activities can alter the abundance of both the sRNA and the mRNA. Such kind of interactions means sRNA expression may show significant correlation with the expression of its mRNA partner. Therefore, co-expression analysis was conducted to highlight sRNAs and mRNAs with correlated expression across all RNAtag-seq conditions.

All 63 replicates from the RNAtag-seq data were included for the co-expression analysis. The varianceStabilizingTransformation (vst) of the DESeq2 package normalised the raw expression coverage of 63 RNAtag-seq replicates (Love et al., 2014). The vst normalisation resulted in similar read distribution across all samples (Figure 4.1a). The biweight mid-correlation (bicor) of the WGCNA package then converted the normalised expression coverage into a correlation coverage. The bicor correlation matrix was power transformed to construct an adjacency network to suppress background noise with low correlation. Scale-fit topology fit and the mean connectivity suggested the optimal soft-threshold power for power transformation was 5 (Figure 4.1b and 4.1c). The adjacency network was an unsigned network, which accounted for negative correlations equally as positive ones. Such an approach helps to identify sRNA-mRNA interactions that led to mRNA degradation. WGCNA then transformed the adjacency matrix into a Topology Overlap Matrix (TOM) to account for network topology such as shared neighbours. Afterwards, unsupervised hierarchical clustering and dynamic tree cutting partitioned genes into eleven co-expression modules based on the TOM. WGCNA named each module as different colours, which were later re-designated as modules I to XI for better clarity (Figure 4.1d).

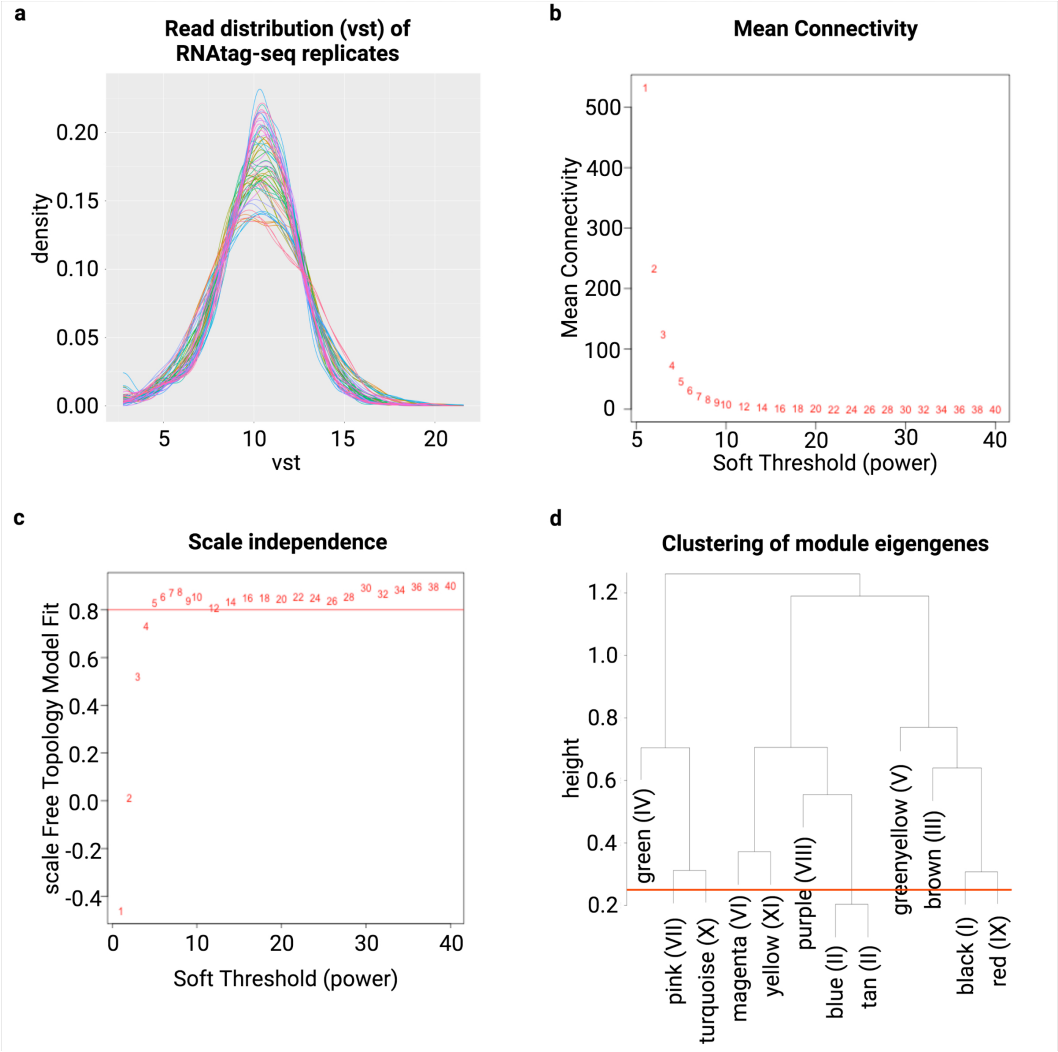

Figure 4.1: **Co-expression analysis by WGCNA.** (a) The read distribution of all RNAtag-seq replicates after vst transformation. The power for soft-threshold power transformation was determined using (b) Mean connectivity and (c) scale-fit topology fit (threshold = 0.8). (d) Average linkage hierarchical clustering of eleven co-expression modules. Originally marked with different colours, the module names were replaced with the roman characters inside the bracket. The red line indicates the default threshold for dynamic tree cut (0.25).

Each co-expression module carry a list of genes that showed significant positive or negative co-expression. Some modules, including module X and module II, carried a larger portion of genes, thus may entail interactions between multiple pathways. Meanwhile, module VIII and module V were relatively small and may only highlight a few specific pathways (Figure 4.2).

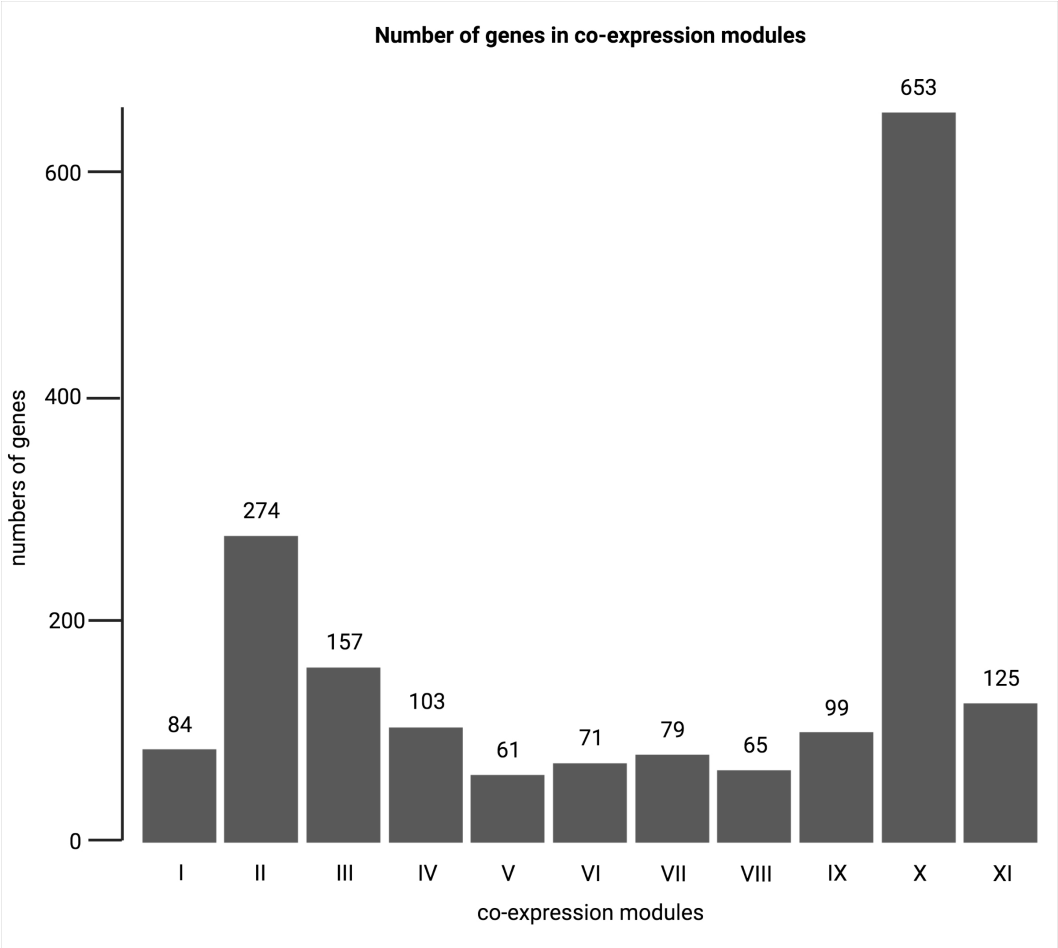

Figure 4.2: **The distribution of genes among co-expression modules.** The numbers above each bar indicate the number of genes in that module.

Enrichment analysis of each co-expression module highlighted the biological pathways represented by each module. Some co-expression modules are predominated by genes from one KEGG pathway, while some modules may reflect the crosstalk between multiple pathways. For instance, modules I and IX are only over-represented with genes from the "Ribosome" pathway, while module VIII is only enriched with ABC transporters genes. These modules may involve sRNA-target interactions that regulate ribosomal biosynthesis or nutrient uptake. In contrast, module X, module XI and module II consist of more genes that lead to statistical over-representation of multiple biological pathways. For example, Module XI consisted of several pathways, with some pathways like the "Citrate cycle (TCA cycle)" pathway being more statistically significant. Some modules represent a variety of biologically related pathways. For example, module II showed statistical enrichment of "Two-component system", "Bacterial chemotaxis" and several metabolite biosynthesis pathways ("beta-Alanine metabolism", "Lysine biosynthesis", "Arginine and proline metabolism" and" Fatty acid biosynthesis". That indicates module II may consist of sRNAs related to the crosstalk between signal transduction, cell motility and metabolites utilisation (Figure 4.3).

Previous study from the host research group (Lam, 2019) highlighted that nutrient-deprived conditions, including the early and late stationary phases and starvation stress, are the main driving force of expression variation. Such pattern agrees with the correlation between co-expression and experimental traits (Figure 4.4), which showed a statistically significant correlation between several nutrient-deprived conditions and co-expression modules. For example, iron-limitation under the early-stationary phase (iron_lim_ES) demonstrates a statistically significant correlation coefficient with all but one co-expression module. Other critical factors for expression change include food processing conditions, such as cold stress (cold), 5% chicken exudate (5%_ce) and NaCl stress (nacl). All of these conditions represent food storage conditions prior to cooking. Cold stress and 5% chicken exudate both represent refrigeration as both conditions involve placing the cells under 4 °C. NaCl is a common food preservative that subjects the cells to hyperosmotic stress that induces water efflux. In contrast, some environmental stress conditions like acid stress (acid), heat stress (heat) and nitrosative stress (GSNO) showed no statistically significant correlation coefficient with any co-expression modules (Figure 4.4). Thus it appears that nutrient deprivation and food storage conditions involve more transcriptional rewiring than other conditions.

The differential expression analysis also showed that statistically overrepresented

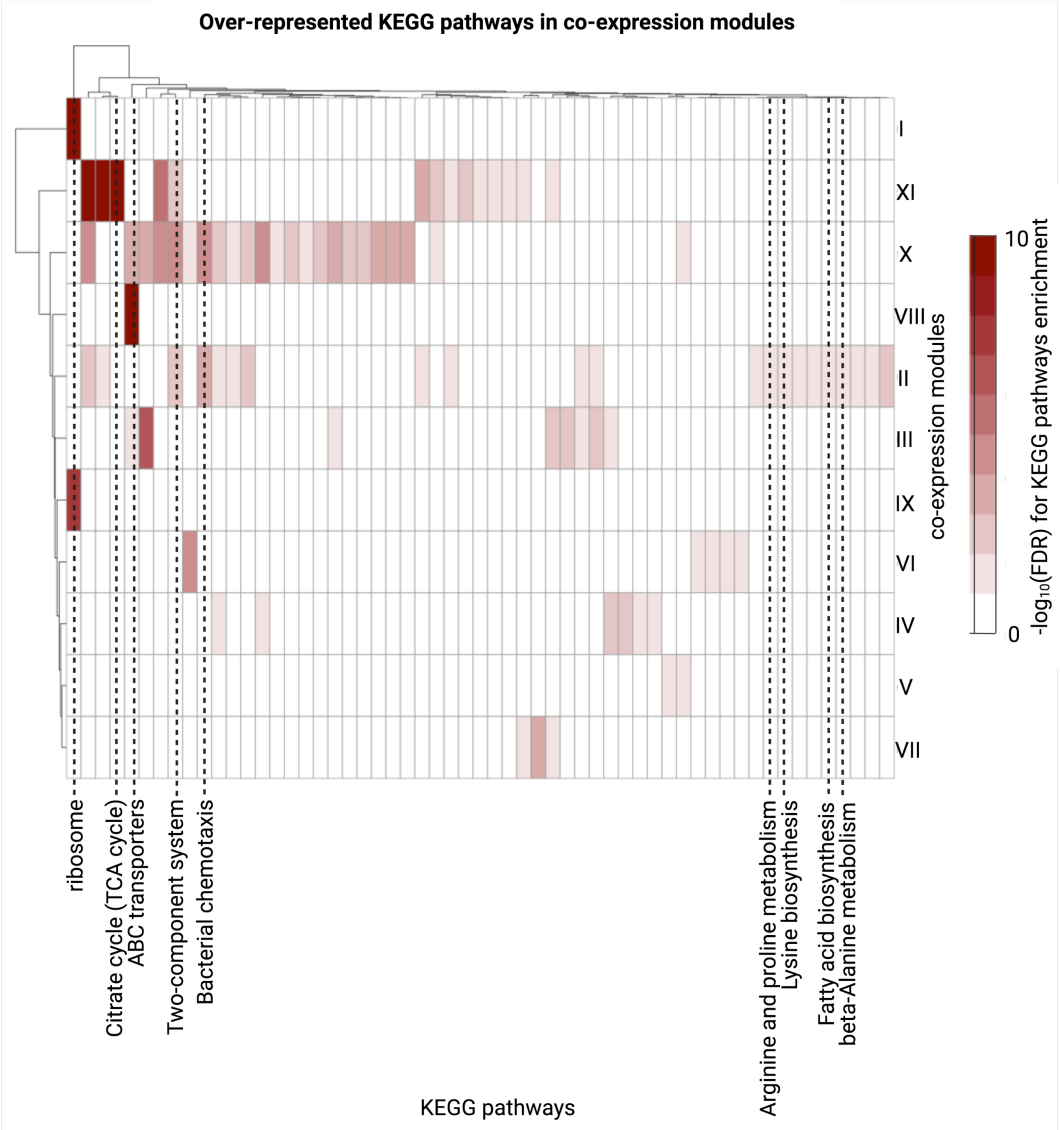

Figure 4.3: **Statistical enrichment of KEGG pathways across co-expression modules.** The colour scale showed the -$\log_{10}$(FDR) of KEGG pathways enrichment. FDR values that are statistically insignificant (> 0.05) are coloured white.

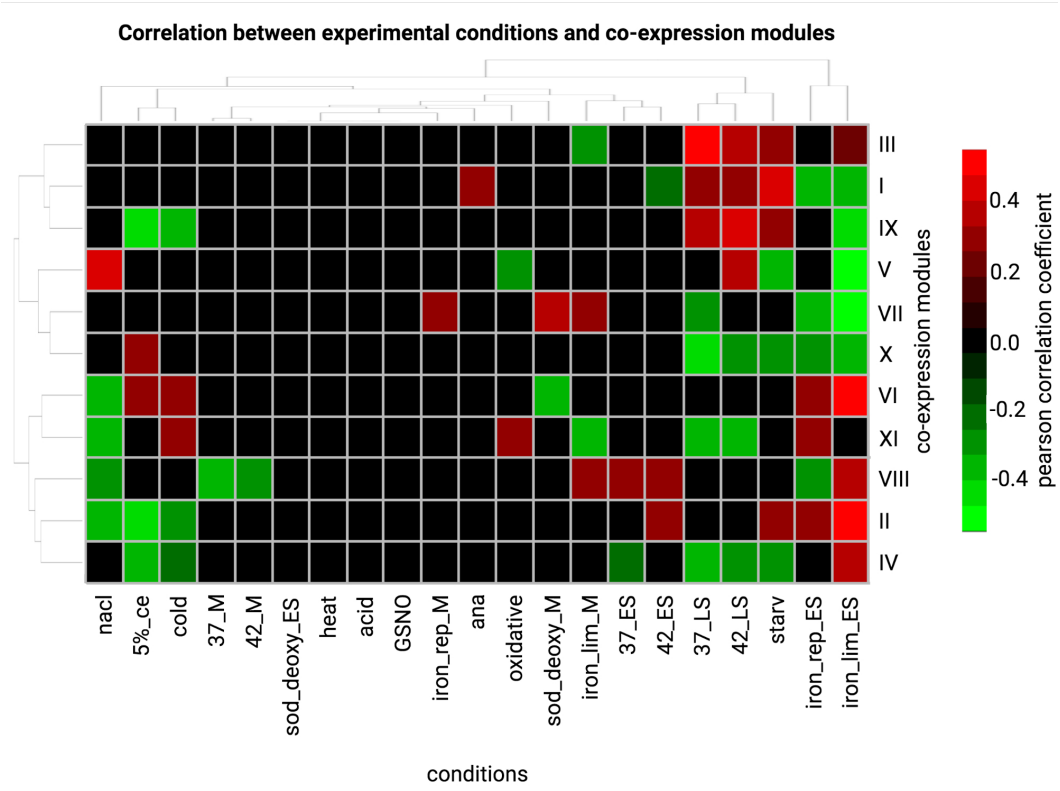

Figure 4.4: **The Pearson correlation between co-expression modules eigengenes and experimental traits.** The eigengene is defined as the first principal component of the gene expression matrix of the respective co-expression module. For correlation coefficients with p-value above 0.05, their boxes are coloured as black.

pathways varied among experimental conditions. Pairwise comparisons related to stationary phases and food storage conditions yielded more enriched pathways than other conditions. For example, the pairwise comparison with the most enriched pathways was with iron limitation under the early-stationary phase (iron_lim_ES) and iron limitation under exponential phase (iron_lim_M) as the control. In contrast, several pairwise comparisons with the fewest enriched KEGG pathways included heat stress (heat) and acid stress (acid), with exponential phase at 37 °C as their controls (Figure 4.5).

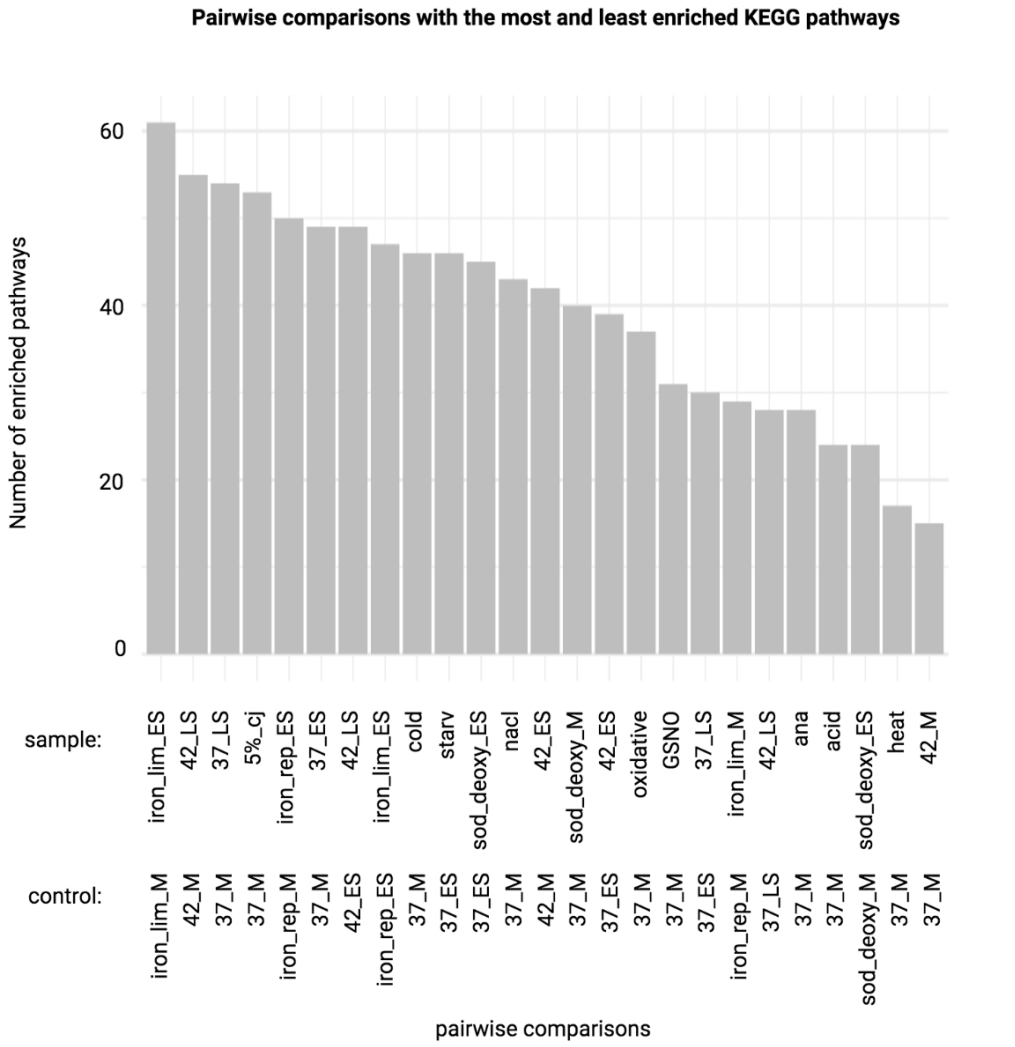

Figure 4.5: **The number of statistically enriched KEGG pathways among pairwise comparisons.** The threshold of statistical threshold is FDR < 0.05.

Another question is which biological pathways are more involved in stress adaptation. This question was addressed by counting the pairwise comparisons that each KEGG pathway exhibited differential enrichment. Some pathways, including "Biosynthesis of amino acids", "ABC transporters" and "Two-component system", were statistically enriched in most pairwise comparisons. In other words, these pathways played a more crucial role in stress adaptations throughout the transmission cycle. In contrast, pathways such as "beta-Lactam resistance" and "Tyrosine metabolism" displayed statistical enrichment in only one or two pairwise comparisons. Hence these pathways may only be involved in adjusting to a few specific environments (Figure 4.6).

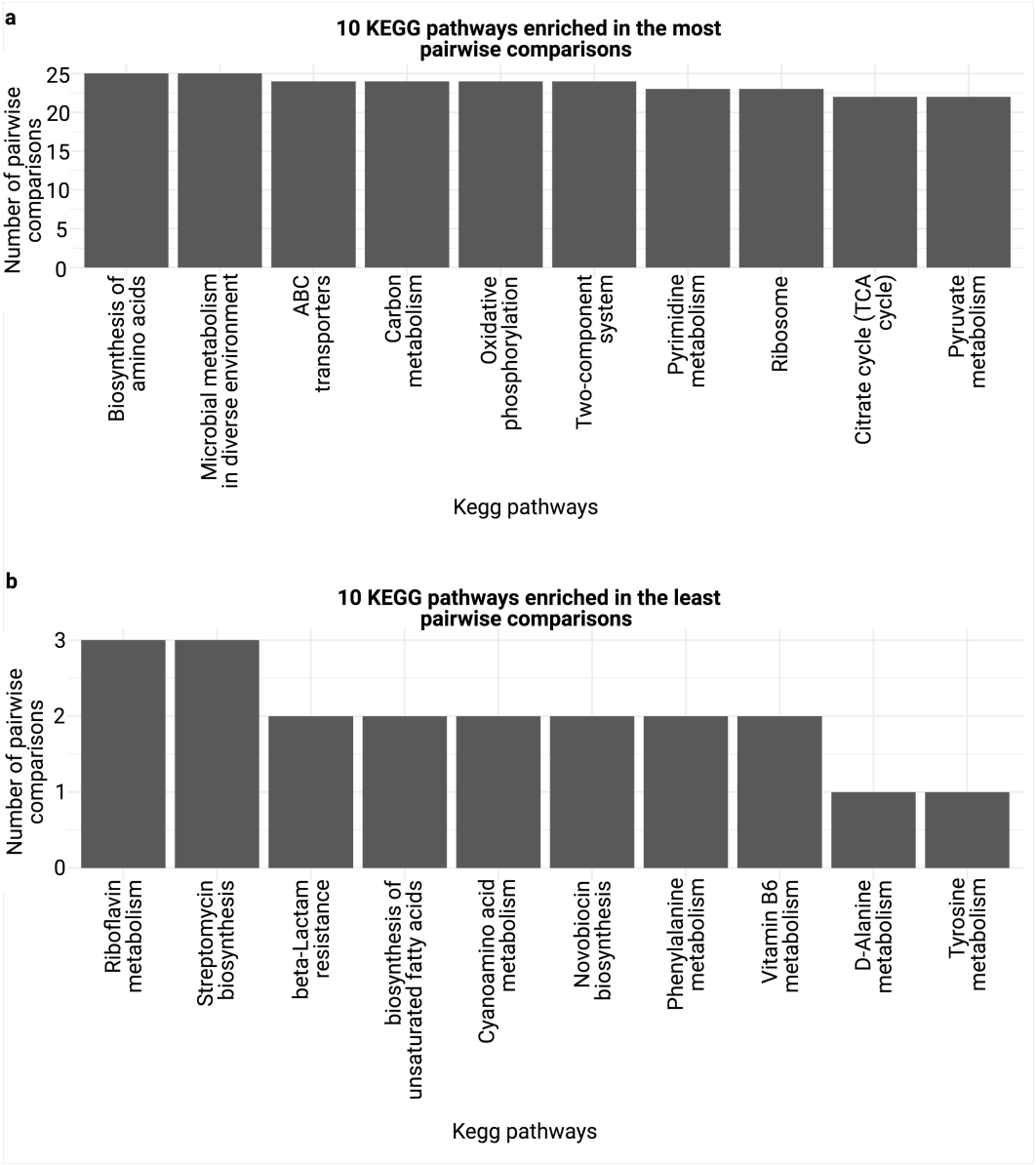

Figure 4.6: **The frequency of KEGG pathway enrichment among RNAtag-seq conditions.** (a) The ten most and (b) least frequently enriched KEGG pathways among all pairwise comparisons.

4.2.2 Overview of sRNA-target network

WGCNA and DESeq2 identified the overall and condition-specific expression patterns, highlighting genes, pathways, and conditions involved in adapting to various environmental stresses. They also helped identify sRNA-target pairs that showed significant co-expression and differential expression in particular conditions. However, such a pattern might result from either direct binding interactions or co-regulation. Therefore, results from co-expression and differential expression were integrated with genome-wide target prediction by IntaRNA to extract sRNA-target pairs that may form direct RNA-RNA interactions. IntaRNA predicts binding energy values between each sRNA-target combination, based on the energy required for unfolding secondary structures and the base-pairing stability. The distribution of the binding energy values between each sRNA and all genes were used to computed the p-values and FDR. That means stable interactions were more likely to display lower p-values and FDR. sRNA-target interactions with p-values $<= 0.05$ and showing significant co-expression and differential expression patterns were rated as high-confidence sRNA-target interactions (Figure 4.8a). As a result, the integration of these three approaches predicted 513 sRNA-target interactions (Figure 4.7).

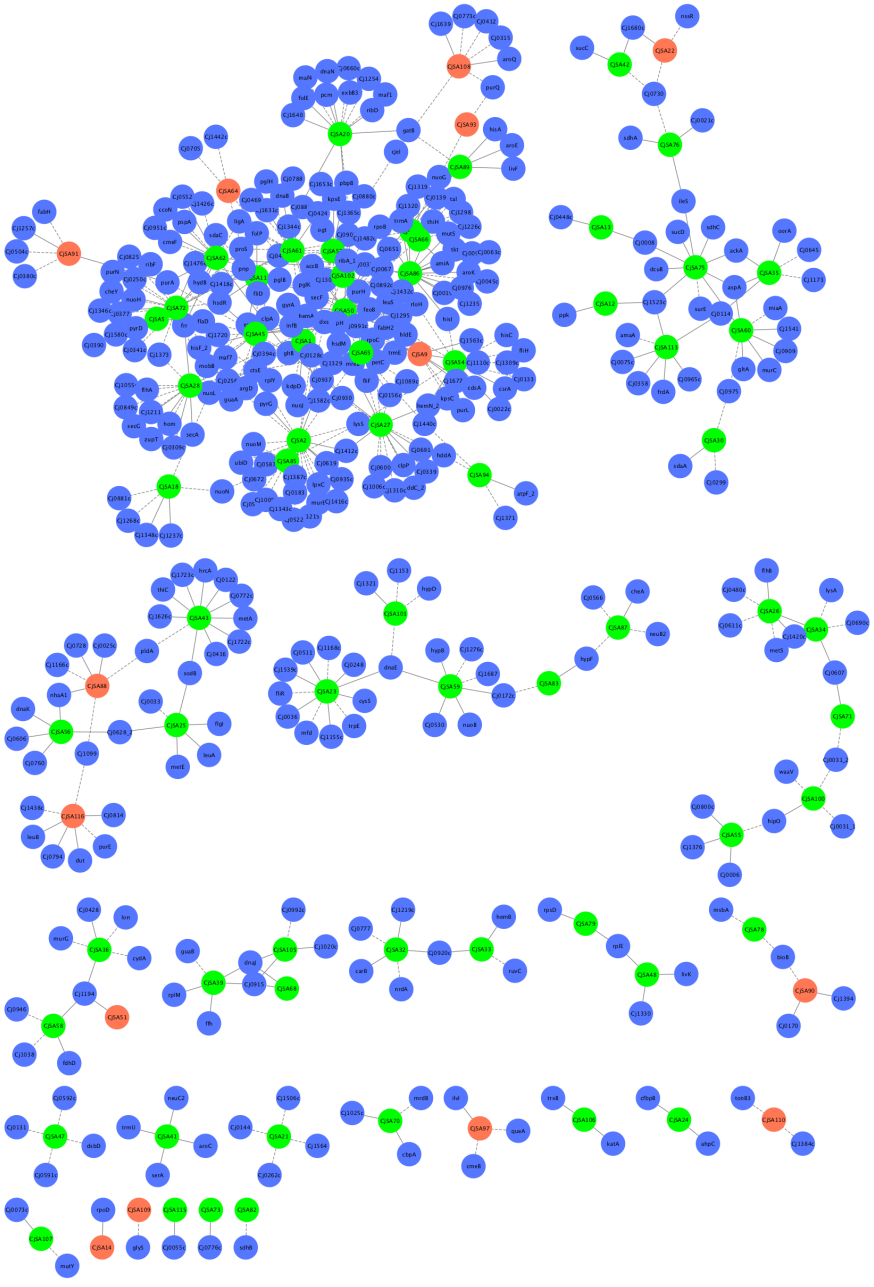

Figure 4.7: **All sRNA-target interactions extracted after integrating WGCNA, DESeq2 and IntaRNA.** Dashed lines and solid lines indicate sRNA-target pairs with negative and positive co-expression. Novel sRNAs, annotated sRNAs and mRNAs are coloured in green, orange and purple, respectively.

Table 4.3: **All sRNA with more than 3 targets from a statistically enriched KEGG pathway.** Only KEGG pathways with less than 100 proteins are shown here in order to focus on pathways with specific biological functions.

Sample	Targets	KEGG pathways	Hits	FDR
CjSA21	4	Bacterial chemotaxis	4	5.08E-8
CjSA21	4	Two-component system	4	4.71E-7
CjSA28	14	Bacterial secretion system	3	2.25E-4
CjSA53	19	Purine metabolism	4	3.58E-3
CjSA53	19	Pyrimidine metabolism	3	1.09E-2
CjSA9	10	Purine metabolism	3	3.76E-3

To further extract sRNA-target interactions that may play an influential regulatory role, KEGG pathways enrichment analysis was conducted on the targets of each sRNA to identify sRNAs that bind explicitly to targets from the same biological pathways. This approach is similar to published tools for analysing RNA-RNA interactomes such as rNAV 2.0, which also searches for sRNA targets statistically enriched for GO, KEGG and UniProt (Bourqui et al., 2017). Remarkably, CjSA21's potential targets were all involved in "Bacterial chemotaxis" and "Two-component system", leading to the lowest FDR values among all predicted sRNA-target interactions. Targets of CjSA28, CjSA53 and CjSA9 were also significantly enriched in "Bacterial secretion system", "Purine metabolism" and "Pyrimidine metabolism". However, only 3 out of 14 CjSA28 targets belonged to the "Bacterial secretion system". Likewise, out of 19 targets of CjSA53, only 4 or less of them belonged to "Purine metabolism" and "Pyrimidine metabolism". Similarly, only 3 of the 10 CjSA9 targets were from "Purine metabolism" (Table 4.3). As a result, they were less statistically significant than CjSA21 targets.

Interestingly, upon setting FDR <= 0.05 as the threshold for IntaRNA prediction, CjSA21 targets were the only ones that showed KEGG pathway enrichment. CjSA21 targets in "Bacterial Chemotaxis" and "Two-component system" had FDR values of 5.65E-10 and 1.16E-08, respectively. That suggested CjSA21 also formed the most stable binding interactions across the entire predicted sRNA-target network. Both IntaRNA prediction (with FDR <= 0.05 as the threshold) and the combined output from WGCNA, DESeq2 and IntaRNA had highlighted CjSA21 as a pathway-specific regulator for signal transduction and chemotaxis. Hence the following section will focus on understanding the regulatory role of CjSA21.

4.3 CjSA21 and chemosensing genes

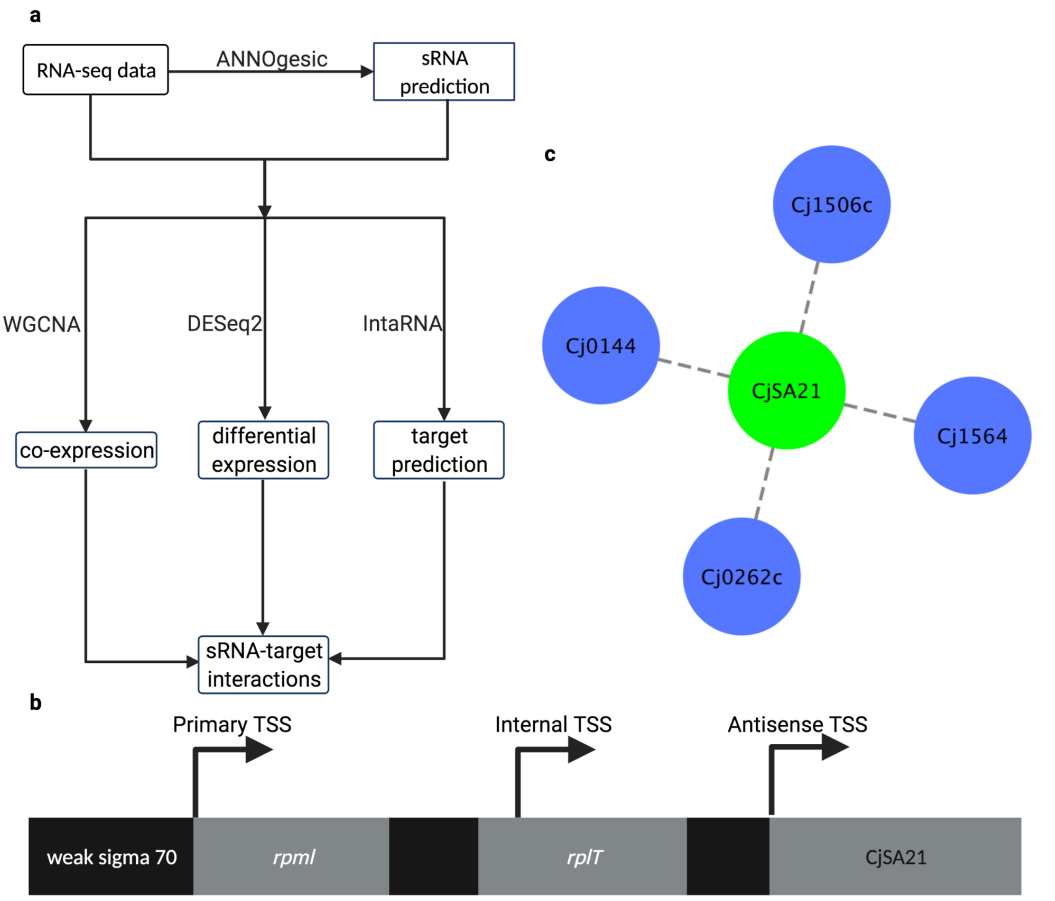

Figure 4.8: **Summary of integrating WGCNA, DESeq2 and IntaRNA.** (a) The integration of co-expression analysis, differential expression analysis and genome-wide sRNA-target prediction to predict sRNA-target interactions. Predicted sRNA-target interactions need to be clustered into the same WGCNA co-expression modules and are connected by an edge in the WGCNA network. Those gene pairs with a statistically significant correlation (FDR < 0.05) will be filtered out. Moreover, genes with negative co-expression need to show opposite differential expression (one gene being differentially upregulated, one gene being differentially downregulated) in at least two pairwise comparisons. Likewise, genes with positive co-expression are required to show matching differential expression (both genes being differentially upregulated or differentially downregulated together) in at least two pairwise comparisons. Last but not least, their predicted binding energy values must result in p-value $<= 0.05$. (b) Transcription start sites and annotated genes in the proximity of CjSA21. (c) Highlighted sRNA-target interactions between CjSA21 and *tlp1-4*. They are connected by dashed lines, which indicate negative co-expression.

The predicted targets of CjSA21 were Cj1506c, Cj0144, Cj1564 and Cj0262c, which correspond to *tlp1*, *tlp2*, *tlp3* and *tlp4*, respectively (Figure 4.8b). The negative co-expression between CjSA21 and *tlp1-4* suggested sRNA-mRNA interactions that reduced the mRNA abundance. Interestingly, *tlp1-4* were transcripts responsible for translating methyl-accepting chemotaxis proteins (MCPs), also known as transducer-like proteins (Tlps), which sense external stimuli and induce chemotaxis. Such an expression pattern suggests CjSA21 might be an inhibitor of chemosensing.

CjSA21 locates downstream of *rpmI* and *rplT*. CjSA21, *rpmI* and *rplT* appear to form an operon together. *rpmI* has a primary TSS that may regulate all three genes in the operon. Nevertheless, an internal TSS and a TSS antisense to Cj0246c are also present inside the operon (Figure 4.8c). The sequencing coverage of CjSA21, *rpmI* and *rplT* (Figure 4.9) showed a more significant portion of reads aligned to CjSA21 under NaCl stress (nacl) when compared to the standard laboratory condition (37_M) (Figure 4.9). The antisense TSS might also influence CjSA21 expression under specific conditions based on the expression pattern.

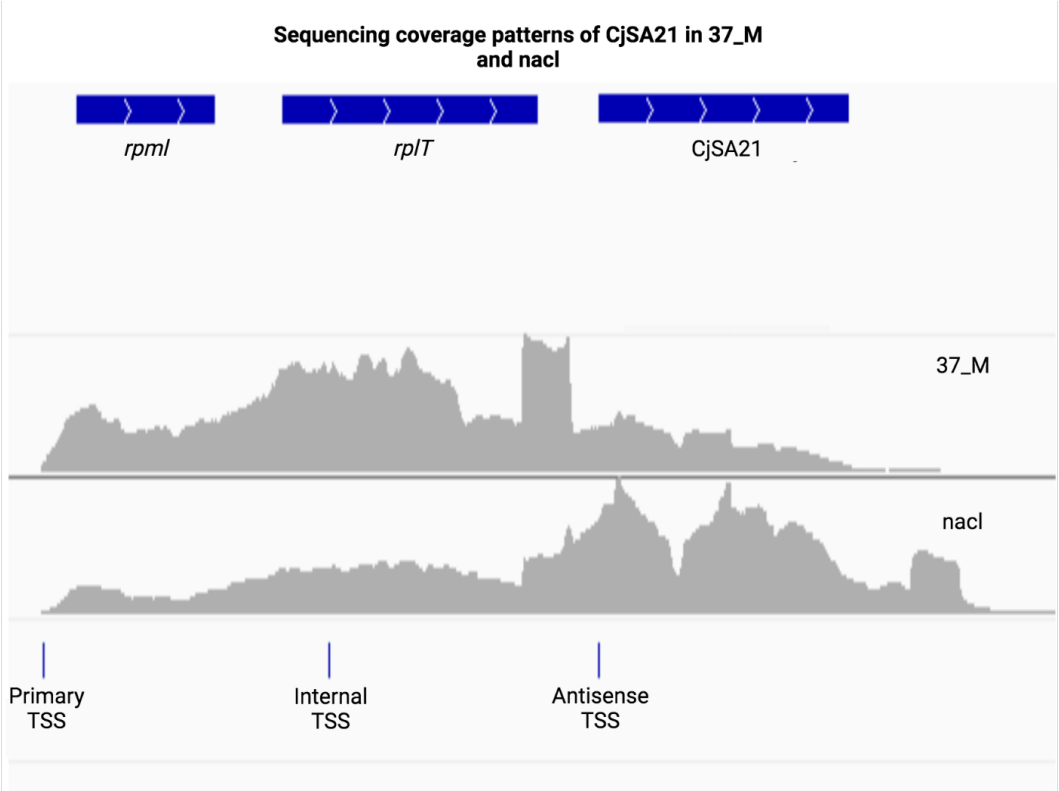

Figure 4.9: **Sense strand sequencing coverage of CjSA21.** The expression coverage is from the standard laboratory condition (37_M) and hyperosmotic stress (nacl) suggested condition-specific transcriptional activity of CjSA21.

4.3.1 CjSA21 belongs to the same co-expression modules as chemotaxis genes

CjSA21 and *tlp1-4* all belong to module II. As mentioned previously, module II is correlated to pathways related to signal transduction, chemotaxis and metabolite utilisation. As mentioned previously, the two-component system is enriched among differentially expressed genes in almost every selected pairwise comparison. Hence module II may provide valuable information on *C. jejuni* stress adaptation.

Gene expression of module II is negatively correlated with food storage conditions, including hyperosmotic stress (nacl), 5% chicken exudate (5%_ce) and cold stress (cold). Module II expression is also positively correlated with nutrient deprivation conditions, such as starvation stress (starv) and those related to the early stationary phase. Those conditions were also conducted under 42 °C (42_ES), iron-limitation (iron_lim_ES) and iron-repletion (iron_rep_ES) (Figure 4.10a and 4.4).

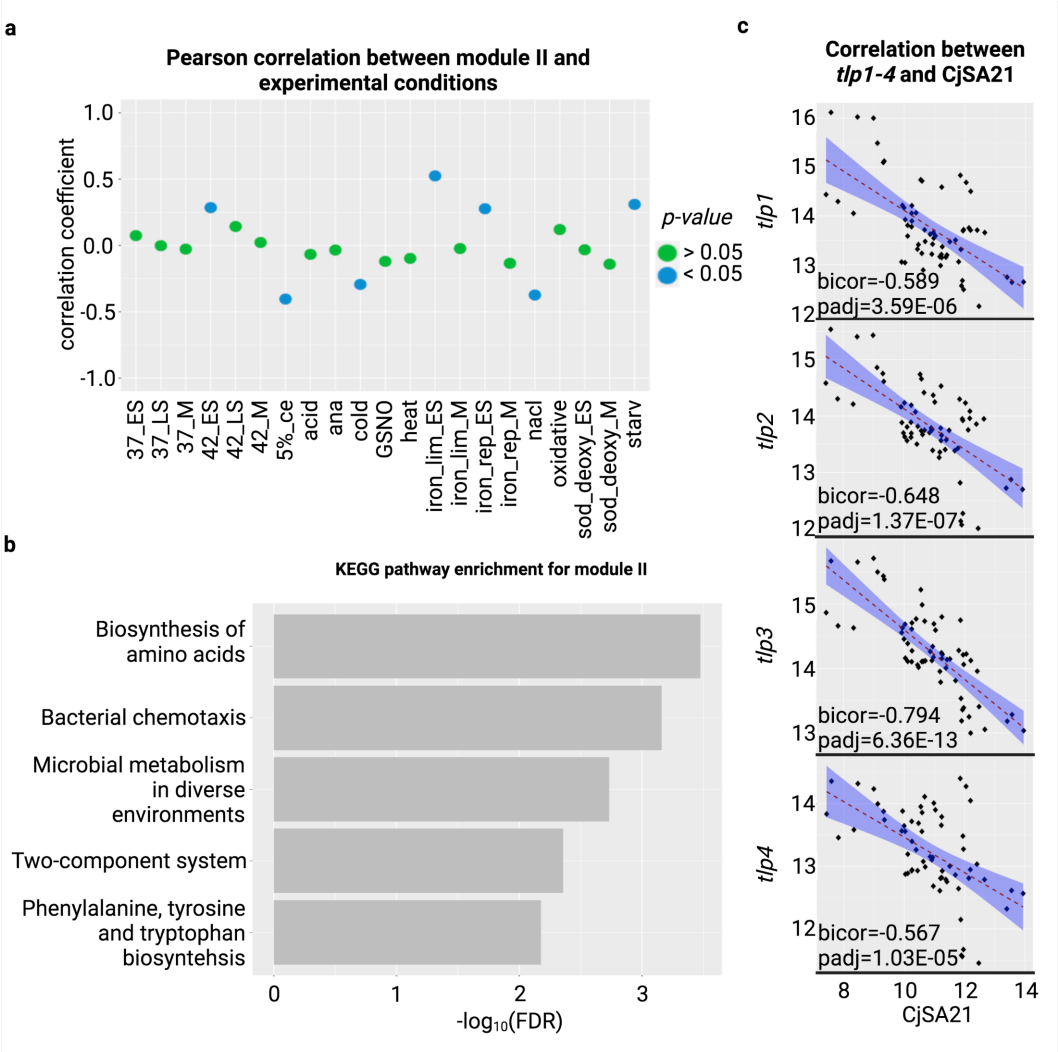

Figure 4.10: **Results of CjSA21 co-expression analysis.** (a) The Pearson correlation coefficient and p-values between the eigengene of modules II and experimental traits. (b) KEGG Pathways enrichment on module II. Only the five pathways with the lowest FDR are shown here. (c) The correlation between CjSA21 and *tlp1-4* across all 63 RNA-Seq replicates.

Due to the presence of *tlp1-4*, module II showed statistical enrichment for KEGG pathways "Bacterial chemotaxis" and "Two-component system". The enrichment of these pathways can also be attributed to the presence of *cheA* and *cheV* in module II. Other KEGG pathways enriched in module II include biosynthetic pathways for amino acids and fatty acids (See Figure 4.10b and 4.11). Hence, module II may illustrate the connection between sensing, migrating and metabolising various metabolites. As WGCNA constructed an unsigned network, each co-expression module consisted of genes that showed significant positive or negative correlations. For instance, the expression of CjSA21 and *tlp1-4* demonstrated moderate to substantial negative correlation, with bicor coefficients between -0.567 and -0.794. All negative correlations led to significant adjusted p-values (padj) less than 0.05 (Figure 4.10c). That suggests CjSA21 may induce degradation of *tlp1-4*.

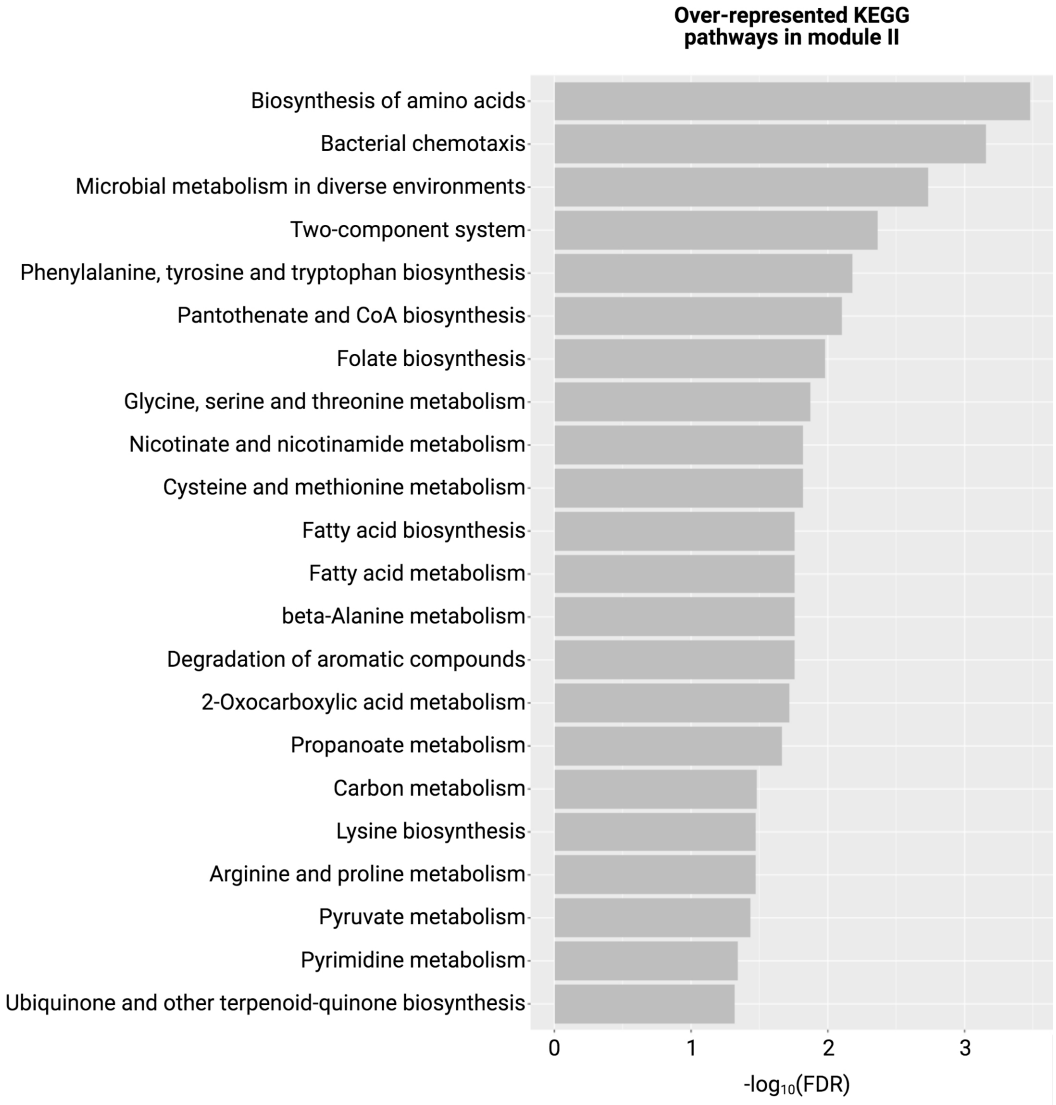

Figure 4.11: **KEGG Pathways enrichment analysis of module II.** All KEGG Pathways with FDR <= 0.05) of module II and less than 100 proteins are shown here.

Genes within module II were further separated into two clusters with opposite expression patterns. The separation assigned *tlp1-4* into cluster 1, alongside other genes from several other pathways, including "Propanoate metabolism", "Lysine biosynthesis" and "Cysteine and methionine metabolism". Genes in cluster 1 showed increased expression levels under the early-stationary phase and other starvation conditions. Hence *tlp1-4* and other metabolic pathways in cluster 1 are responsible for the positive correlation with nutrient deprivation conditions (Figure 4.12 and 4.13).

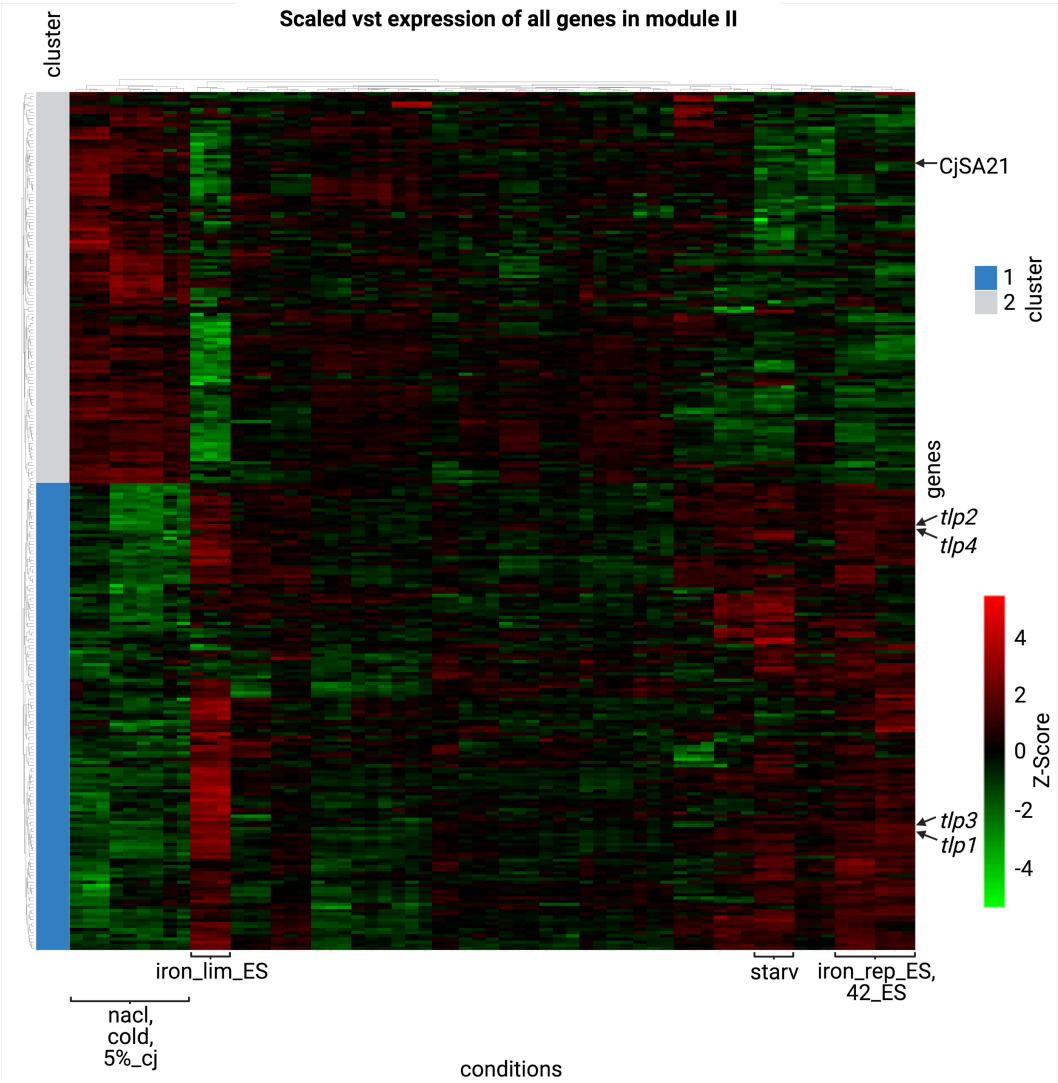

Figure 4.12: **The scaled vst score and complete linkage hierarchical clustering of module II.** The heat map is scaled by rows.

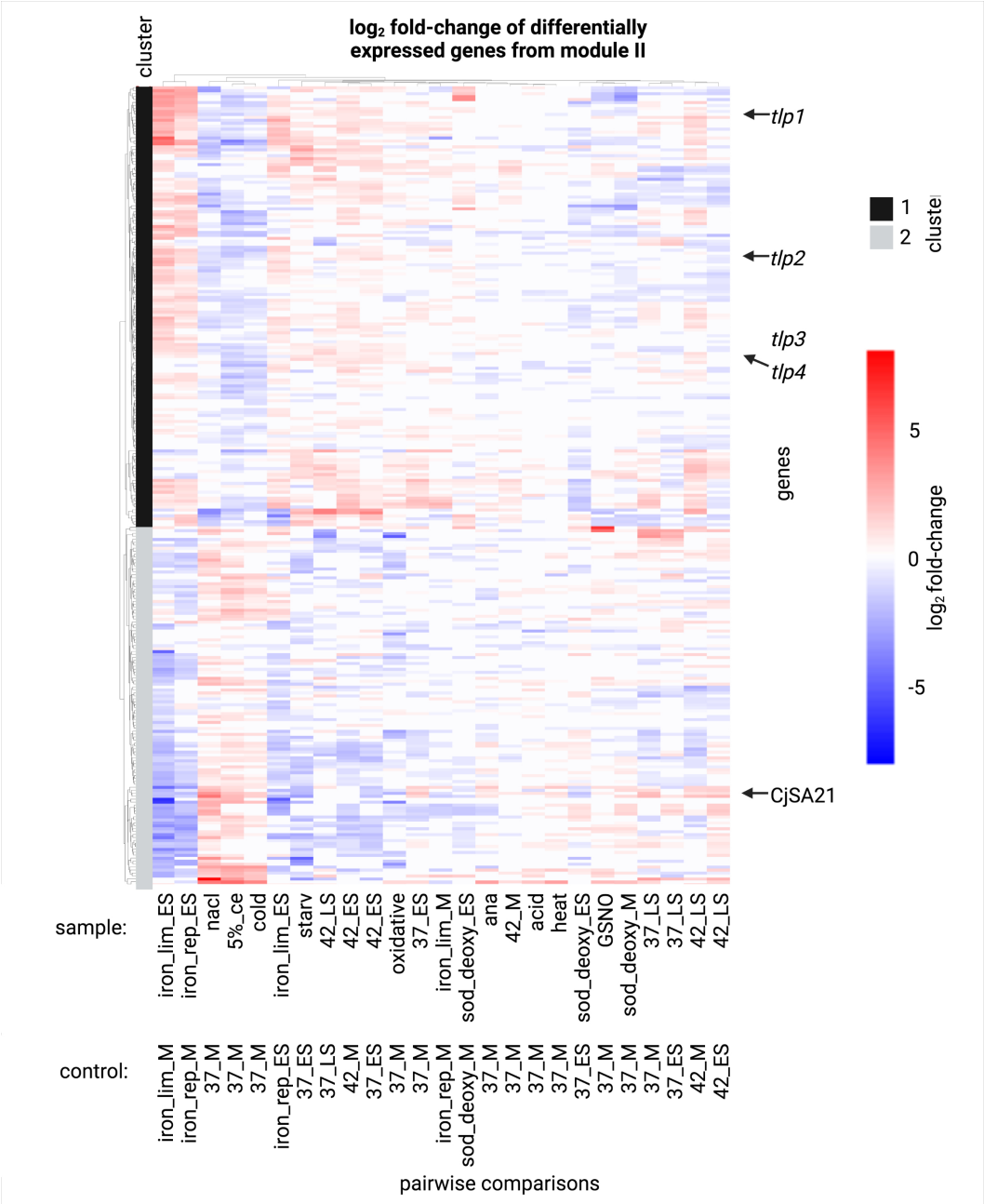

Figure 4.13: **Differential expression of module II genes.** The blue and red boxes showed the $\log_2$ fold-change of differentially expressed genes in module II. White boxes indicate genes with insignificant FDR (> 0.05).

Cluster 2 genes, including CjSA21, showed high expression values and differential upregulation under food storage conditions (cold, 5%_ce and nacl). Apart from CjSA21, cluster 2 also included genes from the Tryptophan biosynthesis (Trp) operon, which resulted in the enrichment of the KEGG pathways "Phenylalanine, tyrosine and tryptophan biosynthesis" and "Glycine, serine and threonine metabolism". The Trp operon was also differentially upregulated in other conditions like heat stress (heat), anaerobic stress (ana) and nitrosative stress (GSNO).

4.3.2 CjSA21 showed an opposite differential expression pattern to *tlp*1-4

Co-expression analyses have shown an overall negative correlation between CjSA21 and *tlp1-4* RNA abundance. Subsequent differential expression analysis was required to understand the condition-specific regulation by CjSA21. Since CjSA21 and *tlp1-4* displayed negative co-expression, the desired conditions were those where CjSA21 and *tlp1-4* show contrasting differential expression across all selected pariwise comparisons (Figure 4.14).

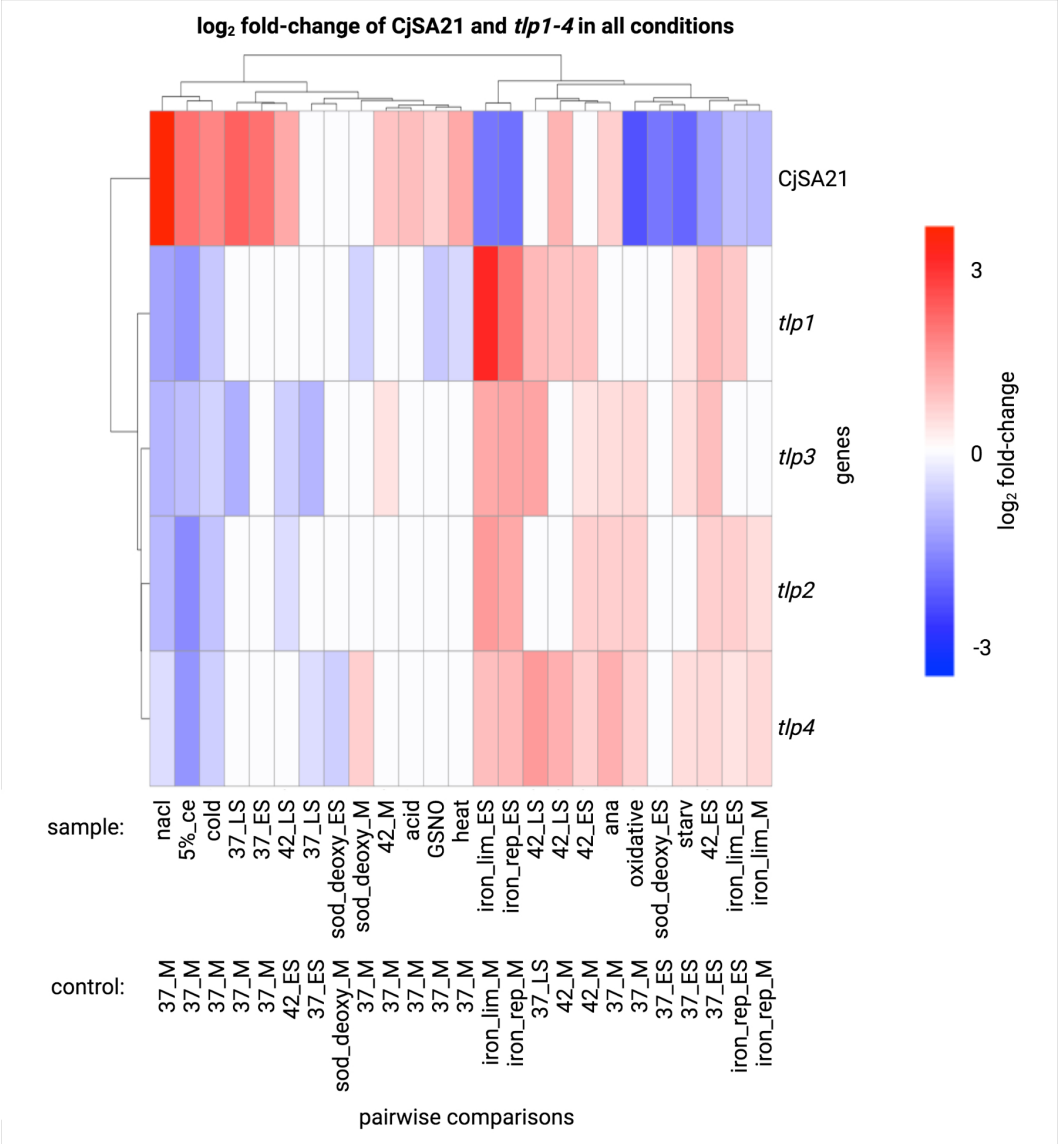

Figure 4.14: **$\log_2$ fold-change of CjSA21 and *tlp*1-4 across all selected pairwise comparisons.** White boxes indicate FDR > 0.05.

In some conditions, CjSA21 exhibited opposite differential expression with some but not all of *tlp1-4*. Notably, an opposite differential expression pattern occurred under the three food storage conditions (cold, 5%_ce and nacl), with the standard laboratory condition (37_M) as their controls. Under these conditions, all four *tlp* genes exhibited significant downregulation, while CjSA21 showed differential upregulation (Figure 4.15a). Among the food processing conditions, DEseq normalised expression of CjSA21 was the most upregulated under hyperosmotic stress (nacl) (Figure 4.15b and 4.16).

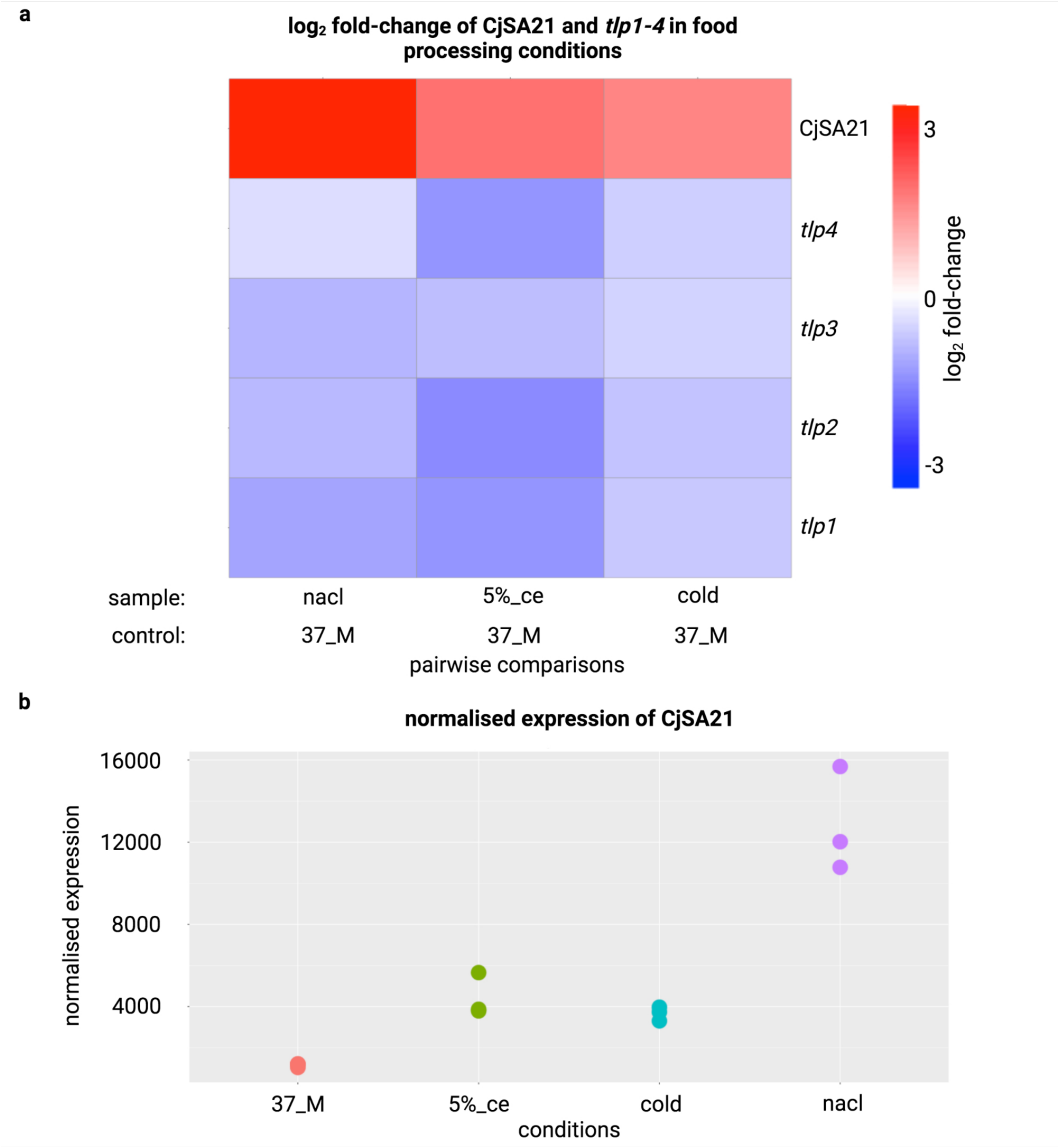

Figure 4.15: **Expression of CjSA21 and *tlp1-4* in food storage conditions.** (a) $\log_2$ fold-change of CjSA21 and *tlp1-4*. The red and blue boxes indicate differential upregulation and differential downregulation respectively. (b) Normalised expression levels of CjSA21 in food storage conditions (cold, nacl and 5%_ce) and the standard laboratory condition (37_M). All expression values were normalised by DESeq2's median of ratios.

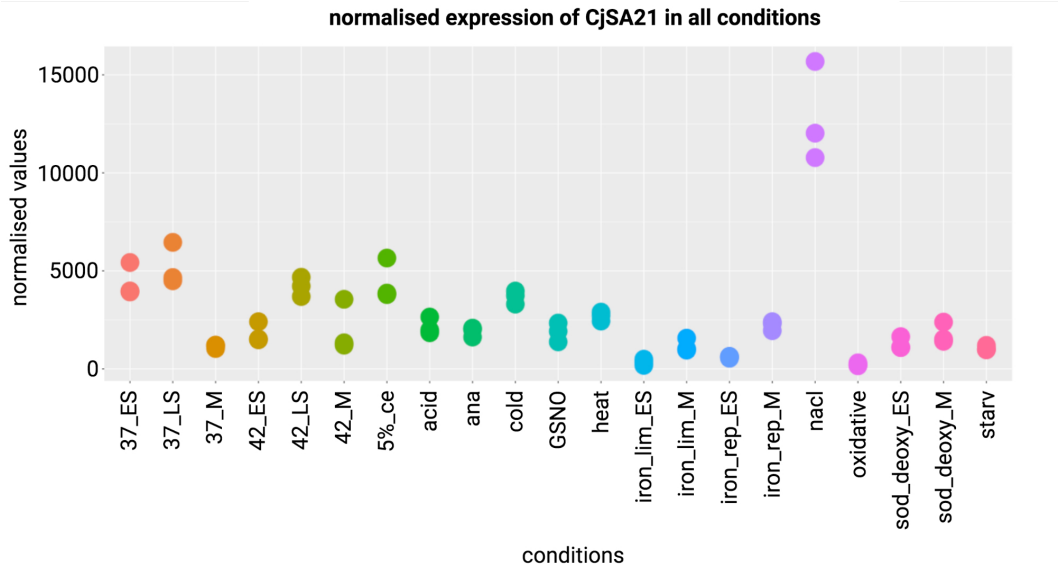

Figure 4.16: **The normalised expression of CjSA21 across 21 experimental conditions.** All expression values were normalised using the DESeq2's median of ratios method.

Unsurprisingly, the "Bacterial chemotaxis" and "Two-component system" pathways displayed statistical enrichment among differentially downregulated genes in the above pairwise comparisons (Figure 4.17a). That further suggests the possibility of CjSA21 inhibiting chemosensing under food storage conditions. Unexpectedly, the "Flagellar assembly" pathway demonstrated statistical over-representation under all three pairwise conditions (Figure 4.17b). That suggests that *C. jejuni* synthesised more flagella under these conditions despite being less sensitive to chemotaxis.

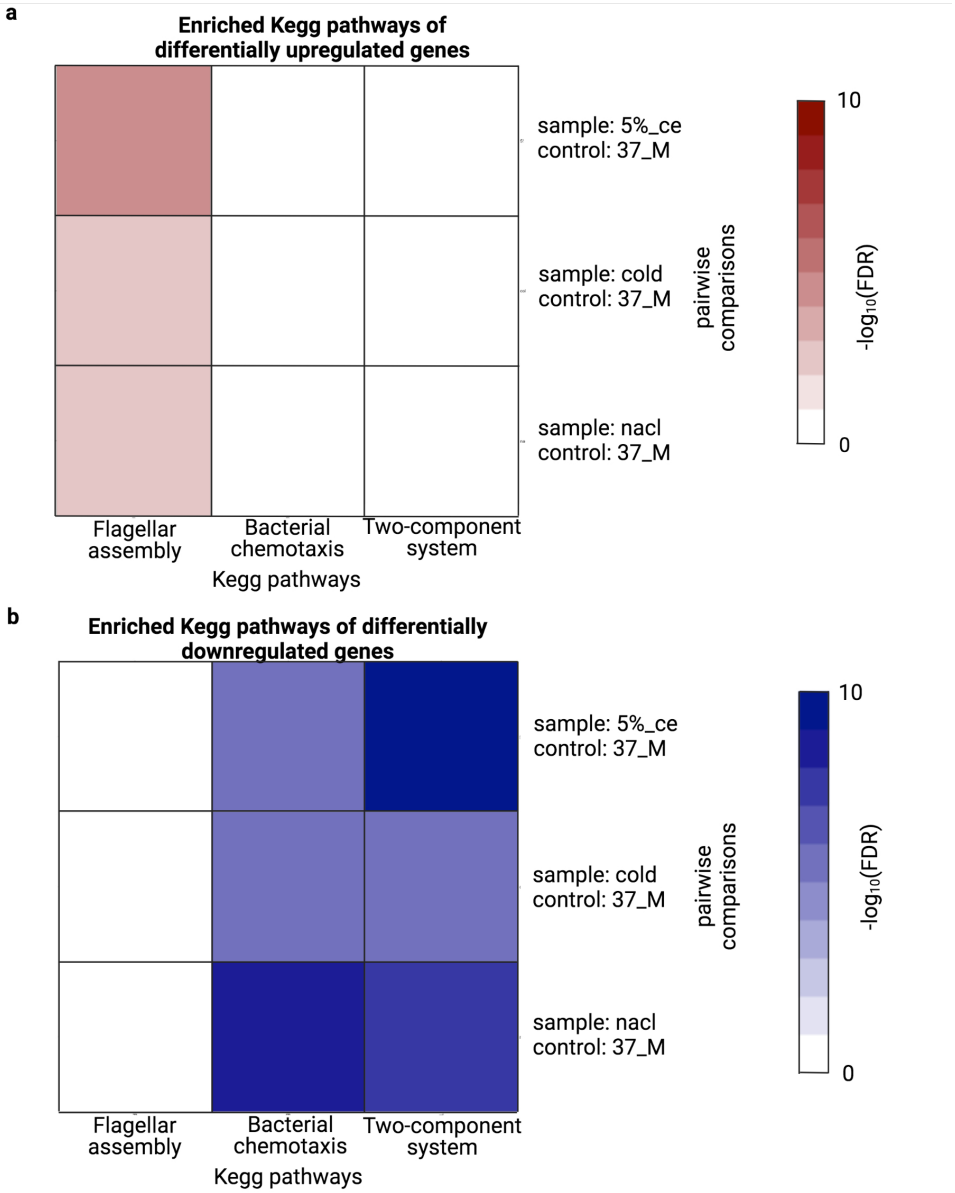

Figure 4.17: **Enrichment of the KEGG pathway "Bacterial Chemotaxis", "Flagellar assembly" and "Two-component system" under food storage conditions.** (a) Statistical enrichment among upreulgated genes. (b) Statistical enrichment among downreulgated genes. White boxes indicate FDR > 0.05.

Similar opposite differential expression pattern was observed under early-stationary phase at 42 °C, iron-limitation and iron-repletion (42°C, iron_lim_ES and iron_rep_ES). Under these conditions, CjSA21 was downregulated while *tlp1-4* was upregulated (Figure 4.14).

4.3.3 IntaRNA predicted significant binding interactions between CjSA21 and *tlps*

IntaRNA calculated the theoretical binding energy values were between CjSA21 and all possible targets, including annotated genes and predicted sRNAs. Among them, 29 targets have predicted *p*-values <= 0.05 (Figure 4.18a). Most of their energy values range between -20 to -40 kcal/mol. Interestingly, the binding energy values between CjSA21 and *tlp1-4* were around -45 kcal/mol, leading to significant FDR (<= 0.05). Other targets with significant FDR were *cetA* and Cj0019c (Table 4.4). Interestingly, these two genes are *tlp9* and *tlp10*, respectively, which are also chemosensors that detect different subtracts from *tlp1-4*. The predicted energy values between CjSA21 and *tlp1-4* and *tlp9-10* were remarkably stable compared to other CjSA21-mRNA interactions, as most binding energy values were > -20 kcal/mol (Figure 4.18b).

Table 4.4: **All IntaRNA-predicted targets of CjSA21**. The FDR threshold is 0.05.

gene id	**gene name**	**binding energies (kcal/mol)**	***p*-value**	**FDR**
Cj0144	*tlp2*	-44.91	1.97E-05	5.71E-03
Cj0262c	*tlp4*	-44.91	1.97E-05	5.71E-03
Cj1506c	*tlp1*	-45.69	1.52E-05	5.71E-03
Cj1564	*tlp3*	-44.91	1.97E-05	5.71E-03
cetA	*tlp9*	-47.73	7.70E-06	5.71E-03
Cj0019c	*tlp10*	-43.2	3.47E-05	8.62E-03

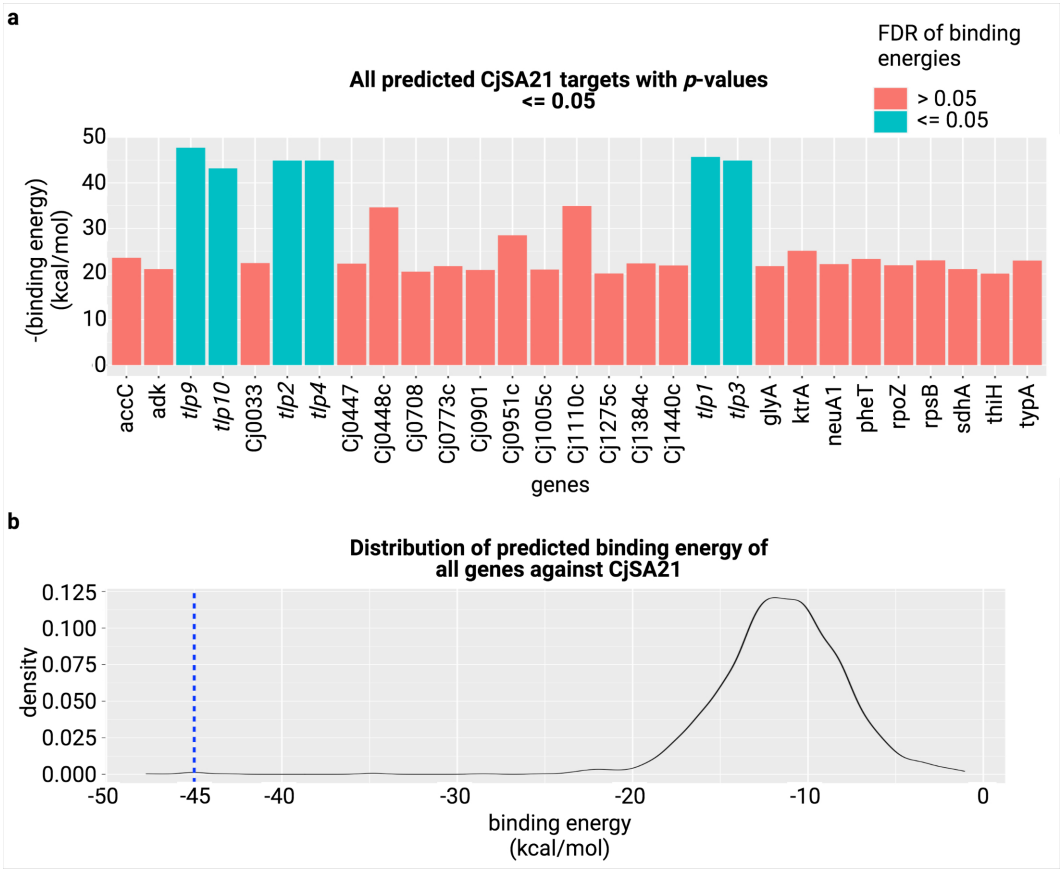

Figure 4.18: **Results of CjSA21 genome-wide target prediction**. (a) IntaRNA-predicted energy values of all CjSA21 targets. Only those with p-values <= 0.05 are shown here. (b) The distribution of IntaRNA-predicted energy values of all CjSA21-target pairs. The blue line indicates -45 kcal/mol (the approximate predicted binding energies between CjSA21 and *tlp1-4*).

IntaRNA predicted that CjSA21 used its 3' end to target *tlp1-4* and *tlp9-10*. In particular, CjSA21 targets *tlp2*, *tlp3* and *tlp4* with identical nucleotides. It also targets (*tlp10*) using a similar sequence, except missing the first two nucleotides (Figure 4.19a).

Secondary structure prediction by RNAfold further shed light on the binding mechanism of CjSA21. The structural prediction located the binding sequences within regions consisting of stem-loops and unpaired nucleotides, with relatively high nucleotide entropy and lower intramolecular base pairing affinities (Figure 4.19b - c). The results suggest CjSA21 may target *tlp1-4* using unpaired regions

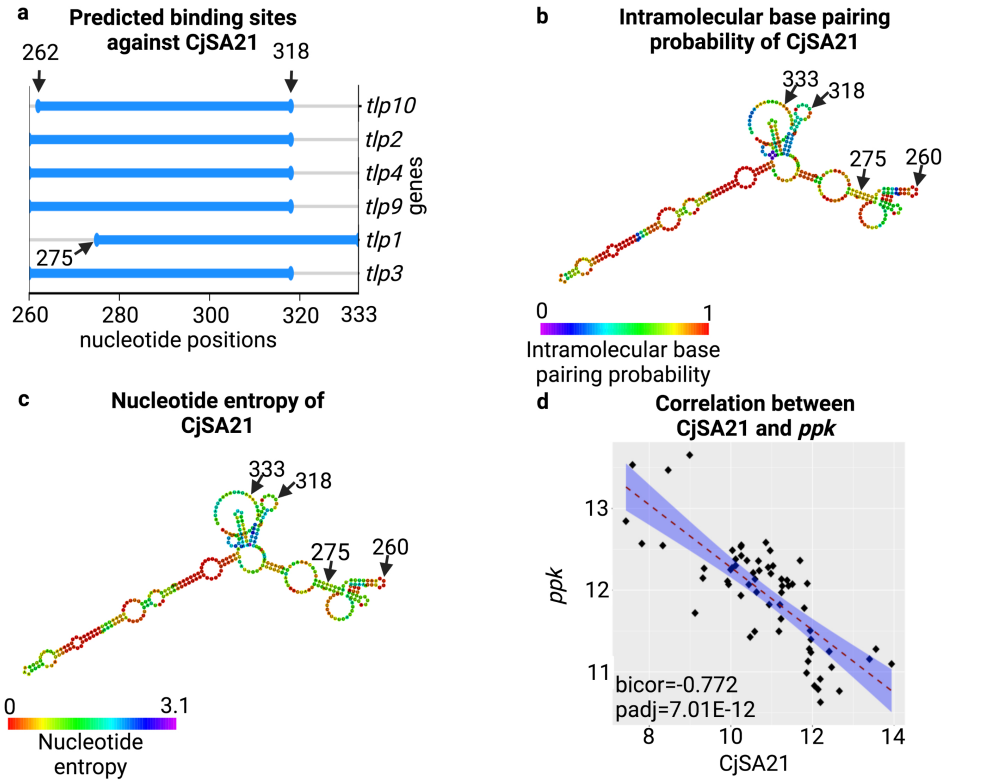

Figure 4.19: **Results of CjSA21 structural prediction.** (a)The distribution of predicted optimal binding sites used by CjSA21. (b - c) RNAfold predicted the secondary structure of CjSA21. The colour keys represent the intramolecular binding probabilities and nucleotide entropies, respectively. (d) Correlation between the expression values of *ppk* and CjSA21.

Despite showing highlighted complementary sequences with CjSA21, neither *tlp10* nor *tlp9* belongs to module II, and their correlation against CjSA21 was less significant (Figure 4.20). Both *tlp9* and *tlp10* also lack clear opposite differential expression patterns under food storage conditions. Both genes showed opposite differential expression with CjSA21 in one of three food processing conditions (Figure 4.21 and Figure 4.22). That suggests CjSA21 may regulate the translation of *tlp9-10* without affecting mRNA abundance.

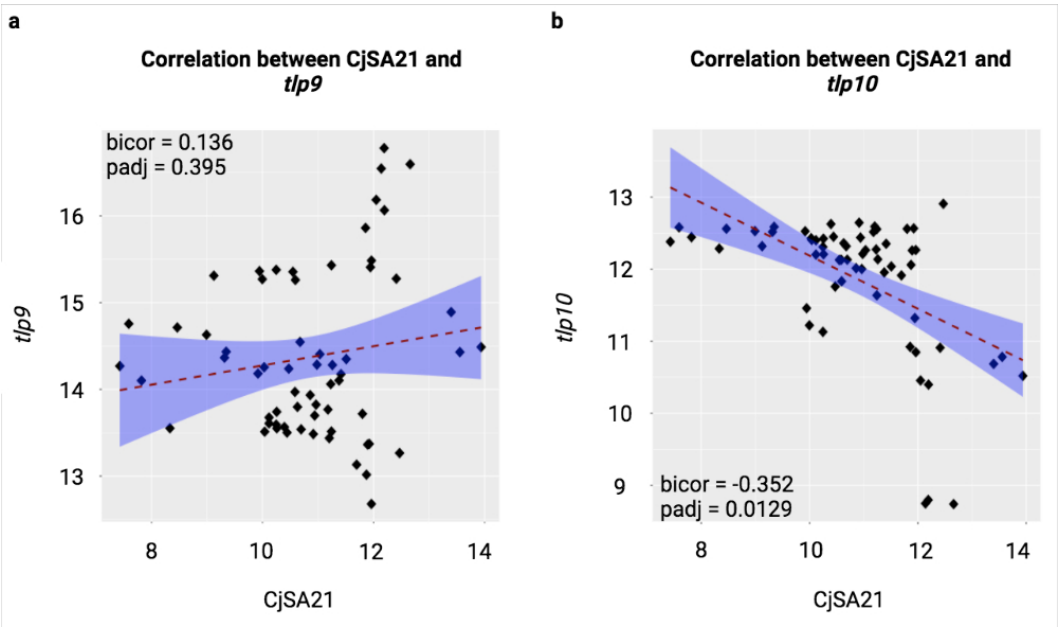

Figure 4.20: **Pairwise correlation between CjSA21 and *tlp9/tlp10*.** Scatter plots that illustrate the correlation between the expression of CjSA21 against (a) *tlp9* and (b) *tlp10*.

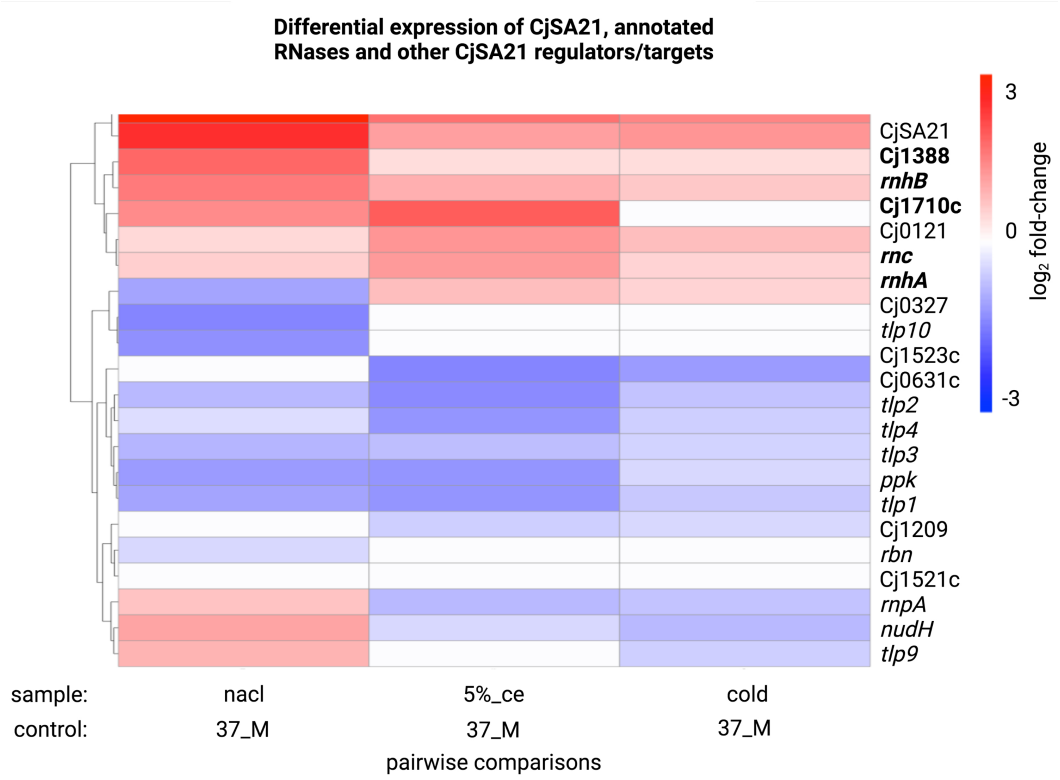

Figure 4.21: **$\log_2$ fold-change of CjSA21, *ppk*, *tlps* and all annotated RNases.** Only pairwise comparisons with 5%_ce, nacl and cold compared against 37_M are shown. RNases differentially expressed in all three conditions are highlighted in bold text.

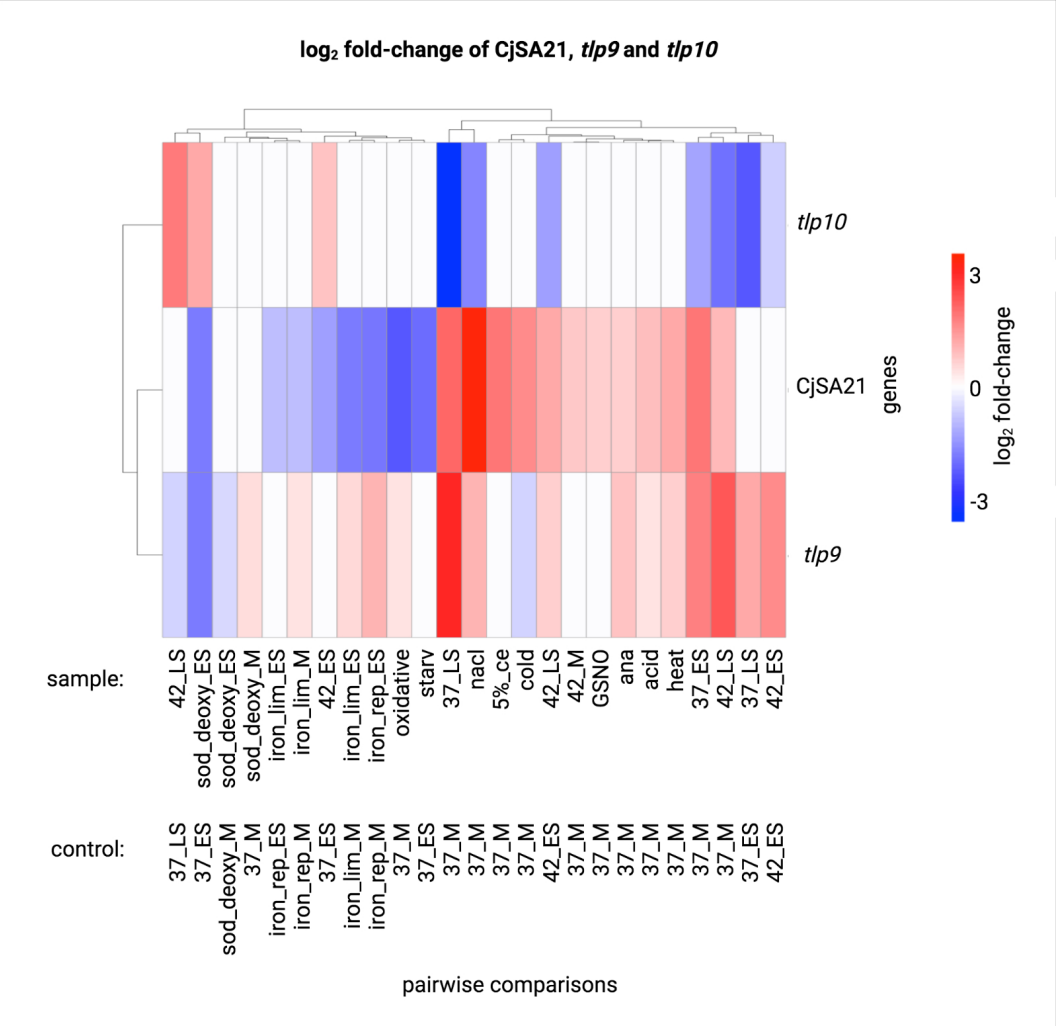

Figure 4.22: **Differential expression between CjSA21 and *tlp9*/*tlp10*.** All $\log_2$ fold-change of CjSA21, *tlp9* and *tlp10*. White boxes indicate FDR > 0.05.

4.3.4 Regulators and RNases that may mediate CjSA21-*tlp*1-4 interactions

In order to further understand the CjSA21 regulatory network, we asked the questions on which other factors promote survival under food storage conditions. One potential candidate for promoting survival under food storage conditions is the Polyphosphate kinase. Studies on *ppk* showed that it is involved in osmotic stress survival and biofilm formation (Candon et al., 2007). If *ppk* is involved in the CjSA21 regulatory network, *ppk* and CjSA21 may display significant co-expression and/or sequence complementarities. The co-expression analysis in this chapter suggested strong negative co-expression between *ppk* and CjSA21 (bicor=-0.792). CjSA21 and *ppk* also revealed opposite differential expression patterns under food storage conditions (Figure 4.19d and 4.21). However, IntaRNA-predicted binding energy between *ppk* and CjSA21 was statistically insignificant, with a p-value above 0.05. That suggests co-regulation between *ppk* and CjSA21 instead of direct binding interactions. Another possibility is that *ppk* inhibit the expression of CjSA21.

Other candidates for further investigations are RNases. Assuming CjSA21 regulates *tlp1-4* by promoting RNase degradation, the RNases involved may demonstrate significant converse correlation with *tlp1-4*. Therefore, the most related RNases are those found in module II, including Cj0121, *rnc*, *rnhA*, Cj1710c. All 4 of them showed negative co-expression (below -0.85) with *tlp1-4* (Figure 4.23 and 4.24).

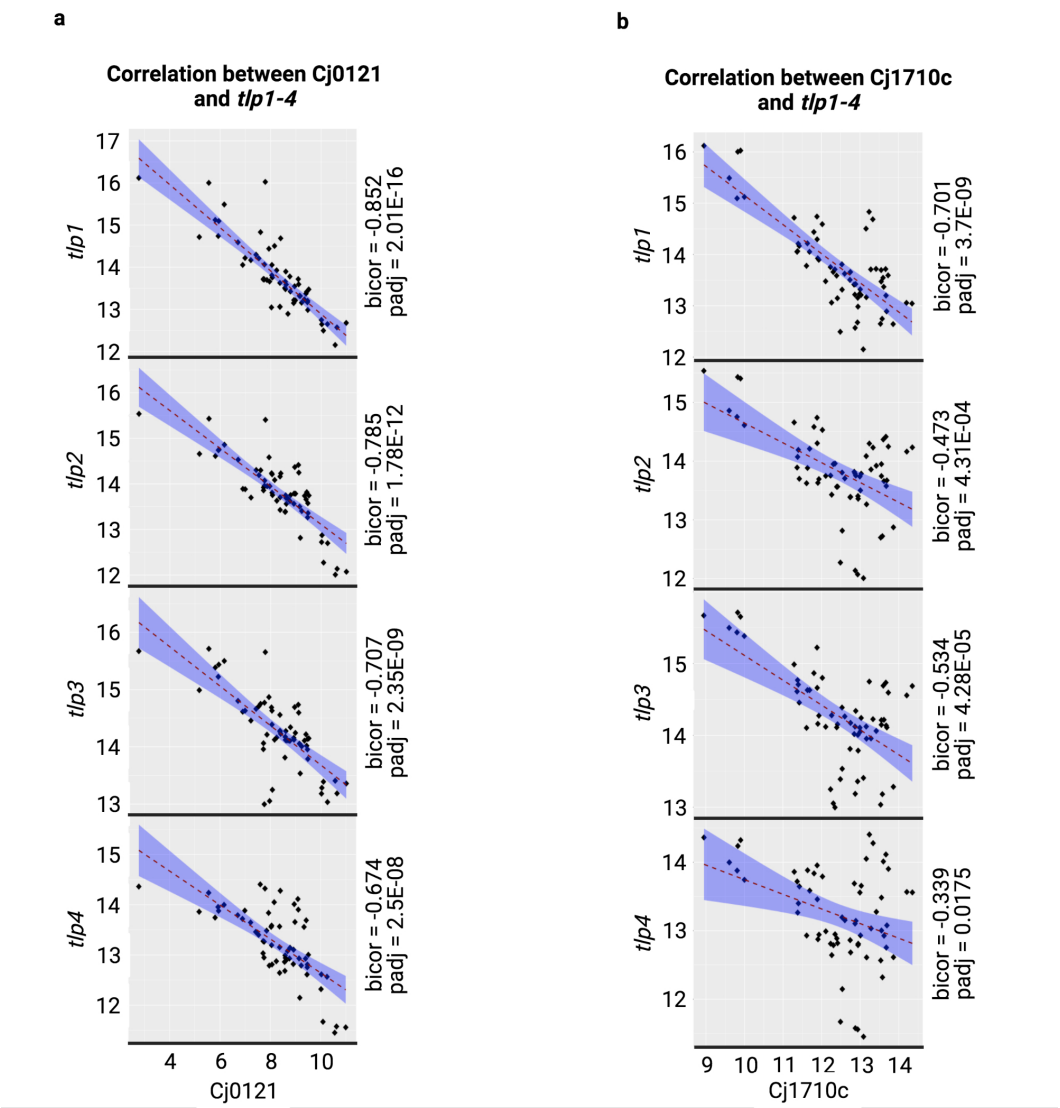

Figure 4.23: **Pairwise correlation between *tlp1-4* and Cj0121 and Cj1710c.** Correlation of vst normalised expression values between *tlp1-4* and (a) Cj0121 and (b) Cj1710c.

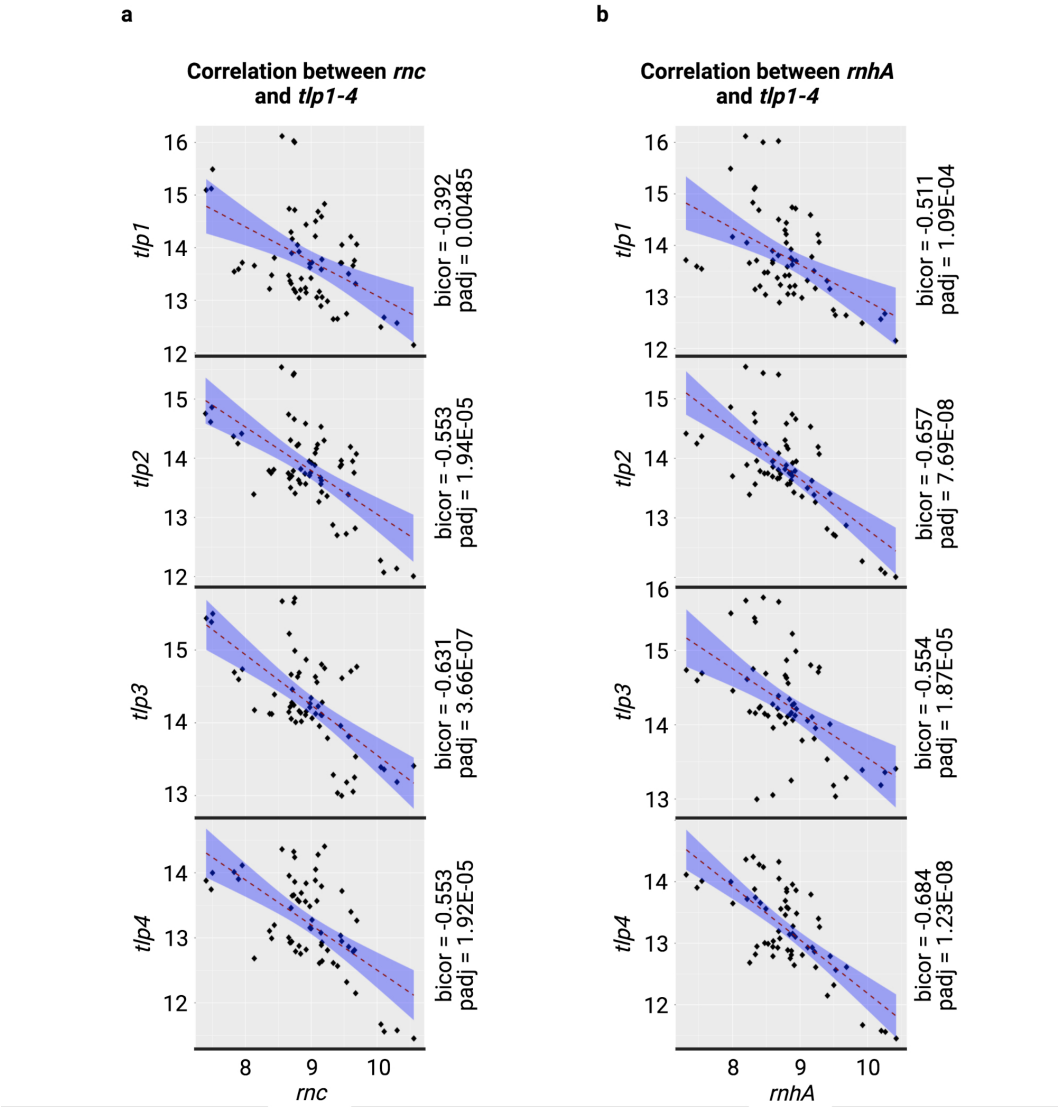

Figure 4.24: **Pairwise correlation between *tlp1-4* and *rnc* and *rnhA*.** Correlation of vst normalised expression values between *tlp1-4* and (a) *rnc* and (b) *rnhA*.

Differential expression showed that *rnc*, *rnhA*, Cj1710c were differentially upregulated in the three food storage conditions (Figure 4.19d). Other likely RNases candidates include Cj1388 and *rnhB*, as they were also differentially upregulated in food storage conditions. However, Cj1388 exhibited no inverse co-expressed with *tlp1-4*. Only *rnhB* had a weak/moderate negative correlation against tlp1-4 (-0.35 to -0.527) (Figure 4.25). These suggest the likely candidates for *tlp1-4* degradation were Cj0121, *rnc*, *rnhA*, Cj1710c and *rnhB*.

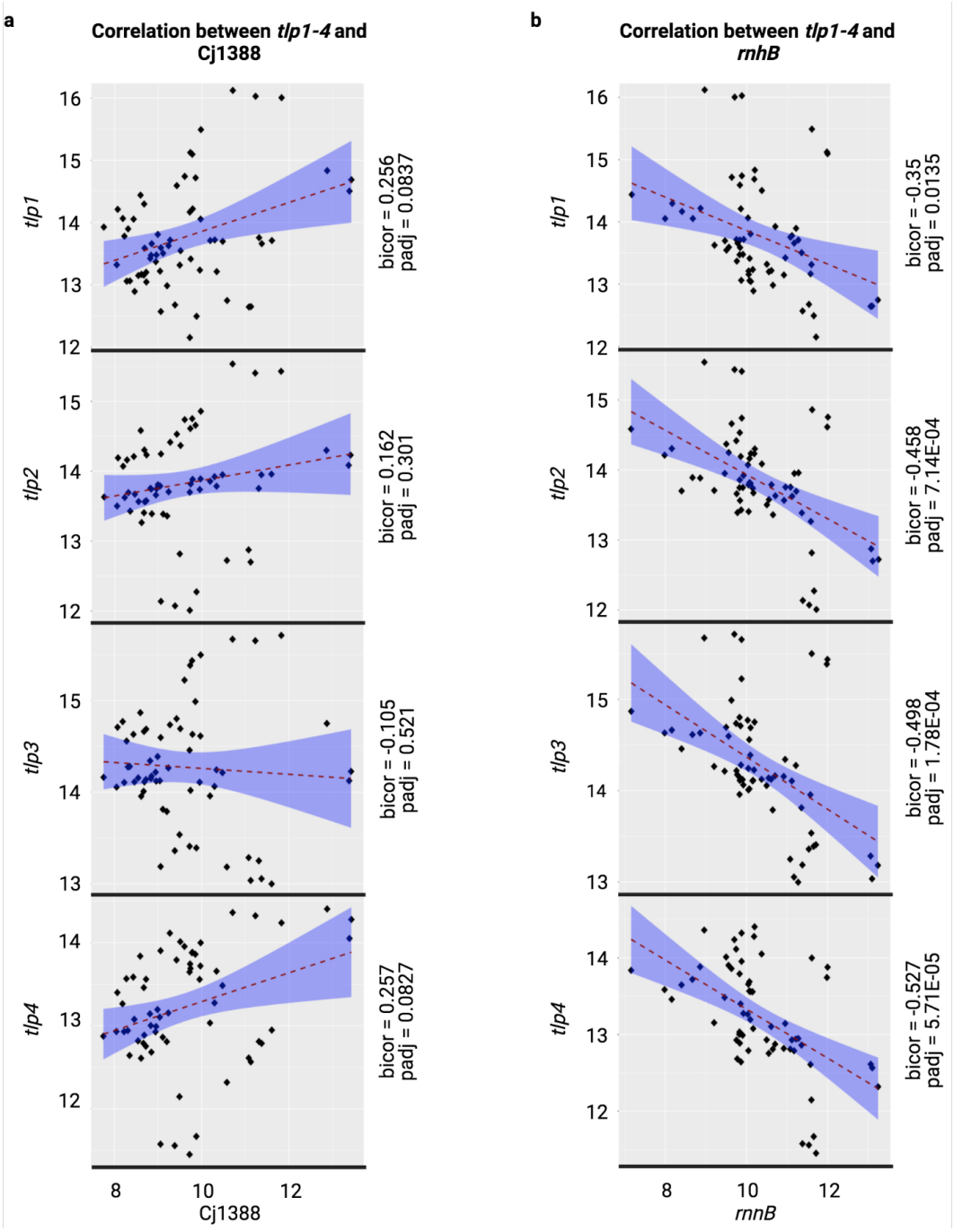

Figure 4.25: **Pairwise correlation between *tlp1-4* and Cj1388 and *rnhB*.** Correlation of expression values between *tlp1-4* and (a) Cj1388 and (b) *rnhB*.

4.4 sRNA-target interactions related to iron transport

4.4.1 The enrichment of iron and oxidative stress response genes in module VIII

CjSA21 was an outstanding sRNA for investigation. All of its predicted targets were involved in chemosensing. Other sRNAs of interest include those regulating processes crucial for stress adaptation. As mentioned earlier, several pathways displayed statistical enrichment in most pairwise comparisons (Figure 4.6a). Moreover, previous in-house analysis (Lam, 2019) highlighted that iron ABC transporter genes displayed the most apparent differential expression across all conditions. Hence the sRNA-mRNA interactions in the co-expression module that carries the most iron ABC transporters might play an important role in *C. jejuni* stress adaptation.

Notably, module VIII exhibited a statistically significant correlation with "ABC transporters" specifically. Module VIII showed a significant positive correlation with iron-limitation/repletion and nutrient deprivation conditions. For instance, module VIII had a positive correlation with iron_lim_ES and iron_lim_M. It also showed a negative correlation with iron_rep_ES, but the correlation coefficient was statistically insignificant with iron_rep_M (Figure 4.26). Interestingly, 37_M exhibited the most negative correlation, possibly because the upregulation of iron transporters occurred in most conditions. In contrast, some other modules positively correlated to iron_lim_ES and iron_lim_M, but not both. That seems to suggest module VIII may represent genes responsible for iron stress response (Figure 4.4).

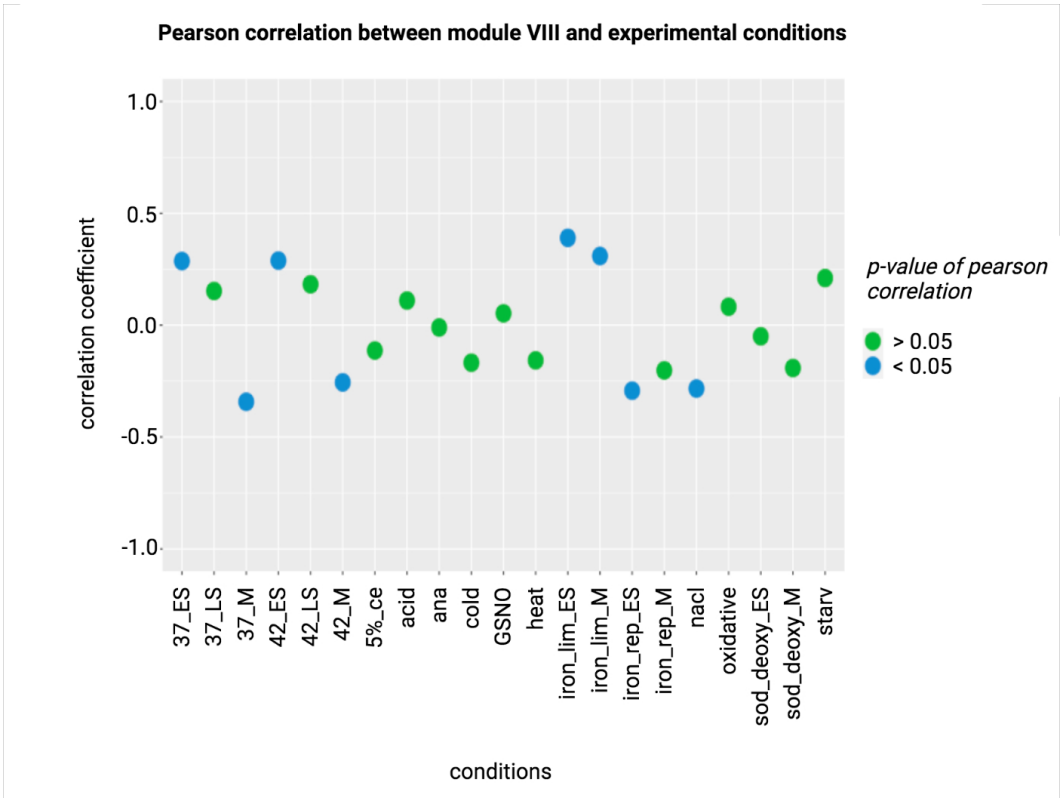

Figure 4.26: **The Pearson correlation between the module VIII eigengene and experimental traits.**

Further functional enrichment analysis of other databases enabled a more in-depth investigation of the biological significance of module VIII. Notably, the results indicated statistical overrepresentation of the Pfam domain "ABC transporters" as well as several TonB-related domains, such as "TonB dependent receptor", "Gram-negative bacterial TonB protein C-terminal", "TonB-dependent Receptor Plug Domain"), suggesting the involvement of TonB membrane receptors responsible for iron siderophore uptake (Figure 4.27). Similarly, InterPro enrichment analysis also highlighted statistical enrichment of features related to ABC transporters and TonB, alongside with "Thioredoxin-like fold" (Figure 4.28).

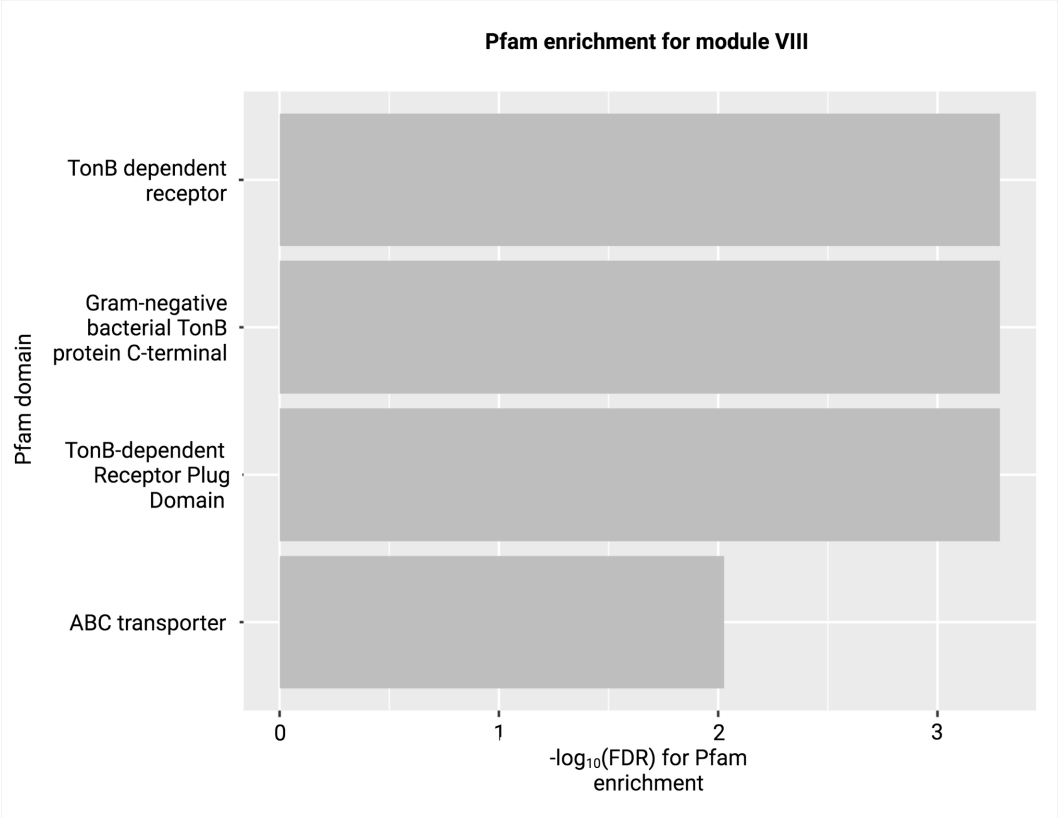

Figure 4.27: **Pfam enrichment of genes in module VIII.** All Pfam domain with with FDR <= 0.05, less than 100 proteins and at least three hits.

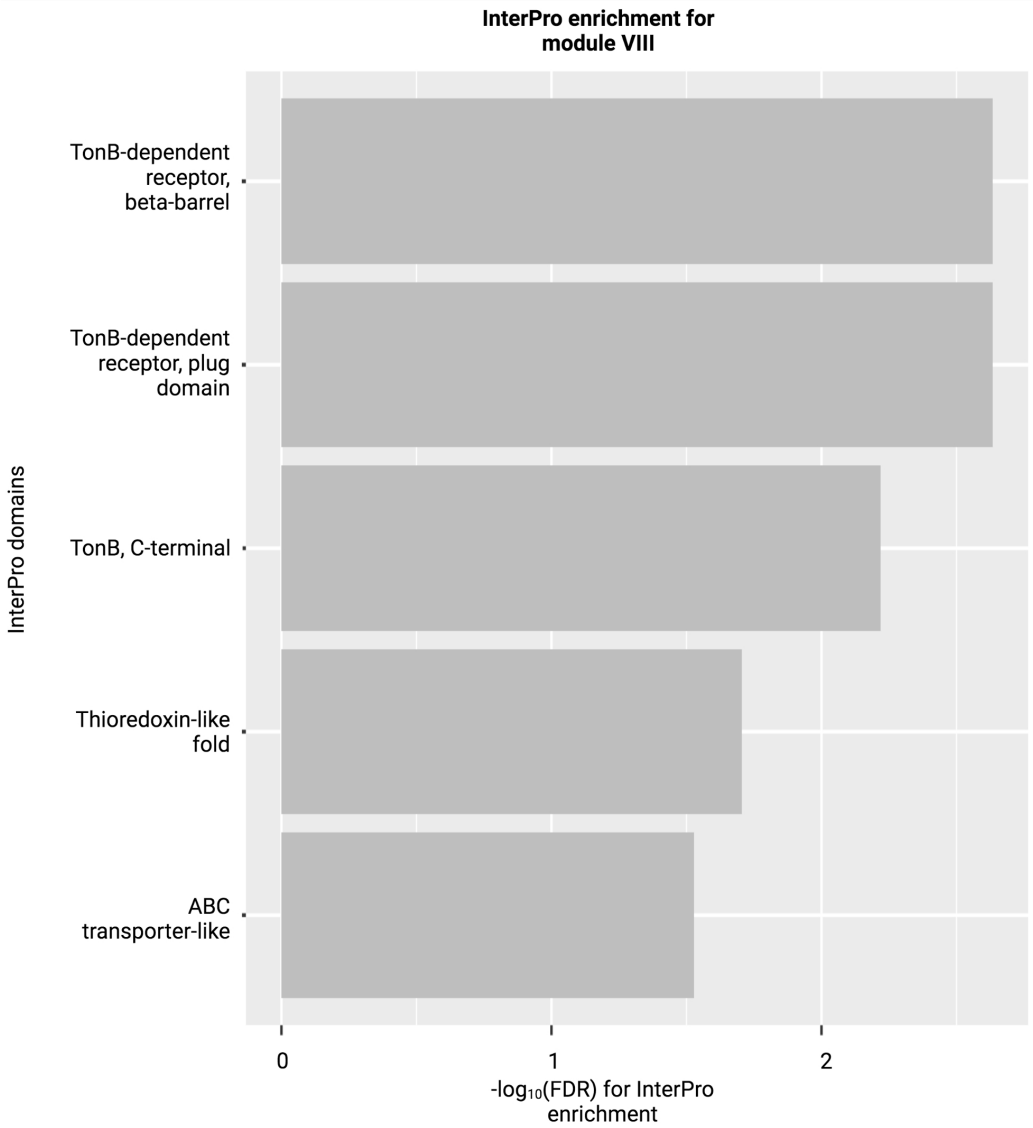

Figure 4.28: **InterPro enrichment of genes in module VIII.** All InterPro domain with with FDR <= 0.05, less than 100 proteins and at least three hits.

The involvement of thioredoxin hinted towards an oxidative stress response linked to the iron stress response. The involvement of oxidative stress response was further supported by Gene Ontology (GO) term enrichment for molecular function, with the most statistically significant terms being "peroxidase activity", "oxidoreductase activity, acting on peroxide as acceptor" and "antioxidant activity". Ribonucleases also appeared to be enriched in module VIII (Figure 4.29).

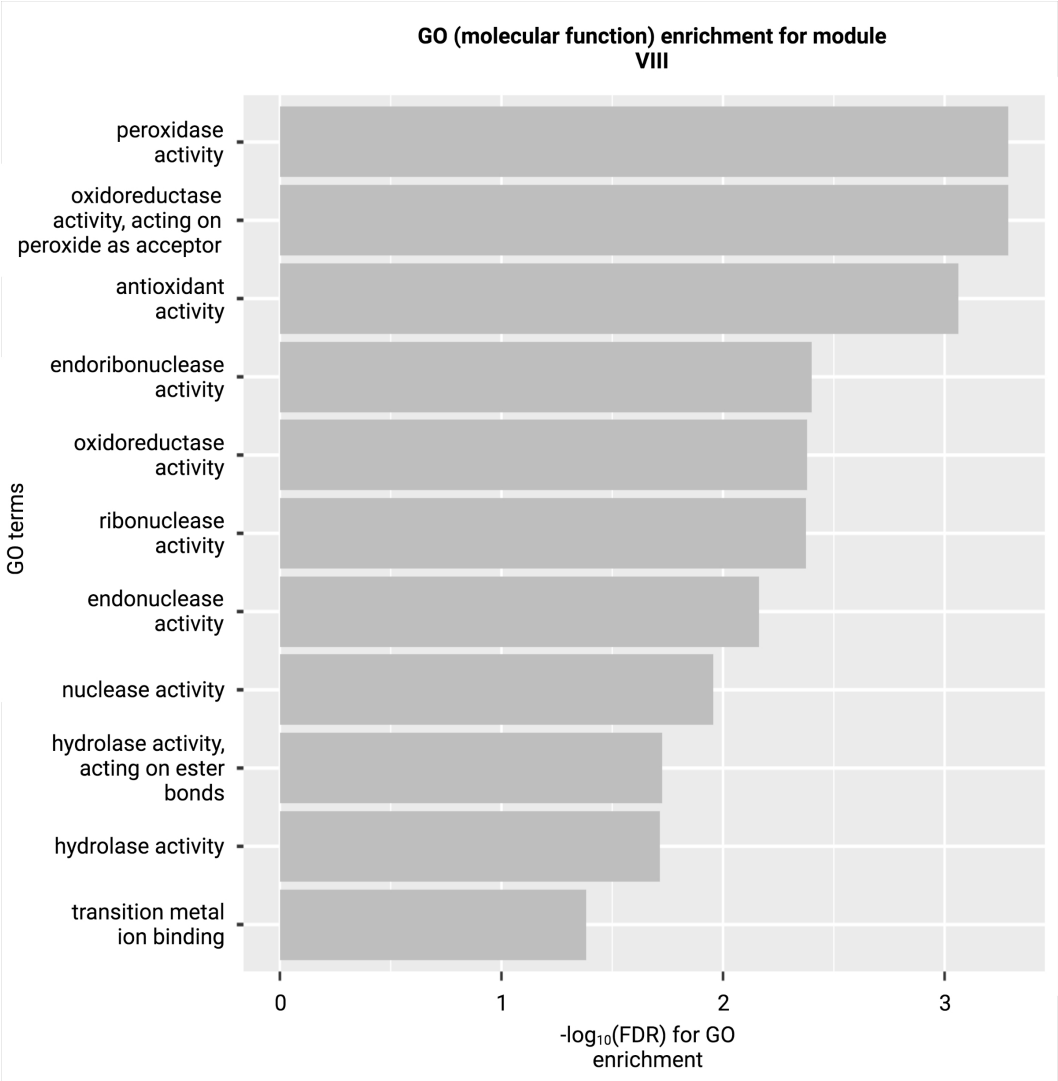

Figure 4.29: **GO (Molecular Functions) enrichment of genes in module VIII.** All GO (Molecular Functions) terms with with FDR <= 0.05, less than 100 proteins and at least three hits.

Since module VIII displayed statistical enrichment of pathways and protein domains involved in iron and oxidative stress responses, one can expect module VIII contain most iron ABC transporter genes. This expectation was confirmed in the distribution of ABC transporter genes among WGCNA co-expression modules. None of those ABC transporter genes belonged to module VII, among the 81 ABC transporters selected from KEGG BRITE cje2000 and DAVID annotation. In comparison, module X comprised most ABC transporter genes, possibly because module X carried the highest number of genes. Nonetheless, module VIII consisted of the most iron ABC transporter genes, including *cfbpABC*, *ceuBDE* (but not *ceuC*), *chuBCD*, *cj1661-1663*) (Figure 4.30). Other annotated iron stress response genes such as *p19*, *cj1658*, *tonB123*, *exbBD* were also partitioned into module VIII, which also included oxidative stress response genes such as *trxB*, *aphC*, *katA*. The presence of iron and oxidative stress response genes suggested the association between module VIII and the crosstalk between iron homeostasis and oxidative stress response (Kim et al., 2015; Palyada et al., 2009).

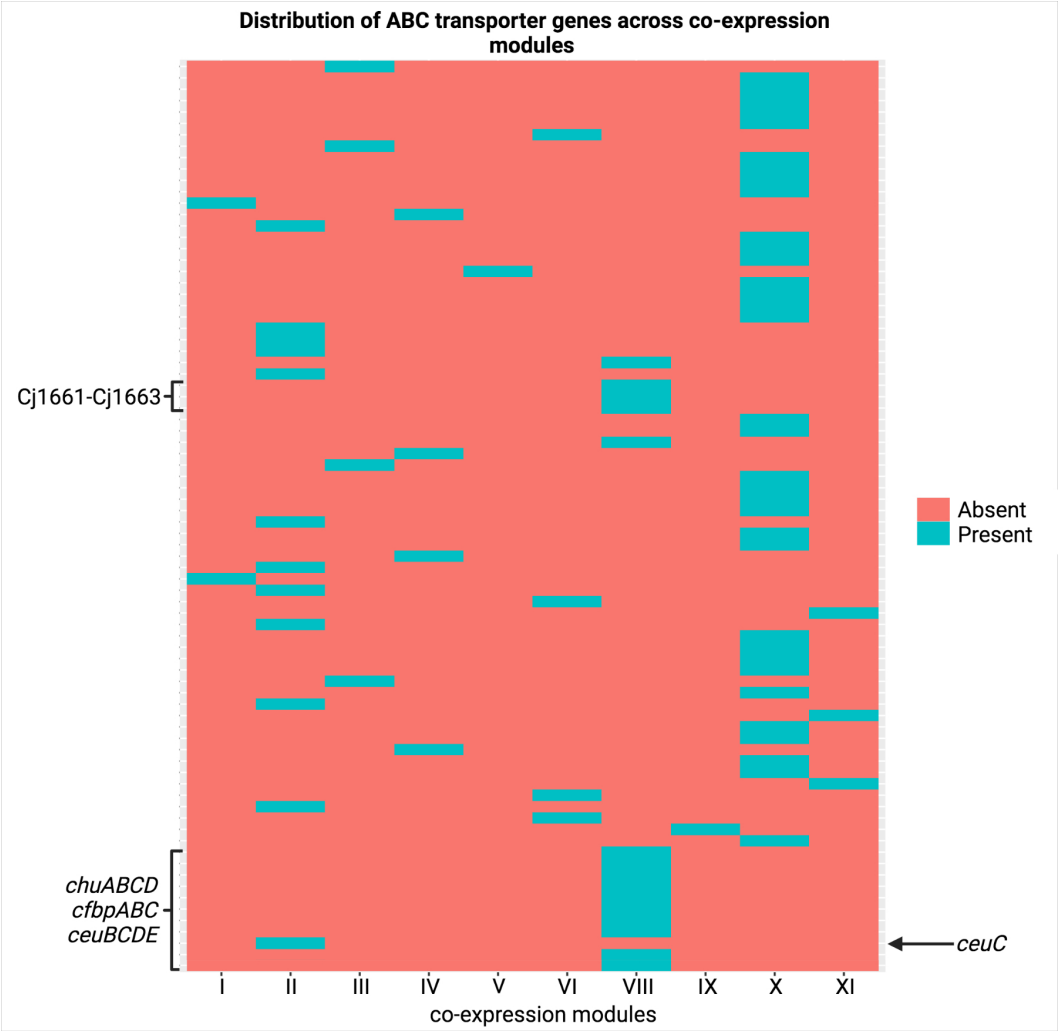

Figure 4.30: **Distribution of all ABC transporter genes across co-expression modules.** Most known iron transporter systems, including Cj1661-Cj1663, *chuABCD*, *cfbpABC* and *ceuBCDE*, all belonged to module VIII.

4.4.2 Key sRNAs in module VIII

The functional enrichment suggested novel and uncharacterised genes (including predicted sRNAs) in module VIII might be associated with iron stress response and oxidative stress response. Module VIII comprised four predicted sRNAs, namely CjSA24, CjSA98, CjSA106 and CjSA110. CjSA24, CjSA98, CjSA106 all appeared to be UTR-derived sRNAs of iron and oxidative stress response genes. That suggested these three sRNAs were found in module VIII either because of their biological functions or physical proximity to the module's genes.

Interestingly, CjSA110 is an intergenic sRNA with no neighbouring genes from module VIII. This putative sRNA shared almost identical genomic coordinates as an annotated sRNA CJnc190 (Dugar et al., 2013). CjSA110 showed weak/moderate yet significant negative correlation (bicor coefficient in between -0.3 to -0.5, FDR < 0.05) with most genes from module VIII, including genes from known iron or oxidative stress response pathways. The weak to moderate correlation value may not be an issue, as eukaryotic miRNA-mRNA also forms a weak or moderate negative correlation (Dai et al., 2019). Indeed, bacterial sRNA-mRNA interactions do not necessarily show a strong correlation as sRNA may only regulate the target's structure without changing the transcript abundance. Moreover, multiple sRNAs or other factors can control the quantity of an mRNA species.

In addition to CjSA110, module VIII also comprised several poorly characterised genes. For example, Cj0343c is a probable integral membrane protein designated as "Function unknown" in KEGG BRITE. Cj0378c contains the Pfam domain PF017 (ferric reductase-like transmembrane component). Cj0378c showed increased expression in avian mucus relative to mammalian mucus (Looft et al., 2019). Interestingly, a DNA microarray analysis that compared the erythromycin-resistant mutant against the wild-type NCTC11168 strain showed a similar expression pattern between Cj378c and PerR (Hao et al., 2013). Similarly, several poorly characterised proteins (Cj1383c, Cj1384c, Cj1627c, Cj1387c) (Looft et al., 2019; Butcher et al., 2015; Palyada et al., 2004; Clark et al., 2014) displayed differential expression in transcriptomic data related to iron limitation, despite not being annotated as iron homeostasis proteins. Other poorly characterised proteins include Cj1386 (Crofts et al., 2018) and Cj1725 (Hepworth et al., 2011; Bronnec et al., 2016).

The extracted sRNA-target interactions linked CjSA110 with *tonB3* and Cj1384c. *tonB3* and Cj1384c belonged to module VIII and showed negative co-expression against CjSA110 and opposite differential expression in several pairwise compar-

isons. IntaRNA prediction also predicted binding energy values that lead to p-values below 0.05. See below for more details.

4.4.3 Co-expression of CjSA110

Further exploration of the co-expression pattern within module VIII would help understand the post-transcriptional regulatory mechanism of CjSA110. Similar to the previous section, module VIII was subdivided into 2 clusters with contrasting expression patterns. CjSA110 and genes including Cj0011c, *rnhB*, *rrc* and *nhaA* belonged to cluster 2, while most genes including *tonB3* and Cj1384c belong to cluster 1. CjSA110 appeared to have the lowest expression under starvation stress and high expression upon exposure to NaCl stress. Moreover, CjSA110 belonged to the same cluster as Cj0011c, *rnhB*, *rrc* and *nhaA*, while also showing an opposite pattern from other genes in module VIII. Genes in cluster 1 displayed high expression under NaCl stress and standard laboratory conditions and low expression under starvation stress.

Genes in cluster 2, including tonB3 and Cj1384c and most iron transport and oxidative stress genes, were characterised by high expression values under iron limiting conditions and early-stationary phases, possibly due to the increased expression of various iron ABC transporters (Figure 4.31 and 4.32).

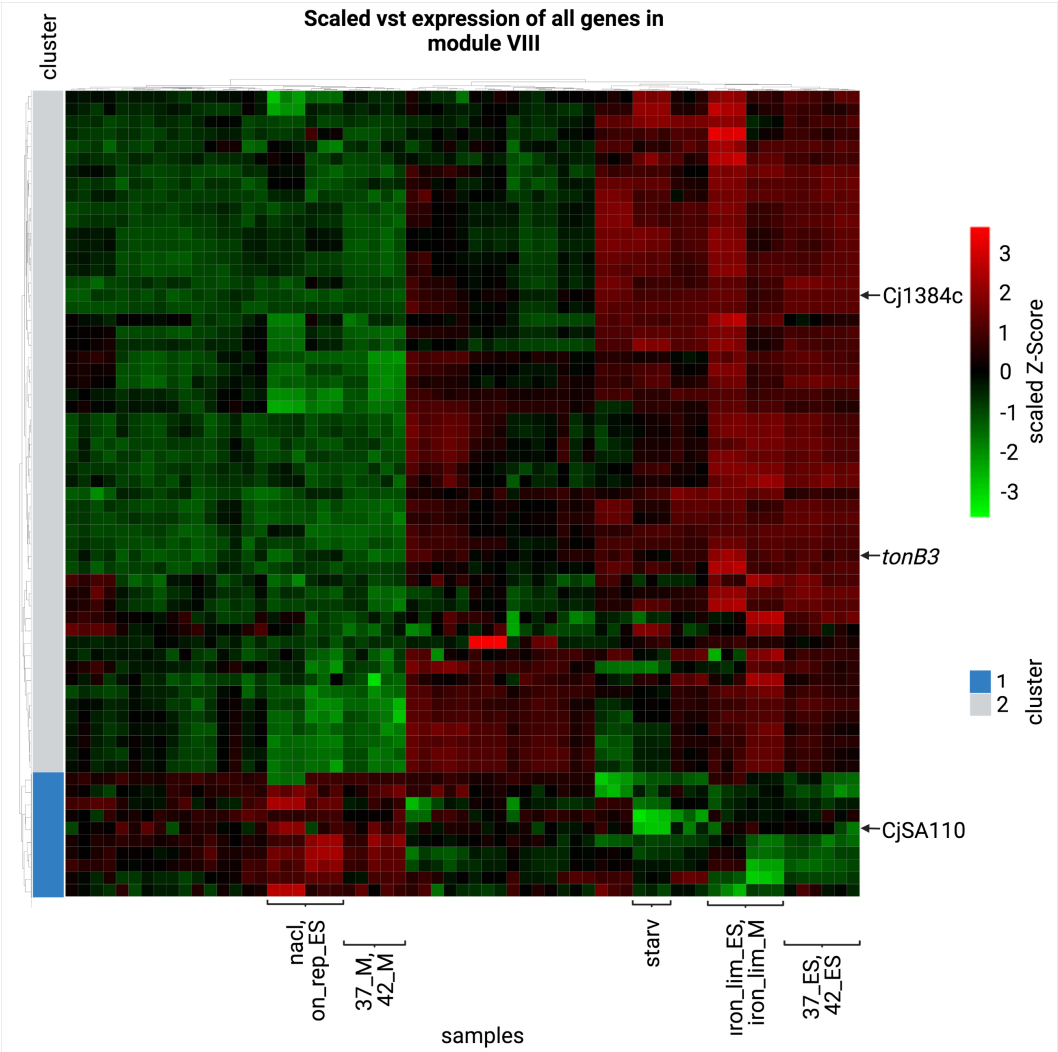

Figure 4.31: **Scaled heat maps for vst expression of all genes in module VIII.** The colour scale is based on the Z-Score of vst normalised values.

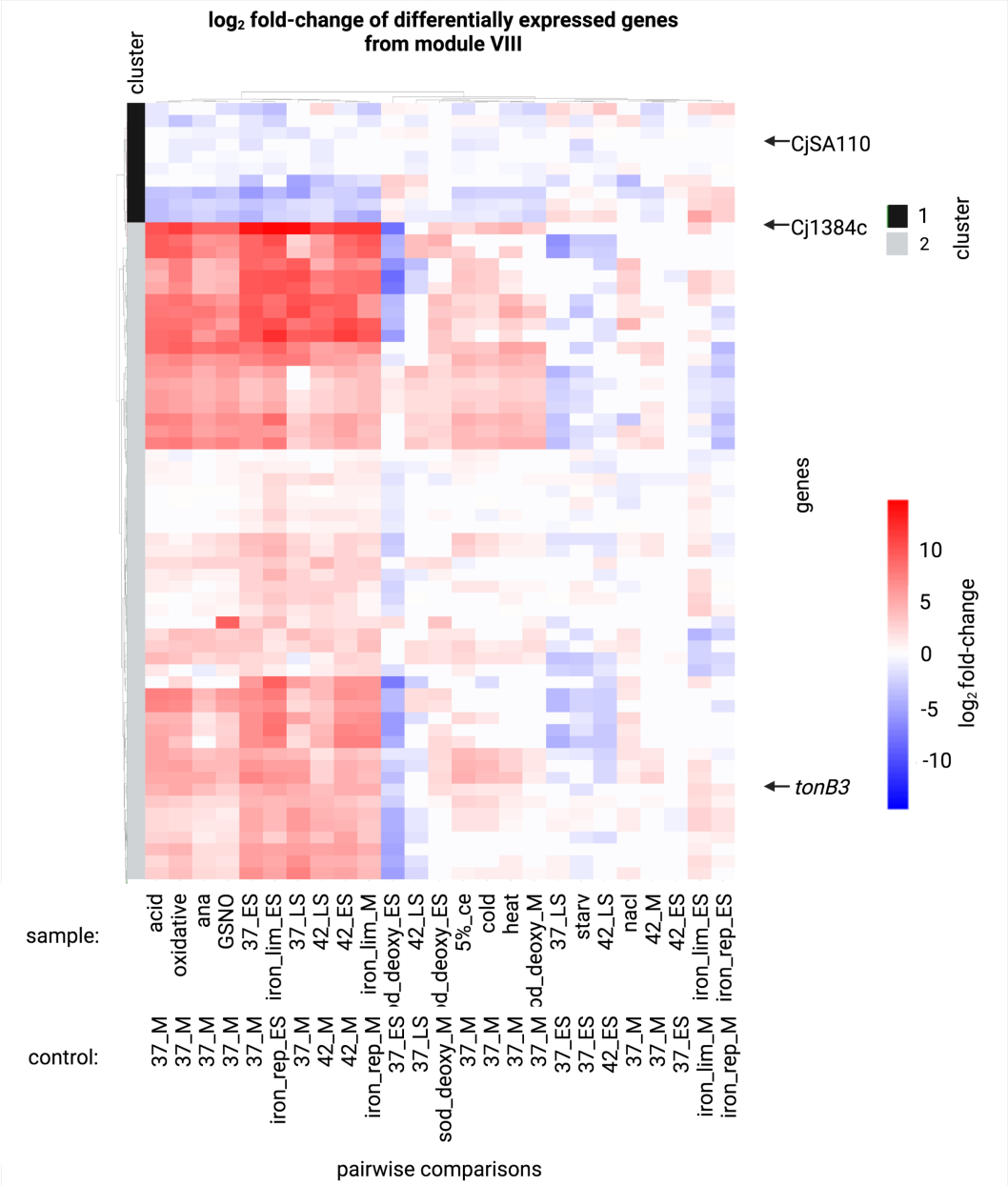

Figure 4.32: **$\log_2$ fold-change expressions and clustering all genes in module VIII.** All boxes with FDR > 0.05 are coloured in white.

As CjSA110 and its predicted targets belonged to different clusters within module VIII, it was not surprising that they displayed opposite expression patterns. Moderately negative yet statistically significant correlation coefficients were observed between CjSA110 and *tonB3* and Cj1384c (Figure 4.33.) That suggests CjSA110 may negatively regulate the abundance of *tonB3* and Cj1384c mRNAs by RNase degradation.

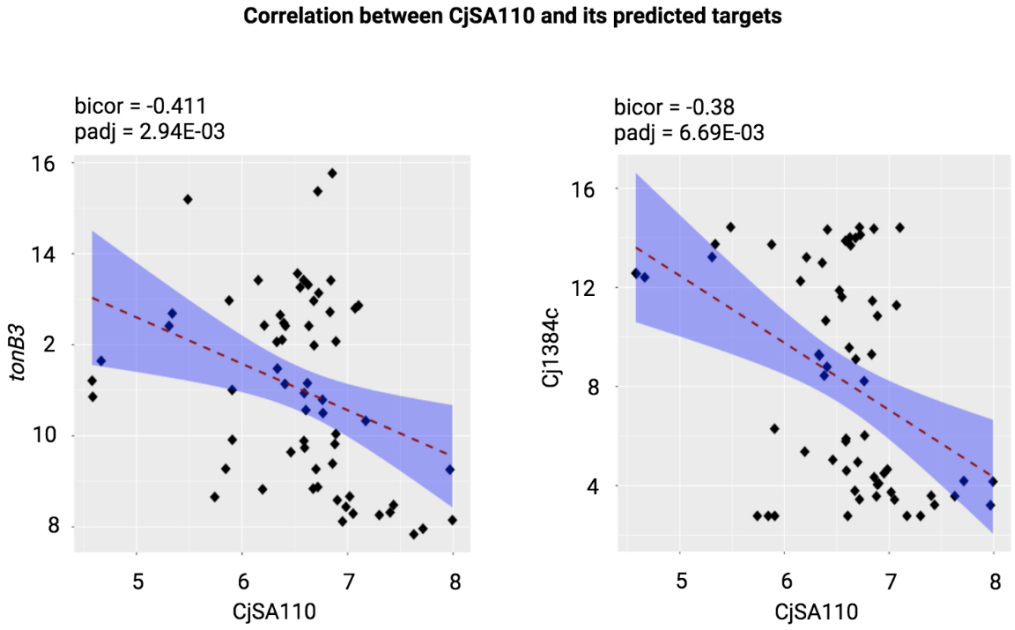

Figure 4.33: **The pairwise correlation between *CjSA110* and *tonB3* and Cj1384c.**

4.4.4 Differential expression of CjSA110

After establishing a negative co-expression between CjSA110 and its targets, the next step was to identify conditions that display converse differential expression patterns. Displaying such a pattern under iron-related conditions would indicate that CjSA110 is directly responsible for iron stress. Alternately, an opposite expression pattern may occur in other pairwise comparisons. This outcome may indicate crosstalk between those conditions and iron limitation, which is reasonable given the association between iron homeostasis and other stress response pathways.

CjSA110 showed the lowest normalised expression in the late stationary phase, early stationary phase and starvation (Figure 4.34). As mentioned before, these conditions were the main driving force of expression variation across the RNAtag-seq dataset. That suggested CjSA110 followed a similar pattern as the overall transcriptomic landscape.

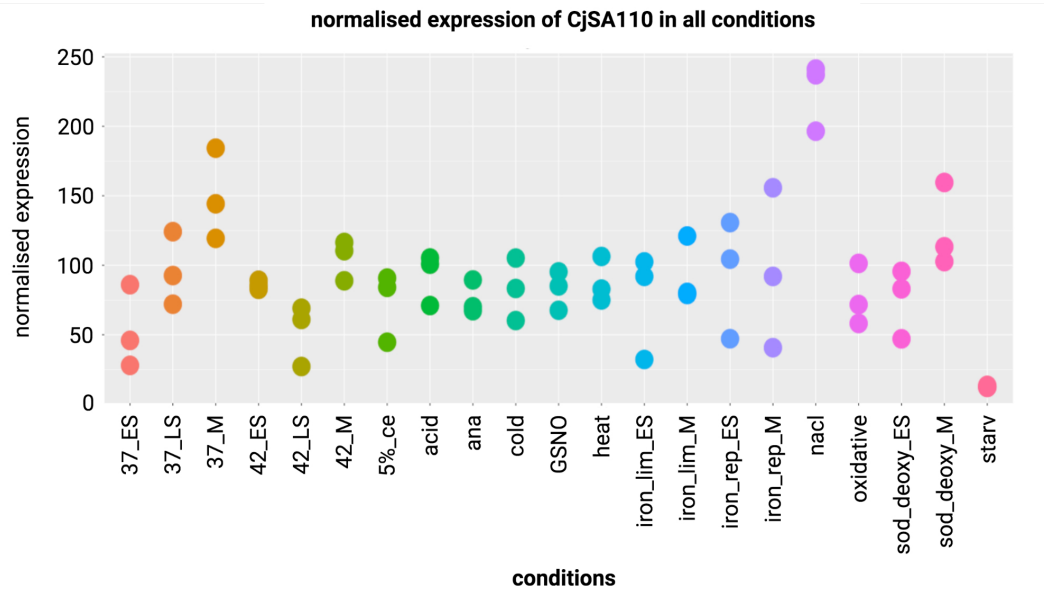

Figure 4.34: **Expression of CjSA110 across RNAtag-seq conditions.** The expression values were normalised using DESeq2's median of ratios.

CjSA110 showed no differential expression in most conditions, except in starv_vs_37_ES. Further investigation revealed that CjSA110 was usually downregulated under most nutrient-deprived and starvation stress. Moreover, CjSA110 is upregulated for nacl_vs_42_LS and nacl_vs_37_ES. In some conditions, CjSA110 shows opposite differential expression to genes from module VIII. Further investigation of differential expression patterns indicated that CjSA110 and *tonB3* and Cj1384c showed opposite differential expression in anaerobic stress, 5% chicken exudate, early-stationary phase under 37 °C, late-stationary phase under 42 °C and cold stress (Figure 4.35). Therefore, these conditions may have relationship with iron stress and oxidiative stress

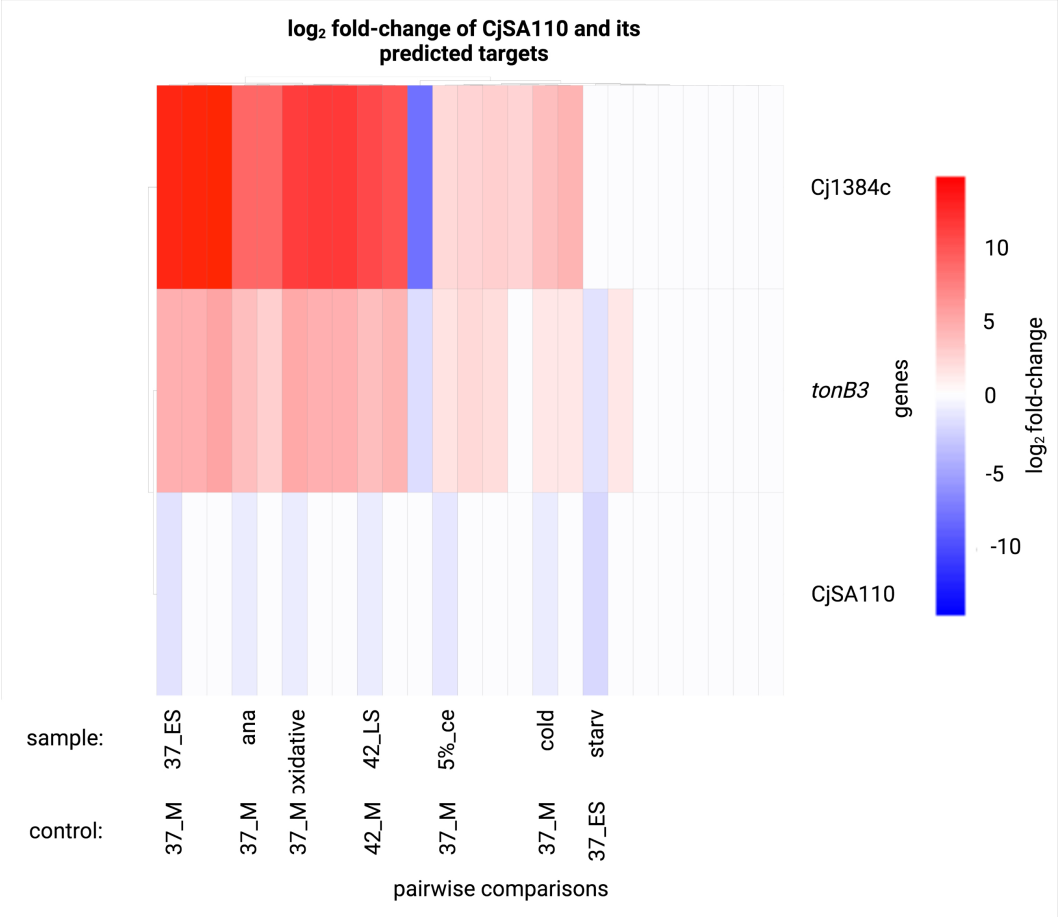

Figure 4.35: **$\log_2$ fold-change expression CjSA110, *tonB3* and Cj1384c.** All boxes with FDR > 0.05 are coloured in white.

Table 4.5: **IntaRNA output of Cj1384c and *tonB3*.**

gene id	**binding energies (kcal/mol)**	***p*-value**	**FDR**
Cj1384c	-15.12	0.00297	0.837
tonB3	-12.77	0.02	0.929

4.4.5 Predicted binding mechanism of CjSA110

For *tonB3* and Cj1384c, the binding energy values led to significant *p*-values but insignificant FDR (Table 4.5). Indeed, only Cj1209 has FDR below 0.05, and the energy above was still at -18.8 kcal/mol. Most other predicted targets with significant *p*-values also lacked insignificant FDR. Even CjSA109, which is antisense to CjSA110, only had an energy level at -11.88 kcal/mol (Figure 4.36a). The total energy distribution explained the statistical results (Figure 4.36b), as most energy values were around -10 kcal/mol. That was different from CjSA21, which comprised predicted targets with distinctive binding energy values from the rest of the targets. Thus, CjSA110 may form many background binding interactions unless RNA chaperones are involved.

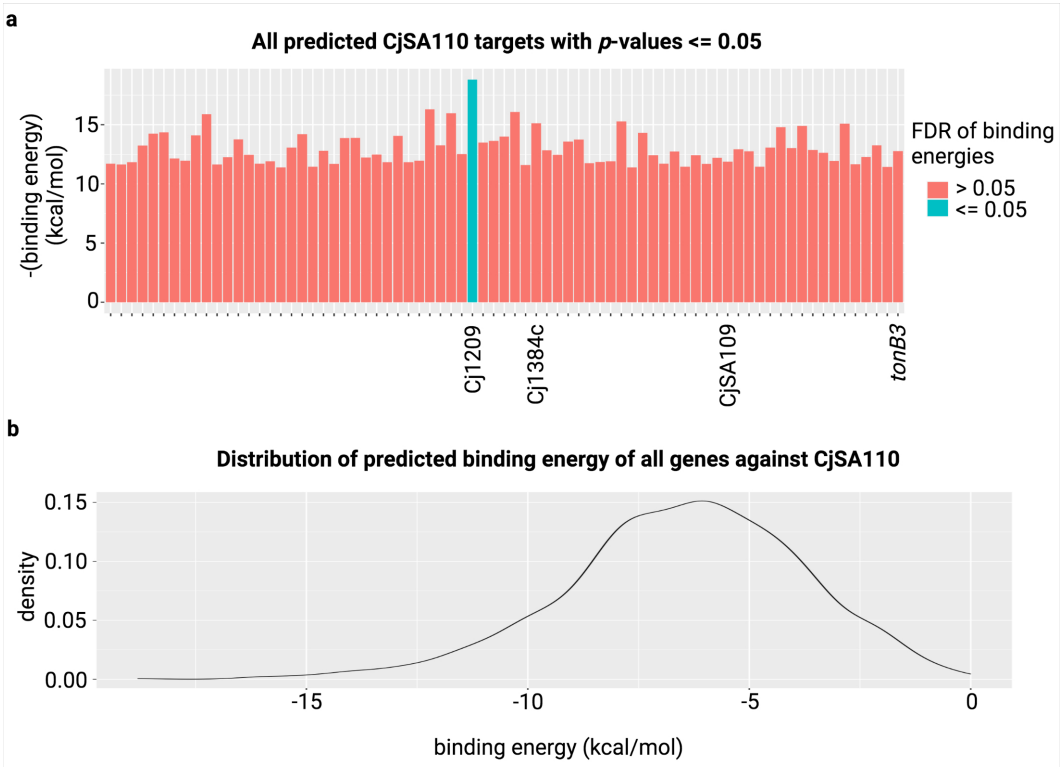

Figure 4.36: **Results of CjSA110 genome-wide target prediction.** (a) Predicted binding energy of all targets with p-values <= 0.05. Those bars coloured in light blue has FDR <= 0.05. (b) IntaRNA binding energy distribution of all sRNA-mRNA pairs.

Most IntaRNA-predicted interactions happened at two ends of CjSA110. For instance, the *tonB3* target site was near the 3' end of CjSA110 and shared similar binding sites as Cj1209 and CjSA109. In contrast, Cj1384c targets near the 5' end of CjSA110 (Figure 4.37). RNAfold computed a secondary structure with stem-loops near both ends with stable base pairing in the centre. Both binding regions consist of some base pairs with high nucleotide entropy and low intramolecular base-pairing probability, especially for most nucleotides at the binding site for *tonB3* (Figure 4.38). That may explain why both regions were predicted as binding sites.

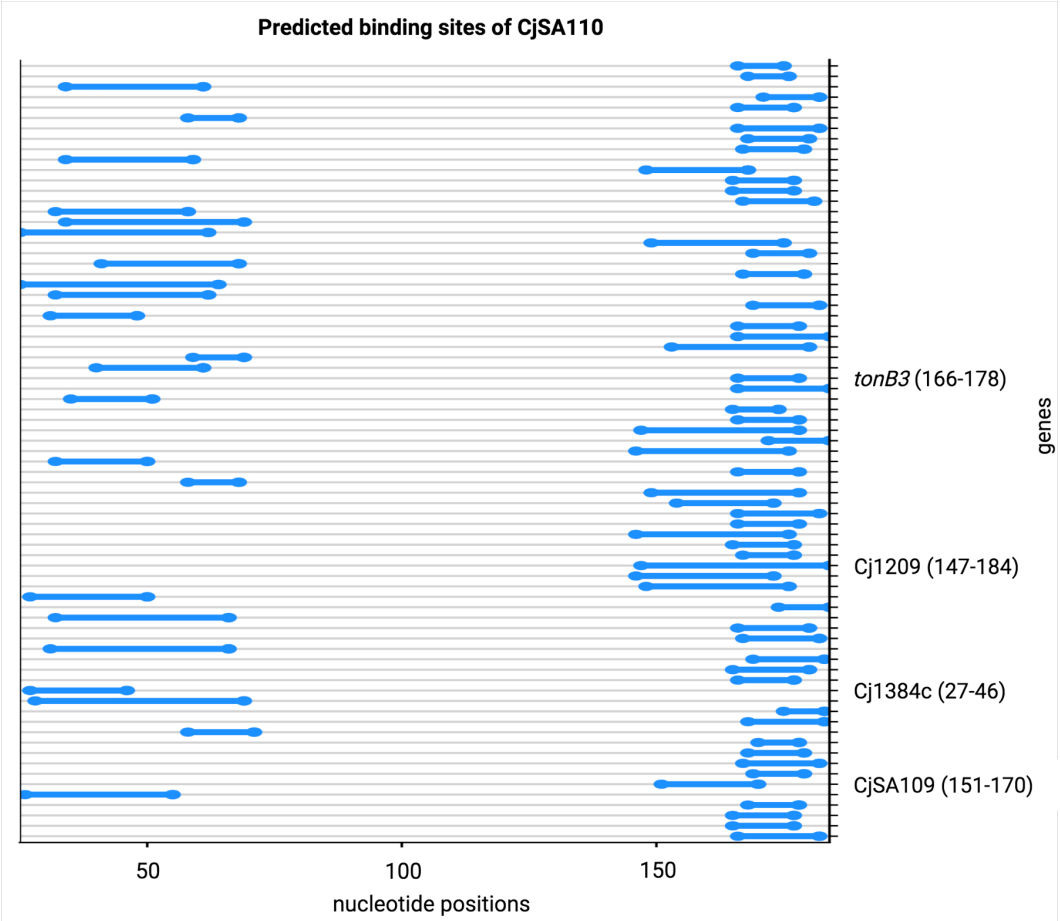

Figure 4.37: **The distribution of optimal binding sites of CjSA110 against targets.** All targets display p-value <= 0.05.

4.4.6 putative RNases involved

Similar to CjSA21, the mode of action of CjSA110 may involve RNase degradation. Therefore, RNases in module VIII might be responsible for post-transcriptional

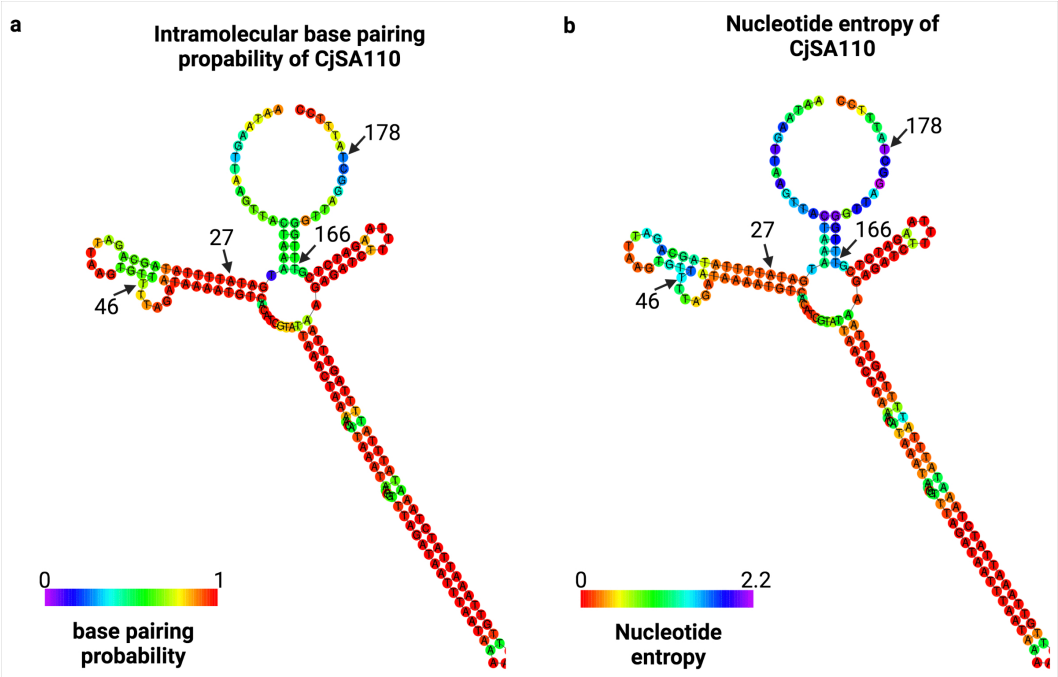

Figure 4.38: **Binding probabilities and entropies of CjSA110 predicted by RNAfold centroid structure.** Those arrows refer to the nucleotide positions of CjSA110.

regulation by CjSA110. RNases in module VIII consisted of *rnhB*, an RNase HI involved in DNA replication. Another RNase *rnc* is an RNase III that digests double-stranded RNA structure and thus might be able to target sRNA-mRNA duplex. They are both likely candidates recruited by CjSA110-tonB3/cj1384c duplex. It is noteworthy that Cj1209, the only IntaRNA-predicted CjSA110 with FDR, is also an RNase Y. RNase Y is similar to RNase E, which can degrade free RNA and regulate the abundance of sRNAs or mRNAs. Interestingly, Cj1209 belonged to module VIII but formed no connection with CjSA110 in the WGCNA network. This observation could be due to Cj1209 expression being strongly correlated with other genes of module VIII, but not CjSA110.

Under conditions where CjSA110 and *tonB3* and Cj1384c showed opposite differential expression, Cj1209 also exhibited a similar differential expression pattern as CjSA110 (Figure 4.39). Further co-expression analysis of Cj1209 revealed a weak positive correlation with CjSA110. The weak correlation between CjSA110 and Cj1209 may explain why they showed no connection in the WGCNA network despite being clustered into the same co-expression module. The result showed a robust negative correlation with *tonB3* (Figure 4.40). This observation suggested Cj1209

was the most likely candidate responsible for degrading *tonB3* and Cj1384c.

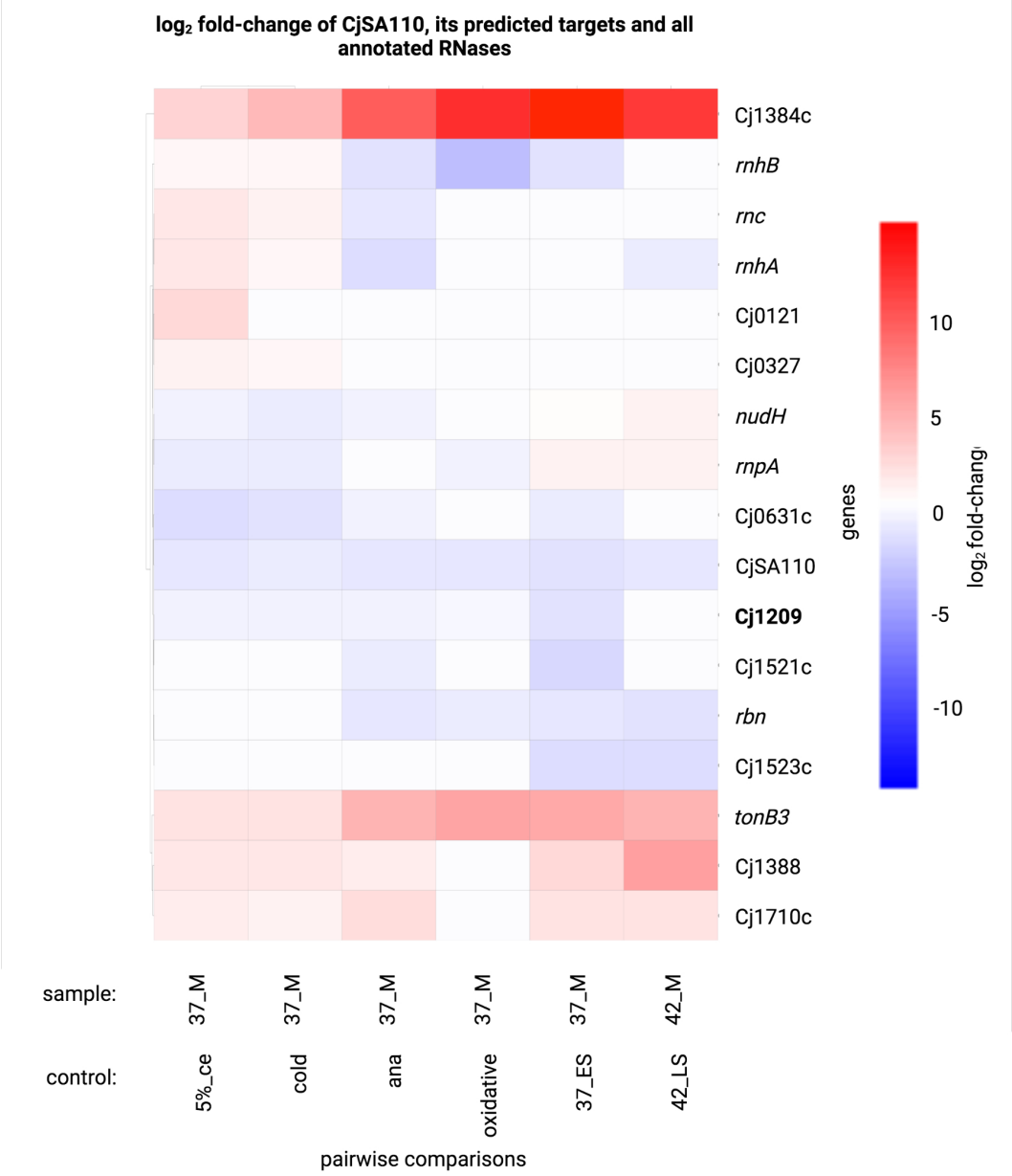

Figure 4.39: **$\log_2$ fold-change expression CjSA110, *tonB3* and Cj1384c and all annotated RNases.** All boxes with FDR > 0.05 are coloured in white.

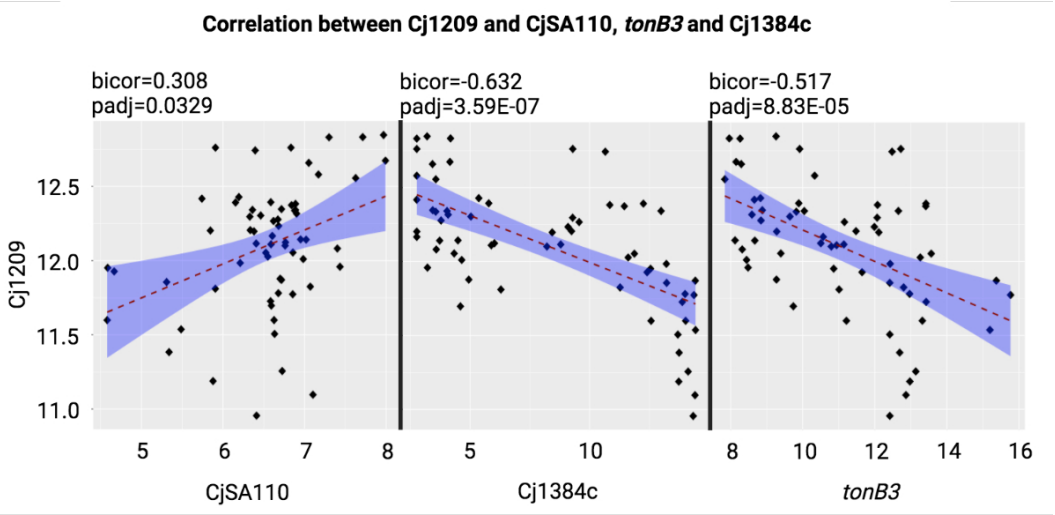

Figure 4.40: **Pairwise correlation between Cj1209 and CjSA110, Cj1384c and *tonB3*.** Cj1209 showed significant negative correlation with both CjSA110 predicted targets, and positive correlation with CjSA110.

4.5 Discussion

4.5.1 Overview of RNAtag-seq data

This chapter has simultaneously explored RNAtag-seq data across 21 conditions, covering growth phases, temperatures, food storage stresses and other environmental stresses. Among the differentially enriched pathways are "ABC transporters" and "Two-component system", which appear in almost every pairwise comparison. The significant enrichment of "ABC transporters" and "Two-component system" across many conditions is reasonable as they are heavily involved in coping with external stresses. In adverse conditions, especially during nutrient shortages, ABC and other membrane transporters increase nutrient uptake. Meanwhile, a two-component system senses environmental stresses, including nutrient deprivation and food storage conditions, to stimulate cellular responses such as altering cell motility or biofilm formation. This chapter has further predicted sRNA-target interactions related to membrane transport and signal transduction and provided additional insights into *C. jejuni* stress adaptations.

4.5.2 Identification of CjSA21 as a chemotaxis inhibitor

CjSA21 inhibits the expression of *tlp1-4*

Tlps are chemoreceptors that sense amino acids, galactose, bile salt, complex sugars, as well as organic ions such as iron and phosphate (Chandrashekhar et al., 2018; Elgamoudi et al., 2021). The *C. jejuni* genome sequences show ten homologues to Tlp and two aerotaxis receptor proteins. These receptors stimulate CheY phosphorylation. Phosphorylated-CheY binds to FliM and mediates a transition from counter-clockwise to the clockwise flagellar rotation.

Group A Tlp receptors comprise the translated products of all CjSA21 targets (Tlp1-4) and Tlp10 generated from *tlp10* mRNA that showed statistically significant sequence complementarity with CjSA21. They share structural similarities to *E. coli* MCP and *H. pylori* family A transducers, including the transmembrane domains and a periplasmic ligand-binding domain. Tlp1 and Tlp 3 are frequently found among *C. jejuni* strains (Clark et al., 2019). The periplasmic ligand-binding domain suggests that Group A receptors sense extracellular signals. Notably, Tlp2-4 share conserved C-terminal domains with 346 identical amino acid residues. That explains why IntaRNA suggested CjSA21 targets *tlp2-4* C-terminal with the same sequence.

Chemoattractants for group A receptors include fucose, pyruvate, fumarate, aspartate, and formate. However, the ligand-binding specificity of most Tlp remains unclear, and their ligand-binding domains are not conserved. Amino acid array hybridisation showed specific binding between Tlp1 and L-aspartate, while yeast two-hybrid system and immunoprecipitation shows protein-protein interactions between Tlp1 and CheV (Hartley-Tassell et al., 2010). However, the lack of aspartate in the Tlp1 crystal structure suggests Tlp1 recognises aspartate indirectly via an unknown periplasmic binding protein (Machuca et al., 2016). Deletion of tlp2 led to reduced chemotaxis towards Asp, pyruvate, phosphate and $FeSO_4$. Transcription of *tlp2* is induced by inorganic phosphate and iron (Chandrashekhar et al., 2018). Meanwhile, Tlp3 is a multi-ligand receptor that interacts with the following chemoattractants and chemorepellents: isoleucine, purine, malic acid and fumaric acid, lysine, glucosamine, succinic acid, arginine and thiamine (Rahman et al., 2014). Tlp3 ligand also includes other hydrophobic amino acids, including alanine, valine and phenylalanine (Khan et al., 2020). Tlp3 and Tlp4 are necessary for sodium deoxycholate chemoattraction. Antibodies blockage and disruption of tlp3/4 reduced chemoattraction towards sodium deoxycholate (Li et al., 2014). Interestingly, deletion of

tlp1 and *tlp3*, but not *tlp2*, showed a 10-fold reduction of invasion of human Coco 205 epithelial cells and chicken embryo intestinal cells. However, none of the above mutations affects chemotactic capacity. That may suggest alternative receptors that target the same chemoattractants (Vegge et al., 2009).

Tlp10 has a bimodal ligand-binding domain and specificity for multiple classes of chemoeffectors. Tlp10 detects aspartate, isoleucine, fumarate, malate, fucose, and mannose for chemoattraction. It also senses arginine, galactose and thiamine as chemorepellents (Elgamoudi et al., 2017). In addition, deletion of *tlp6* and *tlp10* in strain 81-176 decreased chemotaxis toward aspartate and glutamate and TCA intermediates (Chandrashekhar et al., 2015a).

Another translated product of mRNA highly complementary to CjSA21, Tlp9, is a group B receptor. Group B receptors are homologous to family B transducers in *H. salinarum*, consisting of a cytoplasmic protein anchored to a transmembrane region and lacks a periplasmic ligand-binding domain. As *tlp9* locates between *aer1* and *aer2*, it might be an energy taxis system that senses the redox state of the electron transport chain from the PAS-containing Aer2 (Hendrixson et al., 2001; Elliott and DiRita, 2008). Tlp9 might also receive signal transduction from other energy taxis response proteins like CetC and CetZ (Reuter and van Vliet, 2013).

The remaining Tlps, such as Tlp6 and Tlp8, belong to group C. None of the group C *tlp* appeared to be CjSA21 targets. Group C Tlp only carries a signalling domain and is similar to family C transducers in *H. salinarum*. The lack of transmembrane domain suggested that group C receptors might respond to intracellular signals instead of external stimuli (Marchant et al., 2002).

sRNA-mediated regulation of *tlp* translation has been observed in *Helicobacter pylori*. RepG sRNA targets the G-repeats of the leader sequence of the *tlp* gene and regulates *tlpB* translation (Sharma et al., 2010; Pernitzsch et al., 2014). RepG also represses the translation of HP0102, which is downstream of *tlpB* within the same operon. Hp0102 is responsible for smooth LPS production, which confers membrane stress protection and antibiotic resistance (Pernitzsch et al., 2021). Inhibiting *Helicobacter pylori* chemotaxis away from autoinducer-2 (AI-2) promoted biofilm formation. Disruption of *tlpB* and other AI-2 sensing proteins led to more adherent cells. Conversely, the addition of AI-2 increased biofilm dispersal Anderson et al., 2015. In a follow-up computational prediction by agent-based modelling, cells insensitive to AI-2 chemotaxis could form more extensive biofilms, while strains with over-expressed AI-2 formed smaller biofilms Sweeney et al., 2019.

CjSA21 is one of the few predicted sRNAs conserved among all *Helicobacter* and *Campylobacter* strains (see the BLASTn result in Chapter 3). Hence, this chapter might have discovered a new chemotaxis regulatory mechanism among Epsilonproteobacteria. BLASTx search showed that the translated sequences of tlp1-4 and tlp9 match *H. pylori* TlpA, while the tlp10 translated sequence is similar to *H. pylori* TlpB. However, there is no significant nucleotide conservation between *tlp* genes of *C. jejuni* and *H. pylori*. Also, the CjSA21 sequence is absent from the genome sequences of *H. pylori* and other species. That illustrated the possibility of sRNA-mediated Tlp regulation in other species via alternative sRNA and mRNA sequences.

Chemotaxis suppression by cold and hyperosmotic stress

One key observation from our result is that CjSA21 appears to regulate *tlp1-4* expression under food storage conditions. The "Bacteria chemotaxis" pathway was differentially downregulated under chicken exudate, cold, and hyperosmotic stress. Under these conditions, CjSA21 showed differential upregulation. Similarly, *C. jejuni* strain 81-176 showed impaired growth and complete loss of motility under osmotic stress (1.5% NaCl) (Cameron et al., 2012). Other species demonstrate similar gene expression patterns. Salt and low temperature suppress motility and transcriptomic expression of flagella assembly and chemotaxis pathways in *Listeria monocytogenes*. Other downregulated pathways include Branched amino acid metabolism and Other/unknown amino acid-regulated metabolism (Durack et al., 2013). Meanwhile, reducing NaCl concentration increases flagellar biosynthesis and swimming motility in *Escherichia albertii* (Ikeda et al., 2020).

Apart from being less sensitive to external stimuli, the suppression of chemotaxis may reduce biofilm dispersal and favour biofilm formation over a planktonic lifestyle. Under food storage conditions, the flagellar assembly and the chemotaxis pathway experienced upregulation and downregulation, respectively. Despite conferring motility, flagellar activities without chemotaxis may not cause biofilm dispersion. A study of nutrient-induced biofilm dispersion in *Pseudomonas aeruginosa* showed that the dispersion is dependent on a chemotaxis regulator BdlA but not flagellar activity. Mutation of *flgB* did not affect dispersion towards nutrients (Morgan et al., 2006). Svensson et al. (2014) showed that *C. jejuni* 81-176 flagellar facilitates biofilm formation through motility-driven adhesion and the secretion of extracellular DNA.

The presence of 5% chicken exudate may promote additional biofilm formation,

as observed even for the *flaAB* mutant. That suggests the content in the chicken exudate may facilitate cell adhesion even in the absence of flagellar (Brown et al., 2014). That also agrees with proteomic analysis of strain 11168, which showed increased proteomic levels of flagellins, filament cap and basal body in biofilm cells compared to planktonic cells in the stationary phase (Kalmokoff et al., 2006).

The relationship between *C. jejuni* chemotaxis and biofilm formation remains debatable. In some cases, chemotaxis and biofilm formation seem correlated. For example, *C. jejuni* autoinducer 2 (AI-2)-mediated quorum sensing was inhibited by (QS) 2(5H)-Furanone, epigallocatechin gallate, and a citric-based disinfectant. The disruption of quorum sensing coincides with reduced biofilm formation (Castillo et al., 2015). Biofilm formation of strain 81-176 was reduced by Δtlp8, while no significant change was observed was observed in Δtlp6, Δtlp4 and Δtlp10 (Chandrashekhar et al., 2015a). Nonetheless, *C. jejuni* NCTC11168 forms showed chemotaxis towards fucose while also demonstrating less biofilm formation (Dwivedi et al., 2016). In addition, Δ*cheW* and Δ*cheV* exhibited reduced chemotaxis and increased biofilm formation (Tram et al., 2020). Both cheW and cheV exhibited downregulation in hyperosmotic stress, chicken juice and cold stress. The downregulation of *cheW* and *cheV* in these conditions further suggests the likelihood of biofilm formation. In another study in strain 11168, mutation of Cj1564 (Tlp3), which leads to reduced chemotaxis towards chemoattractants such as aspartate and isoleucine, also resulted in increased autoagglutination and biofilm formation (Rahman et al., 2014). All these suggested that different chemotaxis genes/substrates have different effects on biofilm formation, and the effect varies across various strains.

Some biofilm cells may persist under the VBNC state. In cold stress, *C. jejuni* transition from a spiral and more motile form into an inactive, coccoid shape VBNC state for long term survival (Rollins and Colwell, 1986). NCTC11168 strains incubated at 4 °C, biofilm cells of NCTC 11168 V1 and NCTC 11168 V26 and16-2R all enter VBNC at a faster rate than their planktonic counterparts (Magajna and Schraft, 2015).

Biofilm formation under hyperosmotic stress requires further confirmation. In contrary to our suggested model, the addition of NaCl inhibits biofilm formation in strain M129 (Reeser et al., 2007). However, that may not be true for strain NCTC11168. Moreover, the cells were cultured using Muller-Hinton broth (Difco) which was not cation adjusted, unlike the data in this study obtained from cation adjusted MH2 broth.

It remains unclear what signals the transcription of CjSA21. One possibility is that the low temperature and water efflux perturb the cell envelope structure. The cells sense the membrane stress and subsequently signal biofilm formation via CjSA21. The TCS CprRS (Campylobacter planktonic growth regulation) senses cell envelope stress and is essential for osmotolerance and oxidative stress tolerance. A deletion mutant of *cprS* showed reduced survival under 1% NaCl. However, *cprS* mutation leads to increased biofilm formation (Svensson et al., 2009). A previous study suggested that the deletion of the CprRS regulon enhances biofilm formation by inducing envelope stress. *C. jejuni* cannot form biofilm in the absence of CprR or CprR phosphorylation residue Asp52. Meanwhile, deletion of CprS showed improved biofilm formation when subjected to cell envelope stress and reduced expression of *cprR* and *htrA* and *peb4*, which are involved in envelope biosynthesis. The enhanced biofilm formation of *cprS* deletion is rescued by adding Mg^{2+}, which stabilises the cell envelope. This observation suggests that *cprS* deletion reduced CprR phosphorylation which perturbed envelope formation by *htrA* and *peb4*. The envelope stress thus leads to enhanced biofilm formation. Hence CprRS regulon does not directly contribute to biofilm formation but instead does so indirectly by causing envelope damage (Svensson et al., 2015). That also suggested that an alternative sensing mechanism is required to translate envelope stress to increased biofilm formation. Moreover, whether CprRS can sense cold stress or chicken juice remains unclear. Furthermore, if CjSA21 responds to cell envelope stress, then it is unclear why CjSA21 was not particularly over-expressed under both sodium deoxycholate conditions (sod_deoxy_M and sod_deoxy_ES).

Ppk may stabilise CjSA21 targets

Another putative component of CjSA21 signalling is *ppk* (Cj1359) which showed strong negative co-expression against CjSA21. *ppk* belongs to the polyphosphate kinase 1 (PPK1) family and catalyses the reversible poly-phosphate (poly-P) formation from the ATP terminal phosphate. *C. jejuni* 81-176 *ppk1* deletion mutants demonstrated increased biofilm formation (Candon et al., 2007; Gangaiah et al., 2009). Moreover, *ppk* deletion did not affect the motility of *C. jejuni* 81–176, despite showing upregulation of several flagella-associated genes and flagellar glycosylation genes (Chandrashekhar et al., 2015b). Thus *ppk* opposes CjSA21-regulated biofilm formation under food storage conditions by repressing flagellar-mediated cell adhesion.

RNase in module II

The following RNase genes are within module II: Cj0121, *rnc*, *rnhA*, Cj1710c. Apart from Cj0121, all other three RNases showed significant negative co-expression against *tlp1-4* and might be involved in CjSA21-mediated mRNA degradation. Among these candidates, Cj1710c is an RNase J that consists of both RNase E–like endoribonucleolytic and a 5'-to-3' exoribonucleolytic activity responsible for messenger RNA maturation and degradation. Crystal structures of *Thermus thermophilus* RNase J showed similar active site orientation and C-terminal domain architecture as RNase E (Li de la Sierra-Gallay et al., 2008). Interestingly, sRNA-mRNA duplexes can activate RNase E to cleave the mRNA (Bandyra et al., 2012), which can explain the negative correlation between the abundance of CjSA21 and *tlp* mRNAs. Similarly, the deletion of *H. pylori* RNase J (*rnj*) did not affect the abundance of ncRNAs (Redko et al., 2016). Hence, Cj1710c appear to be the most likely candidates that degrade *tlp* mRNAs without degrading CjSA21 together.

Rnc promotes sRNA maturation by cleaving premature sRNA. RNase III deletion mutants of *Streptococcus pyogenes* with RNase III caused differential expression of 6 putative sRNAs with over +2 fold change and p-value $<= 0.05$. This result suggests Rnc's role in sRNA degradation. However, *rnc* abundance appeared to be independent of CjSA21 degradation, as *rnc* was also upregulated by food storage conditions (Rath et al., 2017). Hence, it is more likely that *C. jejuni rnc* promotes maturation of CjSA21 (by cleaving off the upstream regions?) and thus allows CjSA21 to target and destabilise *tlp1-4* mRNAs.

Meanwhile, Cj0121 (YbeY) is a metallo-endoribonuclease that targets single-stranded substrates precisely and plays a role in rRNA processing. YbeY is highly conserved among bacteria and is also responsible for sRNA metabolism. Mutation of Endoribonuclease YbeY increases ReaL sRNA abundance in *Pseudomonas aeruginosa* (Xia et al., 2020b). YbeY also regulates virulence-associated sRNAs in *V. cholerae*. Deletion of *ybeY* leads to an altered abundance of virulence-associated sRNAs such as Qrr1-4, MicX and TarB (Vercruysse et al., 2014). However, YbeY-mediated mRNA degradation has yet to be observed in *C. jejuni*.

Another RNase, RnhB, is an RNase HII involved in DNA replication. Meanwhile, RnhA is an RNaseHI that targets RNA-DNA hybrids and degrades the RNA. None of them appears to be responsible for mRNA degradation.

This analysis discovered several putative sRNAs candidates and their potential regulatory mechanisms, one of which has a strong basis in chemotaxis regulation.

Further experimental confirmation is needed to validate CjSA21's role as a regulator of chemotaxis and stress response under food storage conditions. Figure 4.41 summarises a possible post-transcriptional regulatory mechanism mediated by CjSA21.

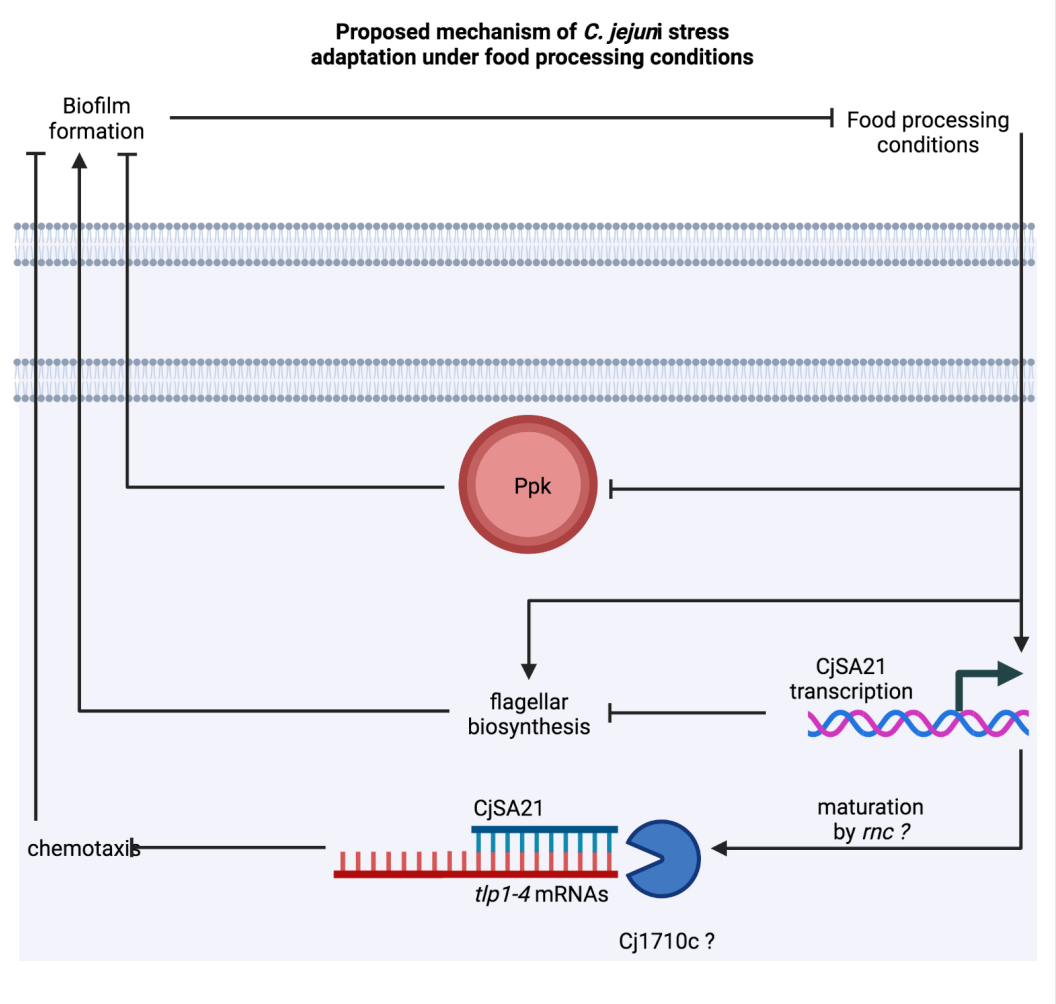

Figure 4.41: **A proposed model of CjSA21 upregulation by hyperosmotic stress, cold stress and chicken exudate.** CjSA21 then binds to *tlp1-4* (and maybe *tlp9* and *tlp10*) and inhibits their translation by either speeding up mRNA turnover or inducing structural rearrangement. Inhibiting *tlp* translation resulted in lower chemosensing. That allows increased biofilm formation, which enhances survival under stress. Cj1710c and *rnc* may promote mRNA degradation and sRNA maturation, respectively. Stress may also promote biofilm formation by upregulating flagellar biosynthesis and downregulating *ppk*. Created in BioRender.com

4.5.3 CjSA110 may regulate iron stress response genes

Previous knowledge about CjSA110 (CJnc190)

CjSA110 shares similar coordinates as CJnc190 and is antisense to CjSA109 (CJnc180). CJnc180 and CJnc190 regulate PtmG to regulate infection in a three-dimensional intestinal tissue model (Alzheimer et al., 2020). SDS-PAGE shows that double deletion of CJnc180 and CJnc190 leads to an increased level of PtmG. The WT level of PtmG was restored upon complementation of CJnc190 alone. Gel-shift assay further confirmed that a predicted C/U-rich loop region of CJnc190 directly block the RBS of PtmG, as a mutation in either one the stem-loop region or the RBS leads to increased PtmG level (Svensson and Sharma, 2021).

In this study, *ptmG* is not within module VIII. *ptmG* (Cj1324, a colonisation factor, in module I) also shows moderate positive co-expression with CjSA14 (Cjnc10), CjSA47 (3' UTR of *rplA*) and CjSA48 (3'UTR of *rplL*). The absence of *ptmG* in module VIII is possible because the regulation is based on translation instead of mRNA turnover. IntaRNA prediction in this study also showed no significant binding between CjSA110 and *ptmG*. Disparities could result from slightly different sRNA boundaries or differences in 5' UTR estimation from previous studies since the 5' UTR here was defined based on the Cappable-seq TSS data.

CjSA110 only shows a weak/moderate correlation with *tonB3* and Cj1384c. Such weak/moderate correlation values may not be an issue, as eukaryotic miRNA-mRNA also form weak/moderate negative correlation (Dai et al., 2019). Indeed, bacterial sRNA-mRNA interactions do not necessarily show a strong correlation as sRNA may only regulate the target's structure without changing the transcript abundance. Another reason for such correlation coefficients is that post-transcriptional regulation by sRNA is multifactorial. For instance, post-transcriptional regulation by CjSA110 might be overridden by other sRNAs or other regulatory factors for iron stress response (e.g. Fur and PerR), thus creating a discrepancy between co-expression patterns and other datasets. In pairwise comparisons such as the late stationary phase against early stationary phase under 37 °C (37_LS _vs_37_ES), iron uptake transporters showed no downregulation despite the lack of differential expression for CjSA110.

RNase degradation in module VIII

It is known that *rnc* is the RNase responsible for CjSA110 maturation. Deletion of *rnc* leads to a higher level of pre-Cjnc180 (the more extended version) and a complete

absence of mature, shorter CJnc190 (Svensson and Sharma, 2021. As an RNaseIII, *rnc* also cleaves double-stranded RNAs, including sRNA-mRNA hybrids.

Another intriguing candidate is Cj1209, which belongs to module VIII and is the only mRNA with a predicted binding energy value against CjSA110 with significant FDR. The weak correlation between CjSA110 and Cj1209 may explain why they were not connected in the WGCNA network despite being clustered into the same co-expression module. This overall pattern suggests the likelihood that Cj1209 is responsible for the degradation of *tonB3* and Cj1384c. That may indicate CjSA110 targets Cj1209 mRNA and stabilises it, leading to more Cj1209 production for degrading *tonB3* and Cj1384c. CjSA110 may also target *tonB3* and Cj1384c and provide a scaffold for degradation. CjSA110 targeting multiple mRNAs may also explain its relatively weak correlations with *tonB3*, Cj1384c and CjSA110.

Predicted binding mechanism of CjSA110

CjSA110 appears to bind to multiple targets using both ends. Both binding regions appear to have relatively low probabilities of forming secondary structures. That corresponds with the *E. coli* sRNA Spot 42. Deletion analysis showed Spot 42 targets multiple messenger RNAs. *In silico* prediction identified 3 binding sites which are partially single-stranded (Beisel and Storz, 2011).

According to the integrative analysis, CjSA110 binds to the coding region or the 3' end of the mRNAs of *tonB3* and Cj1384c. That may further suggest the possibility of sRNA-mediated RNase degradation. 3'-5' degradation can result from sRNA interacting with upstream elements such as the RBS. For instance, RyhB sRNA binds to the *sodB* RBS and induces *in vivo* transcript degradation at over 350 nucleotides downstream. That suggests RBS blockage might be followed by degradation, maybe by recruiting the RNA degradation machinery (Prévost et al., 2011).

Our result suggests binding interaction at the mRNA coding regions or towards the 3' end. While that is contrary to examples that involve physical interactions with the RBS, evidence also suggests that sRNA targets the coding regions or towards the 3' ends. For instance, the distribution of sRNA-mRNA interactions in *E. coli* K-12 MG1655 showed that most sRNA-mediated interactions (both empirical and predicted results) targeted the middle point of the coding regions), including those that are empirical. Their binding energy distribution is mostly around -10 to -20 kcal/mol, based on empirical and IntaRNA predicted values (Tello et al., 2018).

All results together indicate CjSA110 may interact and stabilise Cj1209 mRNA, leading to more translated Cj1209 for degrading *tonB3* and Cj1384c. CjSA110 may also target *tonB3* and Cj1384c and provide a scaffold for degradation.

Crosstalks with other pathways

Fe^{2+} triggers the Fenton/Haber-Weiss reaction, as it reacts with hydrogen peroxide to generate ROS such as the destructive hydroxyl radical. ROS production leads to oxidative stress, which damages [Fe-S] clusters, DNA, Cys/Met residues, and membrane lipid (Jang and Imlay, 2007). Also, under high oxidative stress, Fe^{2+} undergoes oxidisation into insoluble Fe^{3+}, which is more challenging to acquire. While iron is responsible for ROS production, it also acts as cofactors for oxidative stress response enzymes such as superoxide dismutase, superoxide reductase and heme iron (Ueda et al., 2003). Apo-Fur acts as a repressor of the *sodB* gene. Fe^{2+} relieve repression of *H. pylori sodB* (Carpenter et al., 2009). Hence it is reasonable to expect crosstalk between the iron-uptake and the oxidative-stress resistance network.

CjSA110 showed opposite differential expression patterns under various stress conditions other than iron limitation, despite being clustered together with other iron ABC transporters. That suggests CjSA110 might be responsible for the crosstalk between the iron limitation and other stress responses. Notably, CjSA110 and *tonB3* and Cj1384c showed opposite differential expression under oxidative stress (oxidative), as well as cold stress (cold) and 5% chicken exudate (5%_ce). Both cold and 5%_ce involved storing or culturing the cells at low-temperature where oxygen solubility increases. So such an observation may point to crosstalk between the iron stress response and the oxidative stress response. The differential expression pattern at cold stress and chicken exudate agree with (Bronowski et al., 2017), which shows significant upregulation of *katA* and *sodB* in 4 °C. The result illustrated the crosstalk between cold stress adaptation and iron acquisition.

4.5.4 Ways to improve sRNA-target network construction

While this chapter has highlighted a list of high-confidence sRNA-target interactions, several improvements can improve the prediction. For instance, the prediction requires finding an sRNA-target pair from the same co-expression module. The assignment of co-expression modules can be improved further. WGCNA partition genes with similar expression patterns and network topology using hierarchical clustering. However, module assignment by hierarchical clustering is irreversible. As a result, after incorrectly clustering a gene into a co-expression module, it is impos-

sible to reassign that gene into a different module. For example, instead of being assigned to module VIII, PerR was assigned to module II despite being functionally related to oxidative stress and iron homeostasis. PerR shows weak or moderate correlations with module II and module VIII genes. Hence it is arguable that PerR can be reassigned to module VIII instead. In another example, *ppk* is assigned to different modules as CjSA21 despite showing a strong negative correlation. The issue with module assignment may prevent the discovery of some actual sRNA-target interactions. Future work can improve WGCNA module assignment using a recently published k-module algorithm, which adds a step to reassign co-expression modules after the WGCNA hierarchical clustering step (Hou et al., 2021).

Another drawback of this chapter's approach is that it relies on the target transcript abundance changing according to sRNA expression. Such an assumption may not be valid and may miss out on binding interactions that have no impact on transcript stability. Examples include interactions that prevent ribosome access or induce secondary structure rearrangement. Moreover, co-expression can result from co-regulation or indirect transcriptional regulation instead of a direct sRNA-mRNA physical interaction. Although such a problem can be addressed by RNA-RNA binding prediction using IntaRNA, it may not be completely accurate as stable binding energy may not be optimal for sRNA-mRNA interactions. For example, transient and condition-specific binding may favour less stable binding. Therefore, IntaRNA may fail to recognise partially complementary base-pairing mediated by RNA chaperone.

The next chapter will identify more sRNA-target interactions experimentally by RNA crosslinking to address the possible shortfall of *in silico* prediction of sRNA-target interactions. The results will allow a comparison between experimental findings and *in silico* predictions.

4.6 Summary

This chapter has explored the in-house RNAtag-seq data and has identified pathways, conditions and gene clusters that play vital regulatory roles in stress survival. Further integrative analysis has generated a putative global sRNA-target interaction network. CjSA21 stood out for associating with mRNA targets for chemosensing. A detailed analysis of CjSA21 suggested it may inhibit chemosensing under food storage conditions. Gene expression patterns under food storage conditions further hint towards increased biofilm formation, contributing to survival under adverse conditions. This chapter also explored CjSA110 (CJnc190) that suggests crosstalk between iron limitation and other stress responses. While this chapter has established an sRNA-target regulatory network based on bioinformatics analysis, such an approach may miss out on physical interactions that pose no effect on transcript levels or involve partially complementary sequences. Such a limitation has been complemented by *in vivo* RNA crosslinking in Chapter 5.

Chapter 5

RNA crosslinking identifies sRNA-mRNA interactome

5.1 Introduction

Among all validated *C. jejuni* sRNAs, only the interaction involving the CJnc190-CJnc180 antisense pair has been experimentally confirmed (Svensson and Sharma, 2021). While the binding targets of most sRNAs remain unknown, no established high-throughput experimental approaches are available for determining *C. jejuni* sRNA binding activities.

Analysis of RNAtag-seq data highlighted sRNA-mRNA duplexes based on computational prediction (refer to chapter 4). The prediction identified sRNA-mRNA pairs with significant overall co-expression and condition-specific differential expression patterns. Moreover, the computational analysis also included genome-wide target prediction that estimates optimal binding activities with stable base-pairing energy. Despite identifying interactions that induce transcript stabilisation or degradation, such an approach may overlook interactions that only induce structural changes or prevent ribosome association or when the sRNA is degraded together with the mRNA. The assumption of genome-wide target prediction may also neglect *in vivo* interactions with nascent transcripts or binding activities facilitated by RNA chaperons.

This chapter involves performing and optimising *in vivo* RNA crosslinking to identify sRNA-mRNA interactions experimentally. The experimental results will complement the information obtained from the computational prediction. RNA-RNA

pairs identified in this chapter may highlight sRNA-mediated mechanisms that do not alter mRNA stability.

5.1.1 Experimental and bioinformatics challenges for RNA crosslinking

AMT crosslinking and proximity ligation methods are capable of detecting short- and long-range RNA-RNA interactions. Briefly, after culturing cells under specific growth conditions, the psoralen derivative AMT was added to the cells to induce covalent linkage among native RNA-RNA duplexes upon exposure to 365 nm UV light. The crosslinking precedes cell harvesting and total RNA extraction. The RNA samples are then treated with DNase to remove all DNA contamination. Afterwards, rRNA depletion and S1 nuclease digestion minimise rRNA and single-stranded RNA background noise. The processed RNA undergoes proximity ligation to ligate both ends of crosslinked RNA pairs. Proximity ligation produces chimeric RNA species with hybrid sequences of more than one gene. Then, 254 nm irradiation reverses the crosslinked duplex into linear chimeric RNA species. RNase R treatment removes all linear RNA sequences in the samples before the library preparation step that converts the remaining linear chimeric RNA into cDNA for sequencing (Figure 5.1).

This approach has revealed the global RNA-RNA interactome in several experimental models. For example, the LIGR-Seq protocol in human cells has revealed the global interactions between novel small nucleolar (sno) RNAs and mRNAs (Sharma et al., 2016). There are other variants of LIGR-Seq that map intermolecular RNA duplex in eukaryotic cells, with similar steps except for different chimeric RNA enrichment strategies. Examples of such variants include Psoralen Analysis of RNA Interactions and Structures (PARIS) (Lu et al., 2016), and Sequencing of Psoralen crosslinked, Ligated, and Selected Hybrids (SPLASH) (Aw et al., 2017). Similar approaches have also been deployed in *E. coli* (Liu et al., 2017) and in *Bacillus subtilis* (Durand et al., 2021). However, no published studies have applied RNA crosslinking on *C. jejuni*. Practical concerns include how efficiently AMT can permeate through the *C. jejuni* membrane and the choice of rRNA depletion protocol. Furthermore, chimeric reads are challenging to detect due to background noise from rRNA contamination and uncrosslinked RNAs. Hence, a careful examination of AMT uptake and rRNA depletion is required to avoid wasting resources on sequencing uncrosslinked RNA samples or rRNA species.

Sequencing crosslinked RNA species results in chimeric reads that align to multiple

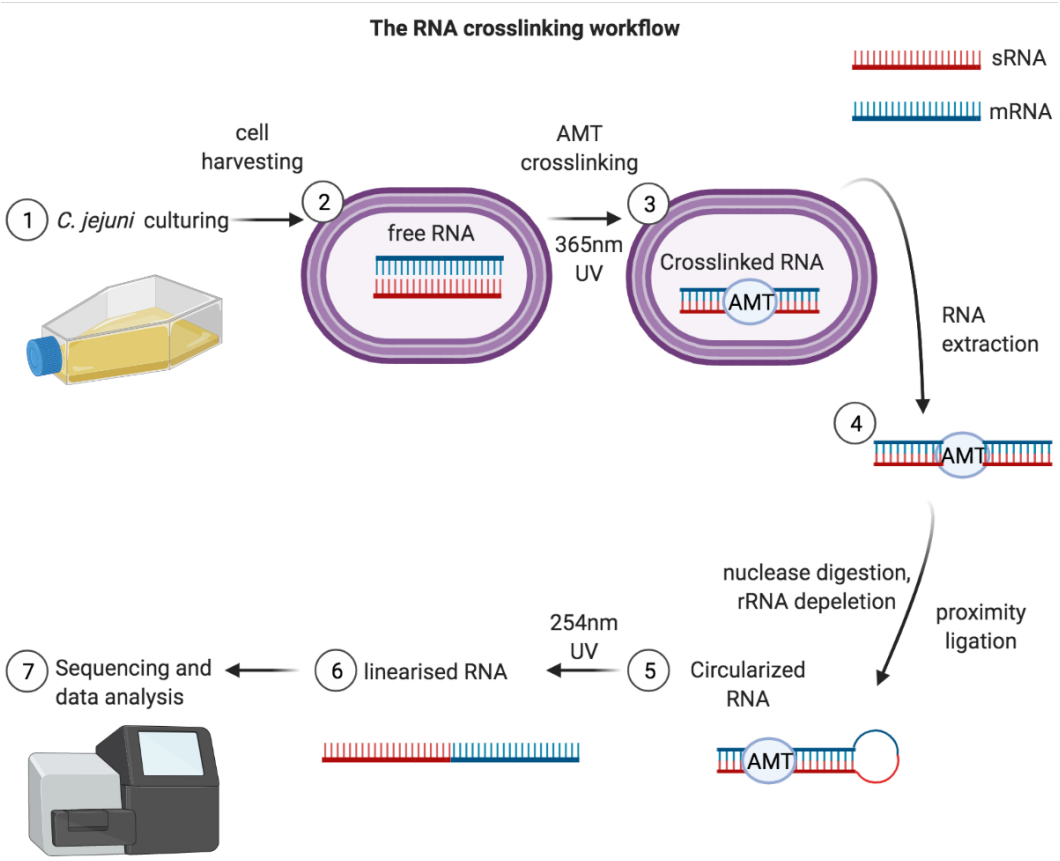

Figure 5.1: **A simplified overview of the RNA crosslinking protocol**. Created in BioRender.com.

genes. Identifying chimeric reads is challenging due to the complexity of mapping the reads to features distant from each other. Some chimeric reads may consist of sequences from two genes with the hybrid junction towards one end. This pattern results in a long sequence for one gene and a short sequence for the other gene. It is easier to map the more extended sequence to its respective gene. However, the shorter sequence is more likely to be clipped out without being aligned elsewhere. Hence, bioinformatics analysis of crosslinking data requires alignment tools with high sensitivity to reads with split sequences.

Various alignment tools have varied sensitivity for detecting chimeric reads. In a study that evaluated the performance of intron-exon junction detection, STAR stood out for being more accurate than other alignment tools such as GSNAP, TopHat2 and HISAT2. That was not surprising as the primary purpose of STAR is to detect intron-exon junctions (Dobin et al., 2013; Boratyn et al., 2019). STAR's chimeric read detection setting enhances chimeric read detection. In addition, replacing STAR's pair-ended alignment mode with single-ended alignment mode can improve the sensitivity of chimeric read detection (`https://github.com/alexdobin/STAR/issues/333`).

Another difficulty with identifying chimeric reads is multi-mapping reads. Some chimeric reads may align to repetitive sequences such as rRNA or conserved coding regions. For example, as mentioned in chapter 4, the three Tlp proteins Tlp2, Tlp3 and Tlp4 all have identical C-terminal amino acid sequences. A read aligned to *tlp2* can originate from *tlp3* or *tlp4* instead. Sample contamination from other species may also introduce RNA species mistaken for hybrid sequences from multiple genes or vice versa. Therefore, BLASTn analysis of identified chimeric reads is necessary to ensure accurate read alignment.

The software featureCounts can quantify the STAR aligned reads mapped to genomic features and extract the number of sRNA-mRNA read pairs for downstream statistical analysis and evaluation. The reads mapped to both an sRNA and an mRNA are extracted for downstream statistical analysis and evaluation. This pipeline has contributed to authorship for a publication on *Bacillus subtilis* (Durand et al., 2021). Several differences exist between this study and Durand et al., 2021. The pipeline for analysing the *Bacillus subtilis* data did not include any demultiplexing and adapter trimming steps. The previous study generated *Bacillus subtilis* Illumina MiSeq data, which generated the data in a fastq format and automatically processed to remove adapters. As the data in this study was generated using the Illumina NextSeq platform, converting the data from BCL to fastq formats was

necessary before demultiplexing and adapter trimming (See the method section). Moreover, the two sets of data consist of different negative controls and numbers of replicates. Hence, the previous pipeline needed some modifications to address the data obtained from this study.

This chapter will discuss optimisation of both experimental protocols and data analysis workflows for identifying RNA interactions in *C. jejuni*. The identified sRNA-mRNA interactions will also be evaluated based on their co-expression and differential expression patterns in our 21 conditions RNAtag-seq dataset using the same analysis workflow as chapter 4. The experimentally obtained data will also be compared with the computational predictions described in chapter 4 to determine identical interactions or discrepancies.

Chapter Aims:

- Apply and optimise RNA crosslinking to *C. jejuni* culture and the subsequent bioinformatics analysis.
- Identify sRNA-mRNA duplexes from crosslinking data.
- Compare the results of RNA crosslinking with computational predictions in chapter 4.

5.2 Optimising the RNA crosslinking protocol and bioinformatics analysis

5.2.1 Cell culturing and RNA extraction

LIGR-Seq aims to detect covalently-linked RNA pairs induced by AMT addition followed by UV exposure (AMT + UV). However, such an approach is susceptible to background noise. Background noise can arise in the absence of either UV illumination or AMT supplementation, as ligases can ligate uncrosslinked RNA close to each other. Consequently, this study detected background ligation by omitting UV light (+AMT) or AMT (+UV) from the protocol. All samples consisted of three replicates.

Table 5.1: **Qubit measurements of extracted total RNA.** The amount of extracted RNA was sufficient for the subsequent steps of RNA crosslinking.

sample	**concentration (ng/ul)**
AU1	2640
AU2	2500
AU3	2270
A1	3650
A2	1140
A3	3210
U1	2580
U2	3930
U3	2680

The extracted RNA concentrations all exceeded 1000 ng/ul. Most samples except A2 had concentrations above 2000 ng/ul (Table 5.1). All samples have sufficient extracted RNA for the subsequent DNase treatment. The DNase treatment removed most DNA contamination and retained RNA with concentrations above 250 ng/ul (Table 5.2). The bioanalyzer results showed limited degradation of rRNA (Figure 5.2 and 5.3). The retained RNA concentration and integrity were both sufficient for downstream processing.

Table 5.2: **Qubit mesurements of DNase treated RNA.**

sample	concentration (ng/ul)
AU1	296
AU2	368
AU3	340
A1	342
A2	538
A3	266
U1	322
U2	324
U3	290

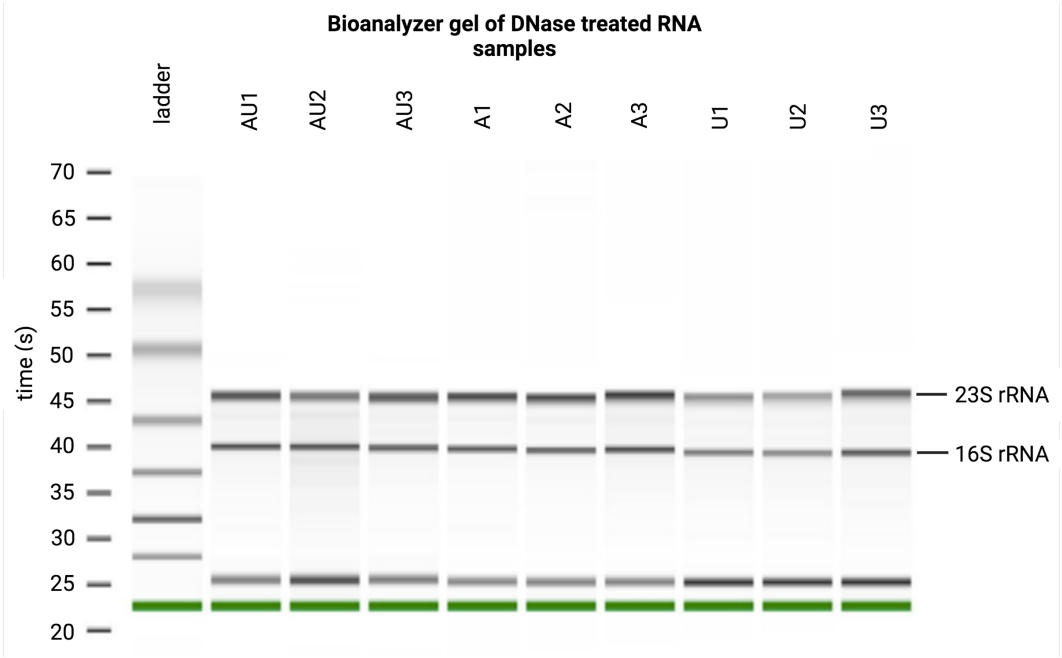

Figure 5.2: **Bioanalyzer gel of all samples after total RNA extraction.** The results indicated limited RNA degradation after DNase treatment, suggesting the samples were in sufficient quality for downstream processing.

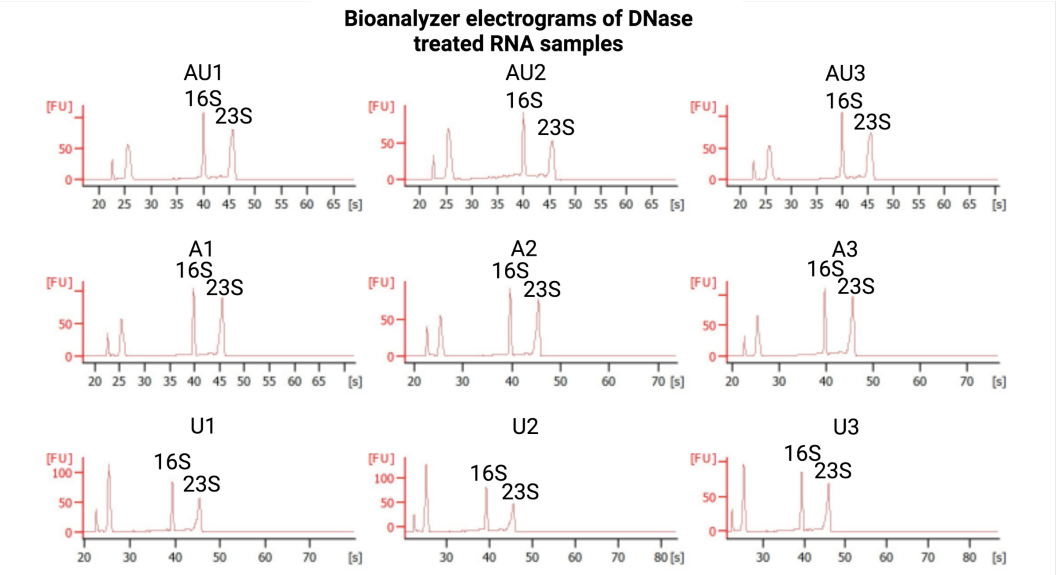

Figure 5.3: **Bioanalyzer electrograms of all samples after total RNA extraction.** Similar to the previous figure, electrograms showed limited RNA degradation after DNase treatment and RNA quality sufficient for downstream processing.

5.2.2 Confirming crosslinking with the Bioanalyzer

Successful crosslinking requires AMT entering the cytoplasm and inducing RNA duplex formation. Not much was known about the ability of AMT to penetrate the *C. jejuni* cell wall and induce crosslinking, as no publication has recorded AMT crosslinking on *C. jejuni* to our current knowledge. Hence, this study applied RNase R digestion on the RNA samples to estimate the efficiency of *in vivo* crosslinking. RNase R selectively degrades linear RNAs. Hence, free RNAs are usually more susceptible to RNase R digestion. In contrast, crosslinked RNA duplex is more resistant to RNase R digestion (Sharma et al., 2016) (Figure 5.4).

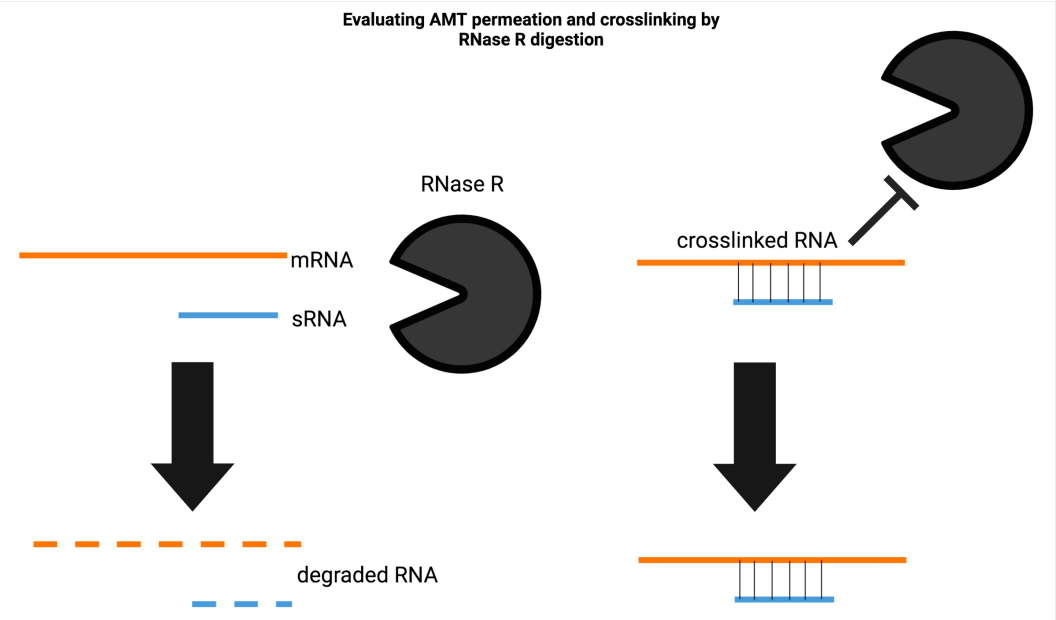

Figure 5.4: **The evaluation of AMT crosslinking efficiency by RNase R digestion.** Crosslinking resulted in RNA duplexes that are more resistant to RNase R degradation. Hence, the resistance against RNase R degration indicate the effectiveness of RNA crosslinking.

Bioanalyzer estimated the extent of RNase R digestion. For the negative control without AMT (+UV), the 16S and 23S rRNA peaks became visibly smaller after RNase R digestion. The 5S rRNA peak, which corresponds to small RNA fragments, became several folds larger than the 16S and 23S rRNA peaks. That suggested that 16S and 23S rRNAs exhibited more degradation in the absence of AMT. In comparison, 16S and 23S rRNA peaks of AMT + UV remained more intact and larger (Figure 5.5). That indicated a heightened resistance to RNase R digestion due to successful crosslinking after AMT penetrated through the *C. jejuni* cell wall.

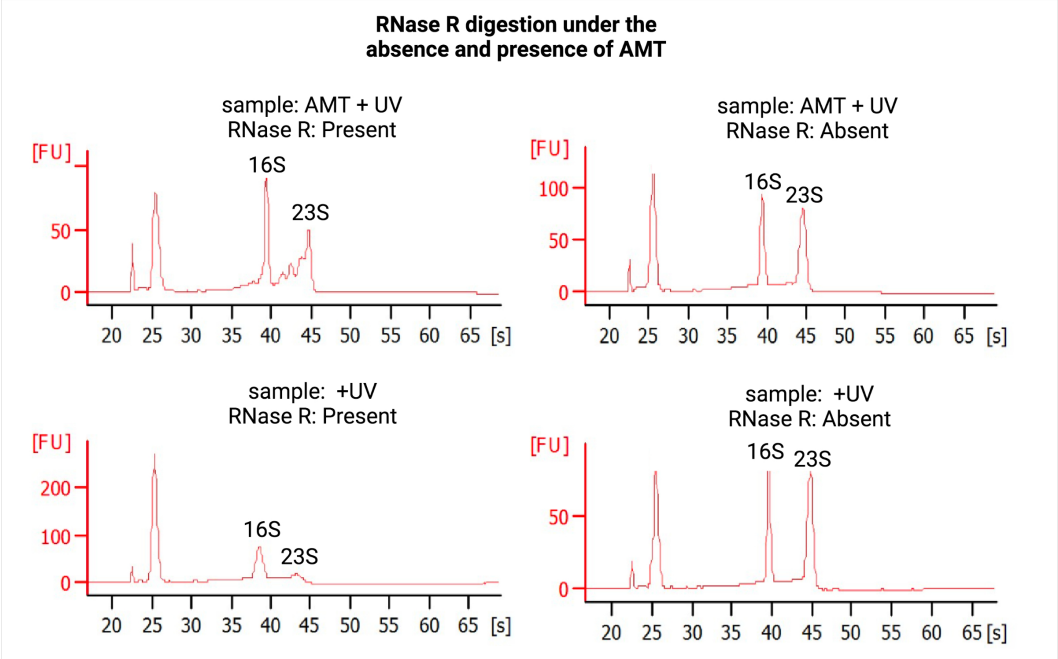

Figure 5.5: **Bioanalyzer analyses of RNase R digestion in the absence and presence of AMT crosslinking.** The crosslinked samples showed increased resistance against RNase R degradation, suggesting the RNA crosslinking step was successful.

5.2.3 Optimisation of the rRNA depletion step

The Illumina Ribo-Zero kit, a gold standard kit for rRNA depletion, was discontinued during this project. Like Ribo-Zero, other stand-alone commercial kits also capture and remove conserved rRNA sequences by hybridisation to DNA probes. Examples include MICROBExpress (Ambion) and RiboMinus (Life Technologies). In a comparative evaluation of rRNA removal kits, RiboMinus showed more efficiency at bacterial rRNA removal and was more cost-effective than MICROBExpress (Petrova et al., 2017). Therefore, RiboMinus emerged as a cheaper and more effective alternative to Ribo-Zero.

Ideally, rRNA depletion would remove as many ribosomal sequences as possible. Hence, the Bioanalyzer results of rRNA depleted samples should significantly reduce the sizes of the 5S, 16S and 23S peaks. After the RiboMinus treatment, all negative controls displayed smaller 16S and 23S peaks. The Bioanalyzer result excluded U1 due to the low amount of remaining U1. Nonetheless, the results on U2 and U3 suggested rRNA depletion was efficient for the +UV samples. However, AU1-AU3 all retained taller rRNA peaks (Figure 5.6 and 5.7). The rRNA removal could be less efficient due to secondary structures induced by crosslinking, which might mask the RiboMinus probe recognition site.

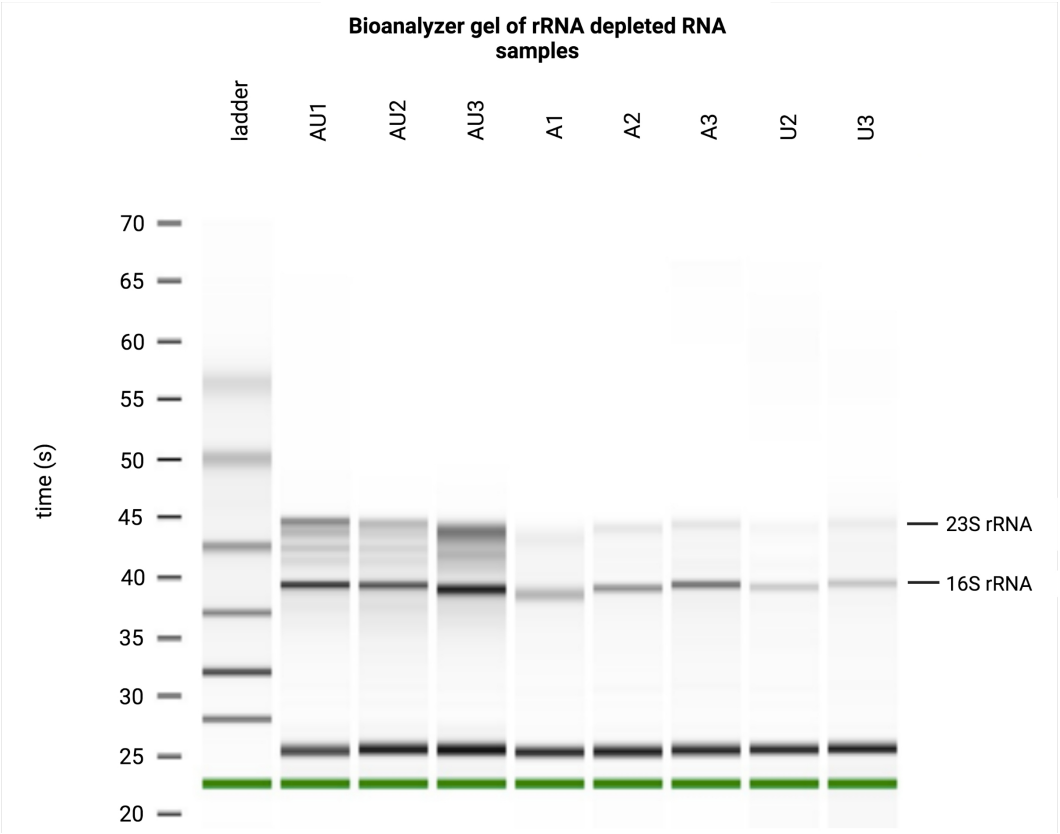

Figure 5.6: **Bioanalyzer gel of all samples after RiboMinus rRNA depletion.** Crosslinked samples showed less rRNA removal than the uncrosslinked samples.

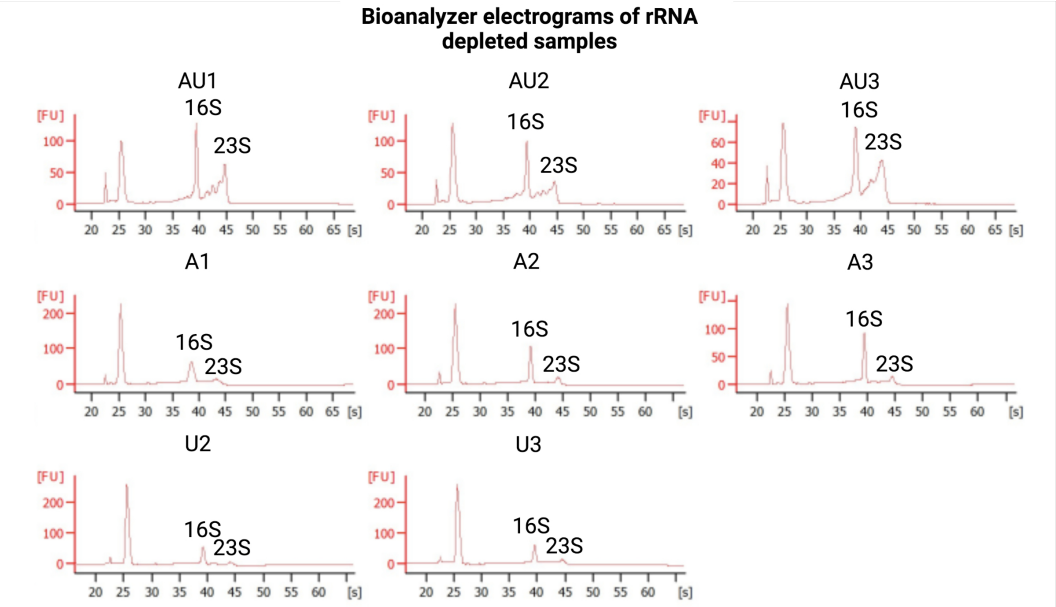

Figure 5.7: **Bioanalyzer electrogram of all samples after RiboMinus rRNA depletion.** Crosslinked samples showed less rRNA removal than the uncrosslinked samples.

RNAseq of the samples processed by the crosslinking protocol provided a more precise estimation of rRNA removal. In order to assess the efficiency of the ribosomal RNA depletion, we looked at the distribution of mapped reads aligning to 5S, 16S, 23S, tRNA or any other genomic feature (sRNA or mRNA). The portion of reads aligned to sRNA or mRNA was particularly crucial, as this study aimed to identify sRNA-mRNA interactions. Hence, the expectation was that depletion of ribosomal RNA would lead to enrichment of reads mapping to mRNA/sRNA. RNAseq results demonstrated the incomplete removal of rRNA. 99.2 % of the reads for AMT + UV were either rRNA or tRNA. In contrast, 5.11 % and 9.13 % of reads aligned to either sRNAs or mRNAs in negative controls (Figure 5.8a). Since chimeric reads tend to be rare, additional rRNA removal steps were crucial to detect more sRNA-mRNA interactions.

Since RiboMinus did not remove rRNA efficiently, other approaches were needed. One alternative was to target the rRNA after cDNA preparation, as cDNA is free from crosslinked secondary structures and RNA degradation. Depletion of Abundant Sequences by Hybridisation (DASH) is an example of a cDNA-based rRNA depletion approach; DASH involves designing sgRNAs targeting human ribosomal cDNA and recruiting Cas9 proteins for cDNA cleavage. PCR cannot amplify

cleaved cDNA fragments, thus only enriching non-ribosomal sequences (Gu et al., 2016). Recent studies have also applied DASH to reduce rRNA contamination from *Salmonella enterica serovar Typhimurium* and *Bacteroides thetaiotaomicron* (Prezza et al., 2020).

Moreover, all commercially available probes were designed based on rRNA sequences of model organisms, instead of specifically for *C. jejuni*. Hence, *C. jejuni* rRNA sequences were not fully complementary to those commercial probes. Such an issue could be overcome by DASH, as it can capture *C. jejuni* rRNA sequences by custom-designed sgRNA. Later, Illumina released Ribo-Zero Plus, a modified version of Ribo-Zero. However, Ribo-Zero Plus became available very late in the project, after all the work optimising RiboMinus and DASH. Hence, instead of replacing the other two methods, Ribo-Zero was included in addition to RiboMinus and DASH to maximise rRNA sequence removal.

The inclusion of DASH and Ribo-Zero Plus led to a higher rate of rRNA removal. Reads aligned to sRNA or mRNA accounted for nearly 10 % of total reads in AMT + UV. About half of the reads mapped to either sRNA or mRNA in both negative controls (Figure 5.8b). Apart from increasing the proportion of sRNA reads and mRNA reads, combining all three rRNA removal steps also increased the percentage of chimeric reads. Most of the reads mapping to ribosomal RNA were expected not to be hybrids. Therefore, ribosomal RNA depletion should result in the enrichment of chimeric reads. Such changes were apparent for AMT + UV, with the proportion of chimeric reads rising from 6.82 % to 44.8 %. More importantly, AMT + UV consisted of a higher fraction of chimeric reads than +AMT and +UV, which carried 28.5 % and 26.2 % of chimeric reads, respectively (Figure 5.9a and 5.9b).

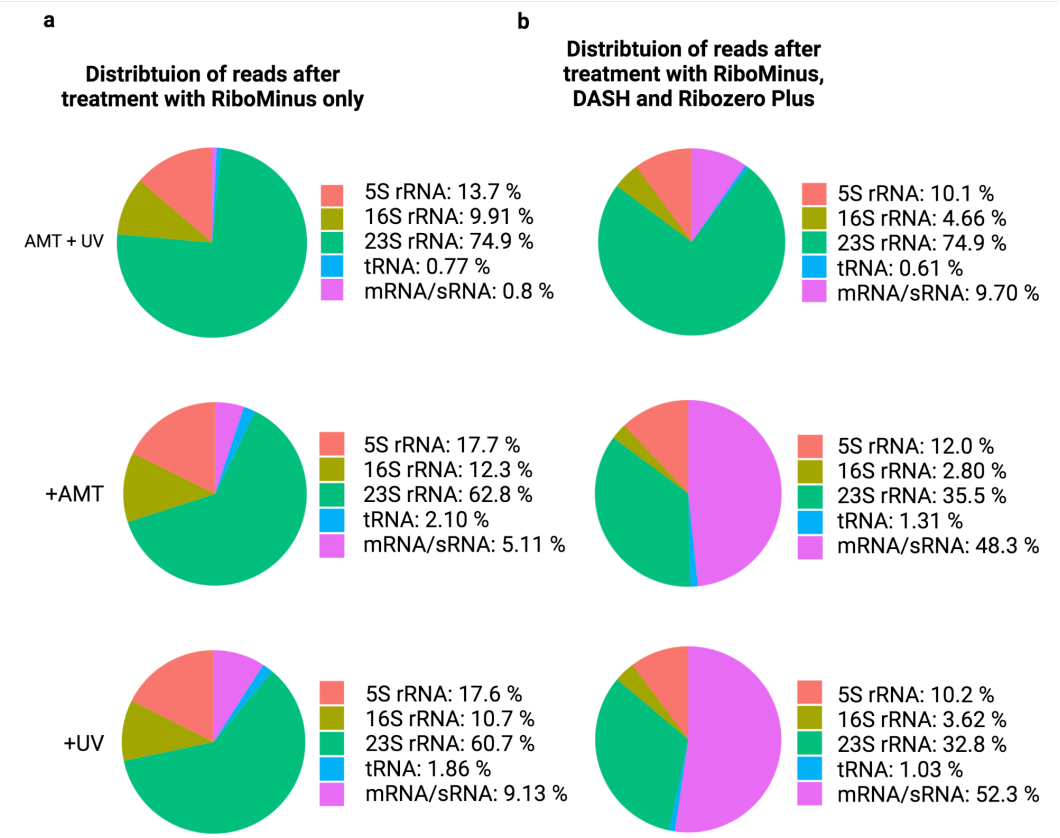

Figure 5.8: **The distribution of RNA read types from samples before and after including DASH and Ribo-Zero Plus.** The addition of DASH and Ribozero Plus improved the enrichment of mRNAs and sRNAs, especially for the crosslinked samples.

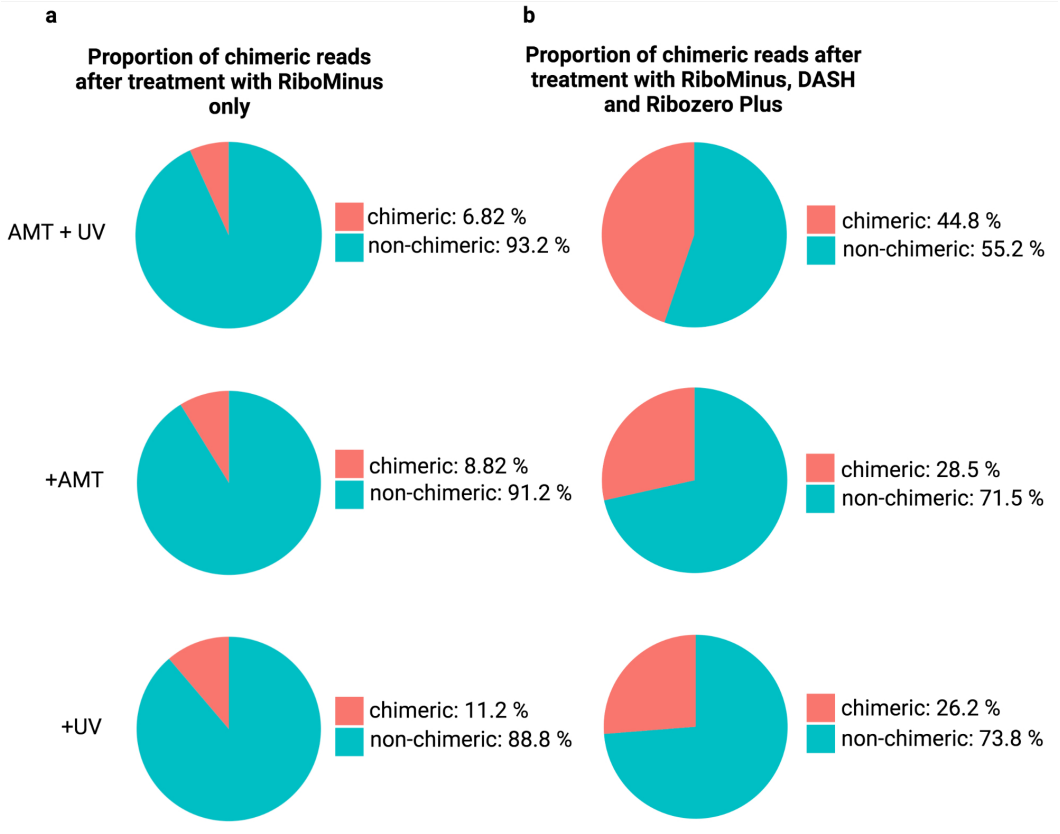

Figure 5.9: **The distribution of RNA read types from samples from samples before and after including DASH and Ribo-Zero Plus.** The addition of DASH and Ribozero Plus increased the portions of chimeric reads, especially for the crosslinked samples.

While adding DASH and Ribo-Zero Plus resulted in more sRNA and mRNA species, it was unclear if that also led to more sRNA-mRNA interactions. Hence, the quality of sequencing data also depends on the abundance of sRNA-mRNA duplexes. Among all chimeric reads, there was an elevation of sRNA-mRNA interactions. AMT + UV showed a 4-fold increase in sRNA-mRNA chimeric reads from 0.03 % to 0.12 %. DASH resulted in about 20 fold increase from 0.04 % to around 0.8 % in both negative controls (Figure 5.10a and 5.10b). However, rRNA-rRNA pairs remained the most abundant interactions among chimeric reads from all samples.

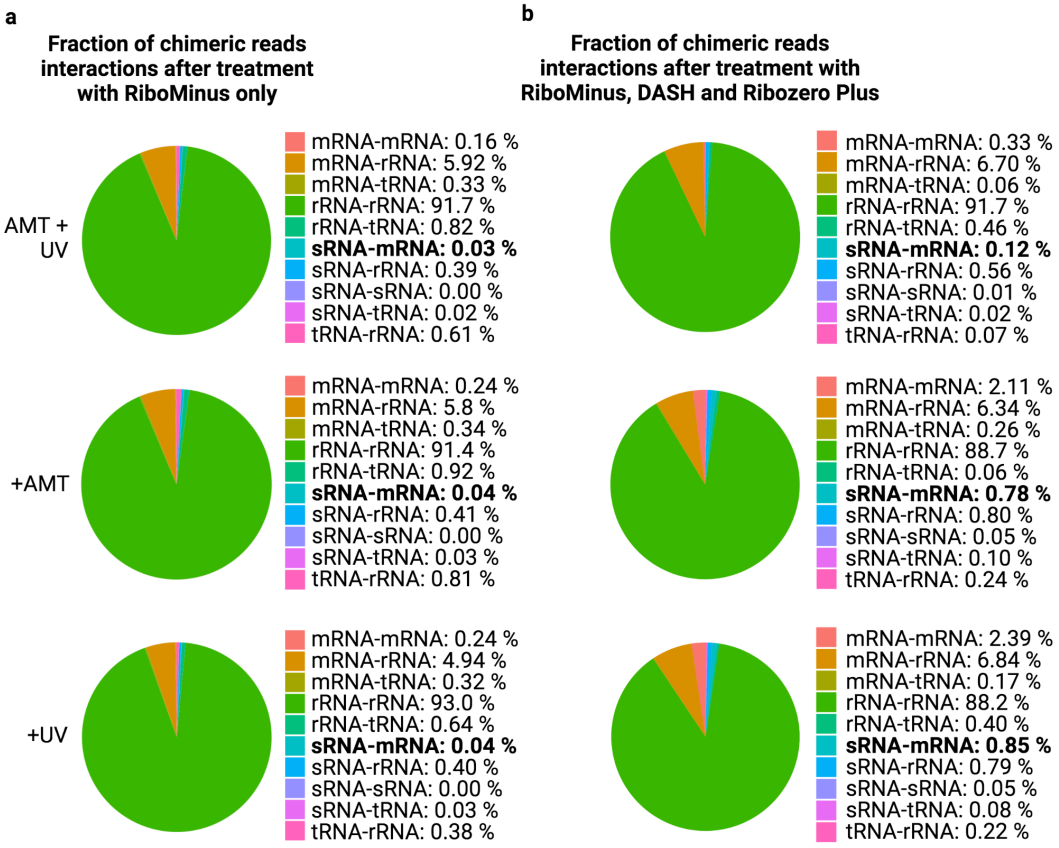

Figure 5.10: **The distribution of RNA read types from samples from samples before and after including DASH and Ribo-Zero Plus**. The addition of DASH and Ribozero Plus increased the portions of sRNA-mRNA duplexes among chimeric reads.

5.2.4 Sequencing data quality control

After treatment from RiboMinus, DASH and RiboZero-Plus, RNAseq data from one of the crosslinked replicates (AU1) displayed millions of unmapped reads, with less than 50 % of reads aligned to the *C. jejuni* genome (Table 5.3). The rest might either be contamination from other species or hybrid sequences created by crosslinking and ligation. In order to distinguish between contamination and chimeric RNA molecules, all unmapped reads were processed by BLASTn. The BLASTn search revealed the most unmapped reads of AU1 aligned to moth species such as *Galleria mellonella*. These contaminations may come from research conducted in the host laboratory, which involved infecting bacteria into *Galleria* worms. However, It was still possible that those unmapped reads were chimeric RNA that shared similar sequences as the *Galleria mellonella*. *De novo* assembly using SPAdes (Bankevich et al., 2012) followed by BLASTn search rules out the possibility of chimeric reads. Almost all assembled reads mapped to *Galleria* genomes but none to *Campylobacter* genomes. Apart from the abundant sample contamination in AU1, AU3 also showed a deficient number of reads, with only 15869 in total. Hence, read counts from all three replicates of each condition were combined to ensure enough reads for statistical analysis and quality control.

Table 5.3: **Number of sequencing reads and alignment rate.** The low alignment rate of AU1 was associated with contamination of reads from *Galleria worms*.

Samples	Sequencing reads	Read 1 alignment (%)	Read 2 alignment (%)
AU1	10473741	47.75	42.77
AU2	5109242	84.07	77.74
AU3	15869	81.49	75.09
A1	13594346	86.81	81.77
A2	5998528	86.95	82.3
A3	6701508	86.56	83.07
U1	13195283	88.23	85.2
U2	2789915	87.36	84.3
U3	10292631	86.53	79.34

5.2.5 PCR duplications

After merging the replicates, the statistical enrichment of chimeric reads were examined using hypergeometric test, just as in Durand et al., 2021. Hypergeometric test assumed that the abundance of crosslinked RNA pairs is directly proportional

to the number of aligned chimeric reads and hence more statistically significant. However, it was unclear if the hypergeometric test was appropriate for this study. Depleting rRNA in *C. jejuni* was more difficult than in *Bacillus subtilis*. Moreover, as mentioned earlier, the sequencing depth for the crosslinked samples was relatively low, especially due to the sample contamination in one of the replicates (AU1). Another replicate (AU3) also had very few sequencing reads. As a result, the number of chimeric reads that represented sRNA-mRNA interactions was less abundant for *C. jejuni*. Due to the low number of non-rRNA chimeric reads, some sRNA-mRNA duplexes mapped to only 1 or 2 reads. Some sRNA-mRNA duplexes were statistically significant even when mapped to only one or two reads. Replacing the hypergeometric test with the binomial test, as suggested by Sharma et al. (2016), produced less significant p-values from duplexes with very few reads compared to the hypergeometric test. Nevertheless, a large number of RNA pairs with only 1 or 2 reads showed significant p-values (Figure 5.11). Hence, both statistical tests were prone to false positives.

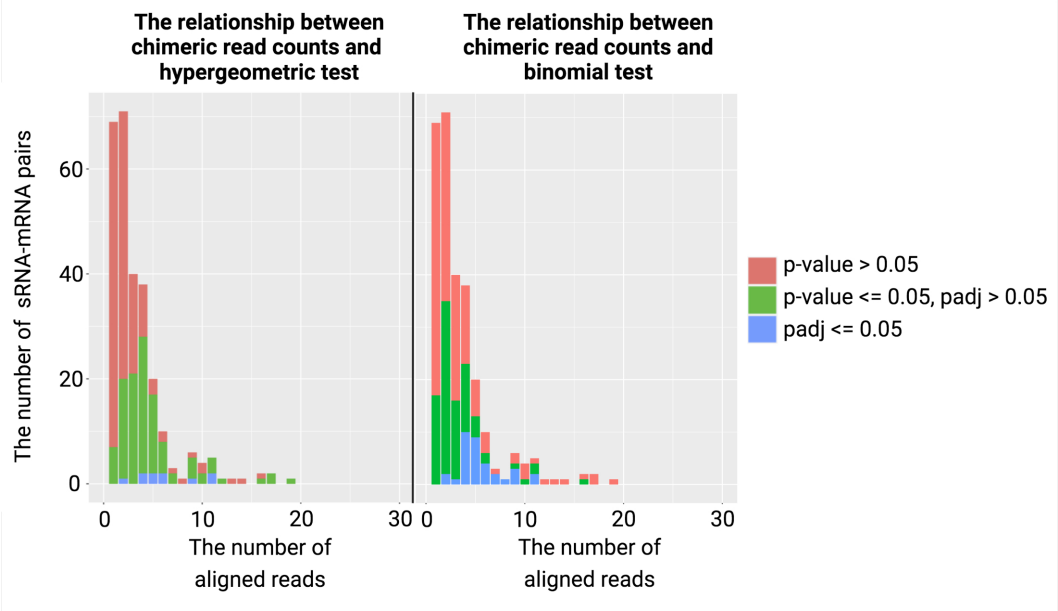

Figure 5.11: **The distribution of *p*-values for chimeric read counts.** The number of chimeric reads aligned to each sRNA-mRNA pair and their statistical significance when the (a) hypergeometic or the (b) binomial test were used.

Another uncertainty with using statistical tests was whether statistical significance represents the *in vivo* abundance of crosslinked RNA molecules. Compared to the RNA crosslinking protocol of Durand et al. (2021), this protocol consisted of an extra step using DASH, which removed rRNA sequences after cDNA synthesis and

enrichment. The reverse transcription reaction may fail to convert some non-rRNA species into cDNA in the crosslinked RNA saturated with rRNA. cDNA conversion of samples saturated with rRNA would yield fewer non-ribosomal species for DASH and the subsequent PCR amplification. After DASH digested most rRNA species, the remaining non-rRNA sequences would be excessively amplified by PCR, leading to more PCR duplicates in the sequencing outputs. As negative controls RNA samples contained fewer rRNA molecules, removing rRNA by DASH would leave more non-rRNA sequences as templates for subsequent PCR amplification. Therefore, there should be fewer PCR duplicates in the negative controls compared to the crosslinked samples. Nonetheless, introducing the DASH treatment prior to the PCR step would still result in plenty of PCR duplicates as DASH may reduce the diversity of PCR templates.

PCR duplicates with Samtools confirmed the hypothesis above. After removing duplicate PCR reads, AU1 - AU2, which were more saturated with rRNA, retained 1 % of reads. AU3 retained about 10 % of all reads but might be an outlier due to the meagre read count. In contrast, all negative controls had more than 1 % of reads. Some replicates such as U2 maintained over 3 % of reads after duplicate removal. Most other de-duplicated negative controls also carried over 2 % of their original reads (Table 5.4).

Table 5.4: **Number of sequencing reads before and after PCR duplicates removal.** The results suggested duplicates reads made up over 95% of all sequencing reads.

sample	**Read 1 before de-duplication**	**Read 1 after de-duplication**	**Read 2 before de-duplication**	**Read 2 after de-duplication**
es-AU1	6615113	45499	6615113	54609
es-AU2	5781680	37292	5383317	42405
es-AU3	18292	1403	16090	1683
es-A1	14974194	191266	14325606	236146
es-A2	6695660	136587	6327947	171005
es-A3	7571508	204685	7213983	252928
es-U1	14740430	229994	14232119	281875
es-U2	3129377	107701	2975123	142727
es-U3	11298545	237916	10322626	309372

PCR duplication occured among sRNA-mRNA pairs aligned to multiple reads. For example, four chimeric reads aligned to Cj0920c shared identical coordinates with perfectly matched sequences. Such duplications led to some reads being more statis-

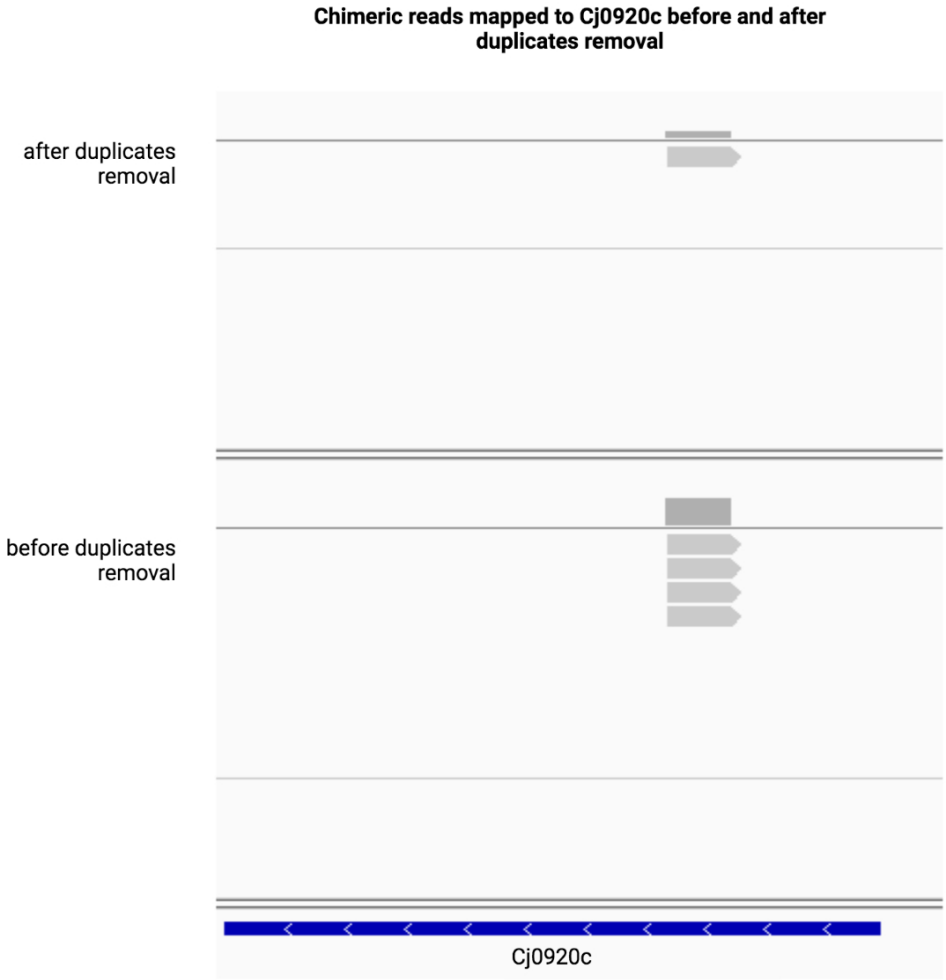

Figure 5.12: **The number of chimeric reads mapped to Cj0920c before and after PCR duplicates removal.** The bottom track was before removing the duplicates, while the top track was after the duplicate removal.

tically significant simply because of PCR duplication instead of being more abundant inside the cells. Other chimeric reads aligned to genes such as *rplN* also exhibited similar duplications (Figure 5.12 and 5.13). To further confirm the extent of PCR duplication, the reads were visualised again after duplicates removal using Samtools. As a result, both Cj0920c and *rplN* were aligned into a single chimeric read (Figure 5.12 and 5.13).

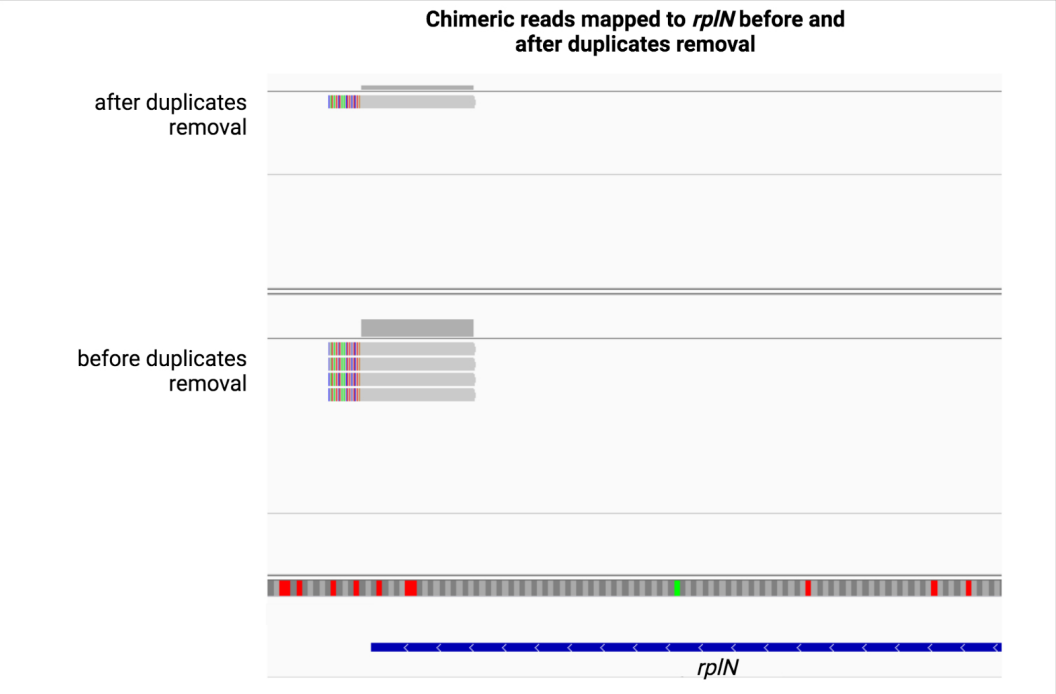

Figure 5.13: **The number of chimeric reads mapped to *rplN* before and after PCR duplicates removal.** The bottom track was before removing the duplicates, while the top track was after the duplicate removal.

The abundnace of PCR duplicates challenged the assumption that statistical significance represented the frequency of *in vivo* RNA-RNA binding events in this dataset. For example, *rnpB* were mapped by three groups of chimeric reads, and each consisted of 4, 2 and 2 reads (Figure 5.14). This result suggested that the chimeric read counts could not accurately represent the *in vivo* quantity of *rnpB*-mRNA interactions. Therefore, statistical tests were not optimal for analysing this dataset after PCR duplicates removal, as most remaining sRNA-mRNA chimeric reads would only have one read.

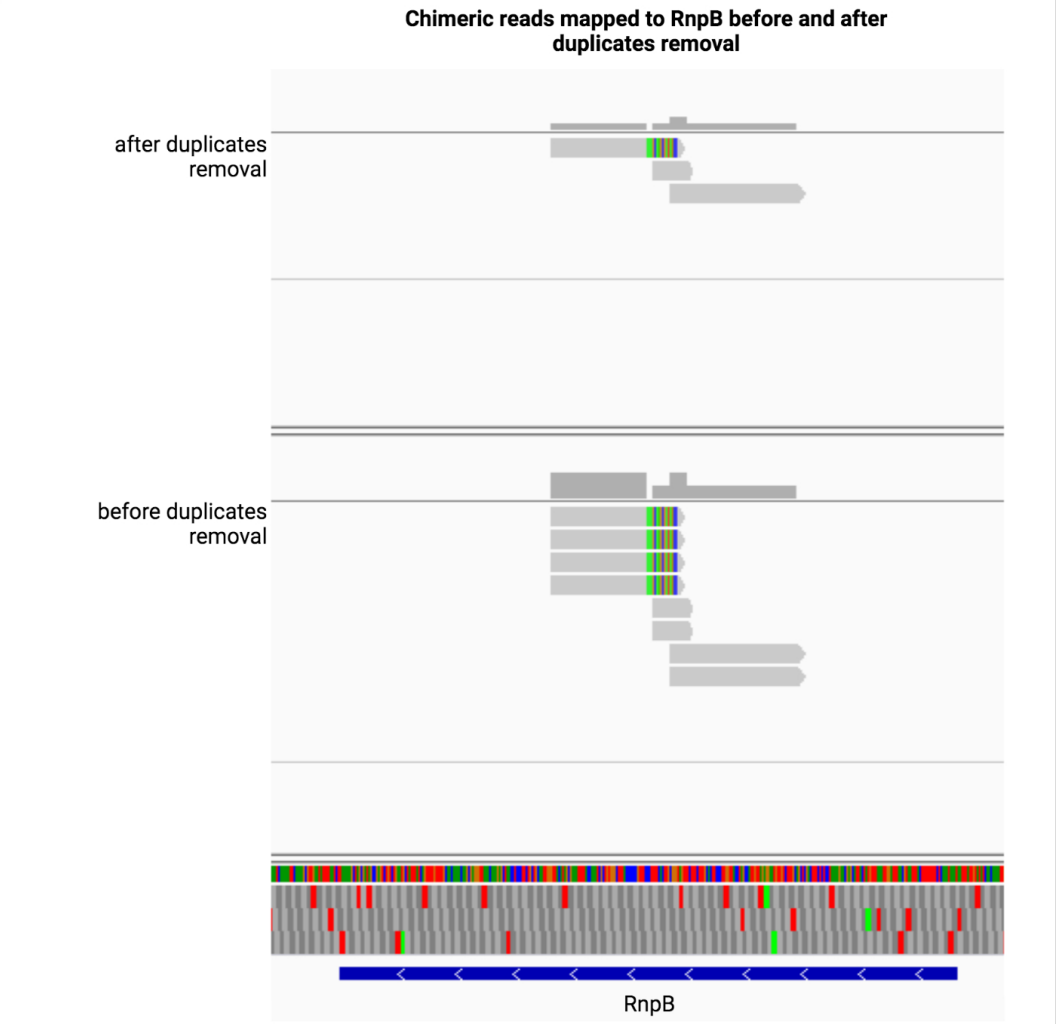

Figure 5.14: **The number of chimeric reads mapped to *rnpB* before and after PCR duplicates removal.** The bottom track was before removing the duplicates, while the top track was after the duplicate removal.

5.3 Understanding extracted sRNA-mRNA chimeric reads

5.3.1 Filtering chimeric reads with negative controls

Low read counts and PCR duplication were likely to hamper statistical analysis, so this study identified background noise using negative controls. sRNA-mRNA interactions in the crosslinked sample (AMT + UV) were filtered against interactions that appeared in either or both negative controls (+AMT and +UV). The crosslinked sample consisted of 285 sRNA-mRNA duplexes. Filtering with negative controls removed 235 (Figure 5.15), leaving 50 extracted pairs for further investigation (Table 5.5).

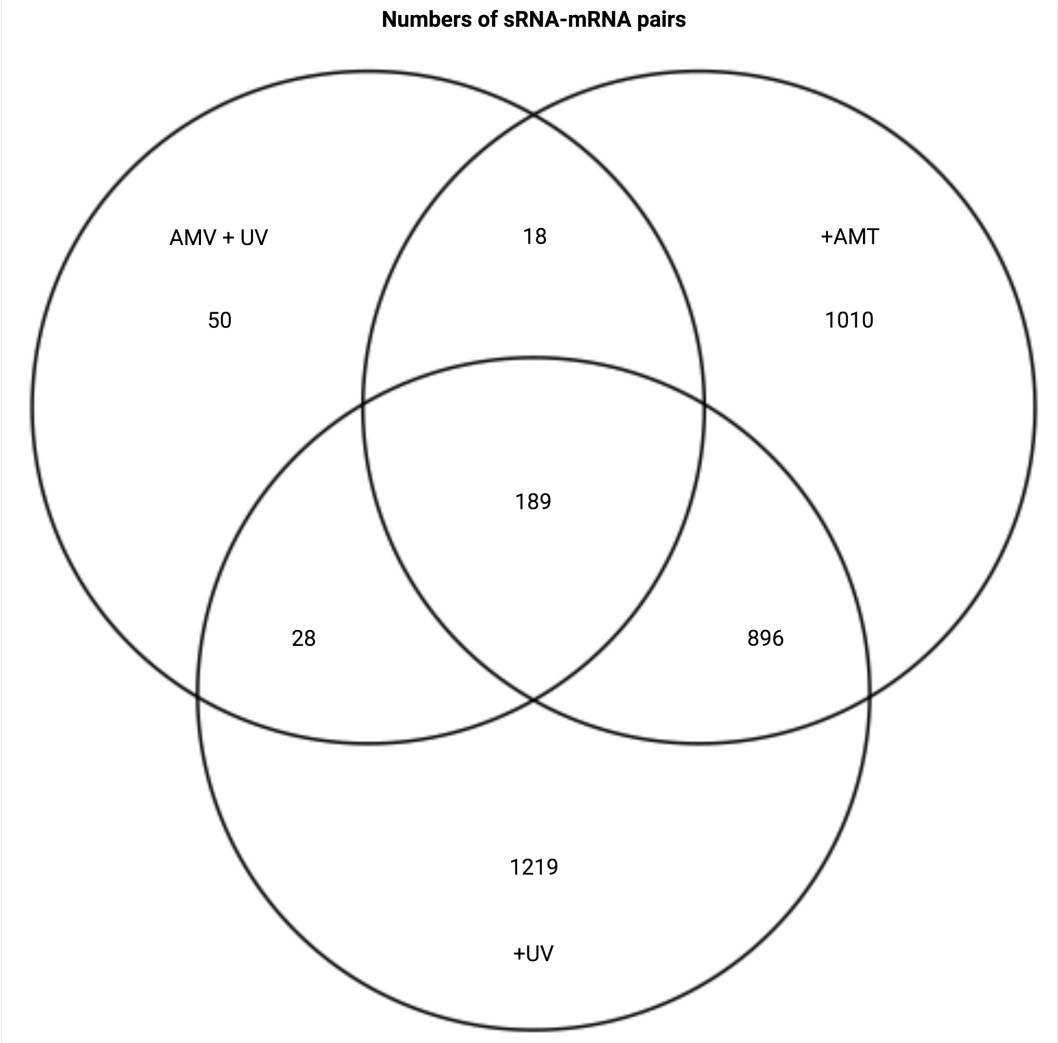

Figure 5.15: **Numbers of sRNA-mRNA duplexes in AMT + UV, +AMT and +UV.** 50 sRNA-mRNA pairs were retained for downstream analysis.

Table 5.5: **All extracted sRNA-mRNA pairs from RNA crosslinking data.** Their respective co-expression and IntaRNA predicted binding affinities are also included.

sRNA	**mRNA**	**bicor**	**padj**	**IntaRNA energy (kcal/mol)**	**IntaRNA *p*-value**	**IntaRNA FDR**
CjSA24	*guaA*	0.316	0.0281	-7.47	0.38	0.95
CjSA24	*rpsE*	0.0913	0.583	-4.12	0.83	0.997
CjSA36	Cj1632c	0.594	2.76E-06	-5.91	0.763	0.943
CjSA39	*atpG*	-0.046	0.794	-9.99	0.346	0.951
CjSA39	*ccoO*	-0.205	0.178	-7.86	0.667	0.951
CjSA39	Cj0262c	0.269	0.0673	-9.89	0.36	0.951
CjSA39	Cj0447	0.136	0.393	-7.33	0.744	0.951
CjSA39	Cj0457c	0.424	0.00202	-7.96	0.652	0.951
CjSA39	Cj0488	-0.244	0.101	-5.33	0.943	0.993
CjSA39	Cj1309c	-0.369	0.00877	-8.43	0.579	0.951
CjSA39	Cj1486c	-0.357	0.0117	-10.5	0.28	0.951
CjSA39	Cj1621	-0.0516	0.768	-8.69	0.539	0.951
CjSA39	Cj1709c	0.193	0.207	-6.42	0.855	0.971
CjSA39	*ffh*	0.309	0.0324	-9.9	0.359	0.951
CjSA39	*fliK*	0.624	5.43E-07	-9.28	0.448	0.951
CjSA39	*fur*	-0.219	0.147	-9.85	0.366	0.951
CjSA39	*galU*	-0.364	0.0097	-10.2	0.316	0.951
CjSA39	*glyS*	-0.333	0.0196	-9.34	0.439	0.951
CjSA39	*guaB*	-0.616	8.80E-07	-11.4	0.186	0.951
CjSA39	*hrcA*	0.302	0.0368	-9.49	0.417	0.951
CjSA39	*hslV*	-0.483	0.000311	-9.47	0.42	0.951
CjSA39	*ilvC*	0.268	0.0687	-7.77	0.681	0.951
CjSA39	*ndk*	0.433	0.00157	-7.53	0.716	0.951
CjSA39	*putA*	-0.0838	0.617	-11	0.228	0.951
CjSA39	*putP*	-0.342	0.0162	-8.6	0.553	0.951
CjSA39	*rbfA*	-0.3	0.0386	-5.37	0.941	0.993
CjSA39	*rplM*	0.762	1.86E-11	-5.87	0.906	0.982
CjSA39	*rpoD*	0.486	0.000276	-12.6	0.0959	0.951
CjSA39	*thiE*	-0.165	0.292	-8.14	0.624	0.951
CjSA49	Cj0760	0.401	0.00379	-9.95	0.0736	0.951

Table 5.5: **All extracted sRNA-mRNA pairs from RNA crosslinking data.** Their respective co-expression and IntaRNA predicted binding affinities are also included.

sRNA	**mRNA**	**bicor**	**padj**	**IntaRNA energy (kcal/mol)**	**IntaRNA *p*-value**	**IntaRNA FDR**
CjSA49	*fliK*	0.347	0.0144	-6.02	0.417	0.954
CjSA51	*gpsA*	0.609	1.26E-06	-11.3	0.216	0.959
CjSA51	*rplN*	-0.101	0.54	-7.34	0.729	0.96
CjSA56	*cheV*	0.589	3.55E-06	-11.5	0.226	0.945
CjSA74	Cj0715	-0.313	0.0299	-7.73	0.844	0.959
CjSA74	Cj0716	-0.403	0.00365	-9.88	0.612	0.913
CjSA74	Cj1026c	-0.161	0.304	-12	0.376	0.913
CjSA74	Cj1207c	-0.0432	0.807	-8.77	0.739	0.925
CjSA74	Cj1360c	-0.222	0.141	NA	NA	NA
CjSA74	Cj1419c	-0.408	0.00319	-13.5	0.257	0.913
CjSA74	*groES*	-0.0168	0.927	-13	0.296	0.913
CjSA74	*rpsU*	0.126	0.432	-11.1	0.473	0.913
CjSA76	*fabD*	-0.0679	0.691	-10.6	0.255	0.952
CjSA78	*fusA*	0.304	0.0357	-10.9	0.295	0.951
CjSA78	*tuf*	0.288	0.0484	-9.92	0.402	0.951
CjSA9	*carB*	0.108	0.51	-13.2	0.017	0.949
CjSA9	Cj0920c	0.143	0.366	-6.06	0.599	0.949
CjSA9	Cj1583c	-0.0424	0.811	-6.07	0.597	0.949
CjSA9	*rp10*	-0.284	0.0517	-5.94	0.618	0.949
CjSA97	*hypB*	0.568	9.58E-06	-5.61	0.89	0.979

None of the extracted crosslinked interactions appeared in the predicted sRNA-target network in chapter 4. This difference indicated different binding mechanisms. A closer look into the co-expression and predicted binding stability showed some crosslinked interactions showed significant co-expression but insignificant IntaRNA binding energy values. Most duplexes with statistically significant co-expression, such as CjSA56-*cheV*, CjSA51-*gpsA* and CjSA97-*hypB* that displayed positive co-expression (Figure 5.16). CjSA74 expression negatively correlated with Cj0715, Cj0716 and Cj1419c. These duplexes might involve suboptimal interactions facil-

itated by RNA chaperon or other RNA binding proteins. The binding activity resulted in the stabilisation or degradation of mRNA targets.

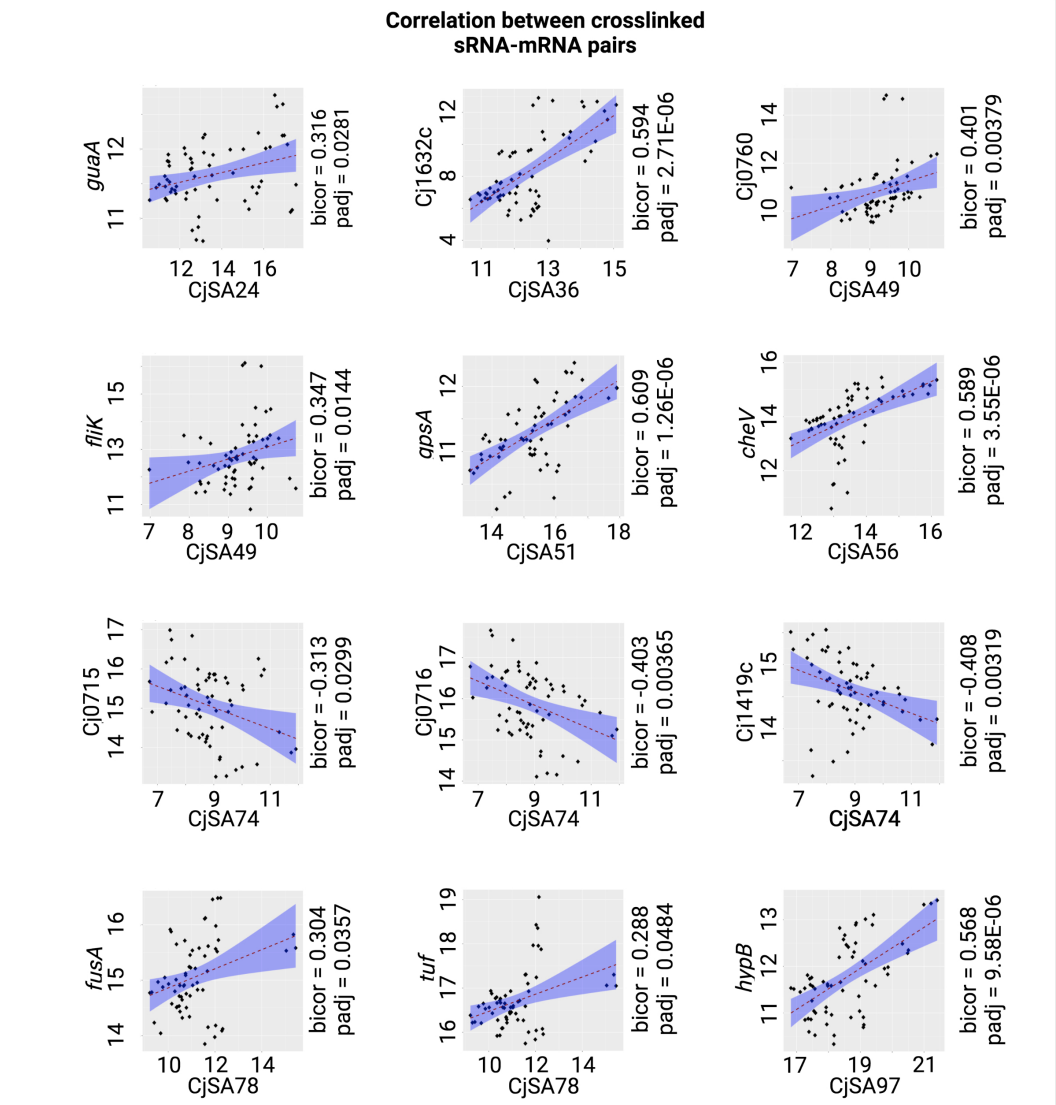

Figure 5.16: **Pairwise correlation of sRNA-mRNA duplexes with statistically significant correlation.** None of them appeared in the computational predictions presented in Chapter 4, suggesting RNA crosslinking and computational prediction had identified two distinct sets of sRNA-mRNA interactions.

CjSA9-*carB* was the only duplex that displayed significant IntaRNA p-value without exhibiting significant co-expression. CjSA9-*carB* may be an interaction with strong affinity and regulate translation initiation by blocking or exposing the RBS. Other interactions showed neither significant co-expression nor predicted binding energy. RNA chaperons might facilitate these interactions to regulate translation without

altering mRNA stability.

5.3.2 Alignment quality and sequence conservation of chimeric reads

Most reads have a MAPQ score of 255, which indicates a unique alignment. While CjSA39 was the most frequently occurring sRNA among all extracted interactions, the MAPQ score for CjSA39 was 1, indicating alignment to three or four regions (Table 5.6). The genomic coordinates of CjSA39 were upstream of a 5S rRNA copy (Figure 5.17a), representing a region that is highly conserved in all three copies of the ribosomal regions present in the C. jejuni genome. That also explains the observed strong co-expression of CjSA39 with ribosomal RNAs such as *rplM* (Table 5.5). Hence the reads mapping to CjSA39 were likely to be multi-mapping across all three 5S rRNA copies.

Table 5.6: **Chimeric reads with MAPQ scores that were not 255.**

sRNA	**mRNA**	**MAPQ**
CjSA39	*ccoO*	1
CjSA39	Cj0262c	1
CjSA39	Cj0447	1
CjSA39	Cj0457	1
CjSA39	Cj0488	1
CjSA39	Cj1309c	1
CjSA39	Cj1486c	1
CjSA39	Cj1621	1
CjSA39	Cj1709c	1
CjSA39	*ffh*	1
CjSA39	*fur*	1
CjSA39	*glyS*	1
CjSA39	*guaB*	1
CjSA39	*hrcA*	1
CjSA39	*hslV*, *rplI*	1
CjSA39	*ilvC*	1
CjSA39	*ndk*	1
CjSA39	*putA*	1
CjSA39	*rbfA*, *infB*	1
CjSA39	*thiE*	1
CjSA74	Cj1419c	3

One of the CjSA74-mediated interactions has a MAPQ score of 3, indicating alignment to two locations (Table 5.6). The ambiguous alignment might come from the conserved sequence from the tRNA-Gly copy upstream (Figure 5.17b). The seven remaining duplexes involving CjSA74 yielded the MAPQ score of 255, suggesting

the reads mapped to unconserved sequence on CjSA74.

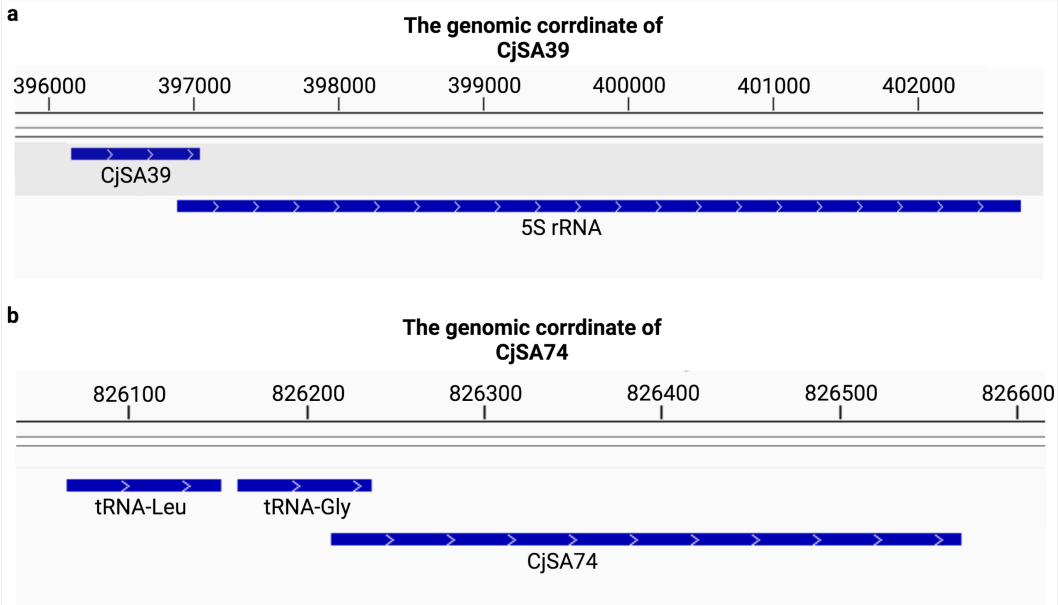

Figure 5.17: **Genomic coordinates of sRNAs and mRNAs involved in multiple alignments.** (a) CjSA39 and a 5S rRNA copy and (b) CjSA74 and a tRNA-Gly copy.

There is a possibility that those chimeric reads were contaminants from other species instead of crosslinked RNA pairs. If they were contaminants, they would be found in genomes from other species and showed only partial identity to the *C. jejuni* genome sequence. Some chimeric read alignments revealed some soft-clipped portions, indicating those regions may originate from contaminants of another species. These possibilities were tested by conservation analysis of all extracted chimeric reads (Table 5.5). 310 chimeric reads were involved in extracted interactions. Most extracted chimeric reads were PCR duplicates. 89 reads showed BLASTn hits for non-*Campylobacter* genomes from various sources, with 71 of them involving CjSA39. A further comparison between those 89 reads and the *C. jejuni* NCTC11186 genome showed almost 100 % identity to the query sequences, including the soft-clipped regions, suggesting they are highly conserved *C. jejuni* features such as rRNA and housekeeping genes.

5.3.3 Crosslinked interactions with low RNAtag-seq expression

All pairs from Table 5.5, except those involving CjSA39, were carried forward for further analysis. Some of the interactions observed might result from background crosslinking and ligation, especially among pairs with highly expressed sRNAs and

mRNAs. Conversely, any duplex with low sRNA and mRNA expression was more likely to represent a true biological association.

During the crosslinking protocol, any RNA molecules that have not crosslinked and are single-stranded are removed through a nuclease digestion step. Hence, the crosslinking dataset cannot accurately estimate gene expression. We, therefore, looked at the transcript abundance derived from the RNAtag-seq dataset (refer to chapter 4). The RNAtag-seq sample grown at 37 °C to early stationary phase (37_ES) (OD_{600} at around 1.575) was the closest to the RNA crosslinking sample. In this sample, among all sRNAs, CjSA97, CjSA24 and CjSA51 exhibited the highest TPM expression, indicating they were more likely to be background noise. Conversely, CjSA74 and CjSA9 exhibited the lowest TPM expression. Among the six CjSA74 targets, Cj0760 (*rpsU*) had the second-lowest TPM. In contrast, Cj0716 and Cj1419c displayed relatively high TPM levels above 1000. Notably, Cj1360c and Cj0715, two of the CjSA74 crosslinked targets, displayed the two highest TPM levels among all crosslinked mRNAs. On the contrary, all five CjSA9-crosslinked targets had TPMs below 1000, with *carB* expression being the third-lowest of all genes and *carB* (a carbamoyl-phosphate synthase large chain) and Cj1583c having a TPM less than 10. These results suggested that CjSA9 interactions were the least likely to originate from background noise (Figure 5.18 and 5.19).

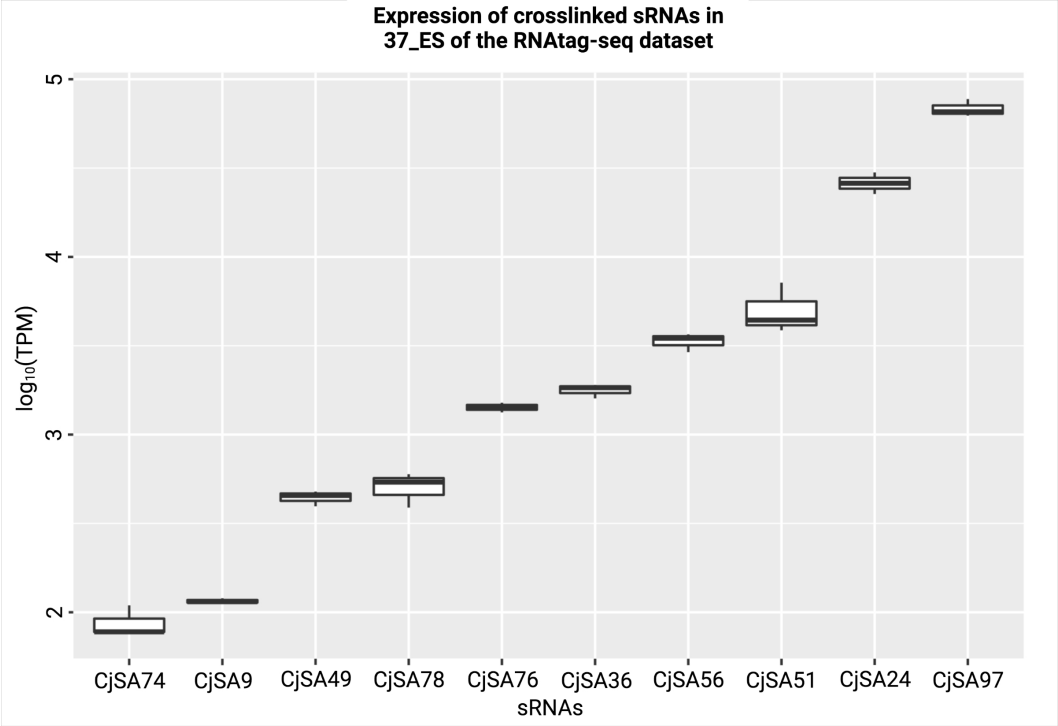

Figure 5.18: **TPM expression of extracted sRNAs in 37_ES of the RNAtag-seq samples.** CjSA74 and CjSA9 showed the lowest expression levels in the 37_ES condition of RNAtag-seq. That suggested that their presence in the crosslinking data was the least likely to be the result of expression background noise.

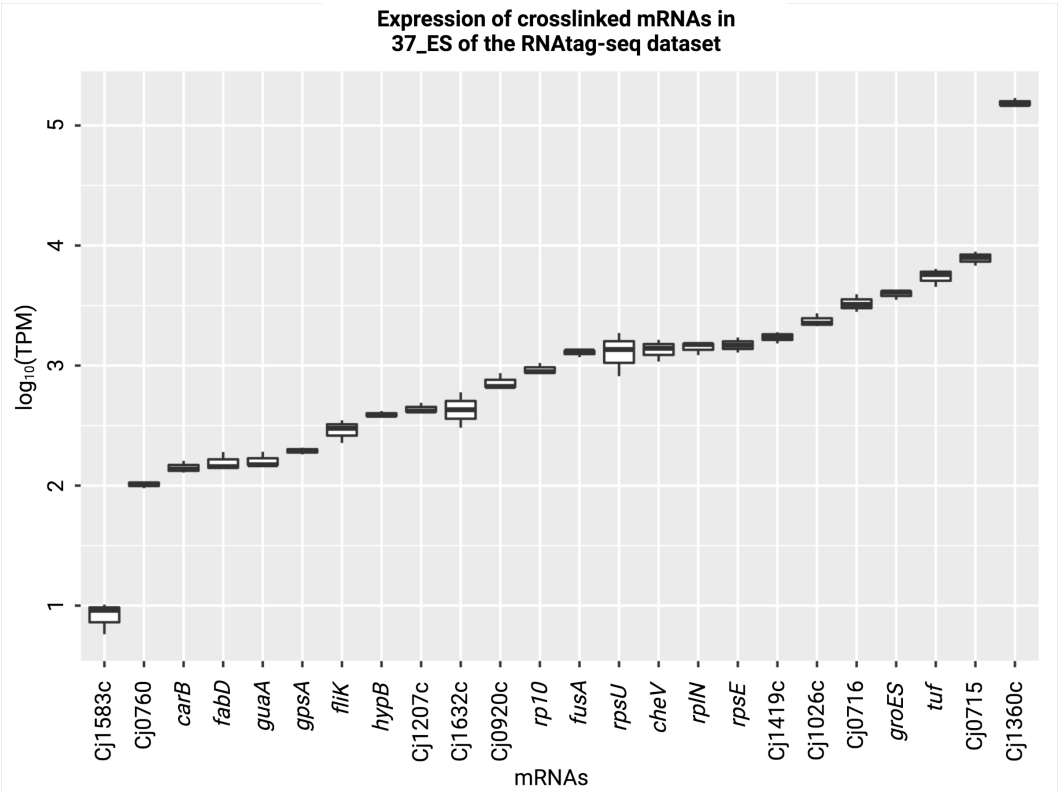

Figure 5.19: **TPM expression of extracted mRNAs in 37_ES of the RNAtag-seq samples.** CjSA9 targets (Cj1583c, *carB*) were among the three mRNA targets with lowest expression levels in the 37_ES condition of RNAtag-seq. This further suggested interactions involving CjSA9 were the least likely to be results of expression background noise under the early-stationary phase.

5.3.4 Common results between crosslinking and computational prediction

When comparing the computational predictions from chapter 4 and results in Table 5.5, CjSA9 was also the most outstanding sRNA. In chapter 4, CjSA9 was one of the four sRNAs with targets overrepresented in the KEGG pathway enrichment analysis. The presence of *rpoC*, *purL* and *purH* resulted in the statistical enrichment of the "Purine metabolism" pathway (Table 5.7), which is an important pathway for *C. jejuni* stress adaptation. For instance, exposure to oxygen and hyperosmotic stress led to frequent genetic variation in two purine biosynthesis genes, *purF* and *apt* (Cameron et al., 2015). Interestingly, one of the targets, Cj1110c, was a signal transduction protein similar to *tlp1-4*. Cj1110c encodes Tlp8 (CetZ), a cytoplasmic sensor responsible for energy taxis. The CetZ orthologue AerC in *A. brasilense* controls chemotaxis away from high redox potential to low redox potential, possibly by monitoring cytoplasmic redox metabolites (Xie et al., 2010). In *C. jejuni*, inactivation of *tlp8* resulted in increased energy taxis, a phenotype opposite to *cetAB* inactivation, suggesting that CetAB and CetZ are two PAS-domain sensors that counteract each other in response to internal metabolic status. Both sensing systems transduce redox signals to CheA and CheY to mediate flagellar rotation (Reuter and van Vliet, 2013).

Table 5.7: **All CjSA9 mRNA partners identified by the computational prediction in chapter 4.** CjSA9 targets identified by computational prediction and RNA crosslinking were completely different.

mRNA	**Description**
rpoC	DNA-directed RNA polymerase β chain
ispH	4-hydroxy-3-methylbut-2-enyl diphosphate reductase
purH	Bifunctional purine biosynthesis protein
purL	Part of the phosphoribosylformylglycinamidine synthase complex in the purines biosynthetic pathway
Cj0991c	Probable oxidoreductase ferredoxin-type electron transport protein
Cj1110c	Probable MCP-type signal transduction protein
cdsA	Phosphatidate cytidylyltransferase
kpsC	Possible polysaccharide modification protein
Cj1563c	Probable transcriptional regulator
Cj1677_1	Putative lipoprotein

In conclusion, the IntaRNA prediction, the abundance of sRNA and mRNA, and the consensus result of computational and experimental approaches all pointed towards CjSA9 as an exciting candidate for further exploration. Hence, the next section will explore CjSA9 in more detail.

5.4 Detailed analysis of CjSA9

It is worth noting that CjSA9 shared similar genomic coordinates as the signal recognition particle RNA (SRP RNA). Bacterial signal recognition particle RNA recognises signal peptide sequences and co-translationally integrates nascent membrane peptides to the cell membrane. The start site of CjSA9 was 66641, which was five nucleotides upstream from the SRP 5' end because of the position of the nearest TSS. The 3' end of CjSA9 was 81 nucleotides downstream of the SRP RNA 3' end (Figure 5.20).

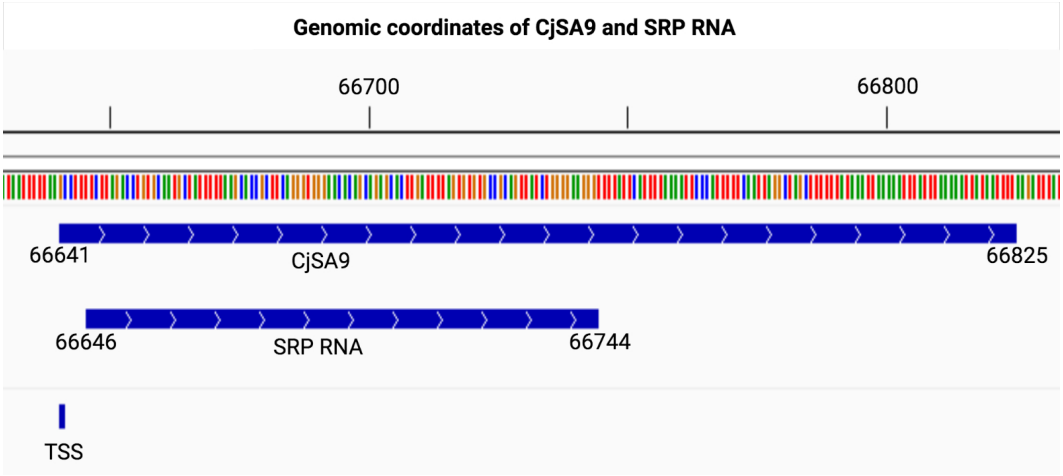

Figure 5.20: **Genomic coordinates of CjSA9 and SRP RNA.** CjSA9 consisted of the whole SRP RNA sequence, with an additional 5 nulcleotide at the 5' end and an extended 81 nuleotides at the 3' end. Only features on the sense strand are shown here.

Given SRP RNA's role in membrane-targeted translation, biological associations of CjSA9 should include transcripts encoding membrane proteins. Among the crosslinked targets, Cj0920c and Cj1583c encoded for a putative ABC-type amino-acid transporter permease protein and putative peptide ABC-transport system permease protein, respectively. Computational prediction also highlighted Cj1677_1 and Cj0991c, transcripts corresponding to lipoprotein and an electron transport chain protein, respectively. Due to the compact *C. jejuni* genome, some targets above belong to the same operon with other membrane protein genes. Hence, the expression of those genes may also be affected by CjSA9. Interestingly, all four crosslinked targets formed continuous transcripts with at least one membrane protein gene. For instance, Cj0920c was in the same operon as *pebC* (Amino-acid ABC transporter ATCP-binding protein), *peb1A* (Bifunctional adhesin/ABC transporter aspartate/glutamate-binding protein) and Cj0919c (probable ABC-type amino-acid transporter permease protein). Meanwhile, Cj1583c was part of the Cj1584c-Cj1580c operon encoding the ABC transporter complex NikZYXWV for nickel uptake. Moreover, *carB* was immediately downstream of *mreC*, an integral component of membrane according to the GO cellular component annotation. Lastly, *rp10* was immediately upstream of Protein translocase subunit SecY that belongs to the inner membrane. The prevalence of neighbouring membrane protein genes agreed with the known biological role of SRP RNA for membrane-associated co-translation.

On the contrary, most computationally predicted targets were not in the same operon as membrane protein genes. Hence, most computational targets may represent interactions irrelevant to the SRP RNA function. The only exception was *kpsC*, which was immediately upstream of *kpsS*, a capsule polysaccharide modification protein annotated as an integral component of the membrane according to GO cellular component.

5.4.1 Expression patterns of CjSA9 and its putative binding targets

To understand the dissimilarity between the targets highlighted by RNA crosslinking and computational prediction and the genomic coordinates of CjSA9 and SRP RNA, the expression pattern and putative binding mechanism of CjSA9 were studied more thoroughly. Once again, our RNAtag-seq data suggested that CjSA9 showed differential upregulation in conditions such as hyperosmotic stress (nacl), oxidative stress (oxidative) and iron repletion in the early stationary phase (iron_rep_ES) (Figure 5.21).

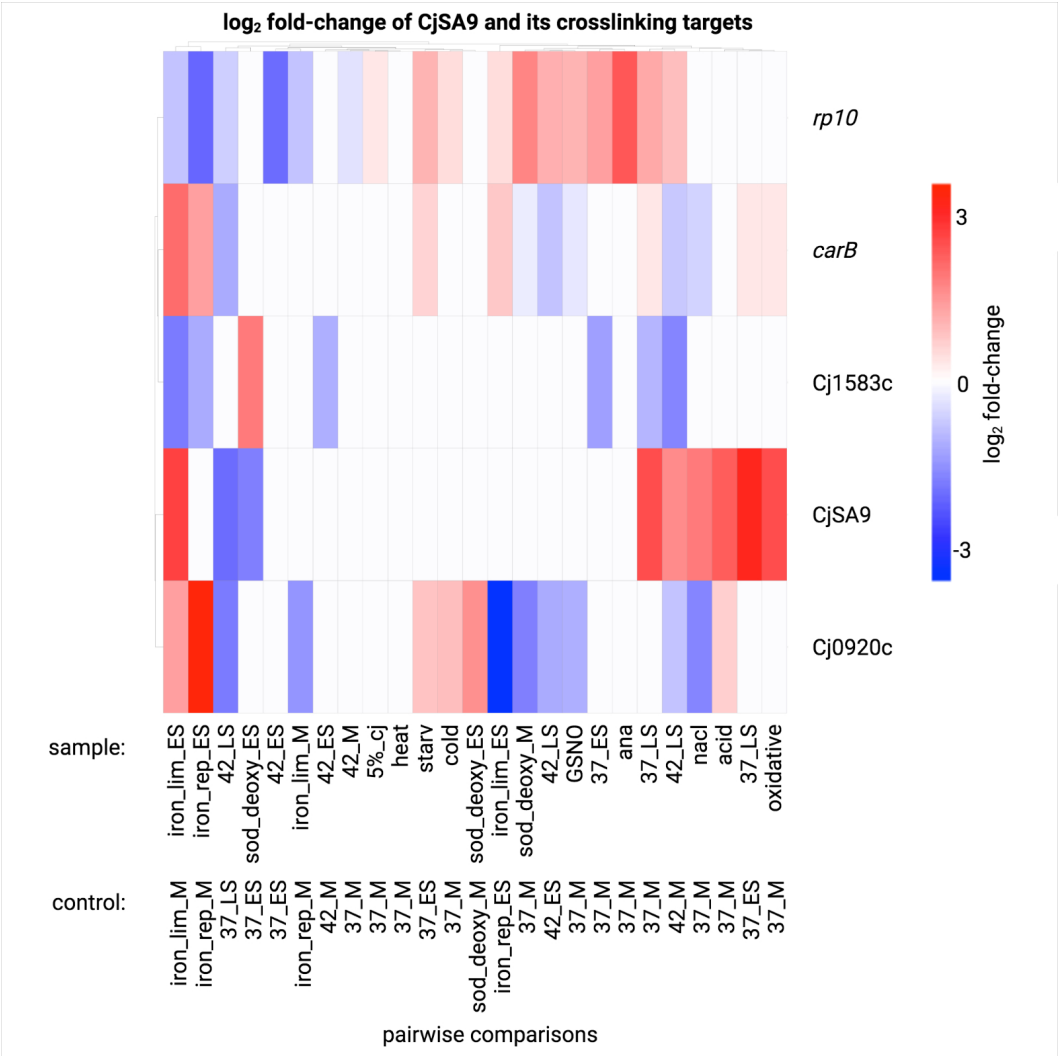

Figure 5.21: **Differential expression of CjSA9 and its crosslinking targets.** No clear differential expression correlation was observed among CjSA9 and its crosslinked targets.

CjSA9 and its computationally predicted targets belonged to co-expression module X, which consisted of the most significant number of genes. In contrast, none of the crosslinking targets was assigned to module X. WGCNA clustered Cj1583c and *rp10* into module III and I, respectively. Both Cj0920c and *carB* were found in module VI (Figure 5.22). As the computational prediction required all targets to show a statistically significant correlation with their sRNA partners, these observations are not surprising. Furthermore, no crosslinked targets statistically correlated with CjSA9 (Figure 5.23).

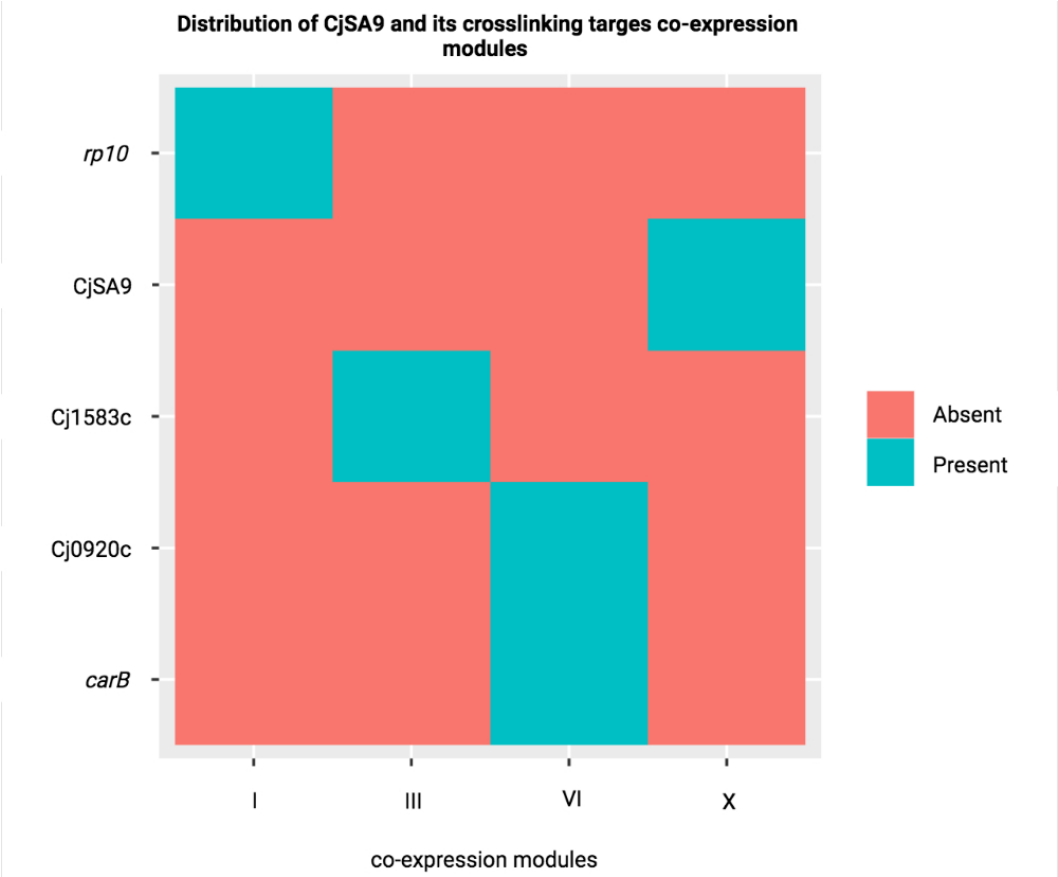

Figure 5.22: **Distribution of crosslinking targets across co-expression modules.** CjSA9 belonged to co-expression module X. None of the four CjSA9 crosslinked targets belonged to module X.

As computational targets all belonged to the same co-expression modules, they should display the same or opposite differential expression patterns with CjSA9 across many conditions. However, that may not be true for crosslinking targets, especially since none of them is from the same co-expression module as CjSA9. Hence,

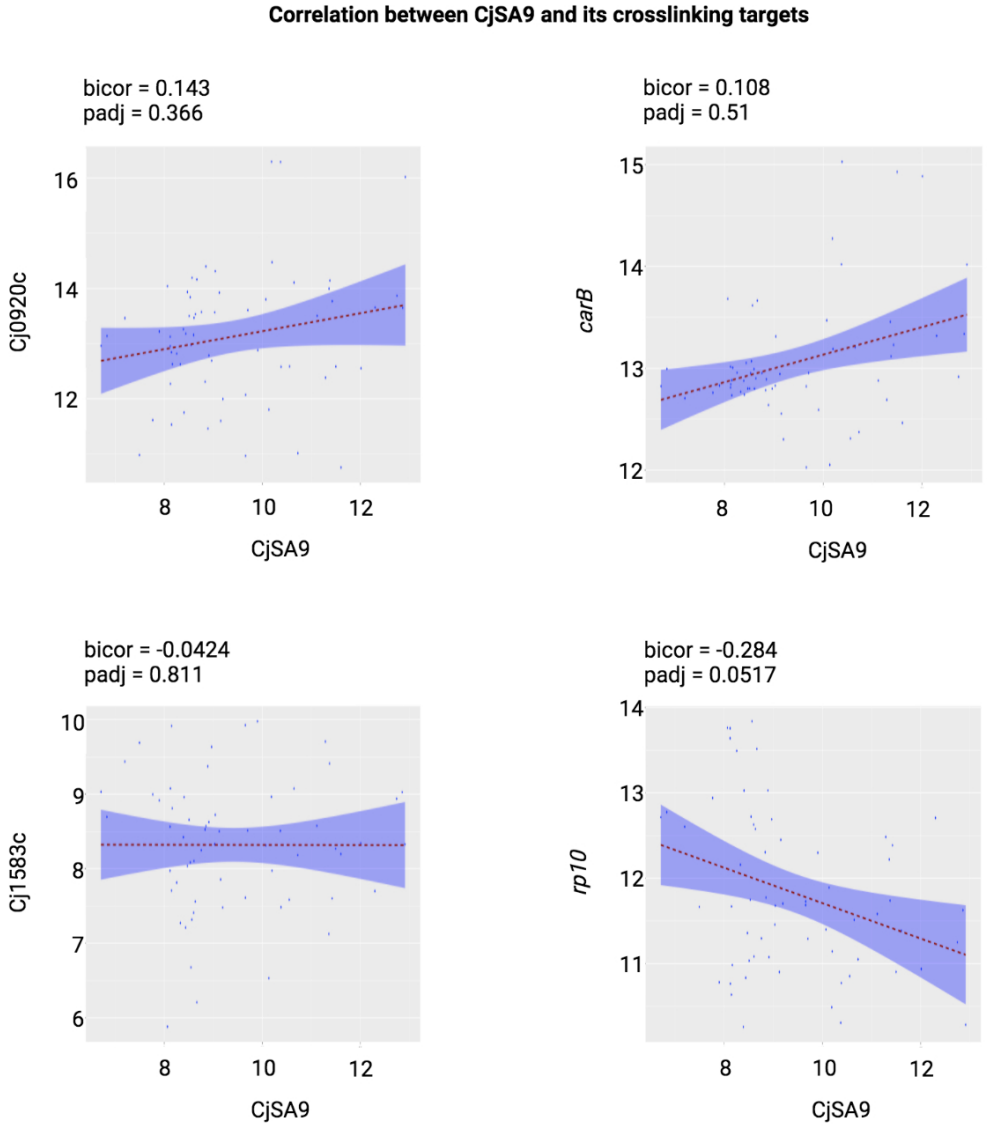

Figure 5.23: **Pairwise correlation between CjSA9 and its crosslinking targets.** No statistically significant (FDR < 0.05) co-expression was observed, suggesting the CjSA9 interaction may regulate mRNA structure instead of stability.

the differential expression pattern of CjSA9, crosslinked targets and computational targets was observed. The result showed that none of the crosslinked targets exhibited clear differential expression patterns compared to CjSA9 (Figure 5.21). On the contrary, computational targets such as purine metabolism genes (*purH*, *purL* and *rpoC*), together with *ispH*, *kpsC* and Cj0991c, showed opposite differential expression in conditions including the late stationary phase, iron limitation and bile stress. Under these conditions, CjSA9 and Cj1677_1, *cdsA*, Cj1110c and Cj1563c had almost identical differential expression patterns (Figure 5.24).

The expression patterns above suggested that, unlike computationally predicted interactions, RNA crosslinking suggested that CjSA9 may not directly affect the transcript abundance of the four crosslinked interactions. Instead, CjSA9 or SRP RNA binding may induce structural rearrangement and regulate translation initiation or elongation. SRP RNA is expected to bind to RNA targets without affecting transcript abundance. For instance, the SRP complex can stall nascent peptides and anchor the ribosomal complex to the membrane before further translation elongation (Mercier et al., 2017).

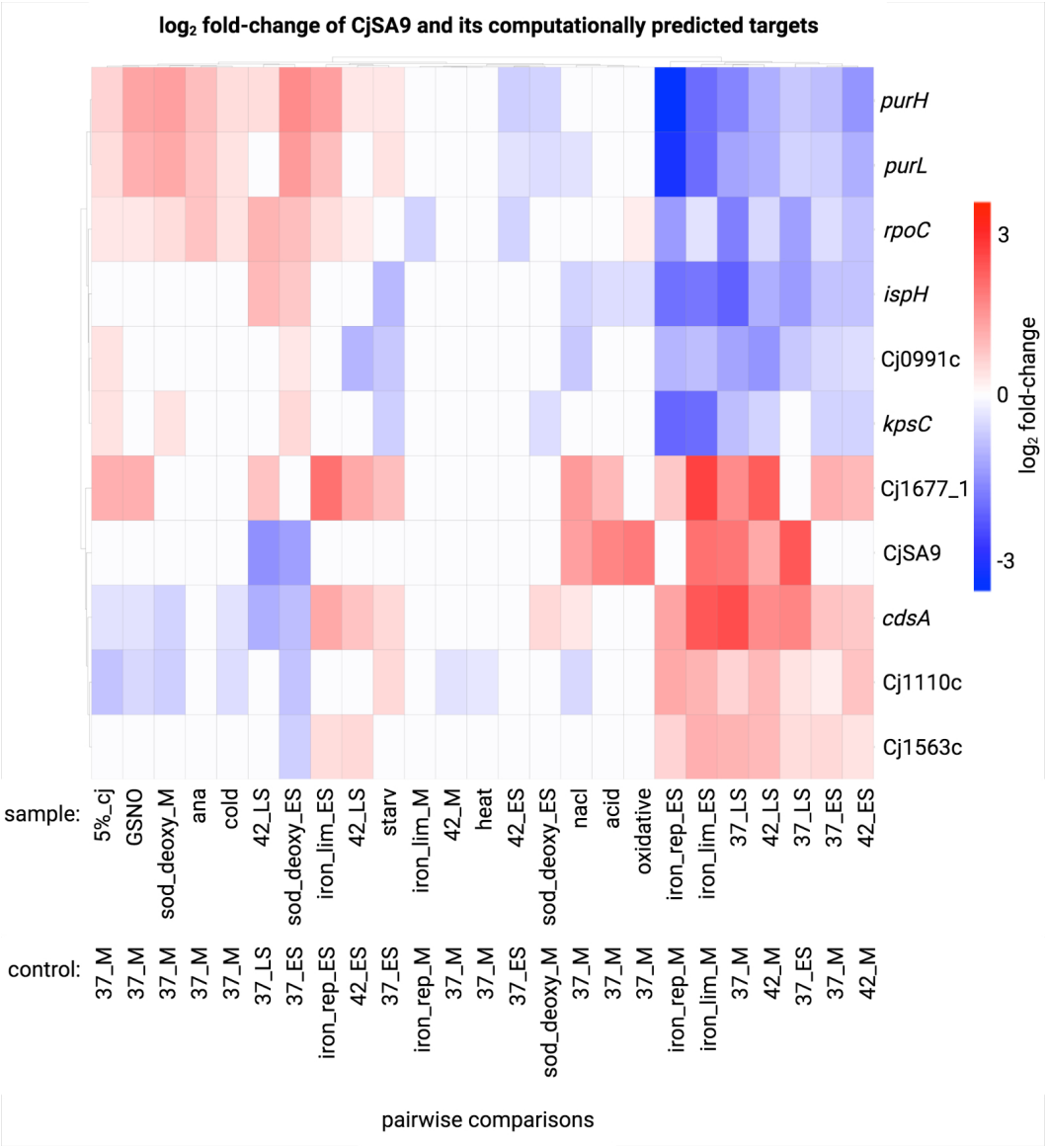

Figure 5.24: **Differential expression of CjSA9 and its computationally predicted targets.** Unlike the crosslinking interactions, computationally predicted targets of CjSA9 showed similar or opposite differential expression patterns.

5.4.2 Binding interactions and mechanisms of CjSA9

Another question to address was whether the crosslinking targets formed stable binding interactions with CjSA9 and if they shared similar interaction sites with other predicted targets. Based on the genome-wide target prediction by IntaRNA in chapter 4, all computationally predicted targets have significant p-values but insignificant FDR. Among the four crosslinking targets, *carB* was the only gene that showed significant p-values with the calculated affinity equalled -13.16 kcal/mol (Table 5.8). In contrast, the other three crosslinked targets displayed binding energies of around -5 or -6 kcal/mol.

Note that *gltX* was the only computationally predicted target with FDR <= 0.05, with a -21.22 kcal/mol binding energy value. Apart from that, none of the other energy values was more stable than -20 kcal/mol. Indeed, most predicted binding energy values for CjSA9 ranged between -15 kcal/mol to 0 kcal/mol (Figure 5.25), suggesting even those statistically significant (p-values <= 0.05) interactions were not exceptionally stable compared to those that were statistically insignificant. The unstable predicted binding energy of CjSA9 may facilitate the biological role of SRP RNA, which form transient interactions with RNA targets to hold off translation until reaching the membrane.

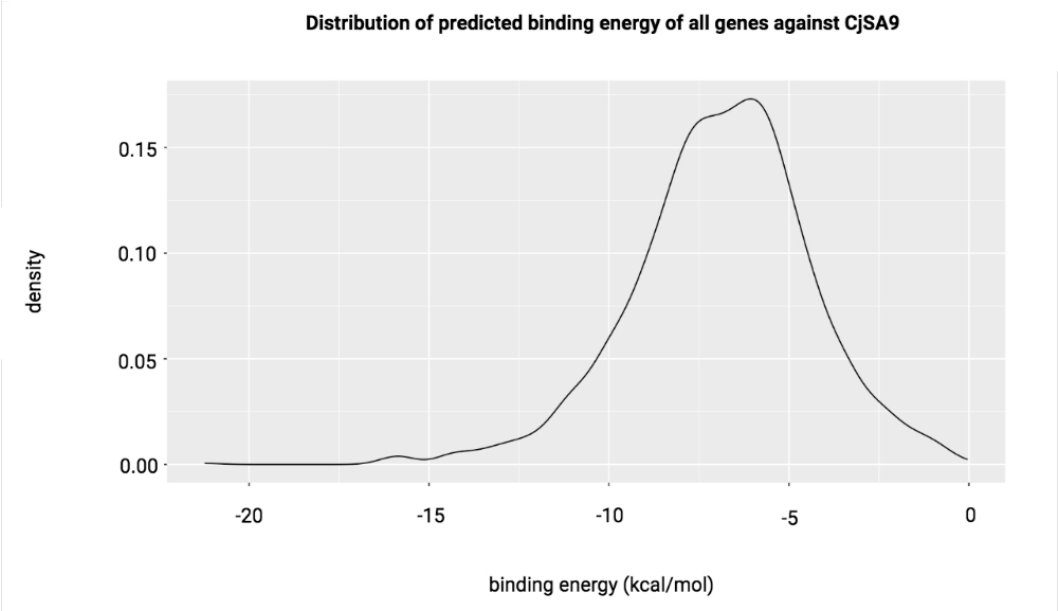

Figure 5.25: **The distribution of IntaRNA-predicted energy values of all CjSA9-target pairs.** Most CjSA9-mRNA interactions showed predicted binding energy values ranged between -15 kcal/mol to 0 kcal/mol.

Table 5.8: **The output of CjSA9 IntaRNA genome-wide target prediction.** Only targets with *p*-values <= 0.05 were shown here.

Gene name	**Binding energy (kcal/mol)**	***p*-value**	**FDR**	**Source**
rpoC	-11.72	0.045293	0.9492869	computational
ispH	-12.33	0.0304162	0.9492869	computational
purH	-12.04	0.0368686	0.9492869	computational
purL	-11.62	0.0482357	0.9492869	computational
Cj0991c	-11.72	0.045293	0.9492869	computational
Cj1110c	-11.65	0.0473366	0.9492869	computational
cdsA	-13.26	0.015765	0.9492869	computational
kpsC	-12.59	0.0254718	0.9492869	computational
Cj1563c	-11.61	0.0485386	0.9492869	computational
Cj1677_1	-12.29	0.0312447	0.9492869	computational
carB	-13.16	0.0169718	0.9492869	crosslinking
Cj1583c	-6.07	0.5974823	0.9492869	crosslinking
Cj0920c	-6.06	0.5990737	0.9492869	crosslinking
rp10	-5.94	0.618137	0.9492869	crosslinking
rrc	-12.43	0.0284259	0.9492869	IntaRNA only
aspA	-12.03	0.0371102	0.9492869	IntaRNA only
rpmA	-15.85	0.0016881	0.4016814	IntaRNA only
Cj0128c	-12.8	0.0219942	0.9492869	IntaRNA only
Cj0141c	-12.88	0.0207805	0.9492869	IntaRNA only
Cj0162c	-13.98	0.0090485	0.8530454	IntaRNA only
frr	-15.82	0.0017398	0.4016814	IntaRNA only
secG	-15.7	0.0019608	0.4016814	IntaRNA only
Cj0264c	-11.77	0.0438791	0.9492869	IntaRNA only
cheA	-14.37	0.0065723	0.749474	IntaRNA only
modA	-12.82	0.0216853	0.9492869	IntaRNA only
cmeA	-11.94	0.0393459	0.9492869	IntaRNA only
dapA_1	-14.1	0.008213	0.8314455	IntaRNA only
Cj0496	-12.02	0.0373532	0.9492869	IntaRNA only
alaS	-13.06	0.0182569	0.9492869	IntaRNA only
htpG	-12.08	0.0359152	0.9492869	IntaRNA only

Table 5.8: **The output of CjSA9 IntaRNA genome-wide target prediction.** Only targets with *p*-values <= 0.05 were shown here.

Gene name	**Binding energy (kcal/mol)**	***p*-value**	**FDR**	**Source**
Cj0541	-13.15	0.0170968	0.9492869	IntaRNA only
proS	-11.57	0.049766	0.9492869	IntaRNA only
Cj0610c	-11.71	0.0455803	0.9492869	IntaRNA only
uvrB	-12.86	0.0210784	0.9492869	IntaRNA only
ackA	-16	0.0014488	0.4016814	IntaRNA only
CjSA60	-16	0.0014488	0.4016814	IntaRNA only
Cj0736	-13.18	0.0167243	0.9492869	IntaRNA only
hrcA	-13.37	0.0145231	0.9492869	IntaRNA only
Cj0777	-12.64	0.024604	0.9492869	IntaRNA only
ilvA	-13.58	0.0123838	0.9492869	IntaRNA only
gltX	-21.22	1.00E-07	0.0001721	IntaRNA only
Cj0846	-13.82	0.0102757	0.884224	IntaRNA only
pabA	-16.13	0.0012658	0.4016814	IntaRNA only
serA	-12.62	0.0249481	0.9492869	IntaRNA only
rpsA	-13.4	0.0141993	0.9492869	IntaRNA only
Cj1004	-13.63	0.0119164	0.9492869	IntaRNA only
cjeI	-11.83	0.0422315	0.9492869	IntaRNA only
mfd	-15.63	0.0021006	0.4016814	IntaRNA only
argS	-14.3	0.0069678	0.749474	IntaRNA only
Cj1199	-13.93	0.0094177	0.8530454	IntaRNA only
kefB	-12.21	0.0329593	0.9492869	IntaRNA only
Cj1241	-14.43	0.0062488	0.749474	IntaRNA only
guaA	-12.7	0.023596	0.9492869	IntaRNA only
Cj1287c	-15.84	0.0017052	0.4016814	IntaRNA only
gltX2	-12.57	0.0258262	0.9492869	IntaRNA only
Cj1302	-13.42	0.013987	0.9492869	IntaRNA only
maf3	-11.63	0.0479344	0.9492869	IntaRNA only
Cj1365c	-11.7	0.0458692	0.9492869	IntaRNA only
Cj1367c	-12.03	0.0371102	0.9492869	IntaRNA only
carA	-14.52	0.0057894	0.749474	IntaRNA only

Table 5.8: **The output of CjSA9 IntaRNA genome-wide target prediction.** Only targets with *p*-values <= 0.05 were shown here.

Gene name	**Binding energy (kcal/mol)**	***p*-value**	**FDR**	**Source**
rloH	-14.31	0.0069101	0.749474	IntaRNA only
hsdR	-12.28	0.0314548	0.9492869	IntaRNA only
sdaA	-12.29	0.0312447	0.9492869	IntaRNA only
Cj1664	-12.72	0.023268	0.9492869	IntaRNA only
Cj1677_2	-12.29	0.0312447	0.9492869	IntaRNA only

Based on IntaRNA, CjSA9 contains two optimal binding sites for targets with *p*-values <= 0.05. One is in position 50 -100, near the annotated 3' end of SRP RNA. The other one is at around position 100 - 150, located after the annotated 3' end of SRP RNA (Figure 5.26). The putative optimal binding sites of the four crosslinking targets are near the 3' end of SRP RNA. The *carB* binding site was longer than the other three binding sites, which explained why the binding energy value for *carB* was more stable than the other three binding energy values (Figure 5.27).

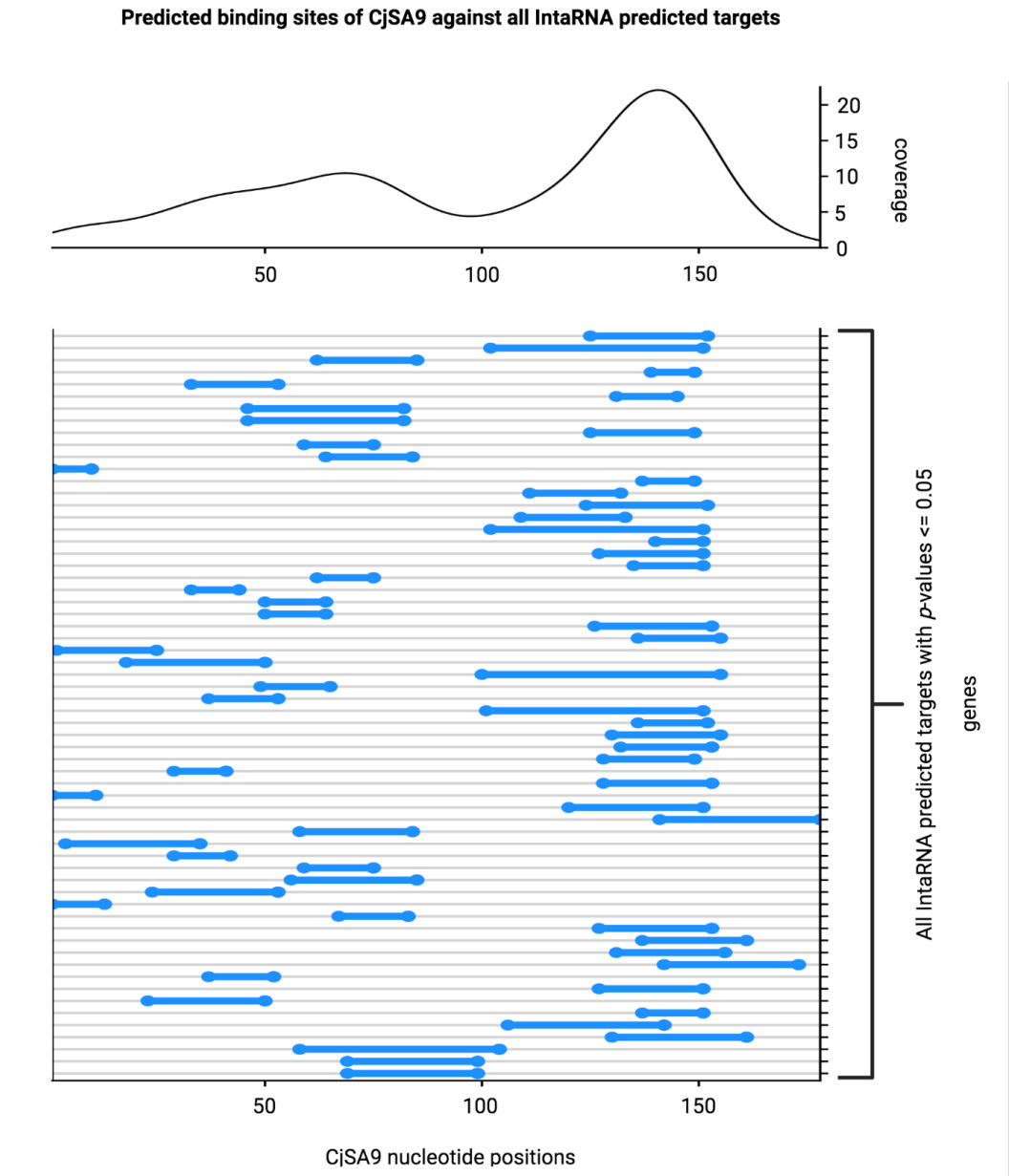

Figure 5.26: **The distribution of predicted optimal binding sites used by CjSA9 against all IntaRNA predicted targets with *p*-values <= 0.05.** The results suggested two putative binding regions, one in between the 50^{th} - 100^{th} nucleotides and the other near the 150^{th} nucleotide.

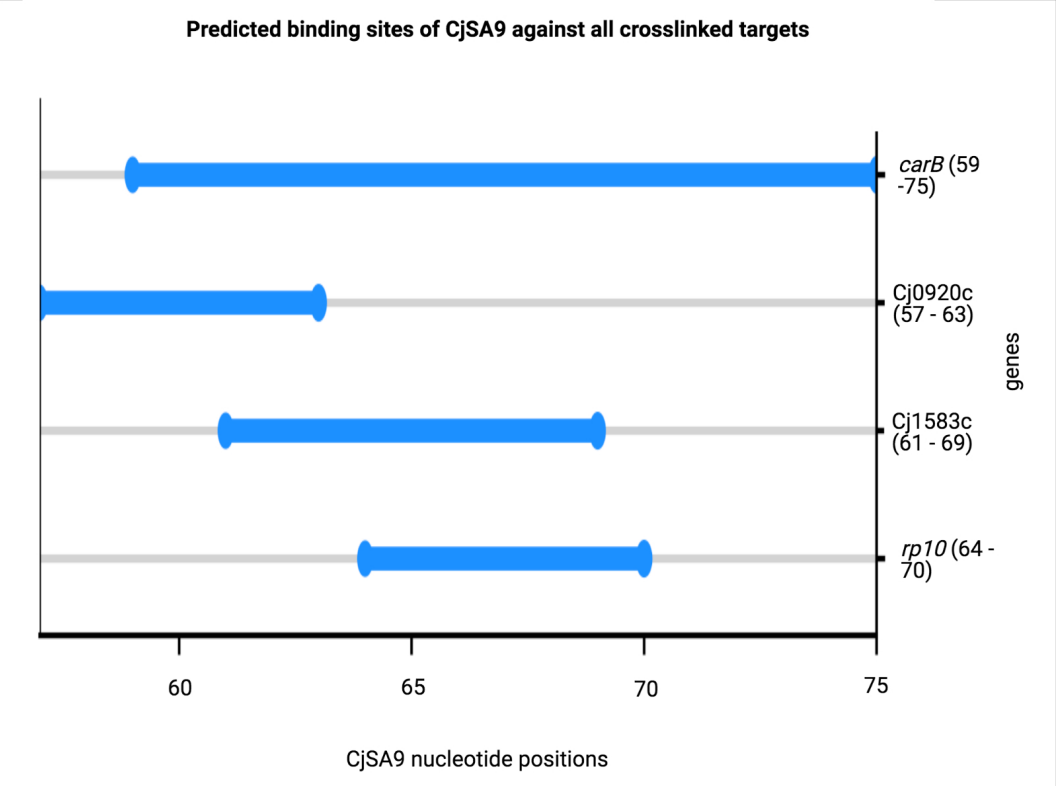

Figure 5.27: **The distribution of predicted optimal binding sites used of CjSA9 against all crosslinking targets.** The numbers in the brackets indicate the CjSA9 nucleotide positions involved in the base pairing.

IntaRNA also predicted the distance between the optimal interaction sites and the target's TSS (Figure 5.28). The predicted interaction sites of CjSA9 and its targets can be compared with the actual position of their mapped chimeric reads. The alignment regions were likely to be located near the ligated ends. The actual positions of ligated ends can be approximated using the locations of soft-clipped bases (shown in rainbow colour in Figure 5.29). A closer look at the chimeric reads mapping to *carB*

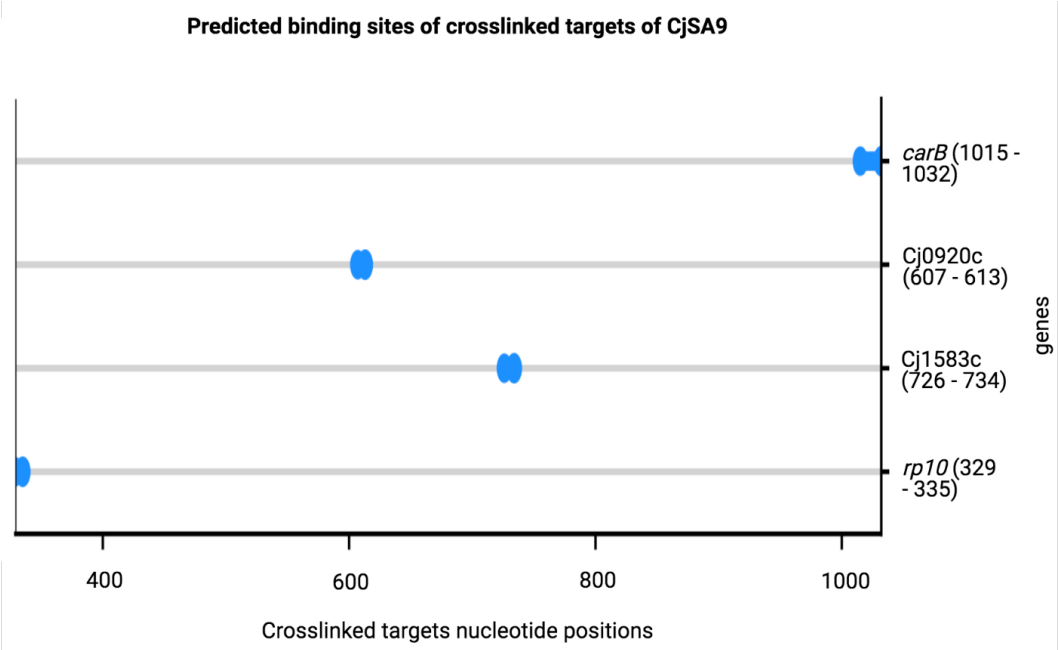

Figure 5.28: **The distribution of predicted optimal binding sites of the crosslinking targets against CjSA9.** The numbers in the brackets indicate the crosslinked targets nucleotide positions involved in the base pairing.

showed that among the aligned paired-end reads, Read 2 was a split read. The first 21 bp of read 2 mapped from the 23th to the 43th nucleotides of CjSA9. The other 55 bp mapped to the middle of *carB*, more than 2000 bp downstream of the 5' end. Read 1 was fully aligned to 76 nucleotides around the middle of *carB* (Figure 5.29a). That may hint towards a chimeric junction near the 43th nucleotide of CjSA9, and somewhere near the middle of *carB*. That was different from the IntaRNA predicted binding site, which was after the 59th nucleotide of CjSA9 and 1015 bp downstream of the *carB* TSS.

Read 2 of CjSA9-Cj0920c was a chimeric read that mapped to CjSA9 twice. One part of read 2 consisted of 34 bp that mapped to the 5' end of CJSA9. The other part included 40 nucleotides aligned near the SRP RNA 3' end. Read 1 mapped to 74

nucleotides around the middle of Cj0920c at about 800 bp downstream of the 5' end (Figure 5.29b). The chimeric read may result from the secondary structure folding of SRP RNA 5' end and 3' end. The result may suggest binding interactions between either end of SRP RNA and the region around the 800^{th} bp of Cj0920c.

Read 1 of CjSA9-Cj1583c was a hybrid of two 38 bp sequences. Both sequences aligned to a similar region on CjSA9, located near the annotated 3' end of SRP RNA. That might result from ligation between the 3' ends of two different SRP RNA molecules. Meanwhile, read 2 was a hybrid read with 47 bp aligned near the SRP RNA 3' end and 24 bp (plus 1 mismatched bp) mapped to the Cj1583c coding region (Figure 5.29c). That suggested a chimeric junction between the 3' end of SRR RNA and the coding region of Cj1583c. Note that the alignment of Cj1583c was slightly upstream of the IntaRNA prediction. The read alignment suggested the interaction site was about 600 bp from the TSS of Cj1583c. However, based on IntaRNA, the interaction took place between the 700^{th} and 800^{th} nucleotides from the TSS.

Read 2 of CjSA9-*rp10* consisted of 49 bp near the annotated 3' end of SRP RNA and the other 27 bp near the 5' end of *rp10*. Read 1 almost aligned completely to the 5' end of *rp10*, except the first 8 bp being soft-clipped and unmapped to anywhere else (Figure 5.29d). That indicated the binding site was near the 3' end of SRP and the 5' end of *rp10*. Note that the crosslinked site of *rp10* was contrary to the target sequence estimated by IntaRNA. IntaRNA suggested the binding site was in between 300^{th} and 400^{th} nucleotides from the TSS.

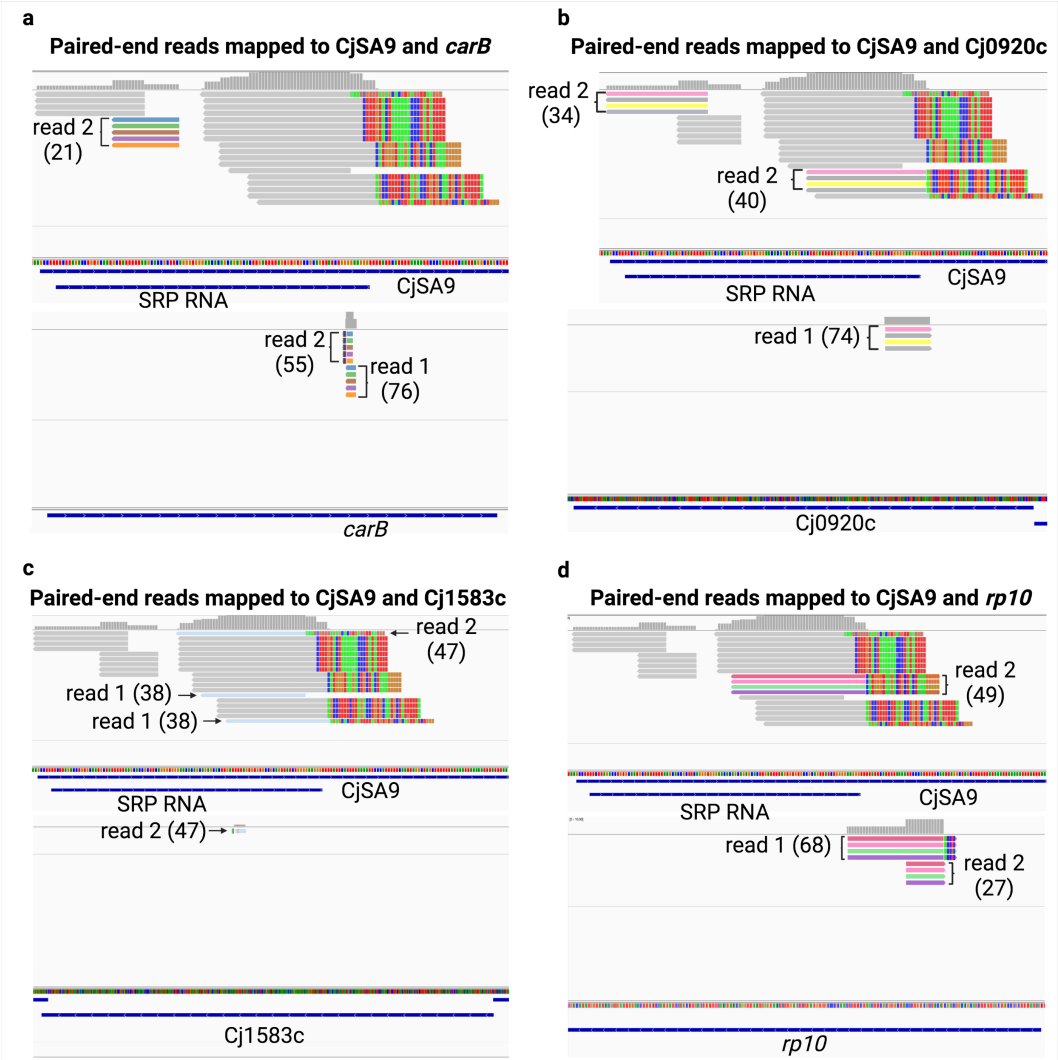

Figure 5.29: **Paired-end reads mapped to CjSA9 and its crosslinked targets.** Those targets were (a) *carB* (b) Cj0920c (c) Cj1583c (d) *rp10*. The numbers in the following brackets were the length of mapping nucleotides.

5.4.3 Expression change and Structural insights of CjSA9 and SRP RNA

As mentioned earlier, IntaRNA suggested two optimal binding sites for CjSA9. One was found around the SRP RNA 3' end, while the other was downstream of the SRP RNA 3' end. Interestingly, many crosslinked sequences above were near the annotated 3' of SRP RNA. These observations suggested several possibilities. First, the CjSA9 transcription signal after the SRP RNA 3' end might result from background noise and non-specific transcriptional readthrough. The second possibility was that CjSA9 represented the pre-processed version of SRP RNA. This process may involve cleavage of the region downstream of the SRP RNA 3' terminal during SRP RNA maturation. During such a process, one can speculate that all three molecules, pre-processed CjSA9, processed SRP RNA and the cleaved 3' end can potentially bind to different sets of targets under certain conditions. For instance, the mature SRP RNA could target all four crosslinking targets, while the cleaved sequence could bind to other targets, including purine metabolism genes including *rpoC*, *purH* and *purL*.

The mapped read coverage of CjSA9 was visualised using the RNAtag-seq samples to determine whether CjSA9 is merely background noise or an independent transcript. Most conditions, such as acid stress (acid), showed limited expression coverage towards the CjSA9 3' end (Figure 5.30a). However, some conditions showed more visible read coverage after the 3' end of the SRP. For instance, a higher level of expression coverage was observed after the SRP RNA 3' end among two out of three replicates of the samples from the early stationary phase under 37 °C (37_ES) (Figure 5.30b). Likewise, expression coverage was observed in SRP RNA 3' end in the samples from the early stationary phase under 42 °C (42_ES) (Figure 5.30c). Hence, the 3' end of CjSA9 may exist as a transcript when the cell has been cultured to the early stationary phase. However, note that the y axis scales of 37_ES and 42_ES were smaller than acid. Hence, the CjSA9 3' end might be more visible simply due to the lower coverage of SRP RNA in the early stationary phase. Therefore, another possibility is that SRP RNA exhibits a low expression level indistinguishable from background noise upon stationary phase entry.

Assuming CjSA9 3' end is an independent transcript from SRP RNA, it should form a secondary structure independently from SRP RNA. Otherwise, SRP RNA and CjSA9 3' end sequences may form a hybrid RNA structure. Hence, the identity of the CjSA9 3' end was examined using the RNAfold computed structure of CjSA9 and SRP RNA. Secondary structure computed by RNAfold revealed two separate

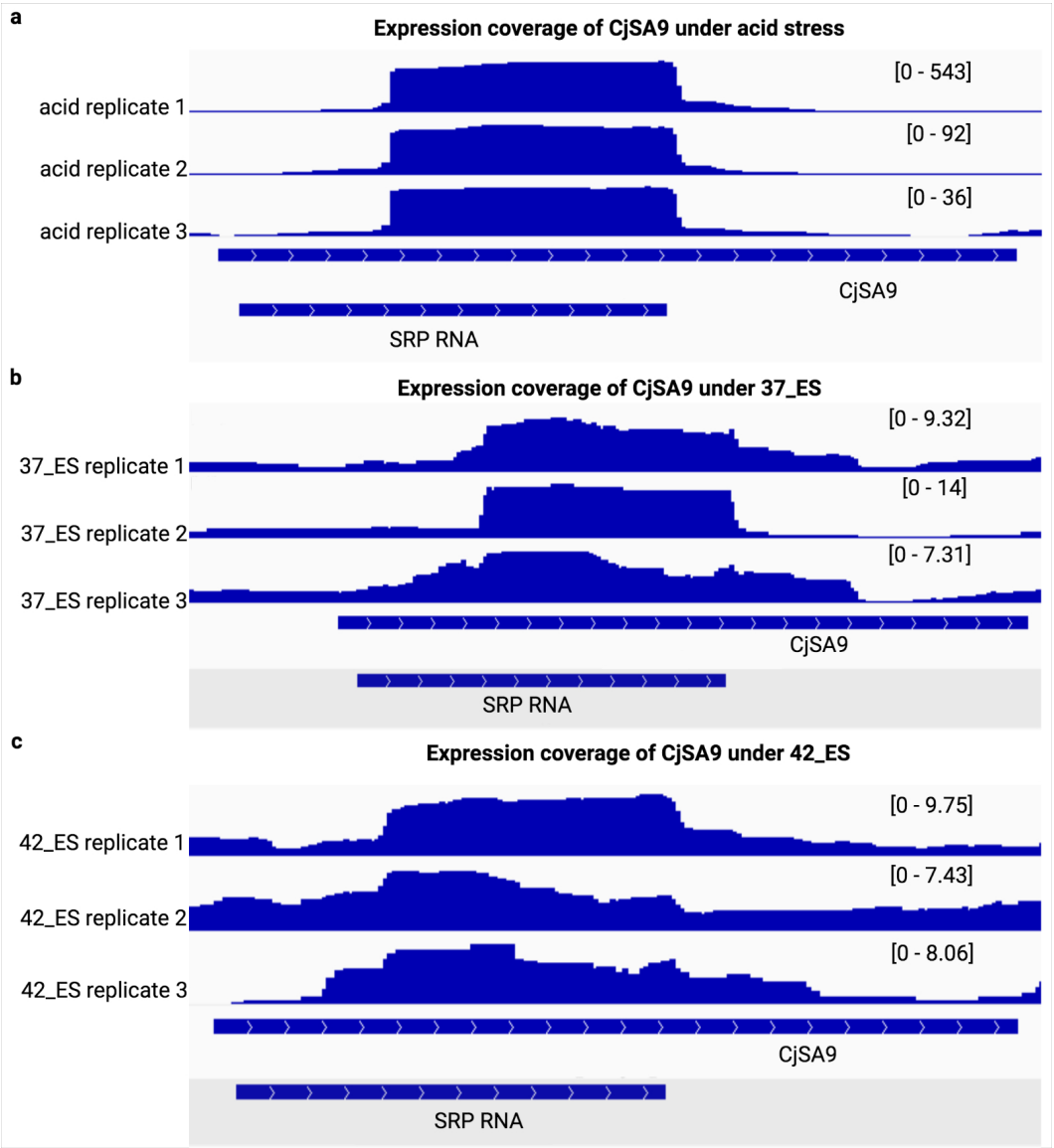

Figure 5.30: **Expression coverage of CjSA9 under RNAtag-seq conditions**. Those conditions were (a) acid, (b) 37_ES and (c) 42_ES. The numbers in the brackets indicate the y-axis scale of the wigsum normalised coverage.

domains within CjSA9. The first domain of CjSA9 formed a linear hairpin structure. Most base pairs have medium to weak intramolecular binding probability and medium to high nucleotide entropy. This pattern indicated flexibility of the hairpin structure, allowing it to unfold and bind to other RNA molecules. The second domain consisted of mostly unpaired nucleotides. The 3' end of SRP RNA was located near the first and second domains interface. Thus, the second domain consisted of only sequence after the SRP 3' end (Figure 5.31a).

The CjSA9 3' end alone (with the SRP RNA sequence removed) displayed an identical structure as when it was attached to the SRP RNA. The structural identity suggested that the CjSA9 3' end function independently from SRP RNA. Notably, the CjSA9 3' end region has a medium secondary structure formation probability and nucleotide entropy, including the region involved in base-pairing according to crosslinking and IntaRNA. That suggested some flexibility for base pairing with other RNA species (Figure 5.31b). The subsequent conservation analysis showed that CjSA9 3' end had a secondary structure distinct from any templates or ncRNA families on the RNACentral and Rfam databases. Such an observation indicates CjSA9 3' end exists as a novel sRNA.

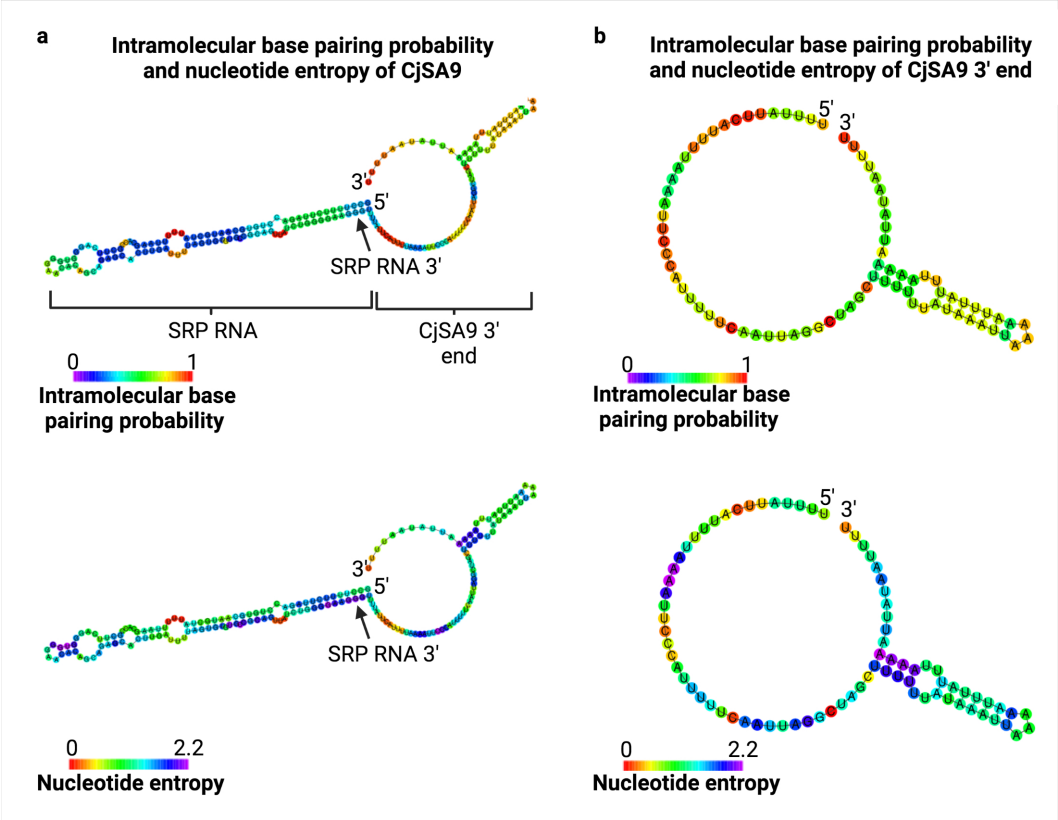

Figure 5.31: **RNAfold prediction of CjSA9 and CjSA9 3' end.** Structures of (a) CjSA9 and (b) CjSA9 3 end with the SRP RNA sequence removed. The result suggested the extended CjSA9 3' end structure was not affected by the SRP RNA sequence.

Removing the extra 3' end region did not affect the structure of SRP RNA (Figure 5.32). The intramolecular binding probability increased after removing the CjSA9 3' end sequence, possibly due to reduced base-pairing competition. This finding suggests that the additional CjSA9 sequence may affect the SRP RNA structural stability and thus its biological function. That also means SRP RNA is more likely to retain its biological function if the CjSA9 3' end was a separate transcript.

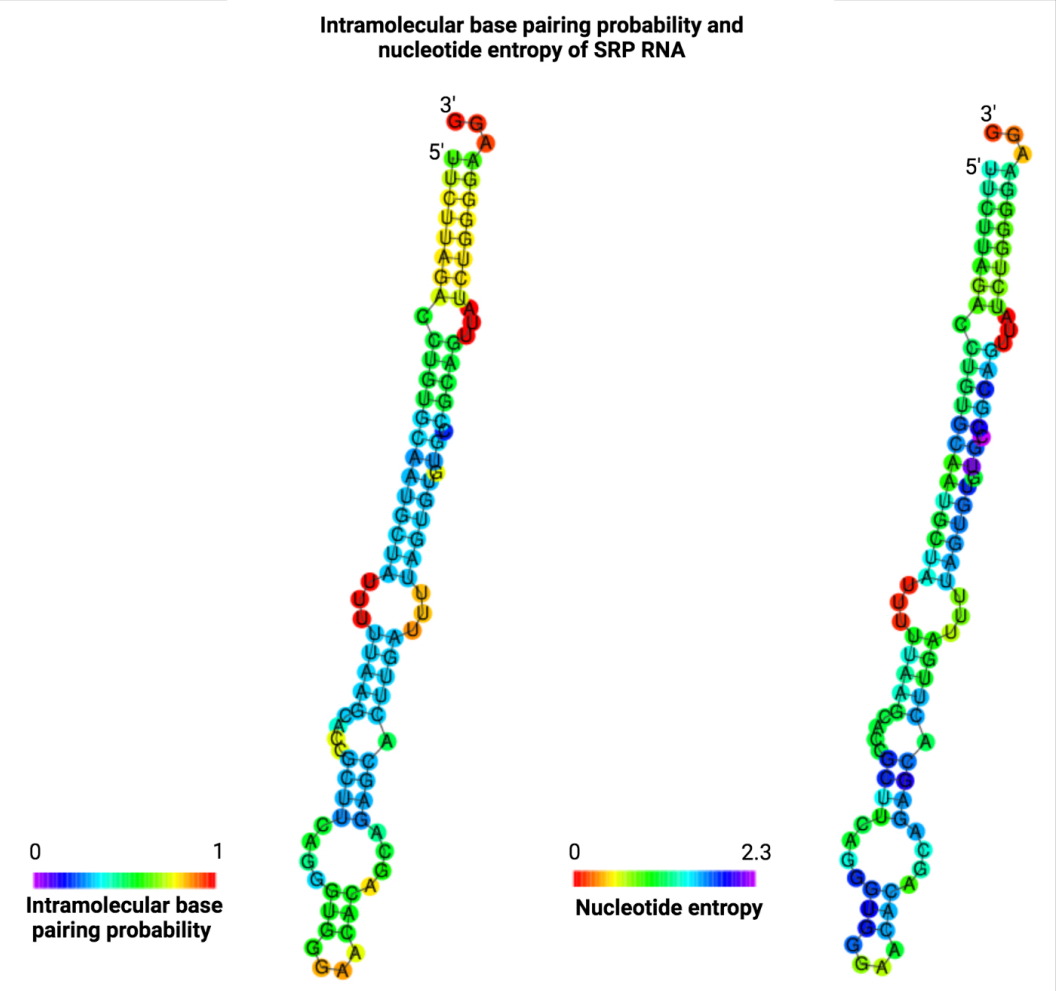

Figure 5.32: **RNAfold structure of SRP RNA.** The linear predicted structure agreed with the published SRP RNA structure, and was not affected by the CjSA9 3' end sequence. The result suggested CjSA9 3' end and SRP RNA may function independently.

5.5 Discussion

5.5.1 Improvement and challenges of rRNA depletion

RiboMinus alone appeared insufficient for rRNA depletion, as demonstrated by the presence of only 0.8 % of mRNAs or sRNAs in the crosslinked samples. The inclusion of DASH and Ribo-Zero Plus increased the percentage of alignment to sRNA or mRNAs to 9.70 % and yielded higher portions of chimeric reads and sRNA-mRNA duplex. Nonetheless, rRNA removal of the crosslinked samples was still less efficient than the negative controls, as demonstrated by the prominence of rRNA-rRNA interactions among chimeric reads. Most rRNA-rRNA interactions may involve alignment to two copies with identical rRNA sequences. In contrast, less than 1 % of chimeric reads were sRNA-mRNA interactions.

As mentioned in the introduction methods, crosslinking may have introduced covalent linkages with rRNA sequences, preventing the rRNA from being captured by the probes (Figure 5.33). Another issue with probes from RiboMinus and RiboZero Plus is that they are not designed explicitly for specific species, leading to lower efficiency at capturing *C. jejuni* rRNA. While DASH bypasses these issues by targeting cDNA using customarily designed sgRNA, the Cas9 digestion step may remove non-rRNA sequences from the cDNA library by off-target cleavage. Another issue with DASH is that it can only enrich non-ribosomal sequences after cDNA library preparation. Reverse transcription reaction may fail to convert some RNA species into cDNA. This drawback is especially problematic for crosslinked samples as they are more saturated with rRNA species. As a result, more non-rRNA species may not be reverse transcribed into cDNA. This issue may account for the observed PCR duplications, which hampers the accuracy of statistical tests.

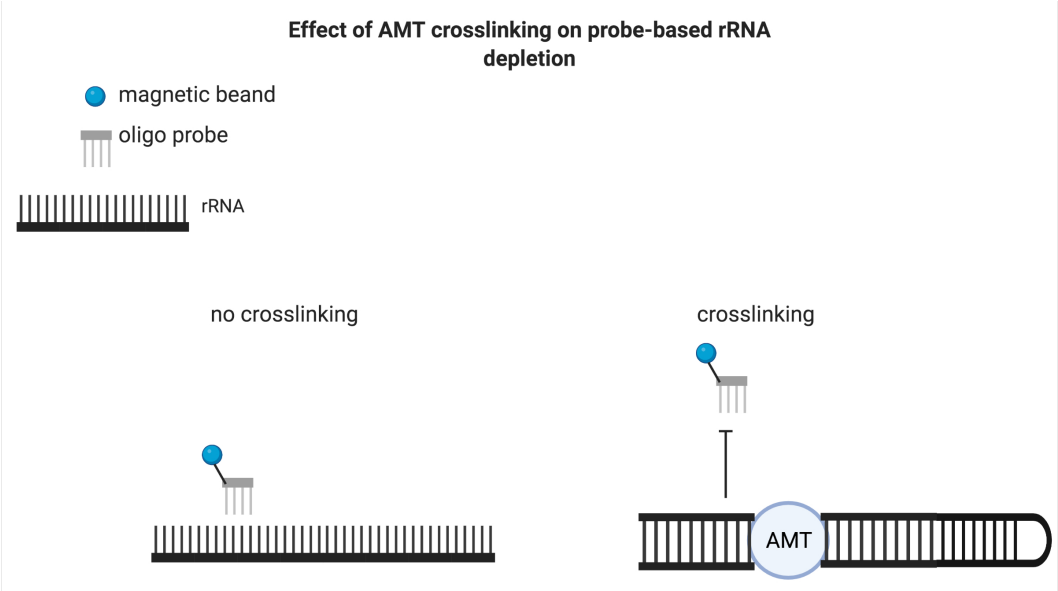

Figure 5.33: **The impact of crosslinked RNA folding on rRNA depletion.** The AMT adduct may prevent the rRNA probe from binding to the rRNA target sequences, thus hampering the efficiency of probe-based rRNA removal. Created in biorender.com.

Adding more probes to the rRNA depletion step prior to cDNA library preparation can improve the diversity of non-ribosomal sequences. Examples of other approaches include RNase H digestion of RNA-DNA hybrids (Huang et al., 2020) and RNAtag-seq (Shishkin et al., 2015). Another possibility is to repeat RiboMinus and Ribo-Zero several times. However, as mentioned before, these methods may also be hindered by crosslinking rRNA secondary structures and the lack of probe sequence specificity. Repeating rRNA depletion steps also exposes the RNA samples to longer processing steps, which leads to more RNA degradation that hampers RNA integrity. RiboMinus and Ribo-Zero probes have also demonstrated limited capacity on targeting 5S rRNA sequences of *C. jejuni*, as suggested by the Bioanalyzer result. Hence, such an approach may result in RNA samples saturated with 5S rRNA. While increasing the sequencing depth may uncover more non-rRNA species, it may also yield more PCR duplicates due to the low abundance of non-rRNA sequences. Instead, it may be worth optimising the reverse transcription step to convert as much sRNA or mRNA into cDNA as possible.

5.5.2 Challenge of Statistical analysis of RNA crosslinking

Due to the low abundance of non-rRNA chimeric reads, few sRNA-mRNA interactions were detected in the crosslinked samples, with most sRNA-mRNA pairs also found in negative controls, suggesting background noise. RNA crosslinking is prone to background noise because RNA pairs without biological associations can still undergo ligation due to their physical proximity. This outcome is especially likely for highly expressed RNA species.

PCR duplication affects the results of statistical analysis. Some published protocols such as PARIS included a PCR duplicates removal step. However, previous studies showed that duplicates removal with Picard or SAMTools led to no or limited improvement of the accuracy of statistical tests (Parekh et al., 2016; Ebbert et al., 2016), possibly because these methods may also remove natural duplicates from highly abundant RNA species. Tools like dupRader aim to distinguish between PCR and natural duplicates by estimating the natural duplication rate and gene expression dependency. However, using dupRader only caused a slight improvement to the subsequent statistical analysis (Sayols et al., 2016). While multiple chimeric reads mapped some sRNA-mRNA pairs, IGV visualisation showed they all showed the exact coordinates and likely resulted from PCR duplication. As a result, this study's statistical power is weaker due to the low abundance of chimeric reads for sRNA-mRNA interactions. Hence, the most appropriate approach was filtering out pairs that appeared in negative controls.

5.5.3 Annotated ncRNA in the filtered results

After filtering using the negative controls, the remaining interactions included predicted sRNAs that shared similar or identical coordinates to published sRNA or ncRNA. Among these are CjSA9 (SRP RNA), CjSA51 (*rnpB*) and CjSA97 (10Sa RNA).

RnpB is a ribozyme that facilitates tRNA maturation by cleaving the tRNA precursor's 5' end. CjSA51 crosslinked with glycerol-3-phosphate dehydrogenase (*gpsA*) and 50S ribosomal protein L14 (rplN) in the RNA crosslinking data. That may hint towards a novel role for *rnpB* in regulating the metabolism of non-tRNA species.

10Sa RNA plays a diverse range of biological functions. Examples of known 10Sa RNA activities include recycling stalled ribosomes, adding proteolysis-inducing tags to unfinished polypeptides, and promoting the degradation of aberrant messenger

RNA. CjSA97 crosslinked an mRNA that encodes probable hydrogenase isoenzymes formation protein (*hypB*) (Zhang et al., 2005). Future work can confirm if CjSA97 affects the translation or stability of aberrant *hypB* mRNA.

5.5.4 Complementing the crosslinking and computational prediction results

Of all predicted sRNAs that appeared in the filtered interactions, CjSA9 stood out as a candidate for further investigation. Besides sharing a similar genomic coordinate as SRP RNA, CjSA9 is the only sRNA highlighted by computational prediction and RNA crosslinking. CjSA9-*carB* also showed the most statistical significance within all extracted duplexes. Additionally, the expression levels of CjSA9 and two of its crosslinked targets are among the lowest of all extracted sRNAs and mRNAs.

SRP RNA is highly conserved among eukaryotes, archaea and bacteria (Rosenblad et al., 2009). Bacterial SRP RNA is responsible for incorporating membrane proteins into the inner membrane co-translationally. Therefore, crosslinking or predicted interactions between membrane protein transcripts and CjSA9 would be particularly noteworthy. Two of the crosslinked mRNA targets, Cj0920c (Müller et al., 2007) and Cj1583c (Hoang et al., 2012), encode an amino acid ABC transporter permease and peptide ABC transporter permease, respectively. Notably, both Cj1583c and CjSA9 showed low expression levels in the early stationary phase in the RNAtag-seq dataset. Cj1583c also showed the lowest TPM value among all extracted mRNA. These results suggested that CjSA9-Cj1583c interaction is more likely to be genuine than background noise. Hence, of all CjSA9 crosslinked interactions, CjSA9-Cj1583c is the most valuable candidate for future experimental validation.

Cj1583c also has significant relevance to food safety. Cj1584c-Cj1580c (*nikZYXWV*) encodes an ABC transporter complex that transports nickel ions as cofactors for hydrogenase activity, which oxidises H_2 to donate electrons to electron transport chains. High nickel concentration represses the expression of Cj1584c–Cj1580c (Howlett et al., 2012). Free cytoplasmic nickel facilitates the periplasmic translocation of CueO, an enzyme that oxidises copper in the periplasm to remove copper toxicity. The Ni-chelator dimethylglyoxime (DMG) abolishes CueO translocation and makes *C. jejuni* more susceptible to copper toxicity. Oral administration by chickens reduced *C. jejuni* colonisation. Hence, further understanding of CjSA9-Cj1583c interactions may provide additional insights into nickel uptake, which is vital to the safety of poultry products (Benoit and Maier, 2021).

Some computationally predicted targets of CjSA9, such as Cj0991c and Cj1677_1 are putative membrane-associated proteins, which agrees with the role of SRP RNA complex in membrane-targeted translation. While CetZ is a predicted cytoplasmic protein that lacks a transmembrane domain, its role as a redox sensor may require direct interaction with membrane proteins such as CheV, maybe via its PAS domain (Parrish et al., 2007). Furthermore, an iLOV reporter assay showed both Tlp5 and Tlp8 localised at the cell poles. Tlp5 accumulated mainly at one pole, while Tlp8 migrated to both poles. In contrast, the iLOV reporter negative control distributed uniformly across the cytoplasm (Elgamoudi and Ketley, 2018). Such localisation patterns further support the polar localisation of Tlp8 mediated by membrane protein interactions and may regulate the activities of membrane proteins. Hence, CetZ may interact and affect electron transport chain activities. Proton motive force is necessary for *C. jejuni* motility. The presence of electron acceptor-donor couples increased the number of mobile cells with faster swimming speed (van der Stel et al., 2017. Therefore, future investigation of CjSA9-Cj0991c may provide additional insight into the association between redox potential and cell motility. However, it remains unclear if targeted translation by CjSA9 plays any role in this process. Perhaps CjSA9 mediates the redox sensing by fine-tuning Cj0991c translation.

CjSA9 also crosslinked with *carB*, which generates a precursor for arginine and pyrimidine (carbamoylphosphate) using ammonia, bicarbonate and ATP (Mclennan et al., 2008). This interaction is the only crosslinked interaction with significant IntaRNA *p*-value. RNAtag-seq data also showed that *carB* has the third-lowest expression level under the early stationary phase, meaning *carB* was likely an actual CjSA9 target instead of background noise. The biological relevance of *carB* to SRP RNA requires further investigation.

UV crosslinking and gel shift assay of *E. coli* demonstrated binding affinity between SRP RNA and ribosomal components, such as 16S rRNA and 23S rRNA (Rinke-Appel et al., 2002). Notably, the protein product of *rp10*, one of the CjSA9 crosslinking targets, is a 50S ribosomal protein L15 that binds to the 23S rRNA. Rp10 mediates the interaction between ribosomes and the translation factors attached to GTF. It remains to be seen if *rp10* mRNA also plays a structural role for *C. jejuni* ribosome.

Further exploration showed no clear expression patterns between CjSA9 and its crosslinking targets. That suggested the binding interactions may regulate translation by inducing secondary structure rearrangement without altering RNA stability. Furthermore, there was a clear distinction between sequencing read align-

ment positions and predicted optimal binding sites. Such discrepancies might indicate biological associations with nascent transcripts, which have different secondary structures than complete transcripts. Moreover, the crosslinked sequences may represent suboptimal interactions rather than those optimal interactions anticipated by IntaRNA. It was also possible that RNA chaperons facilitated the crosslinked interactions. That is especially possible for Cj0920c, Cj1583c and *rp10* as IntaRNA output showed less stable optimal binding energy values to CjSA9, and hence the formation of these duplexes was likely to require the help of RNA chaperons.

RNAfold prediction of SRP RNA showed a single RNA hairpin, which agrees with the published structure. RNA hairpin is involved in the formation of RNA-protein complexes and targeting translating peptides to the membrane (Draper, 1999; Rosenblad et al., 2009). Complex formation with SPR protein and membrane proteins such as *E. coli* Ffh is part of the process that inserts nascent peptides into the membrane upon GTPase activation (Voigts-Hoffmann et al., 2013; Estrozi et al., 2011). Hence, the RNAfold result suggests that CjSA9 retains the hairpin structure of SRP RNA, enabling it to complex with SPR and other proteins.

The involvement of RNA-protein complex formation may explain the discrepancy between IntaRNA prediction and RNA crosslinking binding sites. SRP RNA complexes with SRP protein by the induced-fit mechanism. Upon complexing with proteins, the SRP RNA undergoes structural change and exposes some regions for binding to increase affinity and specificity, as demonstrated by the example between rRNA and S15 (Agalarov et al., 2000; Nikulin et al., 2000) and UTR RNA and U1A protein (Avis et al., 1996; Gubser and Varani, 1996). These interactions occur between the RNA and the amino acid side chains through hydrogen bonds and hydrophobic interactions. The RNA-protein associations result in the exposure of the RNA recognition sequence (Strub et al., 1999; Draper, 1999; Kligun and Mandel-Gutfreund, 2015). Such an induced structural change of SRP RNA has been observed from crystal structure obtained from archeon *Methanococcus jannaschii* (Hainzl et al., 2005). This structural change exposes or covers some part of the predicted secondary structure and enables interactions in regions that are theoretically not optimal for target binding.

CjSA9 starts at a similar position as SRP RNA but is nearly 90 bp longer in length. Genome-wide target prediction suggested one binding sequence at the SRP RNA 3' end position and another at the region after the SRP RNA 3' end. The SRP RNA 3' end targets *rp10*, Cj1583c, Cj0920c and *carB*. RNAfold also suggested that the region after SRP RNA 3' end formed a separate domain from the SRP RNA

sequence. Under most RNAtag-seq conditions, the expression coverage ends at the SRP RNA 3' end. The expression extends beyond the SRP RNA 3' end in some early stationary phase conditions, accounting for the extra length of CjSA9. These all hint towards SRP RNA and the CjSA9 3' end function independently in a condition-specific manner.

One possibility is that 3' end cleavage processes CjSA9 to mature into the annotated SRP RNA. The cleaved CjSA9 3' end degraded less in the early stationary phase. Another possibility is that CjSA9 3' end is a product of transcription readthrough or spurious transcription in the early stationary phase. In either case, CjSA9 3' end can target genes, including those responsible for purine metabolism.

5.6 Summary

This chapter has illustrated the experimental challenges for performing RNA crosslinking in *C. jejuni*. The results have generated a novel list of putative sRNA-mRNA interactions absent from computational prediction. The difference suggests that RNA crosslinking provides a complementary set of interactions that lack changes in transcription expression. Notably, CjSA9 appeared to use two distinct binding sites, with the crosslinking taking place near the SRP RNA 3' end. The CjSA9 3' end may play a more prominent role in the early stationary phase, as suggested by the expression coverage and the computed secondary structure. A biological mechanism was proposed based on these, which provided a novel insight into SRP RNA.

Chapter 6

Discussion and future directions

6.1 Limitation

The current study has evaluated multiple sRNA prediction tools, analyzed condition-specific sRNA activities from 21 conditions, and conducted *in vivo* RNA crosslinking. *C. jejuni* sRNA sequence and activity is difficult to examine due to the absence of known global RNA binding proteins, and condition-specific transcriptomic data is relatively limited. Published RNA-seq studies usually focus on relatively few experimental conditions and are conducted in different laboratories, introducing more technical batch effects. As a result, this study has overcome previous challenges using the in-house Cappable-seq and RNAtag-seq data covering 21 experimental conditions. Apart from computationally predicted sRNA-mRNA interactions, this study has also performed RNA crosslinking to uncover RNA-RNA binding that may not result in transcript abundance variation. This interdisciplinary approach has highlighted a complementary set of sRNA-mRNA duplexes.

ANNOgesic has predicted 96 sRNA from the in-house and published transcriptomic dataset. Further characterisation using *in silico* and molecular biology analysis have suggested biological relevant interactions involving CjSA21, CjSA110 and CjSA9. Despite this, the identity of novel predicted sRNA and the highlighted binding interactions still require experimental validation. No previous publications have described CjSA21 and the CjSA9 3' ends. Moreover, it is unclear whether computational and experimental approaches accurately depict the sRNA-mRNA interactions. Even if those interactions are genuine, their proposed binding mechanisms

and biological functions remain unconfirmed.

In addition to the lack of in-depth validation of highlighted sRNAs and interactions, hundreds of crosslinked and predicted interactions have yet to be studied. Investigating all these interactions is not straightforward due to the number of conditions and replicates involved in the in-house transcriptomic dataset. External researchers and future students in the host research group may not have adequate computational skills and insights into the dataset. Hence, future work by the broader research community may need analytic tools to extract biologically relevant information from the dataset in this study more efficiently.

Constructing a post-transcriptional network for *C. jejuni* also requires a wider range of NGS data. Additional sequencing data will fill in information missing from the RNA-seq data in the current study. The RNAtag-seq analysis in this study focuses on the change in RNA abundance but not the impact on protein translation or the metabolic pathways. Also, none of the RNAtag-seq conditions contains information on the *in vivo* interactions with the gut microbiota. Moreover, the crosslinking experiment is limited to the early stationary phase at 37 °C only. Hence, the current crosslinking data cannot fully complement the rest of the RNAtag-seq conditions.

6.2 Future work

6.2.1 Experimental work

Chapters 4 and 5 have speculated on the binding activities mechanisms of CjSA21, CjSA110 and CjSA9, making them attractive targets for detailed investigation. Future work can conduct northern blot to verify the presence of CjSA21 and the CjSA9 3' end. On the contrary, CjSA110 shares an identical genomic coordinate as an annotated sRNA and hence would not require northern blot verification. Their interactions can be confirmed using a gel-shift assay. Furthermore, deleting or overexpressing these sRNAs and their suggested targets will confirm their proposed biological functions.

Computational prediction and crosslinking results lead to speculation on different regulatory mechanisms. Different experimental methods can distinguish between regulatory mechanisms. For instance, chapter 4 suggests that CjSA21 and CjSA110 bind to their respective targets and alter mRNA stability. On the contrary, chapter 5 indicated CjSA9 crosslinking with four messenger RNA targets. The result suggests

CjSA9 may regulate translation initiation by structural rearrangement of crosslinked targets. In order to identify different mechanisms, northern blot and RT-PCR of the mRNAs can examine the alteration of mRNA stability by CjSA21 and CjSA110. Meanwhile, the impact of CjSA9 on mRNA translation and the levels of protein products can be elucidated using western blot.

Future experiments can also generate RNA-seq data from more conditions. For example, RNA crosslinking can be repeated by culturing the cells to the same conditions used in RNAtag-seq, thus allowing complete parallel comparisons between two sets of expression data. Moreover, both RNA crosslinking and RNAtag-seq can be performed by co-culturing with human intestinal cell lines to understand *in vivo* interactions with the gut microbiota.

6.2.2 R Shiny app

Apart from conducting more experiments, the currently available NGS data also need more thorough analysis. Repeating the RNAtag-seq analysis can be challenging for experimental researchers with limited programming experience, server access or knowledge of the dataset. That includes future students in the laboratory or the wider experimental microbial community. For example, generating gene expression output across all possible combinations of genes, replicates and conditions is extraordinarily time-consuming. Downloading output figures for all combinations will also waste lots of computational memory for data storage, especially for non-biologically relevant outputs. These factors can limit our data's benefits to internal and external users with limited bioinformatics expertise. Hence, one future direction is to build an R Shiny app to make our analysis more accessible and applicable.

A preliminary version has been built, with features that enable users to query different genes, pathways and conditions for gene co-expression and differential expression output. The app can return figures for users to download in response. The app is available here: `https://stephenrengoku.shinyapps.io/RShiny/`. So far, the features of the apps only included co-expression and differential expression outputs from Chapter 4. Possible next steps include adding more features, such as IntaRNA and RNA crosslinking data. Another drawback is that the app can only generate output from the input data parameters in chapter 4. Future improvement will also allow users to repeat gene expression analysis using user-defined input and parameters. Other directions of improvement are to improve the app layout, link to public databases, and make better descriptions. These modifications will make the tool more user-friendly for a wider audience.

6.2.3 Integrating with other NGS dataset

Further exploration of transcriptomic data is still not sufficient as it will likely miss out on mechanisms that do not change RNA abundance or RNA-RNA binding activities. The shortcomings of transcriptomic data can be complemented using other omics datasets. The rapid expansion of NGS technologies enables the faster generation of different omics datasets. Hence, future work can aim to integrate published or in-house *C. jejuni* omics datasets.

Genomics analysis can identify biomarkers responsible for environmental adaptations. For instance, *Campylobacter jejuni* from patient samples displayed nonsynonymous single nucleotide polymorphisms (SNPs) and frameshift mutations. Those genetic changes occurred most frequently among COG groups for cell motility, signal transduction and the major outer membrane protein. Mutations of these gene groups may enable *C. jejuni* to evade host defences (Bloomfield et al., 2021). These biomarkers are similar to sRNA targets identified in this study, such as *tlp1-4* for signal transduction and cell motility. That suggests *Campylobacter jejuni* may adjust its signal transduction and motility at both genomic and transcriptomic levels. Hence, future work can investigate genomic sequences from different sources to identify more biomarkers related to various environments and check if those biomarkers also appear in the interactions depicted in chapters 4 and 5.

A metagenomics dataset helps to understand how the population of *C. jejuni* varies according to the environment. 16S rRNA metabarcoding of faecal microbiota samples from *C. jejuni* infected patients showed a highly varied microbiome composition in terms of both bacterial phyla and genera. These samples also demonstrated varied proportions of *C. jejuni* subpopulations. These suggested that *C. jejuni* evolved after colonizing the gastrointestinal tract in response to different microbiota populations (Bloomfield et al., 2021). Apart from exploring *C. jejuni* abundance, meta-omics data can also explore the variation of sRNA expression. A recent study has applied meta-omics approaches to explore sRNA activities in natural environments. Metagenomic and metatranscriptomics were used to identify anaerobic methane oxidizing archaea group 1 (ANME-1) sRNAs that are highly expressed in the Guaymas Basin sample (Nawaz and Wang, 2021). A similar approach can correlate *C. jejuni* strains abundance and their sRNA expression in the microbiota and other natural environments. For example, our study predicted CjSA110 as a regulator of iron uptake. *C. jejuni* scavenge iron siderophores from other bacteria species in the environment. Hence, future work can explore if CjSA110 expression varies according to the microbiota composition of different patients' samples.

As mentioned, proteomic data from different conditions can elucidate the translational effect. Some published proteomics data covered similar conditions as our RNAtag-seq conditions and thus were valuable for comparisons. Proteomics data are available for exponential and early stationary phases (Turonova et al., 2017) and under bile stress (Man et al., 2020). However, other RNAtag-seq conditions are lacking among published proteomic datasets. Another study modelled infection by obtaining proteomic datasets from *C. jejuni* cells co-cultured with human (INT-407) and porcine (IPEC-1) intestinal cell lines were (Ayllón et al., 2017). Future work can generate proteomic data in conditions covered by other RNAtag-seq conditions. Conversely, RNA-seq data can be generated by co-culturing *C. jejuni* with the cell lines used in the above publication. Genomic, transcriptomic and proteomic data can identify gene groups or mutated pathways or differentially expressed under certain conditions. For example, our RNAtag-seq data showed differential expression of motility and signal transduction genes under food processing conditions. However, some phenotypes are more affected by metabolites such as carbohydrates and lipids. None of the above omics data measures the abundance of those metabolites. Future work can explore biological pathways in more detail by generating metabolomic data.

6.3 Summary

In conclusion, this study has combined computational and experimental methods to predict a list of sRNAs and their mRNA binding activities. CjSA21, CjSA110 and CjSA9 have been explored in detail, with their mechanisms being proposed based on their expression patterns or crosslinking interactions. More than 500 interactions were also proposed and can be investigated further. Future work can experimentally validate some highlighted interactions, develop an app for deeper data analysis, and integrate the current transcriptomic data with other types of NGS data. All these will further expand our understanding of *C. jejuni* stress adaptions by post-transcriptional regulation.

Bibliography

Abuoun, M., Manning, G., Cawthraw, S. A., Ridley, A., Ahmed, I. H., Wassenaar, T. M., and Newell, D. G. (2005). Cytolethal Distending Toxin (CDT)-Negative *Campylobacter jejuni* Strains and Anti-CDT Neutralizing Antibodies Are Induced during Human Infection but Not during Colonization in Chickens. *INFECTION AND IMMUNITY,*, 73(5):3053–3062.

Adams, P. P., Baniulyte, G., Esnault, C., Chegireddy, K., Singh, N., Monge, M., Dale, R. K., Storz, G., and Wade, J. T. (2021). Regulatory roles of *Escherichia coli* 5 ' UTR and ORF-internal RNAs detected by 3 ' end mapping. *eLife*, 10(e62438).

Agalarov, S. C., Sridhar, G., null Prasad, Funke, P. M., Stout, C. D., and Williamson, J. R. (2000). Structure of the S15,S6,S18-rRNA Complex: Assembly of the 30S Ribosome Central Domain. *Science*, 288(5463):107–112.

Akiba, M., Lin, J., Barton, Y. W., and Zhang, Q. (2006). Interaction of CmeABC and CmeDEF in conferring antimicrobial resistance and maintaining cell viability in *Campylobacter jejuni*. *Journal of Antimicrobial Chemotherapy*, 57:52–60.

Alzheimer, M., Svensson, S. L., Ko, F., Schweinlin, M., Metzger, M., Walles, H., and Sharma, C. M. (2020). A three-dimensional intestinal tissue model reveals factors and small regulatory RNAs important for colonization with *Campylobacter jejuni*. *PLoS Pathogens*, 16(2).

Anderson, J. K., Huang, J. Y., Wreden, C., Sweeney, E. G., Goers, J., James Remington, S., and Guillemin, K. (2015). Chemorepulsion from the quorum signal autoinducer-2 promotes *Helicobacter pylori* biofilm dispersal. *mBio*, 6(4).

Andresen, L., Martínez-Burgo, Y., Nilsson Zangelin, J., Rizvanovic, A., and Holmqvist, E. (2020). The Small Toxic *Salmonella* Protein TimP Targets the Cytoplasmic Membrane and Is Repressed by the Small RNA TimR. *mBio*, 11(6):e01659–20.

Apura, P., Saramago, M., Peregrina, A., Viegas, S. C., Carvalho, S. M., Saraiva, L. M., Arraiano, C. M., and Domingues, S. (2020). Tailor-made sRNAs: A plasmid tool to control the expression of target mRNAs in *Pseudomonas putida*. *Plasmid*.

Arias-Carrasco, R., Vásquez-Morán, Y., Nakaya, H. I., and Maracaja-Coutinho, V. (2018). StructRNAfinder: An automated pipeline and web server for RNA families prediction. *BMC Bioinformatics*, 19(55).

Arts, I. S., Vertommen, D., Baldin, F., Laloux, G., and Collet, J.-F. (2016). Comprehensively Characterizing the Thioredoxin Interactome *In Vivo* Highlights the Central Role Played by This Ubiquitous Oxidoreductase in Redox Control. *Molecular & cellular proteomics : MCP*, 15(6):2125–2140.

Askoura, M., Sarvan, S., Couture, J. F., and Stintzi, A. (2016). The *Campylobacter jejuni* ferric uptake regulator promotes acid survival and cross-protection against oxidative stress. *Infection and Immunity*, 84(5):1287–1300.

AU - Meng, X., AU - Cui, W., AU - Meng, X., AU - Wang, J., AU - Wang, J., and AU - Zhu, G. (2021). A non-coding small RNA MicC contributes to virulence in outer membrane proteins in *Salmonella Enteritidis*. *JoVE*, (167):e61808.

Avis, J. M., Allain, F. H., Howe, P. W., Varani, G., Nagai, K., and Neuhaus, D. (1996). Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *Journal of molecular biology*, 257(2):398–411.

Aw, J. G. A., Shen, Y., Nagarajan, N., and Wan, Y. (2017). Mapping RNA-RNA Interactions Globally Using Biotinylated Psoralen. *Journal of Visualized Experiments*, (123):1–10.

Axelsson-Olsson, D., Svensson, L., Olofsson, J., Salomon, P., Waldenström, J., Ellström, P., and Olsen, B. (2010). Increase in Acid Tolerance of *Campylobacter jejuni* through Coincubation with Amoebae. *Applied and Environmental Microbiology*, 76(13):4194–4200.

Ayllón, N., Jiménez-Marín, Á., Argüello, H., Zaldívar-López, S., Villar, M., Aguilar, C., Moreno, A., De La Fuente, J., and Garrido, J. J. (2017). Comparative proteomics reveals differences in host-pathogen interaction between infectious and commensal relationship with *Campylobacter jejuni*. *Frontiers in Cellular and Infection Microbiology*, 7(Article 145).

Azam, M. S. and Vanderpool, C. K. (2020). Translation inhibition from a distance: the small RNA SgrS silences a ribosomal protein S1-dependent enhancer. *Molecular Microbiology*, 00:1–18.

Balbontín, R., Villagra, N., Pardos de la Gándara, M., Mora, G., Figueroa-Bossi, N., and Bossi, L. (2016). Expression of IroN, the salmochelin siderophore receptor, requires mRNA activation by RyhB small RNA homologues. *Molecular Microbiology*, 100(1):139–155.

Ballouz, S., Verleyen, W., and Gillis, J. (2015). Gene expression Guidance for RNA-seq co-expression network construction and analysis : safety in numbers. *Bioinformatics*, 31(13):2123–2130.

Bandyra, K. J., Said, N., Pfeiffer, V., Górna, M. W., Vogel, J., and Luisi, B. F. (2012). The Seed Region of a Small RNA Drives the Controlled Destruction of the Target mRNA by the Endoribonuclease RNase E. *Molecular Cell*, 47:943–953.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., and Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology*, 19(5):455–477.

Barabási, A. L. and Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5:101–113.

Barman, R. K., Mukhopadhyay, A., and Das, S. (2017). An improved method for identification of small non-coding RNAs in bacteria using support vector machine. *Scientific Reports*, 7.

Barnawi, H., Masri, N., Hussain, N., Al-Lawati, B., Mayasari, E., Gulbicka, A., Jervis, A. J., Huang, M. H., Cavet, J. S., and Linton, D. (2020). RNA-based thermoregulation of a Campylobacter jejuni zinc resistance determinant. *PLoS Pathogens*, 16(10):e1009008.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315(March):1709–1712.

Beery, T., Hugdahl, M. B., and Doyle, M. P. (1988). Colonization of Gastrointestinal Tracts of Chicks by *Campylobacter jejuni*. *Applied and Environmental Microbiology*, 54(10):2365–2370.

Beisel, C. L. and Storz, G. (2011). The base pairing RNA Spot 42 participates in a multi-output feedforward loop to help enact catabolite repression in *Escherichia coli*. *Mol Cell*, 41(3):286–297.

Benoit, S. L. and Maier, R. J. (2021). Copper toxicity towards *Campylobacter jejuni* is enhanced by the nickel chelator dimethylglyoxime. *Metallomics*. mfab076.

Bhaduri, S. and Cottrell, B. (2004). Survival of cold-stressed *Campylobacter jejuni* on ground chicken and chicken skin during frozen storage. *Applied and Environmental Microbiology*, 70(12):7103–7109.

Birk, T., Ingmer, H., Andersen, M. T., Jørgensen, K., and Brøndsted, L. (2004). Chicken juice, a food-based model system suitable to study survival of *Campylobacter jejuni*. *Letters in Applied Microbiology*, 38(1):66–71.

Birk, T., Wik, M. T., Lametsch, R., and Knøchel, S. (2012). Acid stress response and protein induction in *Campylobacter jejuni* isolates with different acid tolerance. *BMC Microbiology*, 12(174).

Bischler, T., Tan, H. S., Nieselt, K., and Sharma, C. M. (2015). Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in *Helicobacter pylori*. *Methods*, 86:89–101.

Bloomfield, S. J., Midwinter, A. C., Biggs, P. J., French, N. P., Marshall, J. C., Hayman, D. T. S., Carter, P. E., Mather, A. E., Fayaz, A., Thornley, C., Kelly, D. J., and Benschop, J. (2021). Genomic adaptations of *Campylobacter jejuni* to long-term human colonization. *Gut Pathogens*, 13(72).

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

Bolton, F. J., Hutchinson, D. N., and Coates, D. (1984). Blood-free selective medium for isolation of *Campylobacter jejuni* from feces. *Journal of Clinical Microbiology*, 19(2):169–171.

Boratyn, G. M., Thierry-mieg, J., Thierry-mieg, D., Busby, B., and Madden, T. L. (2019). Magic-BLAST , an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics*, 20(405).

Bordini, M., Zappaterra, M., Soglia, F., Petracci, M., and Davoli, R. (2021). Weighted gene co-expression network analysis identifies molecular pathways and hub genes involved in broiler *White Striping* and *Wooden Breast* myopathies. *Scientific Reports*, 11(1776).

Bossi, L., Schwartz, A., Guillemardet, B., Boudvillain, M., and Figueroa-Bossi, N. (2012). A role for Rho-dependent polarity in gene regulation by a noncoding small RNA. *Genes and Development*, 26(16):1864–1873.

Bourqui, R., Dutour, I., Dubois, J., Benchimol, W., and Thébault, P. (2017). rNAV 2.0: a visualization tool for bacterial sRNA-mediated regulatory networks mining. *BMC bioinformatics*, 18(188).

Bronnec, V., Turo ˇnová, H., Bouju, A., Cruveiller, S., Rodrigues, R., Demnerova, K., Tresse, O., Haddad, N., and Brown, H. L. (2016). Adhesion , Biofilm Formation , and Genomic Features of *Campylobacter jejuni* Bf , an Atypical Strain Able to Grow under Aerobic Conditions. *Frontiers in Microbiology*, 7(Article 1002).

Bronowski, C., Mustafa, K., Goodhead, I. B., James, C. E., Nelson, C., Lucaci, A., Wigley, P., Humphrey, T. J., Williams, N. J., and Winstanley, C. (2017). *Campylobacter jejuni* transcriptome changes during loss of culturability in water. *PLoS ONE*, 12(11).

Brown, H. L., Reuter, M., Salt, L. J., Cross, K. L., Betts, R. P., and van Vliet, A. H. (2014). Chicken juice enhances surface attachment and biofilm formation of *Campylobacter jejuni*. *Applied and Environmental Microbiology*, 80(22):7053–7060.

Brüggemann, H., Montes, D. C., Plourde, L., and Popoff, M. R. (2021). Identification of a non-coding RNA and its putative involvement in the regulation of tetanus toxin synthesis in Clostridium tetani. *Scientific Reports*, 11(4157).

Buchanan, C. J., Webb, A. L., Mutschall, S. K., Kruczkiewicz, P., Barker, D. O. R., Hetman, B. M., Gannon, V. P. J., Abbott, D. W., Thomas, J. E., Inglis, G. D., and Taboada, E. N. (2017). A Genome-Wide Association Study to Identify Diagnostic Markers for Human Pathogenic *Campylobacter jejuni* Strains. *Frontiers in Microbiology*, 8.

Burgess, C. M., Gianotti, A., Gruzdev, N., Holah, J., Knøchel, S., Lehner, A., Margas, E., Esser, S. S., Sela Saldinger, S., and Tresse, O. (2016). The response of foodborne pathogens to osmotic and desiccation stresses in the food chain. *International Journal of Food Microbiology*, 221:37–53.

Burnham, P. M. and Hendrixson, D. R. (2018). *Campylobacter jejuni* : collective components promoting a successful enteric lifestyle. *Nature Reviews Microbiology*, 16.

Bussotti, G., Notredame, C., and Enright, A. J. (2013). Detecting and Comparing Non-Coding RNAs in the High-Throughput Era. *International journal of molecular sciences*, 14:15423–15458.

Butcher, J., Handley, R. A., van Vliet, A. H., and Stintzi, A. (2015). Refined analysis of the *Campylobacter jejuni* iron-dependent/independent Fur- and PerR-transcriptomes. *BMC Genomics*, 16(498).

Butcher, J., Sarvan, S., Brunzelle, J. S., Couture, J.-f., and Stintzi, A. (2012). Structure and regulon of *Campylobacter jejuni* ferric uptake regulator Fur define apo-Fur regulation. *PNAS*, 109(25):10047–10052.

Butcher, J. and Stintzi, A. (2013). The transcriptional landscape of *Campylobacter jejuni* under iron replete and iron limited growth conditions. *PLoS ONE*, 8(11):1–16.

Buzby, J. C., Roberts, T., and Ban Mishu, A. (1997). Estimated Annual Costs of Campylobacter-Associated Guillain-Barré Syndrome. *Agricultural Economic Report No . (AER-756) 33*.

Calvet, J. P. and Pederson, T. (1981). Base-pairing interactions between small nuclear RNAs and nuclear RNA precursors as revealed by psoralen cross-linking *in vivo*. *Cell*, 26(3 PART 1):363–370.

Cameron, A., Frirdich, E., Huynh, S., Parker, C. T., and Gaynor, E. C. (2012). Hyperosmotic Stress Response of *Campylobacter jejuni*. *Journal of Bacteriology*, 194(22).

Cameron, A., Huynh, S., Scott, N. E., Frirdich, E., Apel, D., Foster, L. J., Parker, C. T., and Gaynor, E. C. (2015). High-Frequency Variation of Purine Biosynthesis Genes Is a Mechanism of Success in *Campylobacter jejuni*. *mBio*, 6(5):e00612–15.

Candon, H. L., Allan, B. J., Fraley, C. D., and Gaynor, E. C. (2007). Polyphosphate kinase 1 is a pathogenesis determinant in *Campylobacter jejuni*. *Journal of Bacteriology*, 189(22):8099–8108.

Carpenter, B. M., Gancz, H., Gonzalez-nieves, R. P., West, A. L., Whitmire, J. M., Michel, S. L. J., and Merrell, D. S. (2009). A Single Nucleotide Change Affects Fur-Dependent Regulation of *sodB* in *H . pylori*. *PLoS ONE*, 4(4).

Cassat, J. E. and Skaar, E. P. (2013). Iron in Infection and Immunity. *Cell Host & Microbe Review*, 13:509–519.

Castillo, S., Heredia, N., and García, S. (2015). disinfectant disturb quorum-sensing activity and reduce motility and biofilm formation of *Campylobacter jejuni*. *Folia Microbiol*, 60(1):89–95.

Chandrashekhar, K., Gangaiah, D., and Pina-mimbela, R. (2015a). Transducer like proteins of *Campylobacter jejuni* 81-176 : role in chemotaxis and colonization of the chicken gastrointestinal tract. *Frontiers in Cellular and Infection Microbiology*, 5(Article 46).

Chandrashekhar, K., Kassem, I. I., Nislow, C., Gangaiah, D., Candelero-Rueda, R. A., and Rajashekara, G. (2015b). Transcriptome analysis of *Campylobacter jejuni* polyphosphate kinase (*ppk1* and *ppk2*) mutants. *Virulence*, 6(8):814–818.

Chandrashekhar, K., Srivastava, V., Hwang, S., Jeon, B., Ryu, S., and Rajashekara, G. (2018). Transducer-Like Protein in *Campylobacter jejuni* With a Role in Mediating Chemotaxis to Iron and Phosphate. *Frontiers in Microbiology*, 9(Article 2674).

Chao, Y., Li, L., Girodat, D., Wieden, H.-J., Luisi, B. F., and rg Vogel, J. (2017). *In Vivo* Cleavage Map Illuminates the Central Role of RNase E in Coding and Non-coding RNA Pathways. *Molecular Cell*, 65:39–51.

Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C. M., and Vogel, J. (2012). An atlas of Hfq-bound transcripts reveals 3 UTRs as a genomic reservoir of regulatory small RNAs. *EMBO Journal*, 31(20):4005–4019.

Chao, Y. and Vogel, J. (2016). A 3' UTR-Derived Small RNA Provides the Regulatory Noncoding Arm of the Inner Membrane Stress. *Molecular Cell*, 61:352–363.

Chareyre, S., Barras, F., and Mandin, P. (2019). A small RNA controls bacterial sensitivity to gentamicin during iron starvation. *PLoS genetics*, 15(4):e1008078.

Chaudhuri, R. R., Yu, L., Kanji, A., Perkins, T. T., Gardner, P. P., Choudhary, J., Maskell, D. J., Grant, A. J., and Grant, A. J. (2011). Quantitative RNA-seq analysis of the *Campylobacter jejuni* transcriptome. *Microbiology*, 157:2922–2932.

Chen, C., Qian, X., and Yoon, B. (2018). RNAdetect : Efficient computational detection of novel noncoding RNAs. *Bioinformatics*.

Christiansen, J. K., Nielsen, J. S., and Ebersbach, T. (2006). Functionality of cochlear micromechanics - As elucidated by upward spread of masking and two tone suppression. *RNA*, 12:1383–1396.

Clamp, M., Cuff, J., Searle, S. M., and Barton, G. J. (2004). The Jalview Java alignment editor. *Bioinformatics*, 20(3):426–427.

Clark, C., Berry, C., and Demczuk, W. (2019). Diversity of transducer-like proteins (Tlps) in *Campylobacter*. *PLoS ONE*, 14(3).

Clark, C. G., Chong, P. M., Mccorrister, S. J., Simon, P., Walker, M., Lee, D. M., Nguy, K., Cheng, K., Gilmour, M. W., and Westmacott, G. R. (2014). The CJIE1 prophage of *Campylobacter jejuni* affects protein expression in growth media with and without bile salts. *BMC Microbiology*, 14(70).

COD (2017). Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet Infectious Diseases*, 17(9):909–948.

Concordet, J. P. and Haeussler, M. (2018). CRISPOR: Intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Research*, 46:W242–W245.

Consortium, T. R. (2017). RNAcentral: A comprehensive database of non-coding RNA sequences. *Nucleic Acids Research*, 45(D1):D128–D134.

Crofts, A. A., Poly, F. M., Ewing, C. P., Kuroiwa, J. M., Rimmer, J. E., Harro, C., Sack, D., Talaat, K. R., Porter, C. K., Gutierrez, R. L., DeNearing, B., Brubaker, J., Laird, R. M., Maue, A. C., Jaep, K., Alcala, A., Tribble, D. R., Riddle, M. S., Ramakrishnan, A., McCoy, A. J., Davies, B. W., Guerry, P., and Trent, M. S. (2018). *Campylobacter jejuni* transcriptional and genetic adaptation during human infection. *Nature Microbiology*, 3:494–502.

Dai, R., Xia, Y., Liu, C., and Chen, C. (2019). csuWGCNA: a combination of signed and unsigned WGCNA to capture negative correlations. *bioRxiv*.

Dar, D. and Sorek, R. (2018). Bacterial Noncoding RNAs Excised from within Protein-Coding Transcripts. *mBio*, 9(5).

Darling, A. E., Mau, B., and Perna, N. T. (2010). Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*, 5(6).

Davis, L. and DiRita, V. (2008). Growth and laboratory maintenance of *Campylobacter jejuni*. *Curr Protoc Microbiol*.

Davis, L., Young, K., and DiRita, V. (2008). Genetic manipulation of Campylobacter jejuni. *Current protocols in microbiology*, Chapter 8:Unit 8A.2.1–8A.2.17.

de Bruijn, I. and Verhoeven, K. J. F. (2018). Cross-species interference of gene expression. *Nature Communications*, 9(1):5019.

Desnoyers, G. and Massé, E. (2012). Noncanonical repression of translation initiation through small RNA recruitment of the RNA chaperone Hfq. *Genes and Development*, 26:726–739.

Desnoyers, G., Morissette, A., Prévost, K., and Massé, E. (2009). Small RNA-induced differential degradation of the polycistronic mRNA *iscRSUA*. *EMBO Journal*, 28(11):1551–1561.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21.

Draper, D. E. (1999). Themes in RNA-protein recognition. *Journal of Molecular Biology*, 293(2):255–270.

Dugar, G., Herbig, A., Förstner, K. U., Heidrich, N., Reinhardt, R., Nieselt, K., and Sharma, C. M. (2013). High-Resolution Tra*nscriptome Maps Reveal Strain-Specific Regulatory Features of Multiple* Campylobacter jejuni Isolates. *c*, 9(5).

Dugar, G., Leenay, R. T., Eisenbart, S. K., Bischler, T., Aul, B. U., Beisel, C. L., and Sharma, C. M. (2018). CRISPR RNA-dependent binding and cleavage of endogenous RNAs by the *Campylobacter jejuni* Cas9. *Mol Cell.*, 69(5):893–905.

Dugar, G., Svensson, S. L., Bischler, T., Wäldchen, S., Reinhardt, R., Sauer, M., and Sharma, C. M. (2016). The CsrA-FliW network controls polar localization of the dual-function flagellin mRNA in *Campylobacter jejuni*. *Nature Communications*, 7(11667).

Duqué, B., Rezé, S., Rossero, A., Membré, J.-m., Guillou, S., and Haddad, N. (2021). Quantification of *Campylobacter jejuni* gene expression after successive stresses mimicking poultry slaughtering steps. *Food Microbiology*.

Durack, J., Ross, T., and Bowman, J. P. (2013). Characterisation of the Transcriptomes of Genetically Diverse *Listeria monocytogenes* Exposed to Hyperosmotic

and Low Temperature Conditions Reveal Global Stress-Adaptation Mechanisms. *PLoS ONE*, 8(9):e73603.

Durand, S., Callan-Sidat, A., McKeown, J., Li, S., Kostova, G., Hernandez-Fernaud, J., Alam, M. T., Millard, A., Allouche, D., Constantinidou, C., Condon, C., and Denham, E. (2021). Identification of an RNA sponge that controls the RoxS riboregulator of central metabolism in *Bacillus subtilis* . *Nucleic Acids Research*, 49(11):6399–6419.

Dwivedi, R., Nothaft, H., Garber, J., Kin, L. X., Stahl, M., Flint, A., Vliet, A. H. M. V., Stintzi, A., Szymanski, C. M., and Tg, A. (2016). L-fucose influences chemotaxis and biofilm formation in *Campylobacter jejuni*. *Molecular Microbiology*, 101(4):575–589.

Ebbert, M. T. W., Wadsworth, M. E., Staley, L. A., Hoyt, K. L., Pickett, B., Miller, J., Duce, J., Neuroimaging, D., Kauwe, J. S. K., and Ridge, P. G. (2016). Evaluating the necessity of PCR duplicate removal from next-generation sequencing data and a comparison of approaches. *BMC Bioinformatics*, 17(Suppl 7(239).

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

EFSA (2010). Scientific Opinion on Quantification of the risk posed by broiler meat to human campylobacteriosis in the EU. *EFSA Journal*, 8(1)(1437).

EFSA (2019). The European Union One Health 2018 Zoonoses Report. *EFSA Journal*, 17(12):5926.

Eisenbart, S. K., Alzheimer, M., Pernitzsch, S. R., Dietrich, S., Stahl, S., and Sharma, C. M. (2020). Repeat-Associated Small RNA Controls the Major Virulence Factors of *Helicobacter pylori*. *Molecular Cell*, 80:1–17.

Elgamoudi, B. A., Andrianova, E. P., Shewell, L. K., Day, C. J., King, R. M., Rahman, H., Hartley-tassell, L. E., Zhulin, I. B., and Korolik, V. (2021). The *Campylobacter jejuni* chemoreceptor Tlp10 has a bimodal ligand-binding domain and specificity for multiple classes of chemoeffectors. *Sci. Signal*, 14(eabc8521).

Elgamoudi, B. A., Andrianova, E. P., Shewell, L. K., Day, C. J., King, R. M., Taha, Rahman, H., Hartley-Tassell, L. E., Zhulin, I. B., and Korolik, V. (2017). The C*ampylobacter jejuni* chemoreceptor Tlp10 has a bimodal ligand-binding domain that promotes host colonization. *Sci Signal.*, 14(664).

Elgamoudi, B. A. and Ketley, J. M. (2018). Lighting up my life: a LOV-based fluorescent reporter for *Campylobacter jejuni*. *Research in Microbiology*, 169:108–114.

Elliott, K. T. and DiRita, V. J. (2008). Characterization of CetA and CetB, a bipartite energy taxis system in *Campylobacter jejuni*. *Molecular Microbiology*, 69(5):1091–1103.

Elvers, K. T., Turner, S. M., Wainwright, L. M., Marsden, G., Hinds, J., Cole, J. A., Poole, R. K., Penn, C. W., and Park, S. F. (2005). NssR, a member of the Crp-Fnr superfamily from *Campylobacter jejuni*, regulates a nitrosative stress-responsive regulon that includes both a single-domain and a truncated haemoglobin. *Molecular Microbiology*, 57(3):735–750.

Elvers, K. T., Wu, G., Gilberthorpe, N. J., Poole, R. K., and Park, S. F. (2004). Role of an inducible single-domain hemoglobin in mediating resistance to nitric oxide and nitrosative stress in *Campylobacter jejuni* and *Campylobacter coli*. *Journal of Bacteriology*, 186(16):5332–5341.

Estrozi, L. F., Boehringer, D., Shan, S. O., Ban, N., and Schaffitzel, C. (2011). Cryo-EM structure of the *E. coli* translating ribosome in complex with SRP and its receptor. *Nat Struct Mol Biol.*, 18(1):88–90.

Ettwiller, L., Buswell, J., Yigit, E., and Schildkraut, I. (2016). A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics*, 17(199).

Ezraty, B., Gennaris, A., Barras, F., and Collet, J. F. (2017). Oxidative stress, protein damage and repair in bacteria. *Nature Reviews Microbiology*, 15(7):385–396.

Farhadian, M., Rafat, S. A., Panahi, B., and Mayack, C. (2021). Weighted gene co-expression network analysis identifies modules and functionally enriched pathways in the lactation process. *Scientific Reports*, 11(2367).

Fei, J., Singh, D., Zhang, Q., Park, S., Balasubramanian, D., Golding, I., Vanderpool, C. K., and Ha, T. (2015). Determination of in vivo target search kinetics of regulatory noncoding RNA. *Science*, 347(6228).

Figueroa-Bossi, N., Valentini, M., Malleret, L., and Bossi, L. (2009). Caught at its own game: Regulatory small RNA inactivated by an inducible transcript mimicking its target. *Genes and Development*, 23:2004–2015.

Flint, A., Sun, Y. Q., Butcher, J., Stahl, M., Huang, H., and Stintzi, A. (2014). Phenotypic screening of a targeted mutant library reveals *Campylobacter jejuni* defenses against oxidative stress. *Infection and Immunity*, 82(6):2266–2275.

Flint, A., Sun, Y. Q., and Stintzi, A. (2012). Cj1386 is an ankyrin-containing protein involved in heme trafficking to catalase in *Campylobacter jejuni*. *Journal of Bacteriology*, 194(2):334–345.

Forman, J. J., Legesse-Miller, A., and Coller, H. A. (2008). A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proceedings of the National Academy of Sciences of the United States of America*, 105(39):14879–14884.

Fritsch, T. E., Siqueira, F. M., and Schrank, I. S. (2018). Global analysis of sRNA target genes in *Mycoplasma hyopneumoniae*. *BMC Genomics*, 19(767).

Fröhlich, K. S., Papenfort, K., Fekete, A., and Vogel, J. (2013). A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO Journal*, 32(22):2963–2979.

Gangaiah, D., Kassem, I. I., Liu, Z., and Rajashekara, G. (2009). Importance of polyphosphate kinase 1 for *Campylobacter jejuni* viable-but-nonculturable cell formation, natural transformation, and antimicrobial resistance. *Applied and Environmental Microbiology*, 75(24):7838–7849.

Gao, A. D., Zhang, Y., Liu, R., Fang, Z., and Lu, C. (2019). EsR240, a non-coding sRNA, is required for the resistance of *Edwardsiella tarda* to stresses in macrophages and for virulence Authors:. *Veterinary Microbiology*.

Garber, J. M., Nothaft, H., Pluvinage, B., Stahl, M., Bian, X., Por, S., Enriquez, A., Butcher, J., Huang, H., Glushka, J., Line, E., Gerlt, J. A., Azadi, P., Stintzi, A., Boraston, A. B., and Szymanski, C. M. (2020). The gastrointestinal pathogen *Campylobacter jejuni* metabolizes sugars with potential help from. *Communications Biology*, 3(2).

Gare´naux, A., Guillou, S., Ermel, G., Wren, B., Federighi, M., and Ritz, M. (2008). 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*, 159:718–726.

Gaynor, E. C., Wells, D. H., Mackichan, J. K., and Falkow, S. (2005). The *Campylobacter jejuni* stringent response controls specific stress survival and virulence-associated phenotypes. *Molecular Microbiology*, 56(1):8–27.

Ghodasara, A. and Voigt, C. A. (2017). Balancing gene expression without library construction via a reusable sRNA pool. *Nucleic Acids Research*, 45(13):8116–8127.

González, M. and Hänninen, M. L. (2012). Effect of temperature and antimicrobial resistance on survival of *Campylobacter jejuni* in well water: Application of the Weibull model. *Journal of Applied Microbiology*, 113:284–293.

Gormley, F. J., Bailey, R. A., Watson, K. A., McAdam, J., Avendaño, S., Stanley, W. A., and Koerhuis, A. N. (2014). *Campylobacter* colonization and proliferation in the broiler chicken upon natural field challenge is not affected by the bird growth rate or breed. *Applied and Environmental Microbiology*, 80(21):6733–6738.

Gruber, A. R., Findeiß, S., Washietl, S., Hofacker, I. L., and Stadler, P. F. (2010). RNAZ 2.0: Improved noncoding RNA detection. *Pacific Symposium on Biocomputing 2010, PSB 2010*, 79:69–79.

Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R., and Hofacker, I. L. (2008). The Vienna RNA websuite. *Nucleic acids research*, 36(Web Server issue):70–74.

Gu, W., Crawford, E. D., O'Donovan, B. D., Wilson, M. R., Chow, E. D., Retallack, H., and DeRisi, J. L. (2016). Depletion of Abundant Sequences by Hybridization (DASH): Using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biology*, 17(41).

Gubser, C. C. and Varani, G. (1996). Structure of the polyadenylation regulatory element of the human U1A pre-mRNA 3'-untranslated region and interaction with the U1A protein. *Biochemistry*, 35(7):2253–2267.

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

Gurung, P., Anand, P. K., Malireddi, R. K. S., Walle, L. V., Opdenbosch, N. V., Dillon, C. P., Weinlich, R., and Green, D. R. (2015). A Qrr non-coding RNA deploys four different regulatory mechanisms to optimize quorum-sensing dynamics. *Cell*, 160(0):228–240.

Gustavsen, J. A., Pai, S., Isserlin, R., Demchak, B., and Pico, A. R. (2019). RCy3 : Network biology using Cytoscape from within R. *F1000Research*, 8(1774).

Haddad, N., Burns, C. M., Bolla, J. M., Prévost, H., Fédérighi, M., Drider, D., and Cappelier, J. M. (2009). Long-term survival of *Campylobacter jejuni* at low temperatures is dependent on polynucleotide phosphorylase activity. *Applied and Environmental Microbiology*, 75(23):7310–7318.

Hainzl, T., Huang, S., and Sauer-Eriksson, A. E. (2005). Structural insights into SRP RNA: An induced fit mechanism for SRP assembly. *RNA*, 11(7):1043–1050.

Han, Z., Li, L., Willer, T., Baumg¨artner, W., and Rautenschlein, S. (2019). Adhesion and invasion of Campylobacter jejuni in chickens with a modified gut microbiota due to antibiotic treatment. *Veterinary Microbiology*.

Handley, R. A., Mulholland, F., Reuter, M., Ramachandran, V. K., Musk, H., Clissold, L., Le Brun, N. E., and Van Vliet, A. H. (2015). PerR controls oxidative stress defence and aerotolerance but not motility-associated phenotypes of *campylobacter jejuni*. *Microbiology*, 161:1524–1536.

Hao, H., Yuan, Z., Shen, Z., Han, J., Sahin, O., and Liu, P. (2013). Mutational and Transcriptomic Changes Involved in the Development of Macrolide Resistance in *Campylobacter jejuni*. *Antimicrobial Agents and Chemotherapy*, 57(3):1369–1378.

Hartley-Tassell, L. E., Shewell, L. K., Day, C. J., Wilson, J. C., Sandhu, R., Ketley, J. M., and Korolik, V. (2010). Identification and characterization of the aspartate chemosensory receptor of *Campylobacter jejuni*. *Molecular Microbiology*, 75(3):710–730.

He, J., Liu, K., Hou, X., and Lu, J. (2021). Identification and validation of key noncoding RNAs and mRNAs using co-expression network analysis in pre-eclampsia. *Medicine*, 100(14).

Hendrixson, D. R., Akerley, B. J., and DiRita, V. J. (2001). Transposon mutagenesis of *Campylobacter jejuni* identifies a bipartite energy taxis system required for motility. *Molecular Microbiology*, 40(1):214–224.

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., Mendonca, C., Dorrell, N., Wren, B. W., Wigley, P., Hall, N., and Winstanley, C. (2011). Genomic variations define divergence of water / wildlife-associated *Campylobacter jejuni*

niche specialists from common clonal complexes. *Environmental Microbiology*, 13(6):1549–1560.

Hermansen, G. M. M., Sazinas, P., Kofod, D., Millard, A., Andersen, P. S., and Jelsbak, L. (2018). Transcriptomic profiling of interacting nasal *staphylococci* species reveals global changes in gene and non-coding RNA expression. *FEMS Microbiology Letters*, 365:fny004.

Hickey, T. E., Baqar, S., Bourgeois, A. L., Ewing, C. P., and Guerry, P. (1999). *Campylobacter jejuni* -Stimulated Secretion of Interleukin-8 by INT407 Cells. *Infection and Immunity*, 67(1):88–93.

Hickey, T. E., Veigh, A. L. M. C., Scott, D. A., Michielutti, R. E., Bixby, A., Carroll, S. A., Bourgeois, A. L., and Guerry, P. (2000). *Campylobacter jejuni* Cytolethal Distending Toxin Mediates Release of Interleukin-8 from Intestinal Epithelial Cells. *Infect Immunity*, 68(12):6535–6541.

Hoang, K. V., Wang, Y., and Lin, J. (2012). Identification of genetic loci that contribute to *Campylobacter* resistance to fowlicidin-1 , a chicken host defense peptide. *Frontiers in Cellular and Infection Microbiology*, 2(Article 32).

Hofreuter, D., Novik, V., and Galán, J. E. (2008). Metabolic Diversity in *Campylobacter jejuni* Enhances Specific Tissue Colonization. *Cell Host and Microbe*, 4:425–433.

Holmes, C. W., Penn, C. W., and Lund, P. A. (2010). The *hrcA* and *hspR* regulons of *Campylobacter jejuni*. *Microbiology*, 156:158–166.

Hör, J., Gorski, S. A., and Vogel, J. (2018). Bacterial RNA Biology on a Genome Scale. *Molecular Cell*, 70(5):785–799.

Hou, J., Ye, X., Li, C., and Wang, Y. (2021). K-Module Algorithm : An Additional Step to Improve the Clustering Results of WGCNA Co-Expression Networks. *Genes*, 12(87).

Howlett, R. M., Hughes, B. M., Hitchcock, A., Kelly, D. J., and Kelly, D. J. (2012). Hydrogenase activity in the foodborne pathogen *Campylobacter jejuni* depends upon a novel ABC-type nickel transporter (NikZYXWV) and is SlyD-independent. *Microbiology*, 158:1645–1655.

Huang, Y., Sheth, R. U., Kaufman, A., and Wang, H. H. (2020). Scalable and cost-effective ribonuclease-based rRNA depletion for transcriptomics. *Nucleic Acids Research*, 48(4):e20.

Huergo, L. F., Rahman, H., Ibrahimovic, A., and Day, C. J. (2013). *Campylobacter jejuni* Dps Protein Binds DNA in the Presence of Iron or Hydrogen Peroxide. *Journal Of Bacteriology*, 195(9):1970–1978.

Hughes, R. A., Cogan, T., and Humphrey, T. (2010). Exposure of *campylobacter jejuni* to 6°C: Effects on heat resistance and electron transport activity. *Journal of Food Protection*, 73(4):729–733.

Hughes, R. A., Hallett, K., Cogan, T., Enser, M., and Humphrey, T. (2009). The response of *Campylobacter jejuni* to low temperature differs from that of *Escherichia coli*. *Applied and Environmental Microbiology*, 75(19):6292–6298.

Hull, D. M., van Vliet, A. H. M., Correa, M., and Thakur, S. (2021). Antimicrobial resistance and interspecies gene transfer in *Campylobacter coli* and *Campylobacter jejuni* isolated from food animals , poultry processing , and retail meat in North Carolina, 2018–2019. *PLoS ONE*, 16(2).

Hwang, S., Jeon, B., Yun, J., and Ryu, S. (2011a). Roles of RpoN in the resistance of *Campylobacter jejuni* under various stress conditions. *BMC Microbiology*, 11.

Hwang, S., Jeon, B., Yun, J., and Ryu, S. (2011b). Roles of RpoN in the resistance of *Campylobacter jejuni* under various stress conditions. *BMC Microbiology*, 11.

Hwang, S., Zhang, Q., Ryu, S., and Jeon, B. (2012). Transcriptional Regulation of the CmeABC Multidrug Efflux Pump and the KatA Catalase by CosR in *Campylobacter jejuni*. *Journal of Bacteriology*, 194(24):6883–6891.

Ikeda, T., Shinagawa, T., Ito, T., Ohno, Y., Kubo, A., Nishi, J., Gotoh, Y., Ogura, Y., Ooka, T., and Hayashi, T. (2020). Hypoosmotic stress induces flagellar biosynthesis and swimming motility in *Escherichia albertii*. *Communications Biology*, 3(87).

Ikeda, Y., Yagi, M., Morita, T., and Aiba, H. (2011). Hfq binding at RhlB-recognition region of RNase E is crucial for the rapid degradation of target mRNAs mediated by sRNAs in *Escherichia coli*. *Molecular Microbiology*, 79(2):419–432.

Iosub, I. A., Marchioretto, M., Sy, B., McKellar, S., Nieken, K. J., van Nues, R. W., Tree, J. J., Viero, G., and Granneman, S. (2020). Hfq CLASH uncovers sRNA-target interaction networks enhancing adaptation to nutrient availability. *eLife*, 9(e54655).

Jaakkonen, A., Kivisto, R., Aarnio, M., Kalekivi, J., and Hakkinen, M. (2020). Persistent contamination of raw milk by *Campylobacter jejuni* ST-883. *PLOS ONE*, 15(4):e0231810.

Jagodnik, J., Chiaruttini, C., and Guillier, M. (2017). Stem-Loop Structures within mRNA Coding Sequences Activate Translation Initiation and Mediate Control by Small Regulatory RNAs. *Molecular Cell*, 68:158–170.

Jaishankar, J. and Srivastava, P. (2017). Molecular Basis of Stationary Phase Survival and Applications. *Frontiers in Microbiology*, 8(Article 2000).

Jang, S. and Imlay, J. A. (2007). MICROMOLAR INTRACELLULAR HYDROGEN PEROXIDE DISRUPTS METABOLISM BY DAMAGING IRON-SULFUR ENZYMES. *J Biol Chem*, 282(2):929–937.

Janssen, K. H., Corley, J. M., Djapgne, L., Cribbs, J. T., Voelker, D., Slusher, Z., Nordell, R., Regulski, E. E., Kazmierczak, B. I., Mcmackin, W., and Yahr, L. (2020). Hfq and sRNA 179 Inhibit Expression of the *Pseudomonas aeruginosa* cAMP-Vfr and Type III Secretion Regulons. *mBio*, 11(3):e00363–20.

Janssen, K. H., Diaz, M. R., Golden, M., Graham, J. W., Sanders, W., Wolfgang, M. C., and Yahr, T. L. (2018). Functional analyses of the RsmY and RsmZ small noncoding regulatory RNAs in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 200(11):e00736–17.

Kalmokoff, M., Lanthier, P., Tremblay, T.-l., Foss, M., Lau, P. C., Sanders, G., Austin, J., Kelly, J., and Szymanski, C. M. (2006). Proteomic Analysis of *Campylobacter jejuni* 11168 Biofilms Reveals a Role for the Motility Complex in Biofilm Formation. *Journal of Bacteriology*, 188(12):4312–4320.

Kalvari, I., Nawrocki, E. P., Argasinska, J., Quinones-Olvera, N., Finn, R. D., Bateman, A., and Petrov, A. I. (2018). Non-Coding RNA Analysis Using the Rfam Database. *Current Protocols in Bioinformatics*, 62(1).

Kannaiah, S., Livny, J., and Amster-choder, O. (2019). Spatiotemporal Organization of the *E . coli* Transcriptome : Translation Independence and Engagement in Regulation. *Molecular Cell*, 76:574–589.

Kazantsev, A. V. and Pace, N. R. (2006). Bacterial RNase P : a new view of an ancient enzyme. *Nature Reviews Microbiology*, 4:729–740.

Kelly, A. F., Park, S. F., Bovill, R., and Mackey, B. M. (2001). 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.

Kendall, J. J., Barrero-tobon, A. M., Hendrixson, D. R., and Kelly, D. J. (2014). Hemerythrins in the microaerophilic bacterium *Campylobacter jejuni* help protect key iron – sulphur cluster enzymes from oxidative damage. *Environmental Microbiology*, 16(4):1105–1121.

Khan, M. F., Machuca, M. A., Rahman, M. M., Koç, C., Norton, R. S., Smith, B. J., and Roujeinikova, A. (2020). Structure–activity relationship study reveals the molecular basis for specific sensing of hydrophobic amino acids by the *Campylobacter jejuni* chemoreceptor TLP3. *Biomolecules*, 10(744).

Khanna, M. R., Bhavsar, S. P., and Kapadnis, B. P. (2006). Effect of temperature on growth and chemotactic behaviour of *Campylobacter jejuni*. *Letters in Applied Microbiology*, 43(1):84–90.

Kim, E., Koo, T., Park, S. W., Kim, D., Kim, K., Cho, H. Y., Song, D. W., Lee, K. J., Jung, M. H., Kim, S., Kim, J. H., Kim, J. H., and Kim, J. S. (2017). *In vivo* genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nature Communications*, 8(14500).

Kim, J.-c., Oh, E., Hwang, S., Ryu, S., and Jeon, B. (2015). Non-selective regulation of peroxide and superoxide resistance genes by PerR in *Campylobacter jejuni*. *Frontiers in Microbiology*, 6(Article 126).

Kim, M., Hwang, S., Ryu, S., and Jeon, B. (2011). Regulation of *perR* Expression by Iron and PerR in *Campylobacter jejuni*. *Journal of bacteriology*, 193(22):6171–6178.

Kinoshita-daitoku, R., Kiga, K., Miyakoshi, M., Otsubo, R., Ogura, Y., Sanada, T., Bo, Z., Phuoc, T. V., Okano, T., Iida, T., Yokomori, R., Kuroda, E., Hirukawa, S., Tanaka, M., Sood, A., Subsomwong, P., Ashida, H., Binh, T. T., Nguyen, L. T., Van, K. V., Quy, D., Ho, D., and Nakai, K. (2021). A bacterial small RNA regulates the adaptation of *Helicobacter pylori* to the host environment. *Nature Communications*, 12(2085).

Klancnik, A., Botteldoorn, N., Herman, L., and Možina, S. S. (2006). Survival and stress induced expression of groEL and rpoD of *Campylobacter jejuni* from

different growth phases. *International Journal of Food Microbiology 112*, 112:200–207.

Klancnik, A., Guzej, B., Jamnik, P., Vuckovic, D., Abram, M., and Mozina, S. S. (2009). Stress response and pathogenic potential of *Campylobacter jejuni* cells exposed to starvation. *Research in Microbiology*, 160:345–352.

Klančnik, A., Vučković, D., Jamnik, P., Abram, M., and Možina, S. S. (2014). Stress response and virulence of heat-stressed *Campylobacter jejuni*. *Microbes and Environments*, 29(4):338–345.

Kligun, E. and Mandel-Gutfreund, Y. (2015). The role of RNA conformation in RNA-protein recognition. *RNA Biology*, 12(7):720–727.

Koeppen, K., Hampton, T. H., Jarek, M., Scharfe, M., Gerber, S. A., Mielcarz, D. W., Demers, E. G., Dolben, E. L., Hammond, J. H., Hogan, D. A., and Stanton, B. A. (2016). A Novel Mechanism of Host-Pathogen Interaction through sRNA in Bacterial Outer Membrane Vesicles. *PLoS Pathogens*, 12(6):1–22.

Konkel, M. E., Klena, J. D., Rivera-Amill, V., Monteville, M. R., Biswas, D., Raphael, B., and Mickelson, J. (2004). Secretion of virulence proteins from *Campylobacter jejuni* is dependent on a functional flagellar export apparatus. *Journal of Bacteriology*, 186(11):3296–3303.

Kumar-Phillips, G. S., Hanning, I., and Slavik, M. (2013). Influence of acid-adaptation of campylobacter jejuni on adhesion and invasion of int 407 cells. *Foodborne Pathogens and Disease*, 10(12):1037–1043. PMID: 23952474.

Laconi, A., Drigo, I., Palmieri, N., Carraro, L., Tonon, E., Franch, R., Bano, L., and Piccirillo, A. (2021). Genomic analysis of extra-intestinal *Campylobacter jejuni* and *Campylobacter coli* isolated from commercial chickens. *Veterinary Microbiology*.

Lam, J. (2019). *Condition-dependent transcriptional landscape of Campylobacter jejuni*. PhD thesis.

Langfelder, P. and Horvath, S. (2008). WGCNA : an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(559).

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., Mcwilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007). Clustal W and Clustal X version 2 . 0. *Bioinformatics*, 23(21):2947–2948.

Le, M. T., Porcelli, I., Weight, C. M., Gaskin, D. J. H., Carding, S. R., and van Vliet, A. H. M. (2012). Acid-shock of *Campylobacter jejuni* induces flagellar gene expression and host cell invasion . *European Journal of Microbiology and Immunology*, 2(1):12–19.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C . elegans* Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to & II-14. *Cell*, 75:843–854.

Leonard, S., Meyer, S., Lacour, S., Nasser, W., Hommais, F., and Reverchon, S. (2019). APERO : a genome-wide approach for identifying bacterial small RNAs from RNA-Seq data. *Nucleic Acids Research*, 47(15).

Lertpiriyapong, K., Gamazon, E. R., Feng, Y., Park, D. S., Pang, J., Botka, G., Graffam, M. E., Ge, Z., and Fox, J. G. (2012). *Campylobacter jejuni* type VI secretion system: roles in adaptation to deoxycholic acid, host cell adherence, invasion, and *in vivo* colonization. *PloS one*, 7(8):e42842.

Li, J., Gulbronson, C. J., Bogacz, M., Hendrixson, D. R., and Thompson, S. A. (2018). Liw controls growth-phase expression of *campylobacter jejuni* flagellar and non-flagellar proteins via the post-transcriptional regulator CsrA. *Microbiology*, 164:1308–1319.

Li, W., Wang, L., Wu, Y. U. E., Yuan, Z., and Zhou, J. (2020a). Weighted gene co-expression network analysis to identify key modules and hub genes associated with atrial fibrillation. *INTERNATIONAL JOURNAL OF MOLECULAR MEDICINE*, 45:401–416.

Li, W., Wang, L., Wu, Y. U. E., Yuan, Z., and Zhou, J. (2020b). Weighted gene co-expression network analysis to identify key modules and hub genes associated with atrial fibrillation. *INTERNATIONAL JOURNAL OF MOLECULAR MEDICINE*, 45:401–416.

Li, Y. P., Vegge, C. S., Brøndsted, L., Madsen, M., Ingmer, H., and Bang, D. D. (2011). *Campylobacter jejuni* induces an anti-inflammatory response in human intestinal epithelial cells through activation of phosphatidylinositol 3-kinase/Akt pathway. *Veterinary Microbiology*, 148(1):75–83.

Li, Z., Lou, H., Ojcius, D. M., Sun, A., Sun, D., Zhao, J., Lin, X., and Yan, J. (2014). Methyl-accepting chemotaxis proteins 3 and 4 are responsible for *Campylobacter jejuni* chemotaxis and jejuna colonization in mice in response to sodium deoxycholate. *Journal of Medical Microbiology*, 63:343–354.

Li de la Sierra-Gallay, I., Zig, L., Jamalli, A., and Putzer, H. (2008). Structural insights into the dual activity of RNase J. *Nature Structural and Molecular Biology*, 15(2):206–212.

Liao, Y., Smyth, G. K., and Shi, W. (2014). Sequence analysis featureCounts : an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923–930.

Liaw, J., Hong, G., Davies, C., Elmi, A., Sima, F., Stratakos, A., Stef, L., Pet, I., Hachani, A., Corcionivoschi, N., Wren, B. W., Gundogdu, O., and Dorrell, N. (2019). The *Campylobacter jejuni* Type VI Secretion System Enhances the Oxidative Stress Response and Host Colonization. *Frontiers in Microbiology*, 10(Article 2864).

Lin, J., Akiba, M., Sahin, O., and Zhang, Q. (2005). CmeR functions as a transcriptional repressor for the multidrug efflux pump CmeABC in *Campylobacter jejuni*. *Antimicrobial Agents and Chemotherapy*, 49(3):1067–1075.

Lin, J., Sahin, O., Michel, L. O., and Zhang, Q. (2003). Critical role of multidrug efflux pump CmeABC in bile resistance and in vivo colonization of *Campylobacter jejuni*. *Infection and Immunity*, 71(8):4250–4259.

Liu, T., Zhang, K., Xu, S., Wang, Z., Fu, H., Tian, B., Zheng, X., and Li, W. (2017). Detecting RNA-RNA interactions in *E. coli* using a modified CLASH method. *BMC Genomics*, 18(1):1–11.

Looft, T., Cai, G., Choudhury, B., Lai, L. X., Lippolis, J. D., Reinhardt, T. A., Sylte, M. J., and Casey, T. A. (2019). Avian Intestinal Mucus Modulates *Campylobacter jejuni* Gene Expression in a Host-Specific Manner. *Frontiers in Microbiology*, 9(Article 3215).

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(550).

Lu, L., Wei, R., Bhakta, S., Waddell, S. J., and Boix, E. (2021). Weighted Gene Co-Expression Network Analysis Identifies Key Modules and Hub Genes Associated with *Mycobacterial* Infection of Human Macrophages. *antibiotics*, 10(97).

Lu, Z., Zhang, Q. C., Lee, B., Flynn, R. A., Smith, M. A., Robinson, J. T., Davidovich, C., Gooding, A. R., Goodrich, K. J., Mattick, J. S., Mesirov, J. P., Cech, T. R., and Chang, H. Y. (2016). RNA Duplex Map in Living Cells Reveals Higher-Order Transcriptome Structure. *Cell*, 165(5):1267–1279.

Luethy, P. M., Huynh, S., Ribardo, D. A., Winter, S. E., Parker, C. T., and Hendrixson, D. R. (2017). Microbiota-derived short-chain fatty acids modulate expression of *Campylobacter jejuni* determinants required for commensalism and virulence. *mBio*, 8(3):1–26.

Luo, X., Esberard, M., Bouloc, P., and Jacq, A. (2021). A Small Regulatory RNA Generated from the *malK* 5' Untranslated Region Targets Gluconeogenesis in *Vibrio* Species. *mSphere 6*, 6(3):e00134–21.

Ma, L., Konkel, M. E., and Lu, X. (2021). Antimicrobial-resistance gene transfer of *Campylobacter jejuni* in mono- and dual-species biofilms. *Appl Environ Microbiol*.

Machuca, M. A., Liu, Y. C., Beckham, S. A., Gunzburg, M. J., and Roujeinikova, A. (2016). The crystal structure of the tandem-PAS sensing domain of *Campylobacter jejuni* chemoreceptor Tlp1 suggests indirect mechanism of ligand recognition. *Journal of Structural Biology*, 194(2):205–213.

Magajna, B. A. and Schraft, H. (2015). *Campylobacter jejuni* biofilm cells become viable but non-culturable (VBNC) in low nutrient conditions at 4 °C more quickly than their planktonic counterparts. *Food Control*, 50:45–50.

Mai, J., Rao, C., Watt, J., Sun, X., Lin, C., Zhang, L., and Liu, J. (2019). *Mycobacterium tuberculosis* 6C sRNA binds multiple mRNA targets via C-rich loops independent of RNA chaperones. *Nucleic Acids Research*.

Malik-Kale, P., Parker, C. T., and Konkel, M. E. (2008). Culture of *Campylobacter jejuni* with sodium deoxycholate induces virulence gene expression. *Journal of Bacteriology*, 190(7):2286–2297.

Man, L., Dale, A. L., Klare, W. P., Cain, J. A., Sumer-Bayrakter, Z., Niewold, P., Solis, N., and Cordwell, S. J. (2020). Proteomics of *Campylobacter jejuni* growth in deoxycholate reveals Cj0025c as a cystine transport protein required for wild-type human infection phenotypes. *Mol Cell Proteomics*.

Marchant, J., Wren, B., and Ketley, J. (2002). Exploiting genome sequence: Predictions for mechanisms of *Campylobacter* chemotaxis. *Trends in Microbiology*, 10(4):155–159.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10–12.

Martı, A. and Mackey, B. M. (2005). Physiological changes in *Campylobacter jejuni* on entry into stationary phase. *International Journal of Food Microbiology*, 101:1–8.

Massé, E., Escorcia, F. E., and Gottesman, S. (2003). Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli* Coupled degradation of a small regulatory RNA and its mRNA targets in *Escherichia coli*. pages 2374–2383.

McCaskill, J. S. (1990). The equilibrium partition function and base pair binding probabilities for rna secondary structure. *Biopolymers*, 29(6-7):1105–1119.

McGinnis, S. and Madden, T. L. (2004). BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, 32(WEB SERVER ISS.):20–25.

Mclennan, M. K., Ringoir, D. D., Frirdich, E., Svensson, S. L., Wells, D. H., Jarrell, H., Szymanski, C. M., and Gaynor, E. C. (2008). *Campylobacter jejuni* Biofilms Up-Regulated in the Absence of the Stringent Response Utilize a Calcofluor White-Reactive Polysaccharide. *JOURNAL OF BACTERIOLOGY*, 190(3):1097–1107.

Melamed, S., Peer, A., Faigenbaum-Romm, R., Gatt, Y. E., Reiss, N., Bar, A., Altuvia, Y., Argaman, L., and Margalit, H. (2016). Global Mapping of Small RNA-Target Interactions in Bacteria. *Molecular Cell*, 63(5):884–897.

Melson, E. M. and Kendall, M. M. (2019). The sRNA DicF integrates oxygen sensing to enhance enterohemorrhagic *Escherichia coli* virulence via distinctive RNA control mechanisms. *National Academy of Sciences*.

Mercier, E., Holtkamp, W., Rodnina, M. V., and Wintermeyer, W. (2017). Signal recognition particle binds to translating ribosomes before emergence of a signal anchor sequence. *Nucleic acids research*, 45(20):11858–11866.

Mihaljevic, R. R., Sikic, M., Klancnik, A., Brumini, G., Mozina, S. S., and Abram, M. (2007). Environmental stress factors affecting survival and virulence of *Campylobacter jejuni*. *Microbial Pathogenesis*, 43(2-3):120–125.

Mishu, B. and Blaser, M. J. (1993). Role of Infection Due to *Campylobacter jejuni* in the Initiation of Guillain-Barré Syndrome. *Clinical Infectious Diseases*, 17(1):104–108.

Miyakoshi, M., Chao, Y., and Vogel, J. (2015). Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *The EMBO Journal*, 34:1478–1492.

Miyakoshi, M., Matera, G., Maki, K., Sone, Y., and Vogel, J. (2018). Functional expansion of a TCA cycle operon mRNA by a 3 end-derived small RNA. *Nucleic Acids Research*, pages 1–14.

Monteville, M. R., Yoon, J. E., and Konkel, M. E. (2003). Maximal adherence and invasion of INT 407 cells by *Campylobacter jejuni* requires the CadF outer-membrane protein and microfilament reorganization. *Microbiology*, 149:153–165.

Moon, K., Sim, M., Tai, C.-h., Yoo, K., Merzbacher, C., Yu, S.-h., and Kim, D. D. (2021). Identification of BvgA-Dependent and BvgA-Independent Small RNAs (sRNAs) in *Bordetella pertussis* Using the Prokaryotic sRNA Prediction Toolkit ANNOgesic. *Microbiol Spectr*, 9(e00044-21).

Morgan, R., Kohn, S., Hwang, S. H., Hassett, D. J., and Sauer, K. (2006). BdlA, a Chemotaxis regulator essential for biofilm dispersion in *Pseudomonas aeruginosa*. *Journal of Bacteriology*, 188(21):7335–7343.

Moriano-Gutierrez, S., Bongrand, C., Essock-burns, T., Wu, L., Mcfall-ngai, M. J., and Ruby, E. G. (2020). The noncoding small RNA SsrA is released by *Vibrio fischeri* and modulates critical host responses. *PLoS Biol*, 18(11):e3000934.

Morita, T., Maki, K., and Aiba, H. (2005). RNase E-based ribonucleoprotein complexes : mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes & Development*, 19:2176–2186.

Mortensen, N. P., Schiellerup, P., Boisen, N., Klein, B. M., Locht, H., Abuoun, M., Newell, D., and Krogfelt, K. A. (2011). The role of *Campylobacter jejuni* cytolethal distending toxin in gastroenteritis : toxin detection , antibody production , and clinical outcome. *APMIS*, 119:626–634.

Moser, S., Seth-smith, H., Egli, A., and Kittl, S. (2020). *Campylobacter jejuni* from Canine and Bovine Cases of Campylobacteriosis Express High Antimicrobial Resistance Rates against (Fluoro) quinolones and Tetracyclines. *pathogens*, 9(691).

Mouftah, S. F., Cobo-Díaz, J. F., Álvarez-Ordóñez, A., Mousa, A., Calland, J. K., Pascoe, B., Sheppard, S. K., and Elhadidy, M. (2021). Stress-resistance associ-

ated with multi-host transmission and enhanced biofilm formation at 42°C among hyper-aerotolerant generalist *Campylobacter jejuni*. *Food Microbiology*.

Mourkas, E., Taylor, A. J., Méric, G., Bayliss, S. C., Pascoe, B., Mageiros, L., Calland, J. K., Hitchings, M. D., Ridley, A., Vidal, A., Forbes, K. J., Strachan, N. J. C., Parker, C. T., Parkhill, J., Jolley, K. A., Cody, A. J., Maiden, M. C. J., Kelly, D. J., and Sheppard, S. K. (2020). Agricultural intensification and the evolution of host specialism in the enteric pathogen *Campylobacter jejuni*. *Proceedings of the National Academy of Sciences*.

Mraheil, M. A., Billion, A., Mohamed, W., Mukherjee, K., Kuenne, C., Pischimarov, J., Krawitz, C., Retey, J., Hartsch, T., Chakraborty, T., and Hain, T. (2011). The intracellular sRNA transcriptome of *Listeria monocytogenes* during growth in macrophages. *Nucleic Acids Research*, 39(10):4235–4248.

Mukherjee, S., Banerjee, B., Karasik, D., and Frenkel-morgenstern, M. (2021). mRNA-lncRNA Co-Expression Network Analysis Reveals the Role of lncRNAs in Immune Dysfunction during Severe SARS-CoV-2 Infection. *Viruses*, 13(402).

Müller, A., León-kempis, M. R., Dodson, E., Wilson, K. S., Wilkinson, A. J., Kelly, D. J., Court, F., Bank, W., and Sheffield, S. (2007). A Bacterial Virulence Factor with a Dual Role as an Adhesin and a Solute-binding Protein : The Crystal Structure at 1 .5 Å Resolution of the PEB1a Protein from the Food-borne Human Pathogen *Campylobacter jejuni*. *J. Mol. Biol.*, 372:160–171.

Muraoka, W. T. and Zhang, Q. (2011). Phenotypic and Genotypic Evidence for L-Fucose Utilization by *Campylobacter jejuni*. *JOURNAL OF BACTERIOLOGY*, 193(5):1065–1075.

Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., and Lee, S. Y. (2013). Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nature Biotechnology*, 31(2):170–174.

Nakagawa, R., Ishiguro, S., Okazaki, S., Mori, H., Tanaka, M., Aburatani, H., Yachie, N., Nishimasu, H., and Nureki, O. (2021). Engineered *Campylobacter jejuni* Cas9 variant with enhanced activity. *Research Square*.

Nataro, J. P. and Guerrant, R. L. (2017). Chronic consequences on human health induced by microbial pathogens: Growth faltering among children in developing countries. *Vaccine*, 35:6807–6812.

Nawaz, M. Z. and Wang, F. (2021). Meta-omics approaches reveal unique small rnas exhibited by the uncultured microorganisms dwelling deep-sea hydrothermal sediment in guaymas basin.

Negretti, N. M., Gourley, C. R., Clair, G., Adkins, J. N., and Konkel, M. E. (2017). The food-borne pathogen Campylobacter jejuni responds to the bile salt deoxycholate with countermeasures to reactive oxygen species. *Scientific Reports*, 7(15455).

Negretti, N. M., Gourley, C. R., Talukdar, P. K., Clair, G., Klappenbach, C. M., Lauritsen, C. J., Adkins, J. N., and Konkel, M. E. (2021). The *Campylobacter jejuni* CiaD effector co-opts the host cell protein IQGAP1 to promote cell entry. *Nature Communications*, 1339(12).

Neubacher, N., Tobias, N. J., Huber, M., Cai, X., Glatter, T., Pidot, S. J., Stinear, T. P., Lütticke, A. L., Papenfort, K., and Bode, H. B. (2020). Symbiosis , virulence and natural-product biosynthesis in entomopathogenic bacteria are regulated by a small RNA. *Nature Microbiology*, 5:1481–1489.

Nielsen, J. S., Lei, L. K., Ebersbach, T., Olsen, A. S., Klitgaard, J. K., Valentin-Hansen, P., and Kallipolitis, B. H. (2009). Defining a role for Hfq in Gram-positive bacteria: Evidence for Hfq-dependent antisense regulation in *Listeria monocytogenes*. *Nucleic Acids Research*, 38(3):907–919.

Nikulin, A., Serganov, A., Ennifar, E., Tishchenko, S., Nevskaya, N., Portier, C., Garber, M., Ehresmann, C., Nikonov, S., and Dumas, P. (2000). Crystal structure of the S15–rRNA complex. *Nature structural Biology*, 7(4):273–277.

Noh, M., Yoo, S. M., Kim, W. J., and Lee, S. Y. (2017). Gene Expression Knockdown by Modulating Synthetic Small RNA Expression in *Escherichia coli*. *Cell Systems*, 5:418–426.

Nyati, K. K. and Nyati, R. (2013). Role of Campylobacter jejuni Infection in the Role of Campylobacter jejuni Infection in the Pathogenesis of Guillain- Barré Syndrome. *BioMed Research International*.

Nyati, K. K., Prasad, K. N., Verma, A., Singh, A. K., Rizwan, A., Sinha, S., Paliwal, V. K., and Pradhan, S. (2010). Association of TLR4 Asp299Gly and Thr399Ile polymorphisms with Guillain-Barre syndrome in Northern Indian population. *Journal of Neuroimmunology*, 218:116–119.

Oelschlaeger, T. A., Guerryt, P., and Kopecko, D. J. (1993). Unusual microtubule-dependent endocytosis mechanisms triggered by *Campylobacter jejuni* and *Citrobacter freundii*. *Proc.Natl.Acad.Sci.U.S.A*, 90:6884–6888.

Oh, E., Andrews, K. J., and Jeon, B. (2018). Enhanced biofilm formation by ferrous and ferric iron through oxidative stress in *Campylobacter jejuni*. *Frontiers in Microbiology*, 9(Article 1204).

Oliver, S. (2000). Guilt-by-association goes global. *Nature*, 403:601–603.

O'Mahony, E., Buckley, J. F., Bolton, D., Whyte, P., and Fanning, S. (2011). Molecular epidemiology of *campylobacter* isolates from poultry production units in Southern Ireland. *PLoS ONE*, 6(12).

Pain, A., Ott, A., Amine, H., Rochat, T., Bouloc, P., and Gautheret, D. (2015). An assessment of bacterial small RNA target prediction programs. *RNA Biology*, 12(5):509–513.

Palombo, M., Scarlato, V., and Roncarati, D. (2020). Cooperative Regulation of *Campylobacter jejuni* Heat-Shock Genes by HspR and HrcA. *microorganisms*, 8(1161).

Palyada, K., Sun, Y.-q., Flint, A., Butcher, J., Naikare, H., and Stintzi, A. (2009). Characterization of the oxidative stress stimulon and PerR regulon of *Campylobacter jejuni*. *BMC Genomics*, 10(481).

Palyada, K., Threadgill, D., and Stintzi, A. (2004). Iron Acquisition and Regulation in *Campylobacter jejuni*. *Journal of Bacteriology*, 186(14):4714 – 4729.

Papenfort, K., Hoyos, M., Huber, M., Fo, K. U., and Schiller, F. (2020). Gene autoregulation by 3 ' UTR-derived bacterial small RNAs. *eLife*, 9(e58836).

Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2016). The impact of amplification on differential expression analyses by RNA-seq. *Scientific Reports*, 6(25533).

Park, M., Hwang, S., Ryu, S., and Jeon, B. (2021). CosR Regulation of *perR* Transcription for the Control of Oxidative Stress Defense in *Campylobacter jejuni*. *microorganisms*, 9(1281).

Parkhill, J., Wren, B. W., Mungall, K., Ketley, J. M., Churcher, C., Basham, D., Chillingworth, T., Davies, R. M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev,

A. V., Moule, S., Pallen, M. J., Penn, C. W., Quail, M. A., Rajandream, M. A., Rutherford, K. M., van Vliet, A. H., Whitehead, S., and Barrell, B. G. (2000). The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature*, 403(6770):665–668.

Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., Holloway, E., Kolesnykov, N., Lilja, P., Lukk, M., Mani, R., Rayner, T., Sharma, A., William, E., Sarkans, U., and Brazma, A. (2007). ArrayExpress - A public database of microarray experiments and gene expression profiles. *Nucleic Acids Research*, 35(SUPPL. 1):747–750.

Parmeciano, G., Noto, D., Molina, M. C., and Quiroga, C. (2019). Insights Into Non-coding RNAs as Novel Antimicrobial Drugs. *Frontiers in Genetics*, 10(57).

Parrish, J. R., Yu, J., Liu, G., Hines, J. A., Chan, J. E., Mangiola, B. A., Zhang, H., Pacifico, S., Fotouhi, F., DiRita, V. J., Ideker, T., Andrews, P., and Finley, R. L. (2007). A proteome-wide protein interaction map for *Campylobacter jejuni*. *Genome Biology*, 8(R130).

Patterson-Fortin, L. M., Vakulskas, C. A., Yakhnin, H., Babitzke, P., and Romeo, T. (2013). Dual posttranscriptional regulation via a cofactor-responsive mRNA leader. *Journal of Molecular Biology*, 425(19):3662–3677.

Pavlova, N., Kaloudas, D., and Penchovsky, R. (2019). Riboswitch distribution , structure , and function in bacteria. *Gene*, 708:38–48.

Pernitzsch, S. R., Alzheimer, M., Bremer, B. U., Robbe-saule, M., and Sharma, C. M. (2021). Small RNA mediated gradual control of lipopolysaccharide biosynthesis affects antibiotic resistance in *Helicobacter pylori*. *Nature Communications*, 12(4433).

Pernitzsch, S. R., Tirier, S. M., Beier, D., and Sharma, C. M. (2014). A variable homopolymeric G-repeat defines small RNA-mediated posttranscriptional regulation of a chemotaxis receptor in *Helicobacter pylori*. *PNAS*, 111(4):E501–E510.

Peschek, N., Herzog, R., Singh, P. K., Sprenger, M., Meyer, F., Fröhlich, K. S., Schröger, L., Bramkamp, M., Drescher, K., and Papenfort, K. (2020). RNA-mediated control of cell shape modulates antibiotic resistance in *Vibrio cholerae*. *Nature Communications*, 11(6067).

Petrova, O. E., Garcia-Alcalde, F., Zampaloni, C., and Sauer, K. (2017). Comparative evaluation of rRNA depletion procedures for the improved analysis of

bacterial biofilm and mixed pathogen culture transcriptomes. *Scientific Reports*, 7(41114).

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J. C., and Vogel, J. (2009). Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nature Structural and Molecular Biology*, 16(8):840–846.

Pittman, M. S., Elvers, K. T., Lee, L., Jones, M. A., Poole, R. K., Park, S. F., and Kelly, D. J. (2007). Growth of *Campylobacter jejuni* on nitrate and nitrite: Electron transport to NapA and NrfA via NrfH and distinct roles for NrfA and the globin Cgb in protection against nitrosative stress. *Molecular Microbiology*, 63(2):575–590.

Ponath, F., Tawk, C., Zhu, Y., Barquist, L., Faber, F., and Vogel, J. (2021). RNA landscape of the emerging cancer-associated microbe. *Nature Microbiology*.

Porcelli, I., Reuter, M., Pearson, B. M., Wilhelm, T., and van Vliet, A. H. (2013). Parallel evolution of genome structure and transcriptional landscape in the Epsilonproteobacteria. *BMC Genomics*, 14(1).

Prévost, K., Desnoyers, G., Jacques, J. F., Lavoie, F., and Massé, E. (2011). Small RNA-induced mRNA degradation achieved through both translation block and activated cleavage. *Genes and Development*, 25:385–396.

Prezza, G., Heckel, T., Dietrich, S., Homberger, C., Westermann, A. J., and Vogel, J. (2020). Improved bacterial RNA-seq by Cas9-based depletion of ribosomal RNA reads. *RNA*.

Quereda, J. J., Ortega, Á. D., Pucciarelli, M. G., and García-del Portillo, F. (2014). The *Listeria* Small RNA Rli27 Regulates a Cell Wall Protein inside Eukaryotic Cells by Targeting a Long 5-UTR Variant. *PLoS Genetics*, 10(10):1–11.

Raden, M., Müller, T., Mautner, S., Gelhausen, R., and Backofen, R. (2020). The impact of various seed, accessibility and interaction constraints on sRNA target prediction- a systematic assessment. *BMC Bioinformatics*, 21(15).

Radomska, K. A., Ordoñez, S. R., Wösten, M. M., Wagenaar, J. A., and van Putten, J. P. (2016). Feedback control of *Campylobacter jejuni* flagellin levels through reciprocal binding of FliW to flagellin and the global regulator CsrA. *Molecular Microbiology*, 102(2):207–220.

Rahman, H., King, R. M., Shewell, L. K., Semchenko, E. A., Hartley-Tassell, L. E., Wilson, J. C., Day, C. J., and Korolik, V. (2014). Characterisation of a Multi-ligand Binding Chemoreceptor CcmL (Tlp3) of *Campylobacter jejuni*. *PLoS Pathogens*, 10(1):e1003822.

Raines, D. J., Moroz, O. V., Blagova, E. V., Turkenburg, J. P., Wilson, K. S., and Duhme-Klair, A. K. (2016). Bacteria in an intense competition for iron: Key component of the *Campylobacter jejuni* iron uptake system scavenges enterobactin hydrolysis product. *Proceedings of the National Academy of Sciences of the United States of America*, 113(21):5850–5855.

Raphael, B. H., Pereira, S., Flom, G. A., Zhang, Q., Ketley, J. M., and Konkel, M. E. (2005). The *Campylobacter jejuni* response regulator, CbrR, modulates sodium deoxycholate resistance and chicken colonization. *Journal of Bacteriology*, 187(11):3662–3670.

Rasmussen, J. J., Vegge, C. S., Frøkiær, H., Howlett, R. M., Krogfelt, K. A., Kelly, D. J., and Ingmer, H. (2013). *Campylobacter jejuni* carbon starvation protein A (CstA) is involved in peptide utilization, motility and agglutination, and has a role in stimulation of dendritic cells. *Journal of Medical Microbiology*, 62:1135–1143.

Rath, E. C., Pitman, S., Cho, K. H., and Bai, Y. (2017). Identification of streptococcal small RNAs that are putative targets of RNase III through bioinformatics analysis of RNA sequencing data. *BMC Bioinformatics*, 18(Suppl 1(540).

Redko, Y., Galtier, E., Arnion, H., Darfeuille, F., Sismeiro, O., Coppée, J. Y., Médigue, C., Weiman, M., Cruveiller, S., and De Reuse, H. (2016). RNase J depletion leads to massive changes in mRNA abundance in *Helicobacter pylori*. *RNA Biology*, 13(2):243–253.

Rees, J. H., Gregson, N. A., and Hughes, R. A. C. (1995). Anti-ganglioside GM1antibodies in guillain-barré syndrome and their relationship to *Campylobacter jejuni* infection. *Annals of Neurology*, 38(5):809–816.

Reeser, R. J., Medler, R. T., Billington, S. J., Jost, B. H., and Joens, L. A. (2007). Characterization of *Campylobacter jejuni* biofilms under defined growth conditions. *Applied and Environmental Microbiology*, 73(6):1908–1913.

Regalia, M., Rosenblad, M. A., and Samuelsson, T. (2002). Prediction of signal recognition particle RNA genes. *Nucleic Acids Research*, 30(15):3368–3377.

Reuter, M. and van Vliet, A. H. (2013). Signal Balancing by the CetABC and CetZ Chemoreceptors Controls Energy Taxis in *Campylobacter jejuni*. *PLoS ONE*, 8(1):1–10.

Rhodes, K. M. and Tattershield, A. E. (1982). Guillain-Barre syndrome associated with Campylobacter infection Acquired inhibitor to human factor VIII associated with paraproteinaemia and subsequent development of chronic lymphatic leukaemia. *British Medical Journal*, 285(July):173–174.

Ribardo, D. A. and Hendrixson, D. R. (2011). Analysis of the LIV System of *Campylobacter jejuni* Reveals Alternative Roles for LivJ and LivK in Commensalism beyond Branched-Chain Amino Acid Transport. *Journal of Bacteriology*, 193(22):6233–6243.

Richards, J. and Belasco, J. G. (2019). Obstacles to Scanning by RNase E Govern Bacterial mRNA Lifetimes by Hindering Access to Distal Cleavage Sites. *Molecular Cell*, 74:1–12.

Rinke-Appel, J., Osswald, M., Von Knoblauch, K., Mueller, F., Brimacombe, R., Sergiev, P., Avdeeva, O., Bogdanov, A., and Dontsova, O. (2002). Crosslinking of 4.5S RNA to the *Escherichia coli* ribosome in the presence or absence of the protein Ffh. *RNA*, 8:612–625.

Rivas, E. and Eddy, S. R. (2001). Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*, 2(8).

Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2002). Comparative Genomics of Thiamin Biosynthesis in Procaryotes. *THE JOURNAL OF BIOLOGICAL CHEMISTRY*, 277(50):48949–48959.

Rollins, D. M. and Colwell, R. R. (1986). 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.

Rosenblad, M. A., Gorodkin, J., Knudsen, B., and Zwieb, C. (2003). SRPDB : Signal Recognition Particle Database. *Nucleic Acids Research*, 31(1):363–364.

Rosenblad, M. A., Larsen, N., Samuelsson, T., and Zwieb, C. (2009). Kinship in the SRP RNA family. *RNA biology*, 6(5):508–516.

Ryan, D., Jenniches, L., Reichardt, S., Barquist, L., and Westermann, A. J. (2020). A high-resolution transcriptome map identifies small RNA regulation

of metabolism in the gut microbe *Bacteroides thetaiotaomicron*. *Nature Communications*, 11(3557).

Saha, C., Horst-Kreft, D., Kross, I., van der Spek, P. J., Louwen, R., and van Baarlen, P. (2020a). *Campylobacter jejuni* Cas9 Modulates the Transcriptome in Caco-2 Intestinal Epithelial Cells. *Genes*, 11(1193).

Saha, C., Mohanraju, P., Stubbs, A., Dugar, G., Hoogstrate, Y., Kremers, G.-J., van Cappellen, W. A., Horst-Kreft, D., Laffeber, C., Lebbink, J. H., Bruens, S., Gaskin, D., Beerens, D., Klunder, M., Joosten, R., A Demmers, J. A., van Gent, D., Mouton, J. W., van der Spek, P. J., van der Oost, J., van Baarlen, P., and Louwen, R. (2020b). Guide-free Cas9 from pathogenic *Campylobacter jejuni* bacteria causes severe damage to DNA. *Science Advances*, 6(eaaz4849).

Sahin, O., Yaeger, M., Wu, Z., and Zhang, Q. (2017). *Campylobacter*-associated diseases in animals. *Annual Review of Animal Biosciences*, 5(1):21–42. PMID: 27860495.

Sahni, A., Hajjari, M., Raheb, J., Foroughmand, A. M., and Asgari, M. (2019). The non-coding RNA rprA can increase the resistance to ampicillin in *Escherichia coli*. *Microbial Pathogenesis*.

Sampson, T. R., Napier, B. A., Schroeder, M. R., Louwen, R., Zhao, J., Chin, C.-Y., Ratner, H. K., Llewellyn, A. C., Jones, C. L., Laroui, H., Merlin, D., Zhou, P., Endtz, H. P., and Weiss, D. S. (2014). A CRISPR-Cas system enhances envelope integrity mediating antibiotic resistance and inflammasome evasion. *Proceedings of the National Academy of Sciences*, 111(30):11163–11168.

Santaniello, A., Varriale, L., Dipineto, L., Borrelli, L., Pace, A., Fioretti, A., and Menna, L. F. (2021). Presence of *campylobacter jejuni* and *C. Coli* in dogs under training for animal-assisted therapies. *International Journal of Environmental Research and Public Health*, 18(7).

Sanyal, S. C., Islam, K. M. N., Neogy, P. K. B., Islam, M., Speelman, P., and Hui, M. I. (1984). *Campylobacter jejuni* diarrhea model in infant chickens. *Infection and Immunity*, 43(3):931–936.

Saoud, J., Carrier, M.-c. C., Massé, É., and Faucher, S. P. (2020). The small regulatory RNA Lpr10 regulates the expression of RpoS in *Legionella pneumophila*. *Molecular Microbiology*, 00:1–18.

Sayols, S., Scherzinger, D., and Klein, H. (2016). dupRadar : a Bioconductor package for the assessment of PCR artifacts in RNA-Seq data. *BMC Bioinformatics*, 17(428).

Sedlyarova, N., Shamovsky, I., Bharati, B. K., Gottesman, S., Schroeder, R., and Nudler, E. (2016). sRNA-Mediated Control of Transcription Termination in *E. coli*. *Cell*, 167:111–121.

Sellars, M. J., Hall, S. J., and Kelly, D. J. (2002). Growth of *Campylobacter jejuni* supported by respiration of fumarate, nitrate, nitrite, trimethylamine-N-oxide, or dimethyl sulfoxide requires oxygen. *Journal of Bacteriology*, 184(15):4187–4196.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape : A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Researchch*, 13:2498–2504.

Sharifi, S., Bakhshi, B., and Najar, S. (2021). Significant contribution of the CmeABC Efflux pump in high - level resistance to ciprofloxacin and tetracycline in *Campylobacter jejuni* and *Campylobacter coli* clinical isolates. *Annals of Clinical Microbiology and Antimicrobials*, 20(36).

Sharma, C. M., Darfeuille, F., Plantinga, T. H., and Vogel, J. (2007). A small RNA regulates multiple ABC transporter mRNAs by targeting C / A-rich elements inside and upstream of ribosome-binding sites. *Genes & Development*, 21:2804–2817.

Sharma, C. M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., Chabas, S., Reiche, K., Hackermüller, J., Reinhardt, R., Stadler, P. F., and Vogel, J. (2010). The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature*, 464(7286):250–255.

Sharma, C. M., Papenfort, K., Pernitzsch, S. R., Mollenkopf, H. J., Hinton, J. C., and Vogel, J. (2011). Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Molecular Microbiology*, 81(5):1144–1165.

Sharma, E., Sterne-Weiler, T., O'Hanlon, D., and Blencowe, B. J. (2016). Global Mapping of Human RNA-RNA Interactions. *Molecular Cell*, 62(4):618–626.

Sheng, H., Stauffer, W. T., Hussein, R., Lin, C., and Lim, H. N. (2017). Nucleoid and cytoplasmic localization of small RNAs in *Escherichia coli*. *Nucleic Acids Research*, 45(5):2919–2934.

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., Maiden, M. C. J., and Forbes, K. J. (2009). *Campylobacter* Genotyping to Determine the Source of Human Infection. *Clinical Infectious Diseases*, 48:1072–1078.

Shishkin, A. A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R. P., Rudy, R. F., Patel, M. M., Novod, N., Hung, D. T., Gnirke, A., Garber, M., Guttman, M., and Livny, J. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. *Nature Methods*, 12(4).

Sievers, S., Lund, A., Menendez-Gil, P., Nielsen, A., Mollerup, M. S., Nielsen, S. L., Larsson, P. B., Borch-Jensen, J., Johansson, J., and Kallipolitis, B. H. (2015). The multicopy sRNA LhrC controls expression of the oligopeptide-binding protein OppA in *Listeria monocytogenes*. *RNA Biology*, 12(9):985–997.

Silva, I. J., Barahona, S., Eyraud, A., Lalaouna, D., Figueroa-bossi, N., and Massé, E. (2019). SraL sRNA interaction regulates the terminator by preventing premature transcription termination of *rho* mRNA. *PNAS*.

Sittka, A., Lucchini, S., Papenfort, K., Sharma, C. M., Rolle, K., Binnewies, T. T., Hinton, J. C. D., and Vogel, J. (2008). Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genetics*, 4(8).

Sonnleitner, E. and Bla, U. (2014). Regulation of Hfq by the RNA CrcZ in *Pseudomonas aeruginosa* Carbon Catabolite Repression. *PLoS Genetics*, 10(6).

Sonnleitner, E., Gonzalez, N., Sorger-Domenigg, T., Heeb, S., Richter, A. S., Backofen, R., Williams, P., Hüttenhofer, A., Haas, D., and Bläsi, U. (2011). The small RNA PhrS stimulates synthesis of the *Pseudomonas aeruginosa* quinolone signal. *Molecular Microbiology*, 80(4):868–885.

Sonnleitner, E. and Haas, D. (2011). Small RNAs as regulators of primary and secondary metabolism in *Pseudomonas* species. *Applied Microbiology and Biotechnology*, 91(1):63–79.

Stahl, M., Friis, L. M., Nothaft, H., Liu, X., Li, J., Szymanski, C. M., and Stintzi, A. (2011). L-Fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *PNAS*, 108(17):7194–7199.

Stintzi, A. (2003). Gene expression profile of *Campylobacter jejuni* in response to growth temperature variation. *Journal of Bacteriology*, 185(6):2009–2016.

Stintzi, A., van Vliet, A. H. M., and Ketley, J. M. (2008). Iron Metabolism, Transport, and Regulation. In *Campylobacter , Third Edition*, pages 591–610. American Society of Microbiology.

Stoakes, E. (2017). *Campylobacter jejuni : Deciphering the role of FlhF in flagellar biogenesis*. PhD thesis.

Strakova, N., Shagieva, E., Ovesna, P., Korena, K., Michova, H., Demnerova, K., Kolackova, I., and Karpiskova, R. (2022). The effect of environmental conditions on the occurrence of *Campylobacter jejuni* and *Campylobacter coli* in wastewater and surface waters. *Journal of Applied Microbiology*, 132(1):725–735.

Strub, K., Fornallaz, M., and Bui, N. (1999). The Alu domain homolog of the yeast signal recognition particle consists of an Srp14p homodimer and a yeast-specific RNA structure. *RNA*, 5:1333–1347.

Subramanian, D., Bhasuran, B., and Natarajan, J. (2019). Genomic analysis of RNA-Seq and sRNA-Seq data identifies potential regulatory sRNAs and their functional roles in *Staphylococcus aureus*. *Genomics*, 111:1431–1446.

Sun, D., Chen, J., Wang, Y., Li, M., Rao, D., Guo, Y., Chen, N., Zheng, P., Sun, J., and Ma, Y. (2019). Metabolic engineering of *Corynebacterium glutamicum* by synthetic small regulatory RNAs. *Journal of Industrial Microbiology and Biotechnology*, 46:203–208.

Svensson, S. L., Davis, L. M., Mackichan, J. K., Allan, B. J., Pajaniappan, M., Thompson, S. A., and Gaynor, E. C. (2009). The CprS sensor kinase of the zoonotic pathogen *Campylobacter jejuni* influences biofilm formation and is required for optimal chick colonization. *Mol Microbiol*, 71(1):253–272.

Svensson, S. L., Hyunh, S., Parker, C. T., and Gaynor, E. C. (2015). The *Campylobacter jejuni* CprRS two-component regulatory system regulates aspects of the cell envelope. *Molecular Microbiology*, 96(1):189–209.

Svensson, S. L. and Sharma, C. M. (2021). An RNase III processed , antisense RNA pair regulates a *Campylobacter jejuni* colonization factor. *bioRxiv*.

Sweeney, E. G., Nishida, A., Weston, A., Bañuelos, M. S., Potter, K., Conery, J., and Guillemin, K. (2019). Agent-Based Modeling Demonstrates How Local Chemotactic Behavior Can Shape Biofilm Architecture. *mSphere*, 4(e00285-19).

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., and Mering, C. V. (2019). STRING v11 : protein – protein association networks with increased coverage , supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(Database issue):D607–D613.

Taheri, N., Fällman, M., Wai, S. N., and Fahlgren, A. (2019). Accumulation of virulence-associated proteins in *Campylobacter jejuni* Outer Membrane Vesicles at human body temperature. *Journal of Proteomics*, 195:33–40.

Taheri, N., Mahmud, A. K., Sandblad, L., Fällman, M., Wai, S. N., and Fahlgren, A. (2018). *Campylobacter jejuni* bile exposure influences outer membrane vesicles protein content and bacterial interaction with epithelial cells. *Scientific Reports*, 8(16996).

Taveirne, M. E., Theriot, C. M., Livny, J., and DiRita, V. J. (2013). The Complete Campylobacter jejuni Transcriptome during Colonization of a Natural Host Determined by RNAseq. *PLoS ONE*, 8(8):1–18.

Tello, M., Avalos, F., and Orellana, O. (2018). Codon usage and modular interactions between messenger RNA coding regions and small RNAs in *Escherichia coli*. *BMC Genomics*, 19(657).

Thépault, A., Rose, V., Queguiner, M., and Chemaly, M. (2020). Dogs and Cats : Reservoirs for Highly Diverse *Campylobacter jejuni* and a Potential Source of Human Exposure. *animals*, 10(838).

Thomas, M. T., Shepherd, M., Poole, R. K., Van Vliet, A. H., Kelly, D. J., and Pearson, B. M. (2011). Two respiratory enzyme systems in *Campylobacter jejuni* NCTC 11168 contribute to growth on l-lactate. *Environmental Microbiology*, 13(1):48–61.

Thomason, M. K., Voichek, M., Dar, D., Addis, V., Fitzgerald, D., Gottesman, S., Sorek, R., and Greenberg, E. P. (2019). A rhlI 5' UTR-Derived sRNA Regulates RhlR-Dependent Quorum Sensing in *Pseudomonas aeruginosa*. *mBio*, 10(e02253-19).

Thorvaldsdottir, H., Robinson, J. T., and Mesirov, J. P. (2012). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *BRIEFINGS IN BIOINFORMATICS*, 14(2):178–192.

Tram, G., Klare, W. P., Cain, J. A., Mourad, B., Cordwell, S. J., Day, C. J., and Korolik, V. (2020). Assigning a role for chemosensory signal transduction in Campylobacter jejuni biofilms using a combined omics approach. *Scientific Reports*, 10(6829).

Tree, J. J., Granneman, S., McAteer, S. P., Tollervey, D., and Gally, D. L. (2014). Identification of Bacteriophage-Encoded Anti-sRNAs in Pathogenic *Escherichia coli*. *Molecular Cell*, 55:199–213.

Tribelli, P. M. and López, N. I. (2018). Reporting key features in cold-adapted bacteria. *Life*, 8(8).

Trigui, H., Lee, K., Thibodeau, A., Lévesque, S., Mendis, N., Fravalo, P., Letellier, A., and Faucher, S. P. (2017). Phenotypic and transcriptomic responses of *campylobacter jejuni* suspended in an artificial freshwater medium. *Frontiers in Microbiology*, 8(Aricle 1781).

Tsai, C.-h., Liao, R., Chou, B., Palumbo, M., and Contreras, M. (2015). Genome-Wide Analyses in Bacteria Show Small-RNA Enrichment for Long and Conserved Intergenic Regions. *J Bacteriol*, 197:40–50.

Turonova, H., Haddad, N., Hernould, M., Chevret, D., and Pazlarova, J. (2017). Profiling of *Campylobacter jejuni* Proteome in Exponential and Stationary Phase of Growth. *Frontiers in Microbiology*, 8:913.

Ueda, M., Kinoshita, H., Meada, S. I., Zou, W., and Tanaka, A. (2003). Structure-function study of the amino-terminal stretch of the catalase subunit molecule in oligomerization, heme binding , and activity expression. *Appl Microbiol Biotechnol*, 61:488–494.

Umu, S. U. and Gardner, P. P. (2017). A comprehensive benchmark of RNA-RNA interaction prediction tools for all domains of life. *Bioinformatics*, 33(7):988–996.

Urdaneta, V. and Casadesús, J. (2017). Interactions between bacteria and bile salts in the gastrointestinal and hepatobiliary tracts. *Frontiers in Medicine*, 4(Article 163).

van der Hooft, J. J., Alghefari, W., Watson, E., Everest, P., Morton, F. R., Burgess, K. E., and Smith, D. G. (2018). Unexpected differential metabolic responses of *Campylobacter jejuni* to the abundant presence of glutamate and fucose. *Metabolomics*, 14(144).

van der Stel, A.-X., Boogerd, F. C., Huynh, S., Parker, C. T., van Dijk, L., van Putten, J. P., and Wösten, M. M. (2017). Generation of the membrane potential and its impact on the motility, ATP production and growth in *Campylobacter jejuni*. *Molecular Microbiology*, 105(4):637–651.

Varsaki, A., Murphy, C., Barczynska, A., Jordan, K., and Carroll, C. (2015). The acid adaptive tolerance response in *Campylobacter jejuni* induces a global response, as suggested by proteomics and microarrays. *Microbial Biotechnology*, 8:974–988.

Vegge, C. S., Brøndsted, L., Li, Y. P., Bang, D. D., and Ingmer, H. (2009). Energy taxis drives *Campylobacter jejuni* toward the most favorable conditions for growth. *Applied and Environmental Microbiology*, 75(16):5308–5314.

Vegge, C. S., Jansen van Rensburg, M. J., Rasmussen, J. J., Maiden, M. C., Johnsen, L. G., Danielsen, M., MacIntyre, S., Ingmer, H., and Kelly, D. J. (2016). Glucose metabolism via the entner-doudoroff pathway in *campylobacter*: A rare trait that enhances survival and promotes biofilm formation in some isolates. *Frontiers in Microbiology*, 7(Article 1877).

Velayudhan, J. and Kelly, D. J. (2002). Analysis of gluconeogenic and anaplerotic enzymes in *Campylobacter jejuni*: An essential role for phosphoenolpyruvate carboxykinase. *Microbiology*, 148:685–694.

Vercruysse, M., Köhrer, C., Davies, B. W., Arnold, M. F., Mekalanos, J. J., RajBhandary, U. L., and Walker, G. C. (2014). The Highly Conserved Bacterial RNase YbeY Is Essential in *Vibrio cholerae*, Playing a Critical Role in Virulence, Stress Regulation, and RNA Processing. *PLoS Pathogens*, 10(6).

Versace, G., Palombo, M., Menon, A., Scarlato, V., and Roncarati, D. (2021). Feeling the Heat : The *Campylobacter jejuni* HrcA Transcriptional Repressor Is an Intrinsic Protein Thermosensor. *Biomolecules*, 11(1413).

Vogel, È., Bartels, V., Tang, T. H., Churakov, G., Hu, A., and Slagter-ja, J. G. (2003). RNomics in *Escherichia coli* detects new sRNA species and indicates parallel transcriptional output in bacteria. *Nucleic Acids Research*, 31(22):6435–6443.

Vogel, J. (2020). An RNA biology perspective on species-specific programmable RNA antibiotics. *Molecular Microbiology*, 113:550–559.

Vogel, J. and Luisi, B. F. (2015). Hfq and its constellation of RNA. *nat rev microbiol*, 9(8):578–589.

Voigts-Hoffmann, F., Schmitz, N., Shen, K., Shan, S.-o., Ataide, S. F., and Ban, N. (2013). The structural basis of FtsY recruitment and GTPase activation by SRP RNA Felix. *Mol Cell*, 12(52):5.

Vorwerk, H., Mohr, J., Huber, C., Wensel, O., Schmidt-Hohagen, K., Gripp, E., Josenhans, C., Schomburg, D., Eisenreich, W., and Hofreuter, D. (2014). Utilization of host-derived cysteine-containing peptides overcomes the restricted sulphur metabolism of *Campylobacter jejuni*. *Molecular Microbiology*, 93(6):1224–1245.

Wade, J. T. and Grainger, D. C. (2017). Spurious transcription and its impact on cell function. *Transcription*.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics*, 28(16):2184–2185.

Washietl, S., Hofacker, I. L., and Stadler, P. F. (2005). From The Cover: Fast and reliable prediction of noncoding RNAs. *Proceedings of the National Academy of Sciences*, 102(7):2454–2459.

Waters, S. A., McAteer, S. P., Kudla, G., Pang, I., Deshpande, N. P., Amos, T. G., Leong, K. W., Wilkins, M. R., Strugnell, R., Gally, D. L., Tollervey, D., and Tree, J. J. (2017). Small RNA interactome of pathogenic *E. coli* revealed through crosslinking of RNase E. *The EMBO Journal*, 36:374–387.

Watson, R. O. and Galan, J. E. (2005). Signal transduction in *Campylobacter jejuni*-induced cytokine production Robert. *Cellular Microbiology*, 7(5):655–665.

Weingarten, R. A., Grimes, J. L., and Olson, J. W. (2008). Role of Campylobacter jejuni respiratory oxidases and reductases in host colonization. *Applied and Environmental Microbiology*, 74(5):1367–1375.

Weis, F., Bron, P., Rolland, J.-p., Thomas, D., and Felden, B. (2010). Accommodation of tmRNA – SmpB into stalled ribosomes : A cryo-EM study. *RNA*, 16:299–306.

Westermann, A. J., Förstner, K. U., Amman, F., Barquist, L., Chao, Y., Schulte, L. N., Müller, L., Reinhardt, R., Stadler, P. F., and Vogel, J. (2016). Dual

RNA-seq unveils noncoding RNA functions in host–pathogen interactions. *Nature*, 529(7587):496–501.

WHO (2017). WHO publishes list of bacteria for which new antibiotics are urgently needed, howpublished = https://www.who.int/news-room/detail/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed, note = Accessed: 2010-09-30.

Wilms, C., Noah, J. W., Zhong, D., and Wollenzien, P. (1997). Exact determination of UV-induced crosslinks in 16S ribosomal RNA in 30S ribosomal subunits.

Wilson, T., Mouriño, S., and Wilks, A. (2021). The heme binding protein PhuS transcriptionally regulates the *Pseudomonas aeruginosa* tandem sRNA *prr*F1,F2 locus. *Journal of Biological Chemistry*.

Wösten, M. M., Boeve, M., Koot, M. G., Van Nuenen, A. C., and Van Der Zeijst, B. A. (1998). Identification of *Campylobacter jejuni* promoter sequences. *Journal of Bacteriology*, 180(3):594–599.

Wösten, M. M. S. M., Van Dijk, L., Parker, C. T., Guilhabert, M. R., Van Der Meer-Janssen, Y. P. M., Wagenaar, J. A., and Van Putten, J. P. M. (2010a). Growth phase-dependent activation of the DccRS regulon of *Campylobacter jejuni*. *Journal of Bacteriology*, 192(11):2729–2736.

Wösten, M. M. S. M., Van Dijk, L., Parker, C. T., Guilhabert, M. R., Van Der Meer-Janssen, Y. P. M., Wagenaar, J. A., and Van Putten, J. P. M. (2010b). Growth phase-dependent activation of the DccRS regulon of *Campylobacter jejuni*. *Journal of Bacteriology*, 192(11):2729–2736.

Wösten, M. M. S. M., Wagenaar, J. A., and Van Putten, J. P. M. (2004). The FlgS/FlgR Two-component Signal Transduction System Regulates the *fla* Regulon in *Campylobacter jejuni*. *Journal of Biological Chemistry*, 279(16):16214–16222.

Wright, J. A., Grant, A. J., Hurd, D., Harrison, M., Guccione, E. J., Kelly, D. J., and Maskell, D. J. (2009). Metabolite and transcriptome analysis of *Campylobacter jejuni in vitro* growth reveals a stationary-phase physiological switch. *Microbiology*, 155(1):80–94.

Xia, Y., Lei, C., Yang, D., and Luo, H. (2020a). Identification of key modules and hub genes associated with lung function in idiopathic pulmonary fibrosis. *PeerJ*, 8:e9848.

Xia, Y., Weng, Y., Xu, C., Wang, D., Pan, X., Tian, Z., Xia, B., Li, H., Chen, R., Liu, C., Jin, Y., Bai, F., Cheng, Z., Kuipers, O. P., and Wu, W. (2020b). Endoribonuclease *ybey* is essential for rna processing and virulence in *pseudomonas aeruginosa*. *mBio*, 11:e00659–20.

Xie, Z., Ulrich, L. E., Zhulin, I. B., and Alexandre, G. (2010). PAS domain containing chemoreceptor couples dynamic changes in metabolism with chemotaxis. *Proceedings of the National Academy of Sciences*, 107(5):2235–2240.

Xu, F., Wu, C., Guo, F., Cui, G., Zeng, X., and Yang, B. (2015). Transcriptomic analysis of *Campylobacter jejuni* NCTC 11168 in response to epinephrine and norepinephrine. *Frontiers in Microbiology*, 6(Article 452).

Yap, K. L., Li, S., Muñoz-cabello, A. M., Raguz, S., Zeng, L., Gil, J., Walsh, M. J., and Zhou, M.-m. (2010). Molecular Interplay of the Non-coding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Mol Cell*, 38(5):662–674.

Yoo, S. M. and Na, D.and Lee, S. Y. (2013). Design and use of synthetic regulatory small RNAs to control gene expression in *Escherichia coli*. *Nature Protocols*, 8(9):1694–1707.

Young, K. T., Davis, L. M., and DiRita, V. J. (2007). *Campylobacter jejuni*: Molecular biology and pathogenesis. *Nature Reviews Microbiology*, 5(9):665–679.

Yu, S., Vogel, J., and Förstner, K. U. (2018). ANNOgesic : a Swiss army knife for the RNA-seq based annotation of bacterial / archaeal genomes. *GigaScience*, 7:1–11.

Zhang, A., Wassarman, K. M., Rosenow, C., Tjaden, B. C., Storz, G., and Gottesman, S. (2003). Global analysis of small RNA and mRNA targets of Hfq. *Molecular Microbiology*, 50(4):1111–1124.

Zhang, H., Zhang, Y., Song, Z., Li, R., Ruan, H., and Liu, Q. (2020). sncRNAs packaged by *Helicobacter pylori* outer membrane vesicles attenuate IL-8 secretion in human cells. *International Journal of Medical Microbiology*, 310:151356.

Zhang, J., Zhou, W., Liu, Y., Liu, T., Li, C., and Wang, L. (2018). Oncogenic role of microRNA-532-5p in human colorectal cancer via targeting of the 5'UTR of RUNX3. *Oncology letters*, 15(5):7215–7220.

Zhang, J. W., Butland, G., Greenblatt, J. F., Emili, A., and Zamble, D. B. (2005). A Role for SlyD in the *Escherichia coli* Hydrogenase Biosynthetic Pathway. *Journal of Biological Chemistry*, 280(6):4360–4366.

Zhang, M. G. and Liu, J. M. (2019). Transcription of cis-antisense small RNA MtlS in *Vibrio cholerae* is regulated by transcription of its target gene *mtlA*. *Journal of Bacteriology*.

Zhang, X., Chen, Z., Zang, J., Yao, C., Shi, J., Nie, R., and Wu, G. (2021). LncRNA-mRNA co-expression analysis discovered the diagnostic and prognostic biomarkers and potential therapeutic agents for myocardial infarction. *AGING*, 13.

Zheng, M., Åslund, F., and Storz, G. (1998). Activation of the OxyR Transcription Factor by Reversible Disulfide Bond Formation. *Science*, 279(5357):1718–1722.

Zuker, M. and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(1):133–148.