Graph clustering as a method to investigate riboswitch variation

Matthew Crum

A dissertation submitted to the Faculty of the department of Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School
07/2021

Abstract

Graph clustering as a method to investigate riboswitch variation

Matthew Crum

Advisor: Dr. Michelle M. Meyer, PhD.

Non-coding RNA (ncRNA) perform vital functions in cells, but the impact of diversity across structure and function of homologous motifs has yet to be fully investigated. One reason for this is that the standard phylogenetic analysis used to address these questions in proteins cannot easily be applied to ncRNA due to their inherent characteristics. Compared to proteins, ncRNA have shorter sequence lengths, lower sequence conservation, and secondary structures that need to be incorporated into the analysis. This has necessitated an effort to develop methodology for investigating the evolutionary and functional relationship between sets of ncRNA. In this pursuit, I studied closely related riboswitches. Riboswitches are structured ncRNA found in bacterial mRNA that regulate gene expressions using their two major components: the aptamer and the expression platform. The aptamer of a riboswitch is able to bind a specific small molecule (ligand), and the bound/unbound state of the aptamer influences conformational changes in the expressions platform that can lead to increased or decreased downstream gene expression. Utilizing sequence and structural similarity metrics combined with graph clustering and *de novo* community detection algorithms I have determined a methodology for investigating the functional and evolutionary relationship between closely related riboswitches, and other ncRNA by extension, that are found across a range of diverse phyla.

Acknowledgements

First, I would like to thank my thesis guidance committee and reader: Dr. Michelle M. Meyer, Dr. Babak Momeni, Dr. Peter Clote, Dr. Jeff DaCosta, and Dr. Elena Rivas. I am extremely grateful for my committee's assistance and understanding while I pursued my PhD. The stated purpose of a guidance committee was described to me as a group of experts who simultaneously act to guide and asses a graduate students progress. I am thankful that I had a committee that took the guiding part of this purpose as seriously as the assessment part. I know many graduate students that were not lucky enough to have a committee as supportive and constructive as this committee was. Thank you.

In particular I would like to thank my PI, Dr. Michelle Meyer. Throughout my graduate career, you have been a positive and supportive mentor. While she likely knows the professional and scientific impact she has had on me, I am not sure she is aware of the personal impact she had. Throughout my tenure as a graduate student, I was repeatedly disappointed by the treatment of many graduate students, particularly by their own mentors. I can genuinely say that I chose the best possible boss and mentor, as you repeatedly demonstrated the ability to balance an intense work ethic with treating all members of our lab with integrity and as people. I do not believe I would have had the perseverance to complete my PhD without the support and mentorship you gave me. When so many others let me down, thank you for embodying a researcher I could look up to, professionally and personally.

I would also like to thank the members of my lab, particularly Nikhil Ram Mohan, Daniel Beringer, Elise Gray, Indu Warrier, and Arianne Babina. Each of you had a strong impact on me during my PhD. Danny and Elise, thank you for taking me out and knowing how to get me out of my own head when life became overwhelming. Indu, thank you for always being such a bright spot of sun when the dark bioinformatics room felt oppressive. Arianne, thank you for your guidance regarding wet-lab techniques and our moments bonding over music taste. Nikhil, thank you for being such an incredible mentor and friend; if you choose to be a PI, you will be a fantastic PI and I know you will support and guide the next generation of researchers as you helped guide me.

To my friends and family who were there for me throughout these times, thank you for your support. To my parents, thank you for providing emotional support and always being proud of me, regardless of how small the milestone was. To Defne Surujon, thank you for simultaneously being the graduate student I could most look up to as a researcher and as a human being. To Sergio and TJ, thank you for our weekly game nights, they helped me get through a lot of difficult weeks. To Kevin and Brittany, thank you for your understanding and advice throughout this experience. Kevin, you have always been a role model and I cannot express enough how important to me it was to have someone that supported my efforts in graduate school while simultaneously telling me it is okay to leave. So many people viewed supporting me as telling me that I can persevere and telling I am good enough to get the degree, thank you for telling me I am good enough even if I don't get the degree, Kev.

Finally, I want to thank Samantha Dyckman. These have been the hardest years in either of our lives. I am so thankful that throughout it all we both recognized what is most

important to us and never lost sight of it. Thank you for always making sure we look to the future that we want instead of getting overwhelmed by the present. We celebrated ten years together and got a dog while both of us were struggling professionally. I know that was possible because we both knew that having a future together was more important than any career path either of us would take. Thank you for always reminding me there are more important things in my life and for being the most important part of it.

Table of Contents

List of Figures

List of Tables

Abbreviations

AdoCbl	Cobalamin riboswitch
ADP-CDP	Adenosine/cytidine diphosphate
CM	Covariance model
DNA	Deoxyribonucleic acid
GCV	Glycine cleavage operon
glnA	Glutamine riboswitch
HMM	Hidden Markov model
MFE	Minimum free energy
mRNA	messenger RNA
ncRNA	Non-coding RNA
Nt	Nucleotide
ppGpp	Guanosine tetraphosphate
PRPP	phosphoribosyl pyrophosphate
RBS	Ribosomal binding site
RefSeq	Reference sequence database
RNA	Ribonucleic acid
RFAM	RNA family database
SAH	S-adenosyl homocysteine
SAM	S-adenosyl methionine
SCFG	Stochastic context free grammar
TP	Transport proteins
tRNA	Transfer RNA
UTR	Untranslated region

Introduction

It has long been known that RNA can act as more than a messenger for translating DNA into proteins. However, in recent decades, the true depth and breadth of functions that RNA can perform has become clearer [1–3]. RNAs that play a role in key cellular functions, but do not encode a protein, are called non-coding RNA (ncRNA). ncRNA are vital for the survival of all domains of life, participating in processes from gene regulation to catalysis to protein translation [4–8]. One class of essential ncRNAs is riboswitches, which are found in the 5' untranslated region (UTR) or bacterial mRNA and regulate gene expression at a transcriptional or translational level. A major functional component of riboswitches is the three-dimensional shape they are able to fold into, which allows them to bind a specific ligand and regulate gene expression based on the presence or absence of that ligand within the cell [5,9]. It is this structure that allows riboswitches to perform such vital function. Both secondary and tertiary interactions play a role in ncRNA activity, and structural conservation among functional motifs is often seen across a wide taxonomic range [9,10]. For this reason, studying functional RNA motifs requires a careful consideration of both sequence and structure.

The use of computational tools has greatly expanded the potential of RNA research. As far back as the 1960's, it was demonstrated that secondary structure of an RNA plays a determining role in its function, and altering the secondary structure conformation of an RNA can change its functionality [7]. Soon after, algorithms were developed for identifying the RNA secondary structure which minimize free energy and identifies alternative, sub-optimal folding conformations [11–15]. As RNA secondary structure prediction became more reliable and efficient, implementation of secondary structure within other aspects of the field became easier. This paved the way for the use of probabilistic models for the identification of homologous motifs that can flexibly describe primary sequence and secondary structure consensus of an RNA family [16–19]. As the tools available to researchers expanded, so did our ability to quickly and accurately evaluate RNA based on structure and function.

1.1 Riboswitches

1.1.1 Riboswitch structure

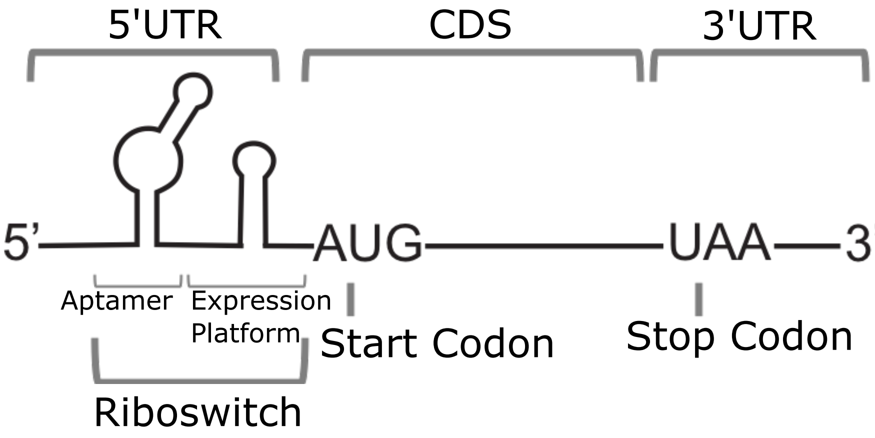

Figure 1.1 ***Anatomy of a riboswitch.*** Representation of a riboswitch, containing the aptamer and expression platform within the 5’ UTR of an mRNA. **Adapted from figure created by Dr. Arianne Babina.*

Riboswitches are regulatory RNA motifs found within the 5’ UTR of bacterial mRNA which regulate downstream gene expression in response to the presence of ligand. They are made up of two major components, the aptamer and the expression platform (Figure 1.1). The aptamer binds a specific ligand, which induces a conformational change in the expression platform immediately downstream. This conformational change is what leads to regulation of the downstream gene [10,20,21]. Bacterial riboswitches facilitate increases or decreases in gene expression at the transcriptional or translational level [10,22,23], while in the few instances of riboswitches in eukaryotes tend to be found in introns of pre-mRNA and are involved in splicing [24].

The overall secondary structure of the riboswitch is vital when studying and characterizing them. This is because the secondary and tertiary structure, and not primary sequence, is what primarily drives riboswitch function. The conformational changes that occur upon ligand binding are what allow riboswitches to regulate gene expression in response to environmental factors. Terminator and anti-terminator formation in the expression platform influences whether there will be complete readthrough of a transcript or early transcriptional attenuation, while stable stem-loop structures that sequester the ribosomal binding site (RBS) function to allow or preclude translation. Within families of riboswitches, the primary sequence can vary, but the structural characteristics of the riboswitch must be maintained in order to retain functionality.

1.1.2 Biological function of riboswitches

Bacteria utilize riboswitches to regulate metabolic pathways and maintain the homeostasis of the cell [25–28]. These regulatory switches have been conserved throughout evolution and can be sensitive to nanomolar concentrations of their ligand [29]. This provides bacteria with fast-acting and energy-inexpensive regulatory mechanisms for maintaining proper metabolic functions. Currently, there are over 40 classes of riboswitches which have been discovered [9]. Included in these are riboswitches which sense amino acids (lysine, glutamine, and glycine)[30–32], nucleotide derivatives (purines, prequeuosine, phosphoribosyl pyrophosphate, and adenosine/cytidine-5'-di-phosphate) [5,28,33–37], tRNAs [38,39], and signaling molecules (cyclic-di-GMP, cyclic-di-AMP, cyclic-GMP-AMP, 5-aminoimidazole-4-carboxamide ribonucleoside-5'-triphosphate, and guanosine tetraphosphate) [40–44]. These, combined with riboswitches which sense ions, enzyme cofactors, and other small molecules, function to monitor essential components of the cellular environment and modulate changes in gene expression.

In response to these ligands, riboswitches regulate pathways involved in synthesis, degradation, and transport of vital molecules within the cell. Some riboswitches are found regulating only a narrow group of genes involved in a specific biochemical pathway, but others can be found regulating a broader range of genes related to a biological function[5,9]. Moreover, there are instances of multiple riboswitches being found in tandem and regulating the same gene [45,46]. This was first characterized in *Desulfovibrio vulgaris* and *Desulfovibrio desulfuricans*, in which a pair of complete TPP riboswitches regulate the thiamine biosynthesis operon (*thiSGHFE*) [47]. Other examples of different riboswitches being found in tandem have been identified since this first example, such as the paired SAM-II and SAM-V riboswitches which transcriptional and translational regulation in response to ligand binding [48]. In the case of the SAM-I and AdoCbl tandem riboswitches and the guanine and PRPP tandem riboswitches, the paired conformation allows for regulation of the downstream gene in response to binding of the ligand for either or both riboswitches [49,50]. A relatively unique iteration of the tandem conformation is pairs of homologous aptamers that regulate a single expression platform. There are only two known riboswitches which fall into this category, the glutamine (glnA) riboswitch and the glycine riboswitch. The glnA riboswitch has been identified with up to three tandem aptamers, but is narrowly distributed to cyanobacteria and the tandem conformations have not been extensively characterized biochemically [32,51,52]. Conversely, the glycine riboswitch is widely distributed and well characterized in both its singleton and paired-aptamer conformations [5,31,53–59]. In the case of the glycine riboswitch, this tandem conformation allows for one dominant ligand-binding aptamer and one primarily providing structural stability, as opposed to a system of paired, cooperative binding of the ligand [57,59,60].

1.1.3 Riboswitch distribution and variation

Riboswitches are found throughout all phyla of bacteria, with riboswitch classes displaying different distribution patterns. Some riboswitches are distributed across a small number of bacterial groups with only a couple hundred (or fewer) representative instances, such as the PreQ1-II/III, Adenine, and SAM-SAH riboswitches [9]. Others are found almost ubiquitously throughout bacteria with over ten thousand representative examples, such as TPP, AdoCbl and SAM-I/IV[9]. The TPP riboswitch also has some members present within eukaryotes, primarily fungi, where it functions in alternative splicing instead of direct regulation of expression levels [24].

On top of the varying levels of prevalence seen across riboswitch classes, the interplay of structural homology and ligand binding also introduces complexities to studying riboswitches. While classifying of *de novo* riboswitch classes relies on the identification of structural homology, there are multiple instances of riboswitches with homologous structures binding distinct ligands [37,50,61,62] and, inversely, of structurally distinct riboswitches binding the same ligand [63,64]. This is highlighted by the ykkC and SAM riboswitches, respectively.

SAM riboswitches are a group of structurally distinct riboswitches which are each able to bind S-Adenosyl Methionine (SAM). There are currently six known families of the SAM riboswitch, and a family of riboswitches called the SAM-SAH riboswitch able to bind both SAM and S-Adenosyl Homocysteine (SAH) [63–66]. Some of these classes have more structural similarity than others, such as the SAM-III and SAM-VI riboswitches[63], but each is different from the other SAM riboswitch families. The SAM-SAH riboswitch is also structurally distinct from the SAM riboswitch families, but is still able to bind SAM. It has been proposed that the SAM-SAH riboswitch is a minimalistic variant of a SAMfamily riboswitch that evolved in an environment lacking SAH, and so lost the ability to distinguish between the two molecules [66]. This group of riboswitches demonstrates that, while structural homology is a major component of riboswitch functionality and classification, that not all riboswitches which bind a ligand are necessitated to appear similar.

The ykkC riboswitches demonstrate a similarly interesting phenomenon, but inverted. While the SAM riboswitches have different structures and bind the same ligands, the ykkC riboswitches show high structural similarity and yet bind distinct ligands. This group of riboswitches was initially categorized as one riboswitch class based on their structural and sequence homology [5]. It was later identified that a large subset of these riboswitches, which regulated guanidine carboxylases and guanidine transporters, bound guanidine and was classified as the guanidine-I riboswitch [62,67]. To clarify this finding, this guanidine-I riboswitch was referred to as the ykkC subtype-1 riboswitch with respect to the original set of ykkC riboswitches, while the remaining ykkC riboswitches were evaluated and grouped into four classes termed ykkC subtype 2A-D. Of the four ykkC subtype 2 riboswitches, three have had their ligand identified. ykkC subtype-2A was shown to bind guanosine tetraphosphate (ppGpp), an alarmone that responds to amino acid starvation conditions, and regulate genes involved in branches chain amino acid synthesis [61]. ykkC subtype-2B was shown to bind phosphoribosyl pyrophosphate (PRPP), a molecule required for *de novo* purine synthesis, and regulate genes involved in purine biosynthesis[50]. ykkC subtype-2C was shown to bind adenosine/cytidine diphosphate (ADP/CDP) and regulate genes that hydrolyze the phosphoester bond of nucleoside diphosphate [37]. Each of these riboswitches show high levels of structural homology and bind distinct ligands, with functional diversity coming down to a few nucleotides' variation involved in ligand identification. Interestingly, as there are multiple guanidine binding riboswitches, the guanidine-I riboswitch can be grouped into a set of riboswitches which are structurally distinct and bind the same ligand *and* a set of riboswitches which are structurally homologous and bind distinct ligands.

In addition to variation across classes, there is also variation within classes[68]. One major example of this is the glycine riboswitch. The glycine riboswitch is unique in that it is often found with tandem, homologous aptamers regulating a single expression platform [31]. However, there are also examples of singleton glycine riboswitches, which contain one aptamer and one expression platform. Singleton riboswitches also contain a ghost aptamer, which is a small stem-loop with no ligand-binding capacity, immediately adjacent to the aptamer [56]. Studies have found that the location of the ghost aptamer, either upstream or downstream of the aptamer, is dependent on what gene is being regulated by the riboswitch [58]. These findings supported studies which identified that the tandem glycine riboswitches tend to have a dominant ligand-binding aptamer, which also coincided with what gene is being regulated [29,58,59]. This culminates in a riboswitch class which has a distinct singleton and tandem variant that regulates glycine cleavage and another singleton and tandem variant which regulates glycine transport.

Despite the relatively straight-forward mechanism of gene regulation that riboswitches represent, there is an impressive amount of variation within and between riboswitch classes. This makes investigating riboswitches more nuanced than it initially appears in many cases. While classifying riboswitches by their paired ligand or structural homology can provide broadly informative groupings, these categorizations reduce the complexity that can often be found upon closer inspection.

1.2 Computation approaches to ncRNA investigation

The study of homologous structured RNA, like riboswitches, relies heavily on the secondary and tertiary shape which the RNA adopts in order to perform its function [5,9,10]. Unlike DNA and protein, in which motifs can be readily identified by investigation of primary sequence alone, structured RNA must have the secondary structure considered to predict and classify motifs. This makes identification and study of classification of structured RNA more complicated than with other biological molecules. However, researchers have generated powerful computational methods for both detecting and studying these motifs.

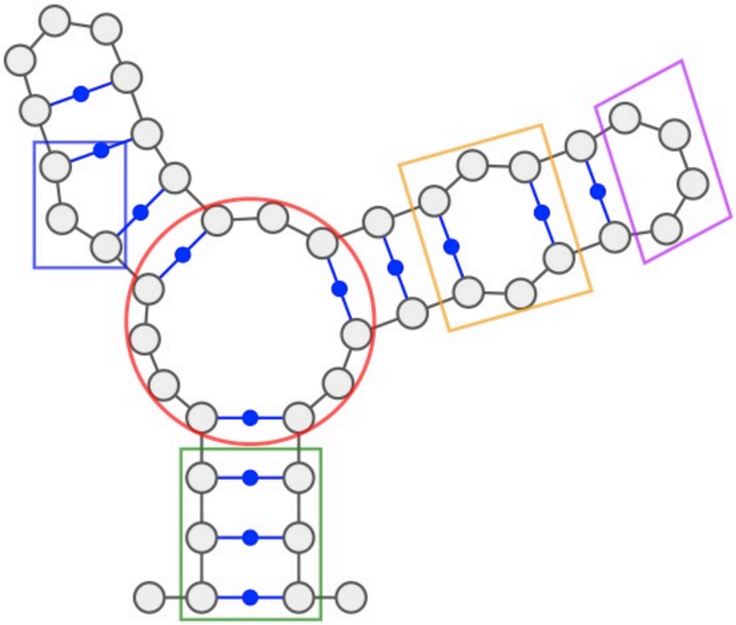

Figure 1.2 ***RNA structure representation.*** Representation of RNA secondary structure showing a stem (green), bulge (blue), loop (pink), internal loop (gold), and multi-stem junction (red).

1.2.1 RNA structure

The function of RNA is highly depended on the three-dimensional shape it takes, which is dictated by energetically favorable pairings of compatible nucleotides within the RNA. Studies of tRNA in the 1960's first made it clear that secondary structure of an RNA plays a determining role in its function, and that alterations to the structural conformation impacts RNA functionality [7]. For this reason, scientists have developed multiple methods and software for predicting secondary structure. The ability to predict secondary structures that optimize these pairings within a given RNA creates the basis for the majority of techniques for computationally investigating structured RNA. The secondary structure of an RNA can be broken into a few components: 1) stems, comprised of paired nucleotides stacked atop each other, 2) loops, the terminal ends of stems which loop back upon themselves, and 3) bulges, which are unpaired gaps within stems (Figure 1.2). There are also pseudoknots that can form in RNA secondary structure, which occur when loops or bulges within a stem pair with distal portions of the RNA. These components can be represented using notations that distinguishes paired and unpaired nucleotides within a sequence, allowing researchers to record RNA secondary structure mathematically or in text.

To determine the most likely secondary structure for an RNA to be found in, it is common to predict the most energetically favorable secondary structure it can take. This is referred to as the minimum free energy (MFE) structure. The algorithms which predict this structure rely on the assumption that the structure that will be most stable is the one that maximizes the number of nucleotide pairings. The Nussinov-Jacobson algorithm, one of the earliest structural prediction algorithms, uses this assumption and recursively iterates through an RNA sequence to identify valid base pairings to identify the structure with the maximize pairings present [15,69]. The Nussinov-Jacobson algorithm lays the foundation for many modern algorithms because its main assumption and process remains valid, making impromments upon the algorithm a matter of increasing computational efficiency and incorporating new insights into the biochemical dynamics [70–72]. One major improvement in identification of the MFE secondary structure is the incorporation of nearest-neighbor thermodynamic parameters for determination of more accurate folding energy [11–14].

These improvements in efficiency and accuracy for identifying MFE structures have proven useful for RNA researchers, but do not fully represent the complexity of *in vivo* RNA folding. RNA can fold into many structures, and while the MFE structure may be the most likely for a given RNA to adopt, it is not the only one it will adopt over time. Secondary structure of an RNA shifts between different conformations depending on the environment and the how energetically favorable different structures are. Particularly for riboswitches, which adopt different conformation in their “On” and “Off” states, it is important to take the range of structural possibilities into account when evaluating their structure. This can be done using a Boltzmann weighted ensemble of the possible secondary structures based on the energetic favorability of each potential secondary structure an RNA can adopt [73]. Using this, researchers are able to better approximate dynamic RNA structure based on the probability that the RNA will adopt different structures with different nucleotide pairings.

1.2.2 Sequence and structure models

The ability to build a model that represents a structured RNA is vital to researching these motifs. Models allow researchers to identify homologous sequences, predict structure, and discover novel classes of structured RNA. These models are primarily built around multiple sequence alignments (MSA) and subsequent (or simultaneous) folding of the sequences. Creating accurate MSAs, however, can be a complex and computationally intensive process, driving the necessity to balance efficiency and accuracy [74]. The two main approaches used for generating MSAs are progressive alignments and stochastic alignments. Progressive alignments focus on aligning the two most similar sequences first and then iteratively adding and aligning other sequences within the set. Stochastic alignments rely on aligning random sequences within the set and determining an optimal

MSA based on the best generated after iterative re-alignments. Two of the most popular tools, Clustal and Muscle, use progressive alignment and a hybrid that uses progressive alignments as a seed for a stochastic method, respectively [75,76]. While Clustal and Muscle align purely on primary sequence, multiple tools have been developed which incorporate secondary structure into the MSA. These tools can be broken into three general categories: align *then* fold, fold *then* align, and fold *and* align simultaneously [77,78]. Tools such as RNAalifold require an MSA and identify a consensus structure within the aligned sequences [79]. Inversely, RNAforester takes secondary structures as input and generates an alignment to identify a consensus structure [80,81]. Software that fold and align simultaneously, such as FoldAlign, Dynalign, RNAmountAlign, and LocARNA, are able to take in unaligned sequences and generate a consensus structure and MSA guided by structural similarity [82–85]. While FoldAlign, Dynalign, RNAmountAlign, and LocARNA perform with comparable accuracy in benchmarking tests, RNAmountAlign has been optimized for speed and performs significantly faster [85].

Using MSA alignments and a consensus structure, a model can be generated to represent a given structured RNA motif. The model can then be used to align new sequences to existing members of the motif based on sequence and structure homology and predict whether a given sequence is a member of the class used to generate the model. Two common methods for generating models for structured RNA are Profile Hidden Markov Models (Profile-HMM) and Covariance Models (CM). Profile-HMMs use a probabilistic model to determine the most likely changes to the aligned sequences and thus determine whether a given sequence is a viable member of the model class [86,87]. CMs function similarly, but use a stochastic context free grammar (SCFG), which are a set of probability-weighted rules able to represent possible RNA structures, to describe the alignment sequences and consensus structure of the input MSA [88,89]. In RNA research, Infernal is the standard for RNA homology searches, allowing investigators to use a CM trained on a MSA to search databases for sequences likely to fit the model based on reported likelihood sores [18,90].

When evaluating structured RNA families, one major obstacle is determining whether the secondary structure is conserved beyond what would be expected if functional structure were not constraining evolution. All RNA sequences will have some secondary structure, and homologous RNA sequences are expected to have some primary sequence conservation. Therefore, to predict whether an RNA family has a functional secondary structure, it becomes necessary to validate that a consensus secondary structure is conserved beyond what would be observed in a non-structured RNA. To address this, software such as R-scape evaluates if there is observed sequence covariation within RNA families that conserves the secondary structure beyond phylogenetic expectation [91,92]. Covariation and conservation within the primary and secondary structure of these families can be visualized using the software R2R [93].

1.2.3 Graph clustering

When studying riboswitches, the use of traditional phylogenetics is enticing, but often insufficient. Compared to protein coding and gene sequences, there are many challenges associated with assessment of riboswitches that are not well addressed by conventional phylogenetic approaches, including: short sequence length, lack of primary sequence conservation, and greater focus on secondary structure for biological function. These factors contribute to the inability of phylogenetic approaches to fully evaluate structured RNA sequences [94]. Short sequence lengths combined with a lack of primary sequence conservation means riboswitches lack sufficient character information to gather evolutionary relationship estimations. The necessity of secondary structure is also an obstacle, as the incorporation of secondary structure into phylogenetic analysis is underdeveloped in the field. Some software, such as RAxML, allow for static structure models to be implemented into a phylogenetic analysis, and modifications to existing algorithms that do not allow for this could be designed to incorporate structure based on sequence alignments. However, these approaches do not account for the dynamic structure and conformational changes riboswitches undergo, nor overcome the lack of information contained within short, highly variable primary sequences. While phylogenetics have been applied to studying the genes regulated by riboswitches in order to gain better evolutionary understanding [95,96], the direct study of riboswitch sequence and structure remains difficult using these methods. To address this, I believe that that use of graph clustering could be used to supplement the pairwise evaluation of riboswitches to gain better insight into variation and evolution of these structured RNA.

Graph clustering is a technique that can be applied to find groupings of highly related objects based on pairwise distance metrics [97–99]. Researchers can investigate how closely related objects are by creating a network of nodes, representing individual objects within a set, and connecting each pair of nodes with a weighted edge corresponding to the pairwise similarity of the two nodes it connects. Each node acts as a charged particle, repelling others, and each edge acts as a spring, pulling connected nodes together. Through a process called thresholding, edges connecting nodes with worse pairwise similarity can be pruned out of the network. This results in an overall topological shifting of the network structure. Highly similar nodes retain the edges connecting them, forming tightly packed clusters, while dissimilar nodes separate. Throughout the thresholding process, clustering of groups of nodes can be observed as they remain close to more highly similar nodes and repel nodes from other, unconnected clusters. We can then evaluate the clustering patterns at thresholds which display high separation and modularity within the network.

Graph clustering can thus be applied to any dataset which can be evaluated based on pairwise similarity. Some of the earliest examples investigating biological questions using pairwise similarity for dataset clustering is the evaluation of gene expression data. In these cases, the vector distance between normalized expression profiles representing different conditions and timepoints was used to perform the clustering [100–102]. The viability of graph clustering has been demonstrated more recently within the field of biology by applying it to various investigations, such metabolic pathways, animal social interactions, and neural systems networks [103]. This laid the groundwork for my research and utilization of this methodology for the investigation of closely related ncRNA motifs, such as riboswitches.

A major strength of graph clustering is the ability to evaluate sets of riboswitches based on a range of different distance metrics. By combining different metrics, riboswitch classes can be evaluated from multiple different angles. This can be done using any metric that calculates pairwise similarity between the riboswitch instances, allowing for flexibility of the analysis based on the needs of a given project. Thus, measures that incorporate purely secondary structure information [104,105], purely sequence information [75,106], or a combination of both [82,83,107–109], may be utilized so that all available information can be captured. Using these metrics allow a researcher to investigate the putative phylogenetic closeness of a given riboswitch or the functional relatedness by comparing the sequence and structural similarity. While these metrics are ideally suited for the study of structured RNA, any number of other distance metrics could similarly be applied to a graph clustering analysis to study riboswitches or other biological data.

Graph clustering provides a methodology for fine-scale comparison of similarity within and between groups of homologous RNAs [99]. My adaptation of graph clustering to ncRNA allows for comparison of conservation between clusters of related ncRNAs. By measuring the density of edges within and between sub-clusters across a range of thresholding cutoff values, researchers can determine relative conservation of the sub-clusters[98]. These clusters can then be validated using *de novo* community detection methods to corroborate the observed clustering based on a set of algorithms [97,110].

While this approach does not yield the same kind of inferences concerning the line of descent as traditional phylogenetic approaches, it does allow researcehrs to assess ncRNA similarity based on any metric that can be represented by a pairwise distance.

Chapter 2: Regulatory context drives conservation of glycine riboswitch aptamers[1]

2.1 Summary

The glycine riboswitch is a ncRNA responsible for the regulation of several distinct gene sets in bacteria. It is found with either one (singleton) or two (tandem) aptamers, each of which directly senses glycine. Which aptamer is more important for gene-regulation, and the functional difference between tandem and singleton aptamers, are long-standing questions in the riboswitch field. Like many biologically functional RNAs, glycine aptamers require a specific 3D folded conformation. Thus, they have low primary sequence similarity across distantly related homologs, and large changes in sequence length that make creation and analysis of accurate multiple sequence alignments challenging. To better understand the relationship between tandem and singleton aptamers, I used a graph clustering approach that allows us to compare the similarity of aptamers using metrics that measure both sequence and structure similarity. My investigation reveals that in tandem glycine riboswitches, one aptamer is more highly conserved than the other, and which aptamer is conserved depends on what gene(s) are regulated. Moreover, I find that many singleton glycine riboswitches likely originate from tandem riboswitches in which the ligand-binding site of the non-conserved aptamer has degraded over time.

---

[1] Adapted from: Crum M, Ram-Mohan N, Meyer MM. Regulatory context drives conservation of glycine riboswitch aptamers. PLoS Comput Biol. 2019;15(12):1–24.

2.2 Abstract

In comparison to protein coding sequences, the impact of mutation and natural selection on the sequence and function of non-coding (ncRNA) genes is not well understood. Many ncRNA genes are narrowly distributed to only a few organisms, and appear to be rapidly evolving. Compared to protein coding sequences, there are many challenges associated with assessment of ncRNAs that are not well addressed by conventional phylogenetic approaches, including: short sequence length, lack of primary sequence conservation, and the importance of secondary structure for biological function. Riboswitches are structured ncRNAs that directly interact with small molecules to regulate gene expression in bacteria. They typically consist of a ligand-binding domain (aptamer) whose folding changes drive changes in gene expression. The glycine riboswitch is among the most well-studied due to the widespread occurrence of a tandem aptamer arrangement (tandem), wherein two homologous aptamers interact with glycine and each other to regulate gene expression. However, a significant proportion of glycine riboswitches are comprised of single aptamers (singleton). Here I use graph clustering to circumvent the limitations of traditional phylogenetic analysis when studying the relationship between the tandem and singleton glycine aptamers. Graph clustering enables a broader range of pairwise comparison measures to be used to assess aptamer similarity. Using this approach, I show that one aptamer of the tandem glycine riboswitch pair is typically much more highly conserved, and that which aptamer is conserved depends on the regulated gene. Furthermore, my analysis also reveals that singleton aptamers are more similar to either the first or second tandem aptamer, again based on the regulated gene. Taken together, my findings suggest

that tandem glycine riboswitches degrade into functional singletons, with the regulated gene(s) dictating which glycine-binding aptamer is conserved.

2.3 Background

The glycine riboswitch provides a unique opportunity to directly assess how riboswitch architecture may change over time or be influenced by which genes are regulated. The glycine riboswitch is commonly found in a tandem conformation where two homologous aptamers interact through tertiary contacts to regulate a single expression platform (tandem) [31]. The tandem glycine riboswitch conformation is well-studied biophysically [29,53,55,59,115–117]. However, there is lack of consensus regarding the mechanism of ligand-binding or which of the tandem aptamers is more essential for ligand-binding to induce gene regulation. Extensive *in vitro* investigation of a tandem glycine riboswitch originating from *Vibrio cholerae* found that ligand binding of the second aptamer (aptamer-2) controlled the expression platform and gene expression, while the first aptamer (aptamer-1) primarily played a role in structural stabilization and aptamer dimerization [29]**.** However, *in vivo* investigation of a tandem glycine riboswitch within *Bacillus subtilis* found that disruption of aptamer-1's binding pocket impeded riboswitch regulation more strongly than disrupting aptamer-2's [59].

A more conventional single-aptamer conformation also appears in nature (singleton), but singleton glycine riboswitches require a "ghost-aptamer" that functions as a scaffold for tertiary interactions similar to those observed in tandem glycine riboswitches

[56]. Glycine riboswitch singletons are divided into two types distinguished by the location of the ghost-aptamer with respect to the ligand-binding aptamer. Type-1 singletons have a ghost-aptamer 3’ of the aptamer, while the ghost aptamer is 5’ of the glycine aptamer for type-2 singletons [56]. The relationship between singleton and tandem glycine riboswitches is not well characterized, and the how and why of tandem vs singleton riboswitch emergence and conservation is a subject of debate. Glycine riboswitches have been identified regulating several different sets of genes (genomic context) [29,31,54,59], and may function as either expressional activators (On-switch) [31] or repressors (Off-switch) [54].

In order to assess the relationship between singleton and tandem glycine riboswitches I used both traditional phylogenetics and graph clustering approaches to examine glycine aptamer sequences across a range of diverse bacterial species. My investigation reveals that genomic context effects which tandem glycine-binding aptamer is more highly conserved. It also demonstrates that singleton riboswitches are more similar to the first or second tandem aptamer based on genomic context. Taken together, my findings suggest strongly that many singleton glycine riboswitches result from degradation of tandems, with the genomic context dictating which glycine-binding aptamer retains ligand responsiveness.

2.4 Methods

2.4.1 Riboswitch Identification

Infernal's cmsearch function was used to query all RefSeq77 bacterial genomes using the RFAM glycine riboswitch covariance model to identify all individual putative glycine riboswitch aptamers (RF00504) [18,118,119]. All hits were filtered based on e-value, with a threshold of e^{-5}. Putative riboswitches were sorted into components of a singleton or tandem glycine riboswitch based on their proximity to any other putative aptamer. Two hits within 100nts of each other were considered to be the two aptamers of a tandem riboswitch; the largest distance between tandem aptamers observed was 32 nts. I then used a set of 30 tandem riboswitches to generate a covariance model which identified both tandem riboswitch aptamers together. The generated model is able to identify tandem glycine riboswitch aptamers, but does not explicitly include the expression platform due to the diversity of mechanisms of action for the glycine riboswitch. This tandem covariance model was used to query the RefSeq77 bacterial database and supplement the current set of putative glycine riboswitches with any tandems that may have been missed by the RFAM covariance model. 2,998 individual riboswitches were identified, 2,216 tandems riboswitches and 782 singleton riboswitches. Singletons were then classified as type-1 or type-2 based on the location of the ghost aptamer, an adjacent stem structure that functions as a scaffold for tertiary interaction with the ligand-binding aptamer. Ghost aptamer location was determined based on conformation to covariance models generated from singleton type-1 and type-2 riboswitches reported in [58]. Of the 782 singleton riboswitches, 342 were characterized as singlet type-1, 125 were characterized as singleton type-2, and 305 were unable to be characterized as one or the other (called singleton type-0). Bedtools was utilized to determine the nearest downstream gene within 500 nucleotides on the same strand, providing a gene that is putatively regulated by each given riboswitch [120]. Genes were binned based on function for determination of genomic context of the glycine riboswitch.

2.4.2 Riboswitch Phylogenetic Analysis

Tandem riboswitches were grouped based on taxonomic origin and genomic context. In order to incorporate secondary structure information, groups were aligned using LocARNA's mlocarna function for *de novo* alignment and folding [84,121,122] and Infernal's cmalign function to align to the tandem covariance model [18,90]. Maximum likelihood phylogenetic trees were generated from these aligned groups using RAxML [123]. Trees from alignments generated by cmalign were run with an accompanying secondary structure file to guide phylogenetic maximum likelihood analysis based on aptamer sequence and structure. In each case, 100 bootstrap replicates were performed and maximum likelihood bootstrap confidence values >=70 are reported.

2.4.3 Graph Clustering and Network Generation

Graphs of aptamer sequences were generated, with vertices representing individual aptamers and edges representing a pairwise similarity metric relating aptamer pairs. Edge weights were then thresholded, resulting in trimmed networks of clustered aptamers containing only edges connecting pairs with higher similarity than the threshold value.

Aptamer networks were generated to determine clustering based on a number of different pairwise metrics. These included partition function (RNApdist) [104,105], sequence and structural similarity (FoldAlign and Dynalign) [82,83,107,108], ensemble expected mountain height (RNAmountAlign) [109], and sequence similarity (Clustal Omega) [75]. Pairwise values were utilized to generate and visualize aptamer networks using the igraph and qgraph R-libraries [110,124]. Optimal visualization thresholds vary between sets relative to the taxonomic diversity represented within them. Clustering of riboswitch groups based on genomic context and aptamer type were compared by network density across a range of thresholds for each distance metric.

Modular clusters were identified using igraph's community identification functions cluster_fast_greedy, cluster_walktrap, cluster_edge_betweenness, and cluster_leading_eigen. These algorhithms differ in that cluster_fast_greedy is a more traditional hierarchical clustering that generates a dendrogram based on pairing nodes connected by increasing strength of edges within the network[111], while cluster_edge_betweenness works using an inverted method that begins with the removal of weak edges from the network[112]. Cluster_walktrap uses a random-walk methodology to identify communities based on the assumption that short random-walks tend to remain in communities due to higher edge density[113]. The cluster_leading_eigen algorithm uses a spectral clustering methodology that first calculates the modularity matrix by subtracting the probability matrix from the adjacency matrix, then identifies the largest positive eigenvalue. The corresponding eigenvector can then be used to separate the nodes based on the sign (positive or negative) of their corresponding element within the eigenvector[114].

Following cluster identification, I performed 100 replicates of a parametric bootstrapping analysis which perturbs 5% of the network and then re-clusters. This analysis perturbs the network by adding/removing edges at random in a 1:1 ratio, resulting in a network that contains the same nodes and an equivalent number of edges, but 5% of the edges connect different nodes. For each iteration, I determined the new clustering for the group of interest using the igraph community detection methods. These were then compared to the original group that had no perturbation. This comparison was done across all 100 iterations and uses Jaccard Similarity Index to calculate similarity of each post-perturbation cluster to the original cluster. Average Jaccard Similarity Index across the 100 iterations was used to determine robustness for the clusters. Cluster composition was also validated by MCL [125] and DBSCAN [126] clustering.

Network density was calculated by determining the percent of total possible pairwise edges remaining for a given set of vertices after edge-trimming based on a distance metric threshold. Inter- and intra-group density calculations represent edge-density within a group and between groups, respectively. The network density measured across a range of thresholds correlates to aptamer similarity with respect to a given distance metric used to weight edges. These generated graphs equate to a flipped cumulative distribution of possible edges and actual edges for a cluster as I threshold the network based on edge weights.

2.4.4 Consensus Structure Generation

I generated Stockholm files for sets of riboswitches using a combination of LocARNA's mlocarna function [121] and alignment to the covariance models using Infernal's cmsearch function [18]. Ralee was then used to perform minor curation of alignments and VARNA was implemented for secondary structure visualization throughout the process [127,128]. R2R was then used to generate consensus structures based on these Stockholm files [93]. The "#=GF R2R SetDrawingParam autoBreakPairs true" flag was used to allow for breaking of base pairs in instances where aptamer stems were not highly conserved.

2.5 Results

2.5.1 Glycine riboswitches within the Bacillaceae and Vibronaceae families cluster based on genomic context.

My first goal was to computationally investigate the differences observed between the *V. cholerae* and B. *subtilis* tandem riboswitches by conducting a comprehensive sequence analysis of glycine aptamers. To identify glycine riboswitch aptamers, I used the RFAM covariance model RF00504 to search RefSeq77 [118,119,129]. Identified aptamers within 100 nucleotides (nts) of each other were considered to be part of a tandem riboswitch. A tandem aptamer covariance model was created using infernal and trained from this identified set and used to search RefSeq77 to supplement the dataset [18,90]. In total, 2,998 individual riboswitches were identified, 2,216 tandem riboswitches and 782 singleton riboswitches. Each was classified by genomic context based on the RefSeq annotated function of the putatively regulated gene. This dataset does not include the variant glycine riboswitches identified in a previous study [68], as the vast majority of these examples are present in metagenomic data, and therefore are not within RefSeq77.

To determine whether the functional differences observed between the glycine riboswitches from *B. subtilis* and *V. cholerae* are reflective of detectable sequence variation/divergence across their respective families, I first performed a phylogenetic analysis on examples from the dataset found within the *Bacillaceae* and *Vibrionaceae* bacterial families. I gathered sequences spanning both aptamers from 48 *Bacillaceae* and 37 *Vibrionaceae* tandem riboswitches. Within this set, all 37 *Vibrionaceae* riboswitches regulate transport proteins (TP), while 41 *Bacillaceae* riboswitches regulate glycine cleavage system (GCV) and the remaining 7 regulate TP (**Table 2.1**). I then utilized RAxML to generate a phylogenetic tree (**Figure 2.1A**) from these sequences and the consensus secondary structure from alignment to the tandem covariance model using Infernal. The tree shows clustering within the group of *Bacillaceae* riboswitches regulating GCV, as well within *Bacillaceae* riboswitches regulating TPs. However, there is a clear phylogenetic separation of the two groups of *Bacillaceae* riboswitches, splitting them into distinct clades. Furthermore, the clade representing *Bacillaceae* riboswitches regulating TP more closely clustered with the clade of riboswitches from *Vibrionaceae* regulating TP, although each set forms a distinct group. This finding suggests that genomic context may play a prominent role in the evolution of tandem glycine riboswitches.

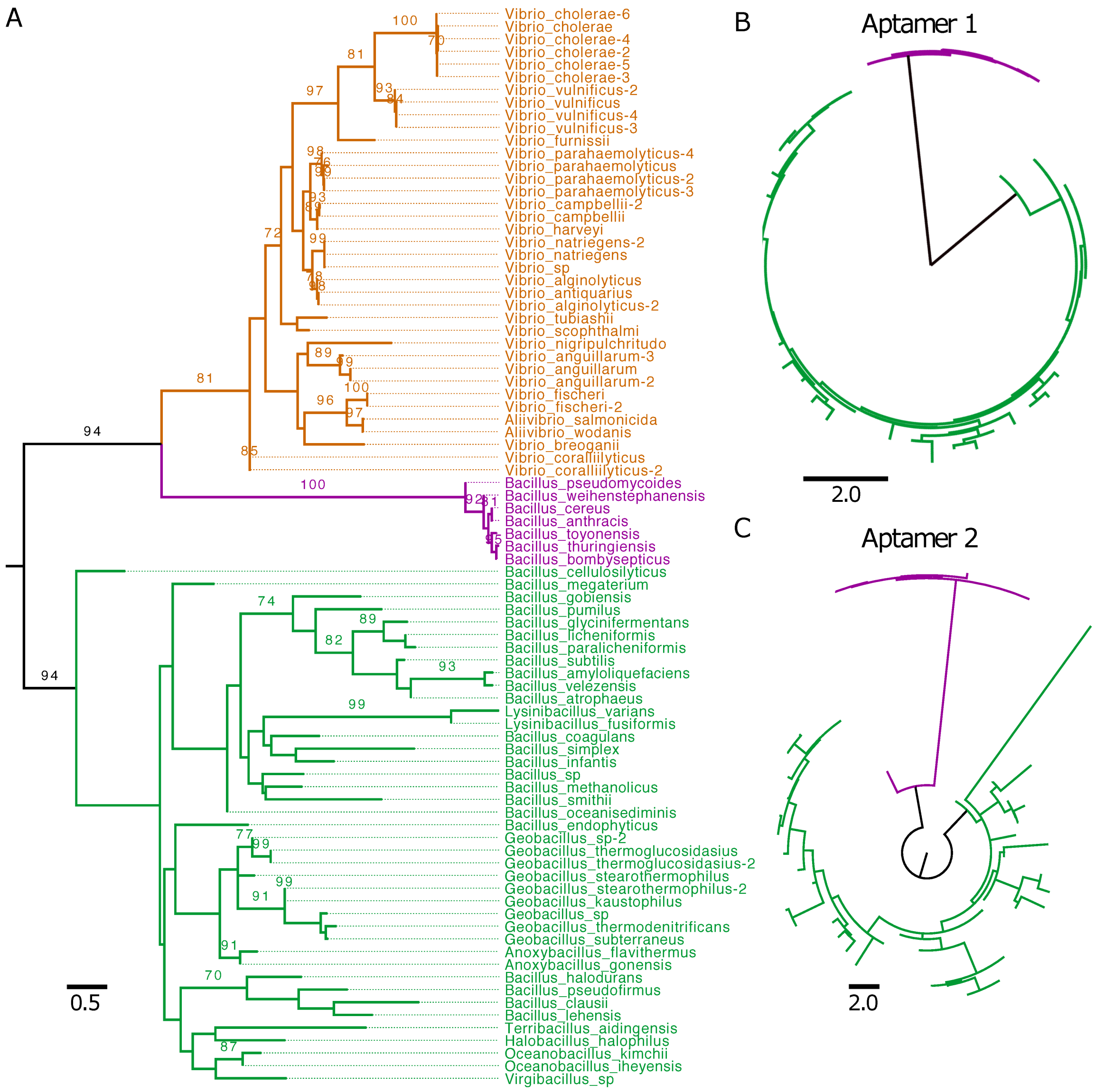

Figure 2.1: ***Phylogenetic comparison of tandem riboswitches across Bacillaceae and Vibrionaceae.***

A) 48 *Bacillaceae* and 37 *Vibrionaceae* tandem riboswitches were clustered based on aptamer sequence and structure across both aptamers of the riboswitch. After phylogenetic clustering, individual aptamers were colored based on the class of gene being regulated and the bacterial family of origin (*Vibrionaceae* TP are orange, *Bacillaceae* TP are purple, *Bacillaceae* GCV are green). Clusters have been labeled with the bacterial family and gene class being regulated. Bootstrap support values are displayed for 100 replicates when >=70. B,C) Phylogenetic clustering of 48 tandem riboswitches, separated into aptamer-1 (B) and aptamer-2 (C), taken from the *Bacillaceae* family and colored according to class of gene being regulated (GCV are green, TP are purple). All trees are midpoint rooted.

2.5.2 Bacillaceae tandem riboswitches display different patterns of aptamer conservation based on genomic context.

To investigate whether the evolutionary pressure driving divergence of tandem glycine riboswitches regulating GCV and TP occurs evenly across both aptamers, or is specific to a single aptamer, I split tandem riboswitches within the *Bacillaceae* family into individual aptamers (**Table 2.2, 2.3**). This provided us with two groups: one containing all aptamer-1's (first aptamer) and one containing all aptamer-2's (second aptamer). I then generated a phylogenetic tree for each set to determine whether the phylogenetic divergence seen within the riboswitch set is explained by variances within one specific aptamer or is present in both (**Figure 2.1B, C)**. Both aptamer-1 and aptamer-2 sets display clear clustering based on genomic context. This indicates that divergence of tandem glycine riboswitches in differing genomic contexts cannot be fully explained by variation within the first or second aptamer alone. Moreover, within the GCV context, it appears that aptamer-1 is more highly conserved than aptamer-2, as indicated by the shorter branch lengths across the clade (**Figure 2.1B, C)**.

2.5.3 Genomic context dictates aptamer clustering in Bacillaceae and Vibrionaceae.

To better understand how the homologous aptamers of the tandem glycine riboswitch have diverged, I broadened the taxonomic scope and focused my investigation onto the individual aptamer domains of the glycine riboswitch. However, the shorter sequence length of the individual glycine aptamers confounded the analysis. Thus, relative conservation of aptamer-1 to aptamer-2 in different genomic contexts was investigated using graph clustering of all *Bacillaceae* and *Vibrionaceae* aptamers within a given genomic context, excluding identical aptamer pairs coming from different strains of the same species. This set comprised of 84 pairs of aptamer-1 and aptamer-2 from *Bacillaceae* regulating GCV and 36 pairs from *Vibrionaceae* regulating TP (**Table 2.4, 2.5**). The number of TP riboswitches was reduced by one in this analysis compared to the previous, as one of the riboswitches was no longer unique within the set when evaluating only the individual aptamer sequences.

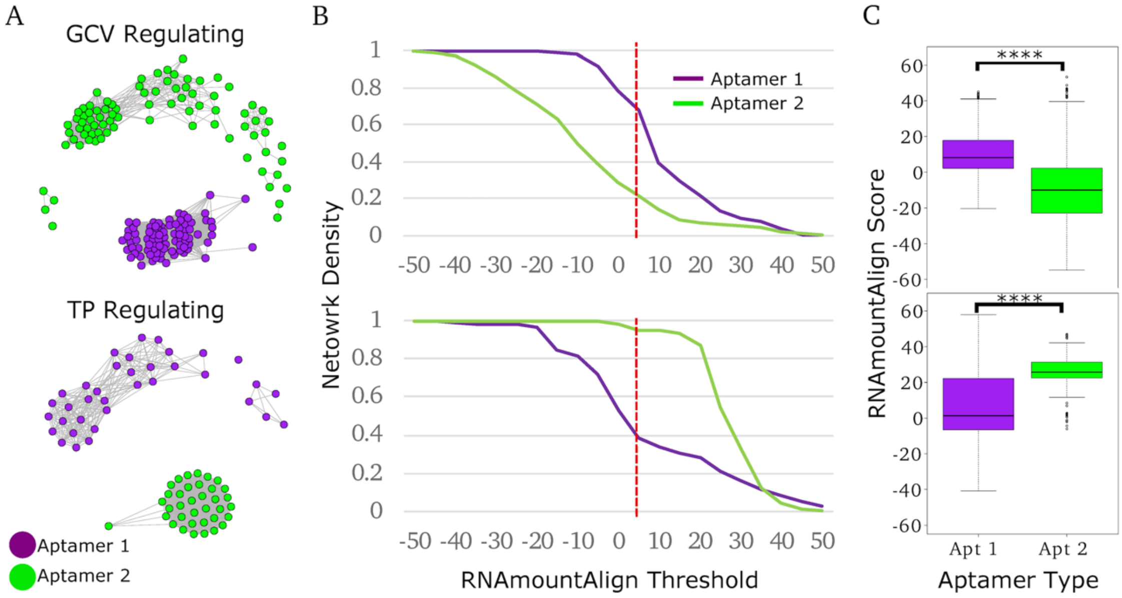

Figure 2.2: ***Clustering of tandem riboswitch aptamers across Bacillaceae and Vibrionaceae.***
A) Bacillaceae and Vibrionaceae tandem riboswitch aptamers were clustered using RNAmountAlign as a distance metric (threshold of 5). All represented Bacillaceae riboswitches regulate GCV (top), while Vibrionaceae riboswitches regulate TP (bottom). Aptamers are colored based on aptamer type, purple for aptamer-1 and green for aptamer-2. B) Network density was calculated for each aptamer in both networks across a range of RNAmountAlign thresholds. Dotted red line indicates the RNAmountAlign threshold (5) at which the networks in A were visualized. C) Box blots represent all pairwise edge-weights within each aptamer type. **** p-value < $2x10^{-16}$.

I generated networks comprised of vertices corresponding to individual glycine riboswitch aptamers with edges weighted based on the pairwise RNAmountAlign distance score [109]. RNAmountAlign was chosen as the primary metric for edge-weighting in this work due to its implemented use of primary sequence information and ensemble mountain distance of secondary structure to generate a pairwise score more quickly and efficiently than other software. After weighting with RNAmountAlign, edges were trimmed if they were below a selected RNAmountAlign threshold, thus altering the topology of the network from completely pairwise to containing clusters of aptamers whose similarity is greater than the threshold. Thresholding was done across a range of RNAmountAlign scores to identify conserved aptamer groups which retained tight clustering (**Figure 2.2A**). Each network corresponds to a specific genomic context, TP or GCV. I find that within these contexts, aptamers group based on their position within the tandem arrangement (aptamer-1 vs. aptamer-2). Network density, defined as the fraction of edges present within a group compared to the total number of edges in the non-thresholded network, was calculated across a range of RNAmountAlign score thresholds and used to gauge relative conservation of each aptamer type for each genomic context (**Figure 2.2B**). Differing cluster densities between aptamer types revealed that genomic context effects which aptamer is more highly conserved: aptamer-1 is more highly conserved in riboswitches regulating the GCV, while aptamer-2 is more highly conserved in those regulating TP. A Wilcoxon rank-sum analysis of all intra-group edges was performed as well to validate these findings (**Figure 2.2C**). I obtain very similar findings using a variety of alternative distance metrics calculated using Dynalign [83,107], FoldAlign [82,108], Clustal Omega [75], and RNApdist [104,105] (**Figures 2.3, 2.4**).

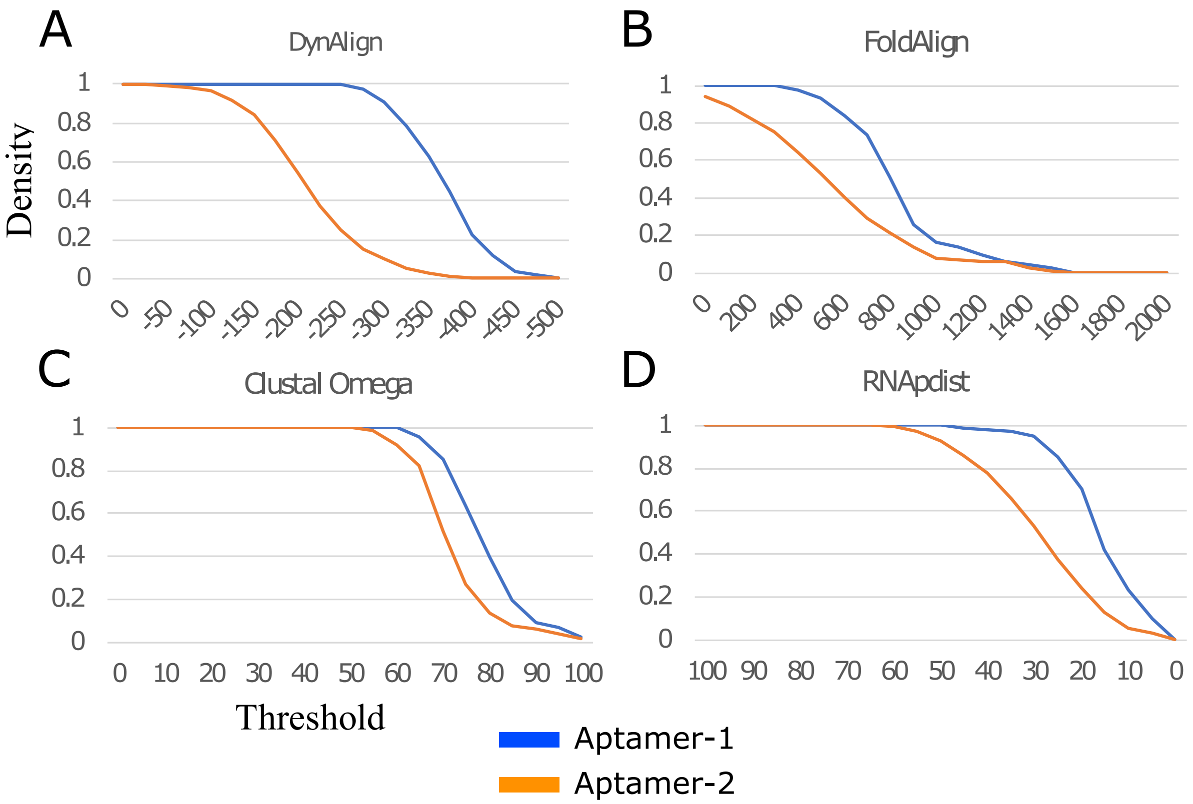

Figure 2.3: ***Clustering of Bacillaceae tandem riboswitch aptamers using Dynalign, FoldAlign, Clustal Omega, and RNApdist.***
A) Dynalign intra-edge density across a range of -500 to 0 (x-axis reversed to display decreasing density).
B) FoldAlign intra-edge density across a range of 0 to 2000.
C) Clustal Omega intra-edge density across a range of 0 to 100.
D) RNApdist inter-edge density across a range of 0 to 100 (x-axis reversed to display decreasing density).

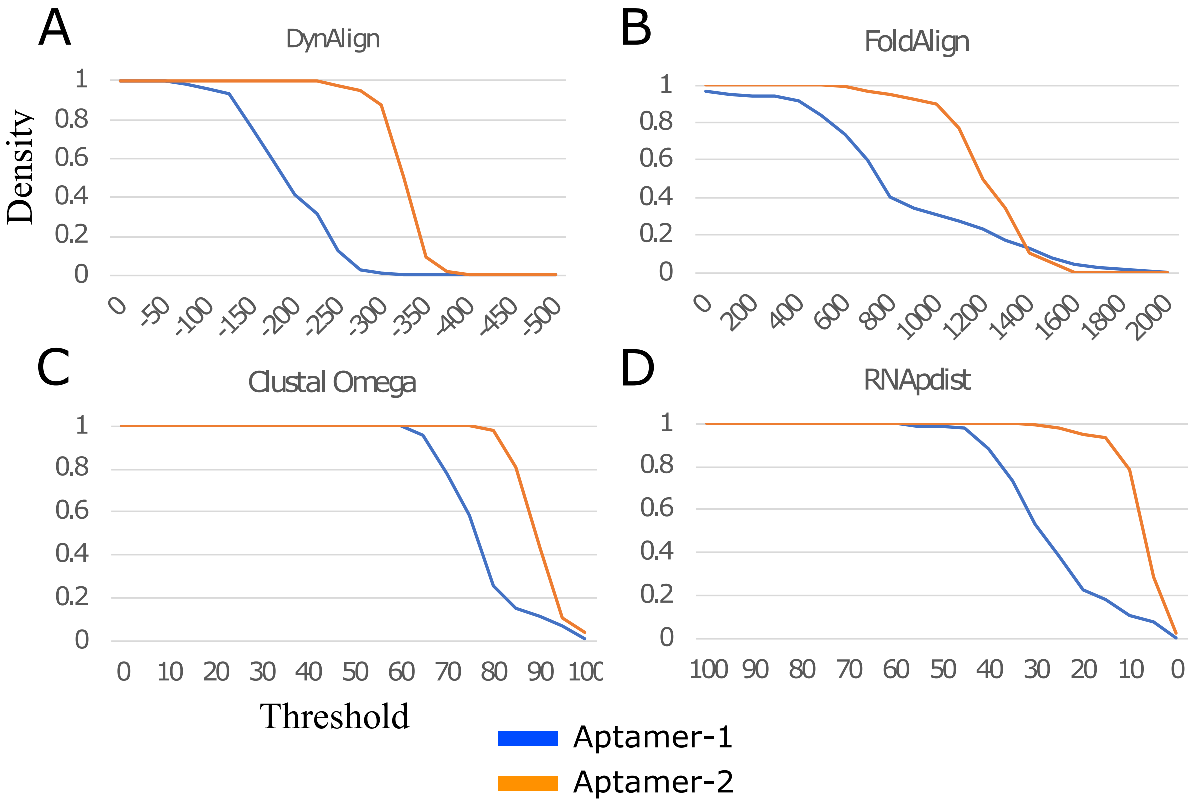

Figure 2.4: ***Clustering of Vibrionaceae tandem riboswitch aptamers using Dynalign, FoldAlign, Clustal Omega, and RNApdist.***
A) Dynalign intra-edge density across a range of -500 to 0 (x-axis reversed to display decreasing density).
B) FoldAlign intra-edge density across a range of 0 to 2000.
C) Clustal Omega intra-edge density across a range of 0 to 100.
D) RNApdist inter-edge density across a range of 0 to 100 (x-axis reversed to display decreasing density).

2.5.4 *Bacilli* class of bacteria shows clustering of singleton and tandem aptamers together.

To assess the relationship between singleton and tandem riboswitches, I implemented graph clustering of individual aptamers from both tandem and singleton glycine riboswitches. I first categorized singleton aptamers within the dataset (includes all bacteria, refseq77-microbial) into singleton type-1 or singleton type-2 based on whether the ghost aptamer was found 3' or 5' of the glycine aptamer. Of 782 singleton riboswitches, 342 were characterized as singlet type-1, 125 as singleton type-2, and 305 were unable to be conclusively characterized as one or the other (designated singleton type-0) (**Table 2.6**). I found that singleton type-1 riboswitches regulate GCV 93% of the time, while 90% of the singleton type-2 riboswitches regulate TP. This context dependent appearance of singleton riboswitches agrees with previous findings and gives confidence in my singleton annotation pipeline [58].

I then implemented graph clustering on a set containing all glycine riboswitch aptamers from *Bacilli*, excluding identical aptamers coming from different strains of the same species, totaling 436 aptamers (**Figure 2.5A**) (**Table 2.7**). This set from the *Bacilli* class was selected for its representation of both GCV (58%) and TP (31%) regulating riboswitches. Remaining riboswitches were labeled as regulating genes involved in glycine metabolism (Gly_Met) that is not part of the GCV operon or as Other. Using four distinct *de novo* community detection algorithms available in R (see methods) I identified modular communities within the set. Communities were selected based on each groups' core cluster, which was present within all community detection algorithms utilized. Aptamers that were found to be grouped with the core cluster in at least half of the community detection algorithms were subsequently added to the cluster. Cluster stability was verified using 100 replicates of parametric bootstrapping (**Table 2.8**) (see methods), as well as comparison to

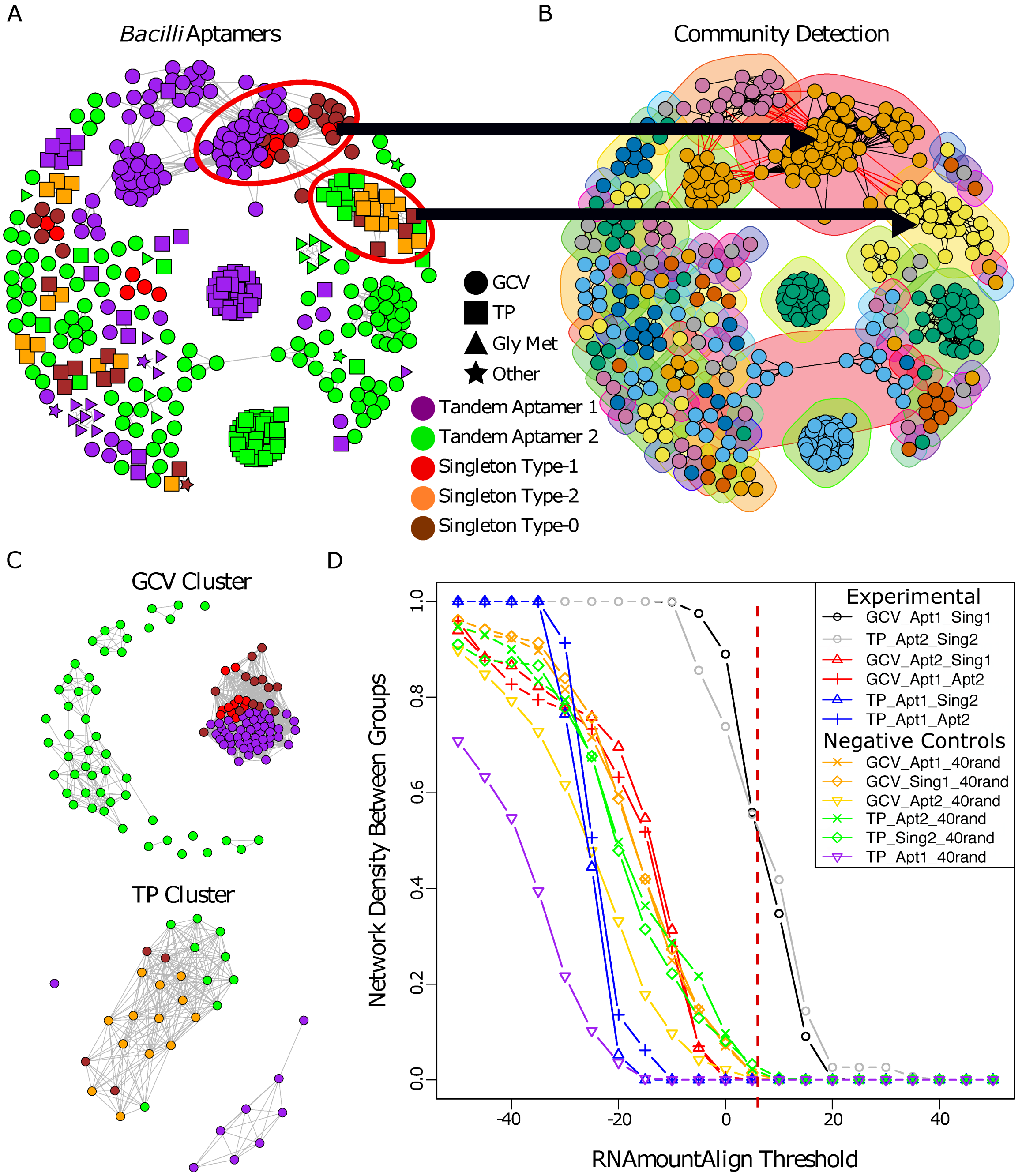

Figure 2.5: ***Clustering of glycine riboswitch aptamers identified within the Bacilli class of bacteria.***
A) Aptamers within the *Bacilli* bacterial class were identified and clustered based on RNAmountAlign pairwise similarity (visualized at threshold of 12). B) Sub-clusters (communities) were identified using the four community detection functions within R's igraph package. Two communities were identified that contain two different aptamer types: aptamer-1 and singlet type-1, and aptamer-2 and singlet type-2 that regulate GCV and TP respectively. Network shows visualization of the community detection algorithm cluster_fast_greedy (as implemented by R). Node colors correspond to distinct clusters detected. C) The two sub-clusters containing different aptamer types were parsed from the overall network, the tandem aptamers' partners were added to the set (as an out group within the same context), and graph clustering was visualized (RNAmountAlign threshold of 5). D) Edge density between aptamer groups was calculated for networks generated across a range of RNAmountAlign edge-weight thresholds. Dotted red line indicates the RNAmountAlign threshold (5) at which the networks in (C) were visualized.

MCL [125] and DBSCAN [126] clustering output. This resulted in clusters comprised of a highly conserved core set and aptamers that closely grouped with them. Most communities contain either aptamers derived from a tandem arrangement or singleton aptamers. However, two communities included both singlet and tandem derived aptamers. The first contains singleton type-1 aptamers and aptamer-1 of tandem riboswitches, all regulating GCV. The second includes singleton type-2 aptamers and aptamer-2 of tandem riboswitches, all regulating TP (**Figure 2.5B**).

Members of both mixed communities were extracted and networks were generated for each as described above (**Figure 2.5C**). For aptamers originally part of a tandem arrangement, the paired aptamer was included to assess relative conservation (**Table 2.9, 2.10**) of the singlet aptamers to each tandem aptamer type. I determined relative conservation between aptamer types by calculating the network density of edges connecting each aptamer type (inter-edge density) (**Figure 2.5D**) across a range of RNAmountAlign thresholds. I observe that singleton type-1 aptamers regulating GCV are most similar to aptamer-1 of tandem riboswitches in the same context and conversely that singleton type-2 aptamers regulating TP are most similar to aptamer-2 of tandem riboswitches in the same context. The inter-edge density between singleton type-1 aptamers and tandem aptamer-1's regulating GCV is comparable to that seen between singleton type-2 aptamers and tandem aptamer-2's regulating TP (**Figure 2.5D**). These two groupings also represent the highest conservation across aptamer types within their networks, with other pairings being comparable to inter-edge density measurements with a random set of 40 aptamers (**Table 2.11**). Using Dynalign, FoldAlign, Clustal Omega, and RNApdist as distance metrics yields similar findings (**Figures 2.6, 2.7**).

To further investigate the similarities between these aptamers, I generated consensus structures of the riboswitches found within each genomic context using a combination of tools (see methods). Consensus structures of riboswitches regulating GCV show tandem aptamer-1 and singleton type-1 aptamers have high conservation of the P2 and P3 stems, as well as the binding pocket, while the P1 stem of tandem aptamer-2 shows high conservation with the singleton type-1 ghost aptamer (**Figure 2.8A, B**). This conservation of the ghost aptamer P1 stem correlates with the region required for tertiary interactions of tandem and singleton riboswitches [56]**.** This is observed within riboswitches regulating TP as well, except the aptamer of singleton type-2 and tandem aptamer-2 are the conserved aptamers (**Figure 2.8C, D**). I then used R-scape to identify significantly covarying nucleotide pairs throughout the structures, which revealed highly correlating patterns between the singleton aptamer and corresponding tandem aptamer within each context (**Figure 2.8A-D)**. Interestingly the ghost aptamers did not show significant covariation, and this was reflected in the lack of significant covariation within the P1 stem of the corresponding tandem aptamer in each context. This could be due to the ghost aptamer, and corresponding tandem P1 stem, undergoing A-minor interactions with the P3 internal-loop of the adjacent aptamer during ligand binding. The importance of these interactions in the riboswitches structural stability could be necessitating high conservation at a nucleotide level that precludes covarying mutations within the ghost aptamers.

Together, my results indicate three things about *Bacilli* glycine riboswitch aptamers within each genomic context: 1) one tandem aptamer shows high conservation to the singleton aptamer, 2) conservation between the alternative tandem aptamer and singleton aptamers is no greater than conservation of the singleton to a random set of glycine riboswitch aptamers, and 3) ghost aptamer location correlates with the less conserved tandem aptamer. This fits with a model wherein these singleton riboswitches are the result of tandem riboswitch degradation, and which aptamer to be conserved and which to be degraded is dependent on genomic context. If the situation was reversed and tandems were the result of duplication events of singleton riboswitches, I would expect higher conservation between the singleton aptamer and both tandem aptamers compared to a random set of glycine riboswitch aptamers. However, I only observe such conservation with one tandem aptamer in each genomic context.

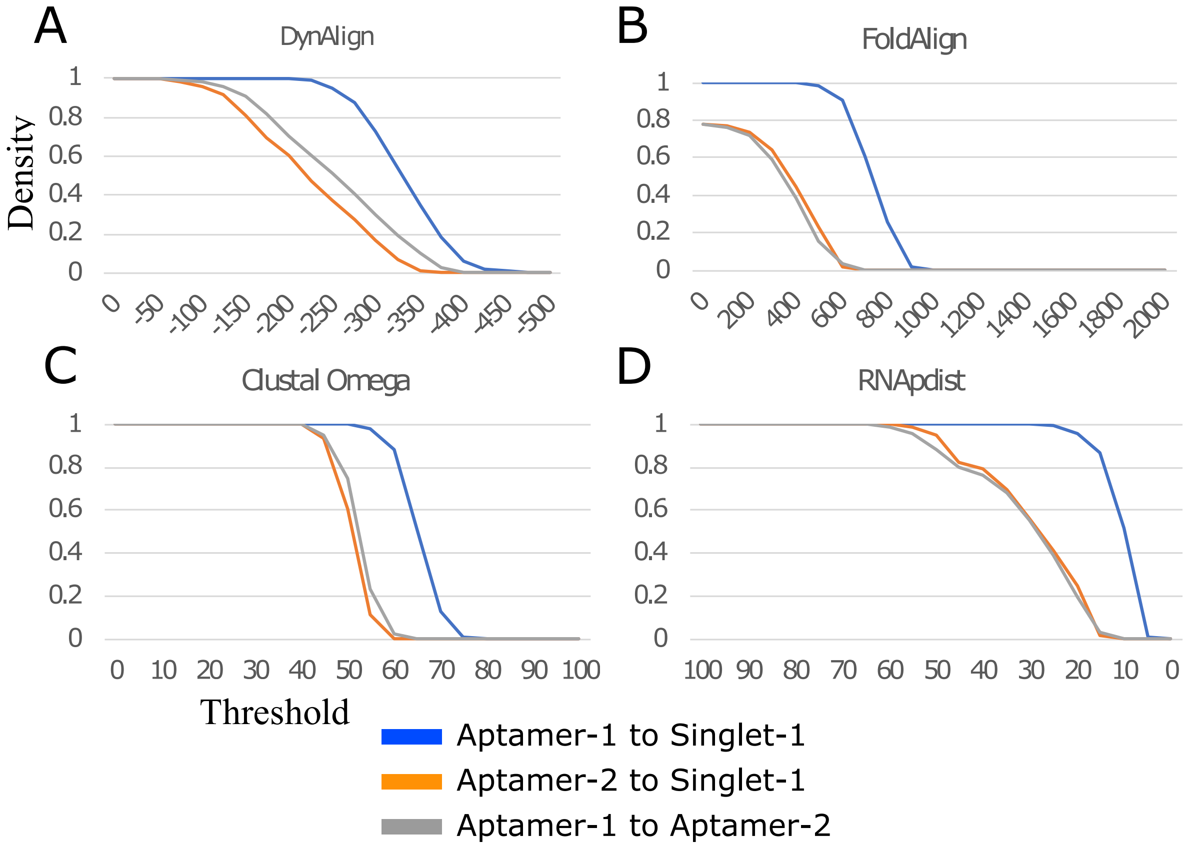

Figure 2.6: ***Clustering of Bacilli aptamer-1 and singleton type-1 aptamer subset using Dynalign, FoldAlign, Clustal Omega, and RNApdist.***
A) Dynalign inter-edge density across a range of -500 to 0 (x-axis reversed to display decreasing density).
B) FoldAlign inter-edge density across a range of 0 to 2000.
C) Clustal Omega inter-edge density across a range of 0 to 100.
D) RNApdist inter-edge density across a range of 0 to 100 (x-axis reversed to display decreasing density).

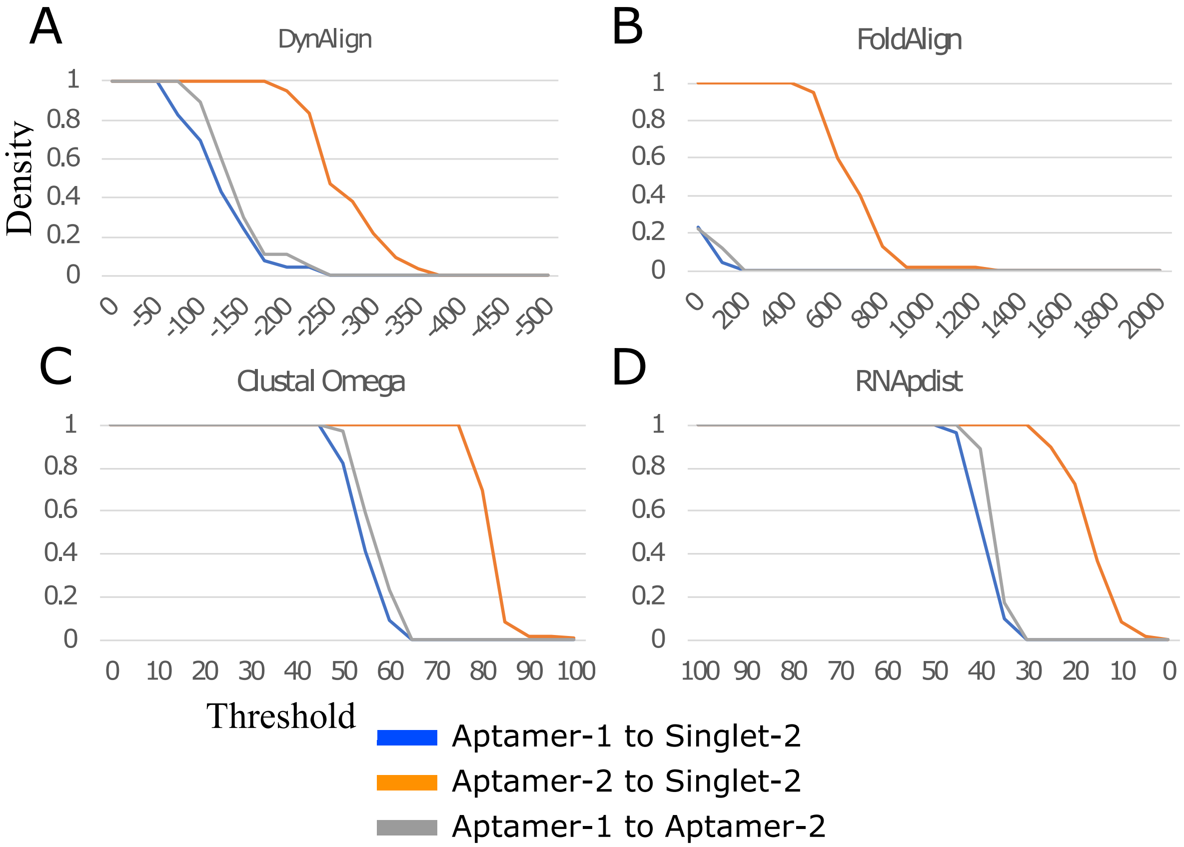

Figure 2.7: ***Clustering of Bacilli aptamer-2 and singleton type-2 aptamer subset using Dynalign, FoldAlign, Clustal Omega, and RNApdist.***
A) Dynalign inter-edge density across a range of -500 to 0 (x-axis reversed to display decreasing density).
B) FoldAlign inter-edge density across a range of 0 to 2000.
C) Clustal Omega inter-edge density across a range of 0 to 100.
D) RNApdist inter-edge density across a range of 0 to 100 (x-axis reversed to display decreasing density).

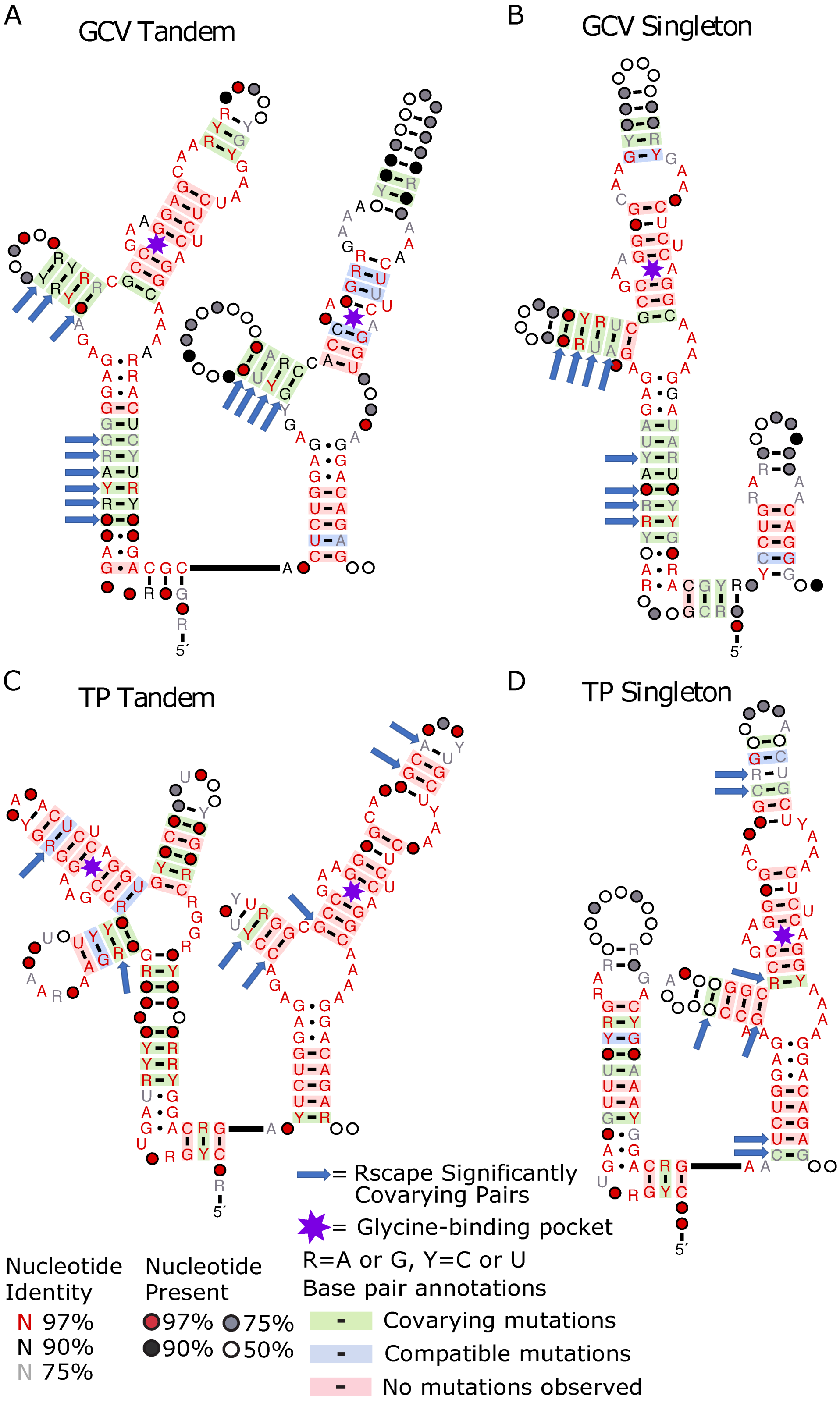

Figure 2.8: Consensus structures of Bacilli riboswitches within a given genomic context display conservation between tandem and singleton aptamers. Consensus secondary structure of the singleton and tandem riboswitches delineated by the genomic context. Conservation of nucleotides generated using R2R with the individual covariance models. Significant covariation of base pairs detected using R-scape. Tandem (A) and singleton (B) riboswitches regulating GCV. Tandem (C) and singleton (D) riboswitches regulating TP.

2.5.5 *Actinobacteria* riboswitches display similar clustering pattern observed in *Bacilli*.

To determine whether these patterns are observed within other clades of bacteria, I gathered all glycine aptamers within the dataset in the *Actinobacteria* phylum (distantly related to both the *Vibrio* and *Bacilli* classes analyzed previously), excluding identical aptamers coming from different strains of the same species, totaling 606 aptamers (**Table 2.12**). I then evaluated all aptamers within the set in the same manner as the *Bacilli* investigation. Within this phylum, glycine riboswitches primarily regulate GCV (74%) or other genes involved in glycine metabolism (22%). I identified a group of 34 conserved aptamers corresponding to riboswitches regulating GCV (**Figure 2.9A**) and utilized *de novo* community detection algorithms to validate my observation (**Figure 2.9B**). Cluster stability was verified using 100 replicates of parametric bootstrapping (**Table 2.8**) (see methods), as well as comparison to MCL and DBSCAN clustering output. The aptamers within this group comprised primarily singleton type-1 aptamers and tandem aptamer-1 sequences, with five singleton type-0 and two singleton type-2 aptamer sequences accounting for the remainder. The singleton type-2 aptamers within the set may be misclassified aptamers or examples of singleton aptamers which do not conform to the patterns observed for other investigated aptamers. I performed graph clustering on the group, with paired tandem aptamer-2s included as an out-group, to investigate conservation of aptamer types (**Figure 2.9C**) (**Table 2.13**). I then calculated the edge densities within and between singleton type-1 aptamers, tandem aptamer-1s, and tandem aptamer-2s, which demonstrate a clear conservation between singleton type-1 aptamers and tandem aptamer-

1s (**Figure 2.9D**). These findings fit my conclusions drawn from the *Bacilli* class of bacteria. Using Dynalign, FoldAlign, Clustal Omega, and RNApdist as distance metrics yield similar findings (**Figure 2.10 Fig**).

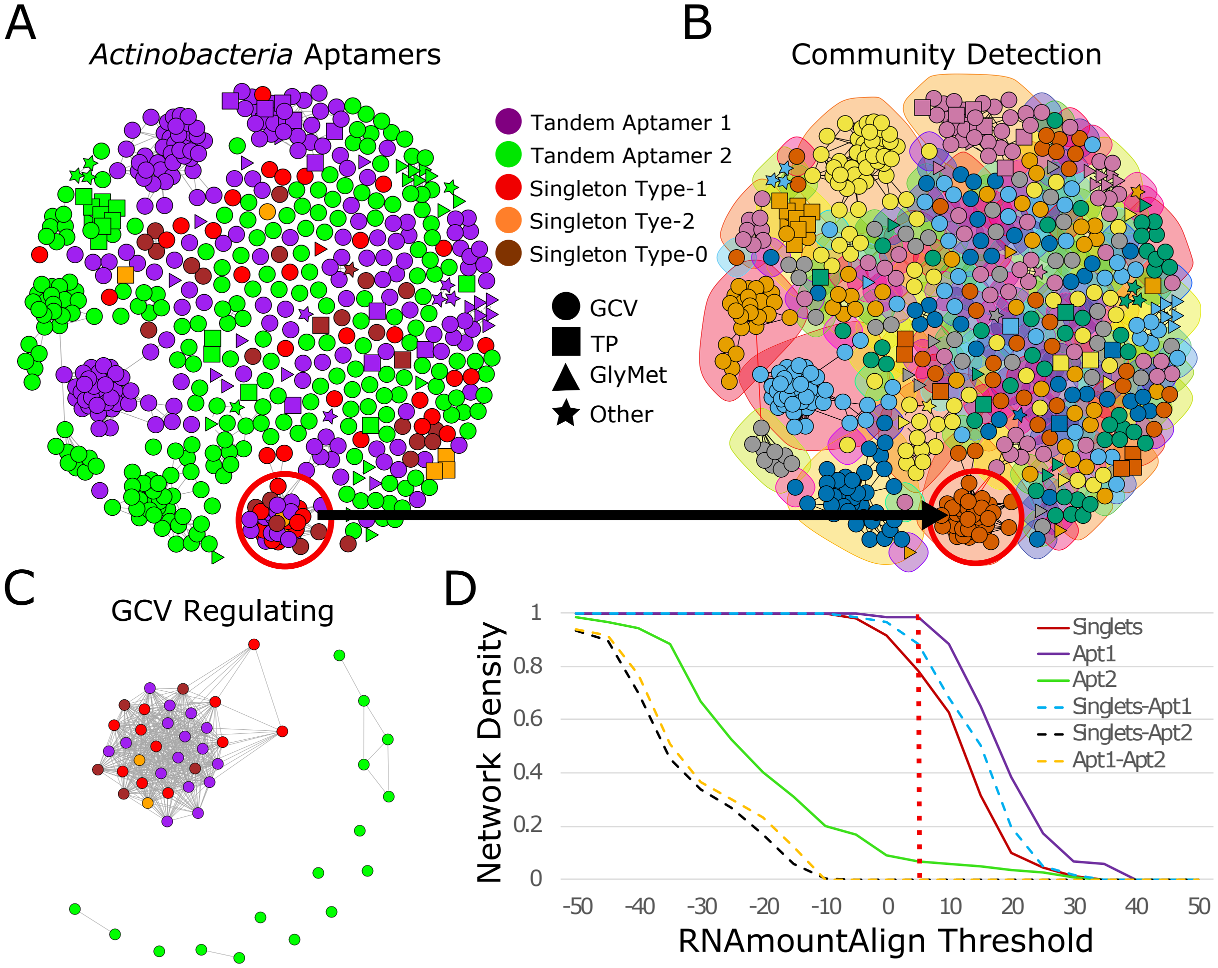

Figure 2.9: ***Clustering of glycine riboswitch aptamers identified within the Actinobacteria phylum of bacteria.***
A) Aptamers within the *Actinobacteria* bacterial phylum were identified and clustered based on RNAmountAlign pairwise similarity (visualized at threshold of 12). B) Sub-clusters (communities) were identified using the four community detection functions within R’s igraph package. One community containing primarily two different aptamer types: aptamer-1 and singlet type-1 was identified. Display visualization uses the community detection algorithm cluster_fast_greedy. Node colors correspond to distinct clusters detected. C) The community containing different aptamer types was parsed from the overall network, the tandem aptamers’ partners were added (as an out group within the same context), and graph clustering was visualized (RNAmountAlign threshold of 5). D) Edge density between aptamer groups was calculated for networks generated across a range of RNAmountAlign edge-weight thresholds. Solid lines correspond to edge density within a group and dashed line correspond to edge density between the two indicated groups. Dotted red line indicates the RNAmountAlign threshold (5) at which the networks in (C) were visualized.

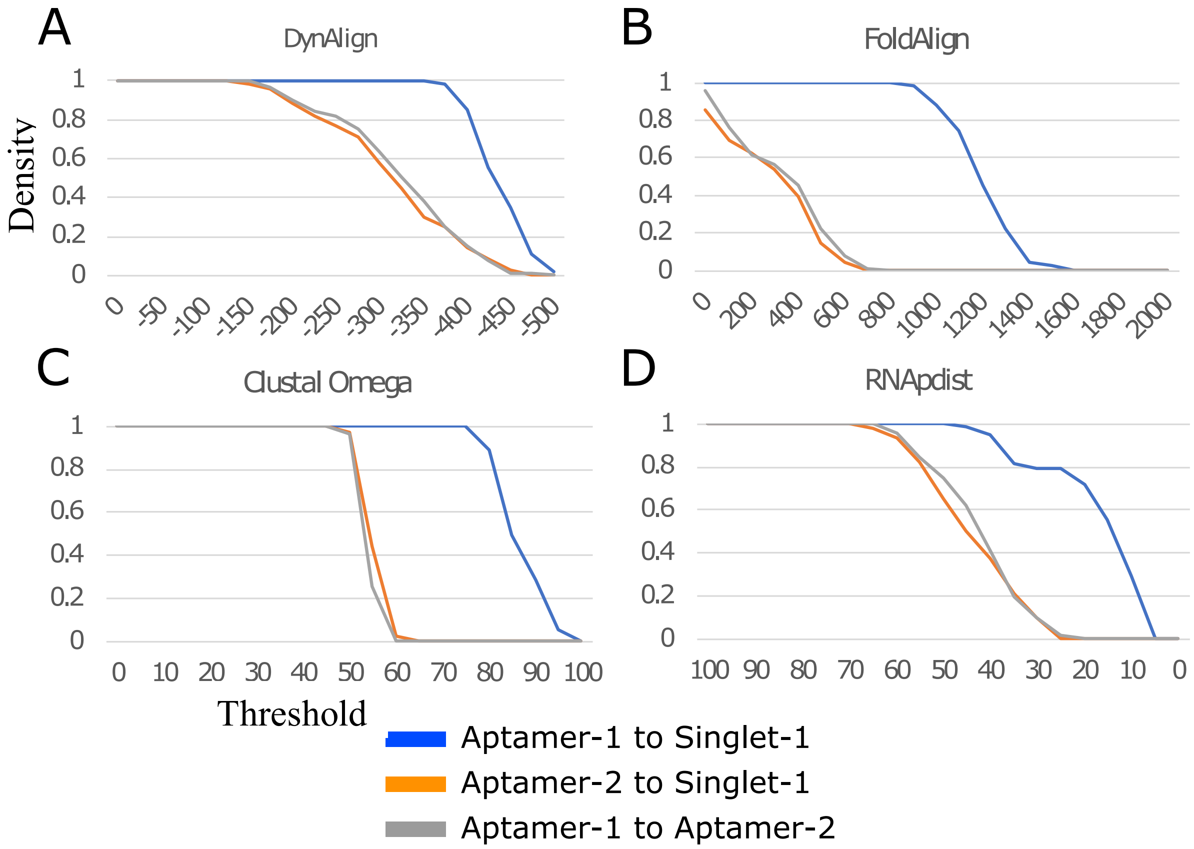

Figure 2.10: ***Clustering of Actinobacteria aptamer-1 and singleton type-1 aptamer subset using Dynalign, FoldAlign, Clustal Omega, and RNApdist.***
A) Dynalign inter-edge density across a range of -500 to 0 (x-axis reversed to display decreasing density).
B) FoldAlign inter-edge density across a range of 0 to 2000.
C) Clustal Omega inter-edge density across a range of 0 to 100.
D) RNApdist inter-edge density across a range of 0 to 100 (x-axis reversed to display decreasing density).

2.5.6 Clustering based on genomic context is observed throughout entire bacterial kingdom.

To determine whether clustering patterns observed within the *Bacilli* class and *Actinobacteria* phylum are reflected throughout the rest of the bacterial kingdom and can be observed among randomly selected aptamers, I randomly selected 150 distinct glycine riboswitch aptamers each for the GCV and TP genomic context (**Table 2.14, 2.15**). The selection retained comparable numbers of each aptamer type and excluded singleton type-0 aptamers. Singleton type-1 and type-2 aptamers are underrepresented in the TP and GCV regulating sets, respectively, because each aptamer type has few instances within that genomic context. Despite the diverse taxonomic range represented within this dataset, the generated networks display clustering patterns which align with my previous observations: a tendency towards clustering of singleton type-1 aptamers with tandem aptamer-1s when regulating GCV, and clustering of singleton type-2 aptamers with tandem aptamer-2s when regulating TP (**Figure 2.11)**. Inter-edge density graphs of the aptamers shows similar trends to those seen within the *Bacilli* class and Actinobacteria phyla. Using Dynalign, FoldAlign, Clustal Omega, and RNApdist as distance metrics yield similar findings (**Figure 2.12, 2.13**).

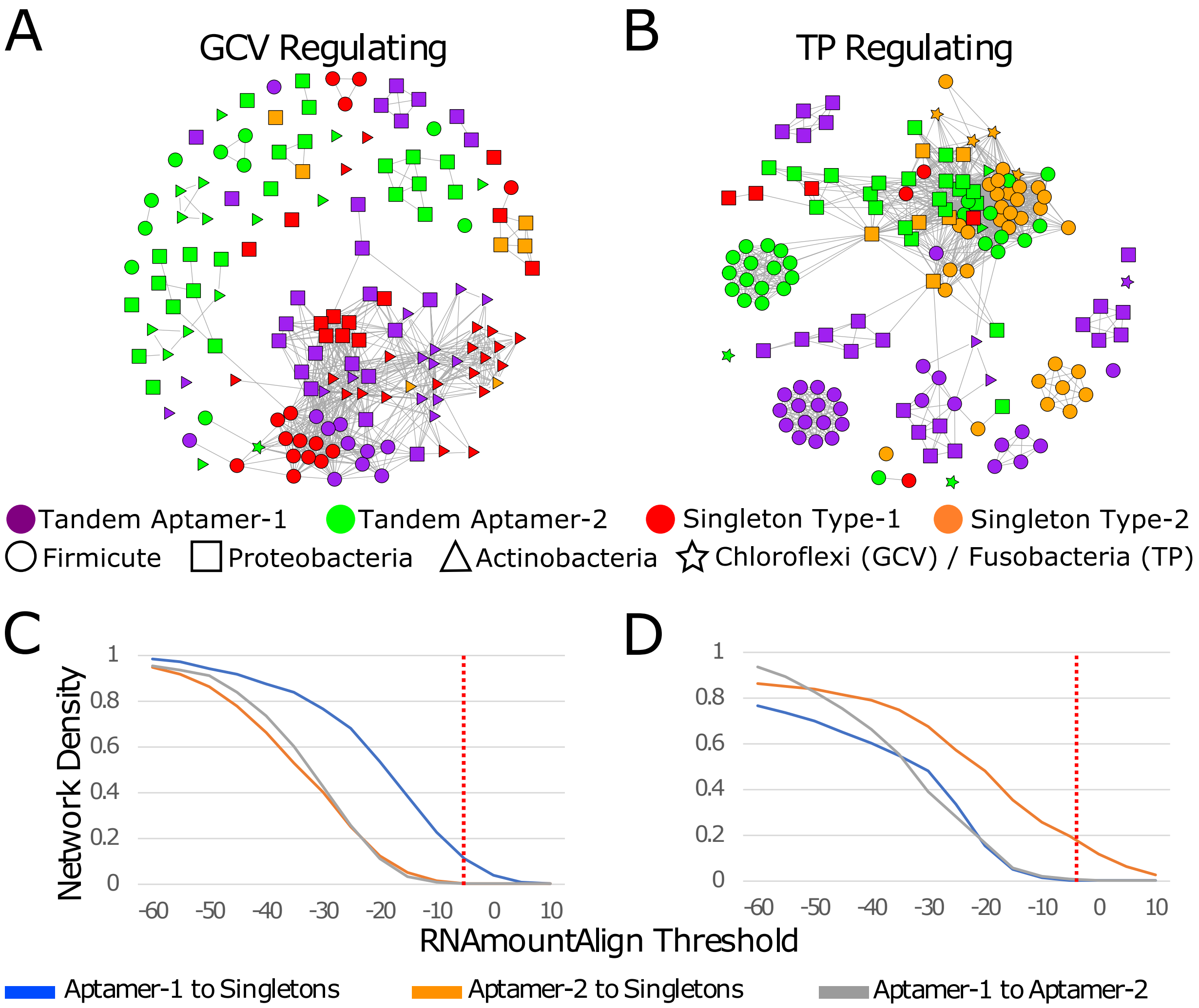

Figure 2.11: ***Clustering of random glycine riboswitch aptamers across the bacterial kingdom.***
Network visualization of 150 randomly chosen aptamers regulating GCV (A) and TP (B) were clustered based on RNAmountAlign pairwise similarity (threshold -5). Only singletons that could be classified as type-1 or type-2 were included in this set. Inter-edge density was calculated between aptamer types across a range of RNAmountAlign thresholds for the GCV regulating set (C) and the TP regulating set (D). Dotted red line on graphs indicate the threshold at which the clusters were visualized in Figure 5. Only singletons that could be classified as type-1 or type-2 were included in this set.

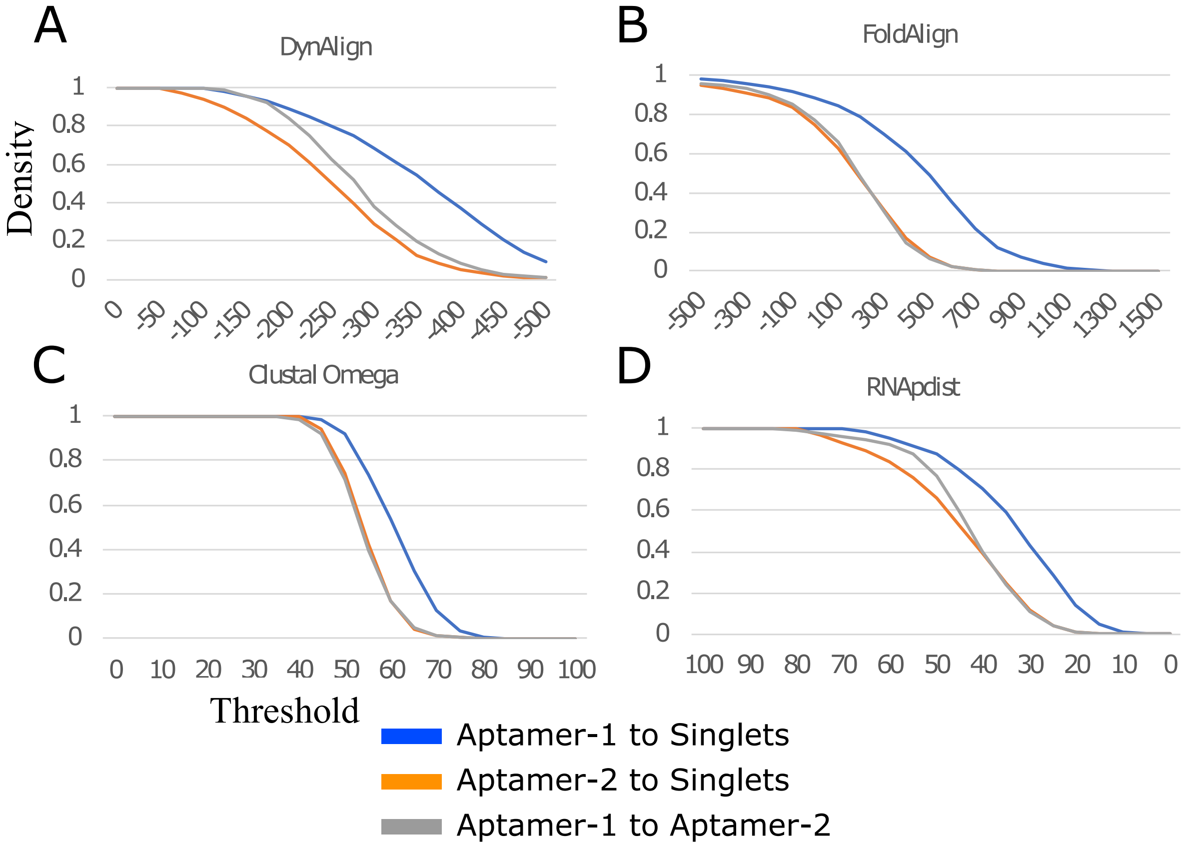

Figure 2.12: ***Clustering of random riboswitch aptamers regulating GCV using Dynalign, FoldAlign, Clustal Omega, and RNApdist.***
A) Dynalign inter-edge density across a range of -500 to 0 (x-axis reversed to display decreasing density).
B) FoldAlign inter-edge density across a range of -500 to 1500.
C) Clustal Omega inter-edge density across a range of 0 to 100.
D) RNApdist inter-edge density across a range of 0 to 100 (x-axis reversed to display decreasing density).

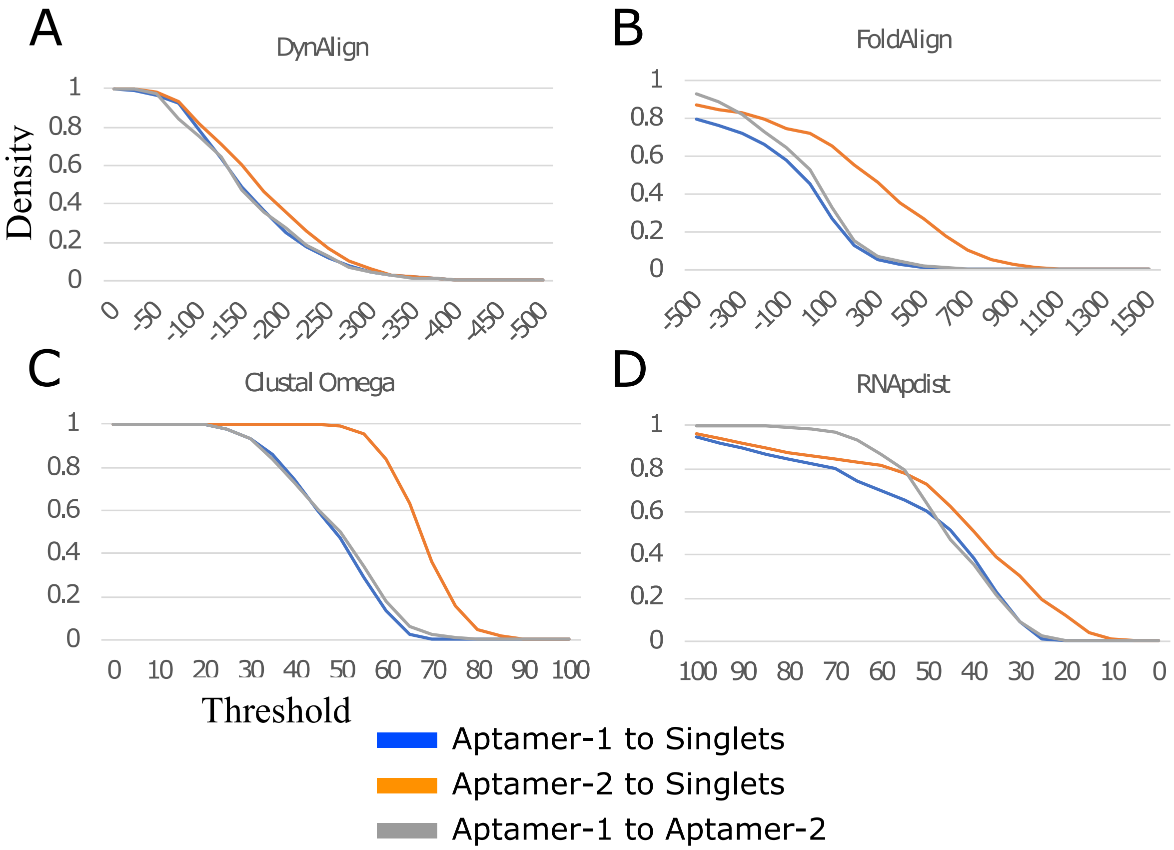

Figure 2.13: ***Clustering of random riboswitch aptamers regulating TP using Dynalign, FoldAlign, Clustal Omega, and RNApdist.***
A) Dynalign inter-edge density across a range of -500 to 0 (x-axis reversed to display decreasing density).
B) FoldAlign inter-edge density across a range of -500 to 1500.
C) Clustal Omega inter-edge density across a range of 0 to 100.
D) RNApdist inter-edge density across a range of 0 to 100 (x-axis reversed to display decreasing density).

2.6 Discussion

The tandem aptamers of the glycine riboswitch have fascinated RNA biologists since their identification in 2004 [5,31]. Extensive work has assessed whether the two homologous aptamers of the tandem glycine riboswitch functioned cooperatively [53,55,115,116], which tandem aptamer was more important for ligand binding [29,59], and what, if any, benefit a tandem conformation provided over the singleton glycine riboswitch [56]**.** In this work I use graph clustering analysis to investigate a similarly divisive question: what is the evolutionary relationship between tandem and singleton glycine riboswitches? While it may appear intuitive to believe that the tandem riboswitches that have been identified are the result of a duplication of identified singleton riboswitches, the findings point towards most singleton riboswitches being the result of tandem riboswitch degradation.

Phylogenetic evaluation of *Bacillaceae* and *Vibrionaceae* tandem riboswitches revealed that genomic context impacts riboswitch evolution. This is illustrated by *Bacillaceae* riboswitches which regulate TP grouping more closely with *Vibrionaceae* riboswitches regulating TP than with *Bacillaceae* riboswitches regulating GCV. Further investigation into the individual aptamers of *Bacillaceae* tandem riboswitches regulating GCV compared to TP showed that both aptamers individually show the same pattern of divergence.

Taking this analysis a step farther using graph clustering, I was able to determine that genomic context dictates which aptamer within a tandem glycine riboswitch is more highly conserved: aptamer-1 is more highly conserved in riboswitches regulating GCV, while aptamer-2 is more highly conserved in those regulating TP. These findings provide an elegant answer to a contradiction within the field in which investigations of diverse glycine riboswitch homologs yielded different results for whether ligand-binding of the first or second aptamer is more important for functionality [29,59]. The results align with both studies' findings: the aptamer identified as the essential binding partner for regulation in each study is the aptamer found in the study to be more highly conserved within that genomic context. With the results of these previous studies combined with this new perspective provided by the data, it is reasonable to conclude that a difference in genomic context has driven glycine riboswitches to conserve different primary ligand-binding aptamers. Widespread horizontal transfer of the riboswitch with its accompanying gene could account for these findings. To investigate this possibility, I generated gene trees for aminomethyltransferases and symporters preceded by glycine riboswitches. From these trees, there is limited evidence of horizontal transfer of these genes (**Table 2.16, 2.17**).

My observation that tandem glycine riboswitch evolution is affected by genomic context led us to question the impact of genomic context on singleton glycine riboswitches. I extended my network analysis to singleton riboswitches, which provided valuable insight into the relationship of tandem and singleton glycine riboswitches. Clustering of singleton and tandem aptamers from the *Bacilli* and *Actinobacteria* clades revealed that singleton aptamers are more similar to the first or second tandem aptamer based on genomic context: singleton type-1 aptamers regulating GCV are more similar to aptamer-1 of tandems regulating GCV, while singleton type-2 aptamers regulating TP are more similar to aptamer-2 of tandems regulating TP. This similarity of singletons to one tandem aptamer within a genomic context is highlighted by the fact that the singleton aptamers show no higher similarity with the other tandem aptamer than with a random set of 40 glycine riboswitch aptamers. This is observed within both GCV and TP regulating riboswitches and directs us towards the conclusion that singleton riboswitches are the remnants of degraded tandem aptamers.

I propose a model for the evolutionary path of the glycine riboswitch in which tandem riboswitches become singleton riboswitches by undergoing degradation of one aptamer into a ghost aptamer which retains regions relevant to tertiary interaction (**Figure 2.14**). In this model, the aptamer which is conserved is dependent on genomic context. The different conservation of tandem aptamers based on genomic context also fits recent studies that demonstrate a high likelihood that whether a glycine riboswitch is regulating TP or GCV is predictive of whether they are an on or off switch [31,54,58,59]. This fits a logical model for cellular response to high concentrations of glycine as a toxin [59,130–134]: genes responsible for glycine degradation become upregulated and those involved in glycine uptake become downregulated. In this way riboswitches in each genomic context protect the cell from the glycine toxicity as concentrations increase. This difference in regulation functionality accounts for riboswitches in different genomic contexts diverging, culminating in conservation of different aptamers and ultimately the formation of singleton riboswitches. It is possible that some singletons may have arisen from deletion of the middle section of a tandem riboswitch, leaving the 5’ half of aptamer 1 and the 3’ half of aptamer 2, resulting in a singleton. However, this scenario seems unlikely because it does

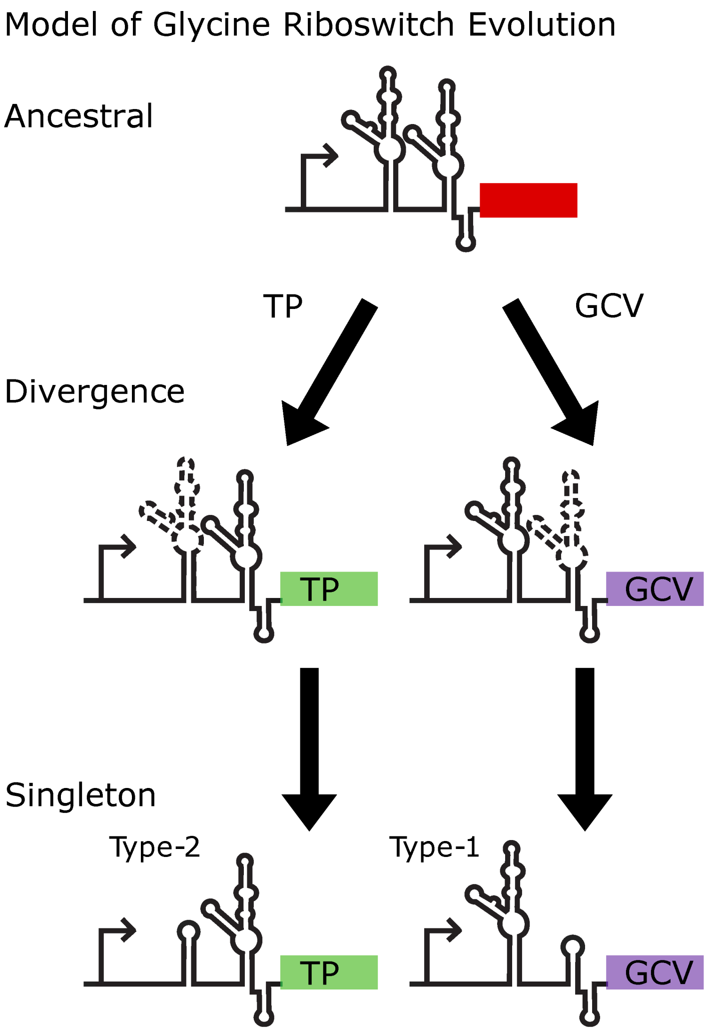

Figure 2.14: ***Model of glycine riboswitch evolution***. Model proposed for the evolution and divergence of the glycine riboswitch. In this model a progenitor tandem riboswitch conserves one of the tandem aptamers based on the genomic context of the riboswitch, while the other slowly degrades down to the minimalistic components required for tertiary interaction to drive gene regulation. In this way, tandem glycine riboswitches may degrade into functional singleton tandem riboswitches.

not account for the ghost aptamer, which is important for structural stability of the glycine riboswitch.

While I have used graph clustering to unravel a specific discrepancy arising in existing experimental data, my approach may be used more broadly to assess how other riboswitches and other ncRNAs evolve and change over time. Variations in homologous riboswitch aptamers have demonstrated functional consequences. There are a range of variant riboswitch classes that interact with differing ligands [68,135,136], the most compelling of which are the homologous *ykkC* riboswitches [5,137] which include at least five subclasses each of which binds a distinct ligand and regulate a distinct set of genes [37,50,61,62]. There are also examples where structurally distinct riboswitches interact with the same or very similar ligands, such as the seven riboswitch classes involved in regulating S-adenosylmethionine concentration [63,138]. These RNAs include the SAM/SAH riboswitch which has been proposed to be a minimalistic form of a SAM riboswitch that evolved in organisms which readily degrade SAH [66]. The approaches I developed circumvent the limitations of traditional phylogenetic methods for assessing ncRNA similarity and enable identification of patterns in aptamer conservation that may point toward differences in biological function across diverse organisms.

Chapter 3: Clustering validation using ykkC riboswitches

3.1 Abstract

While the ability of computational methods to classify homologous ncRNA families has increased drastically in recent decades, researchers still lack systematic methods for evaluating the diversity and subtypes within those families. Investigating diversity within ncRNA classes is crucial to understanding how they diverge and evolve new functionality. However, nuanced evaluation of homologous ncRNA raises problems not often encountered when dealing with datasets of other homologous molecules, such as proteins. ncRNA can vary greatly in their distribution and often lack high primary sequence homology, as the structural components driving function are often not conserved in sequence. To investigate variation within ncRNA classes, an approach that focuses on the pairwise relationship between individual instances within an identified class is needed. This would allow for grouping more highly related instances based on a range of similarity metrics. Graph clustering provides this analysis structure, allowing researchers to study sub clusters within a ncRNA class that could potentially have functional and/or evolutionary relevance. Here, I present that graph clustering is able to distinguish between various ykkC subtype riboswitches, as well as perform local structural evaluation of the P3 stem-loop to find that the structural homology reflects similarity between the ligand of each riboswitch.

3.2 Background

Riboswitch classes are primarily distinguished based on the specific ligand that is bound by the members. While many riboswitch classes have had their ligand identified, there are many more that have not been classified which are likely to play important biological roles [139]. These orphan riboswitches are often looked at as untapped sources of diversity and insight for functional RNA research [68]. However, variation within already identified classes of riboswitches hold this same potential. Therefore, investigating the diversity within and between riboswitch classes is vital to gaining a deeper understanding of riboswitches as a whole, including what factors drive divergence, the functional impact of the variation, and implementation in synthetic systems.

In recent years, studies have highlighted the functional importance of diversity within and across defined riboswitch classes. The ykkC riboswitches share high structural homology and were previously characterized as a single class based on the similarity, but were subsequently discovered to be composed of multiple classes that bind distinct ligands [5,37,50,61,137]. On the other hand, the SAM family of riboswitches represent a set of distinct riboswitch structures that each bind a common ligand [63]. Evidence suggests that the SAM-SAH riboswitch may be a descendent of a SAM riboswitch that lost specificity in a SAH-limited environment [66]. The glycine riboswitch has been demonstrated to have undergone divergence across genomic contexts to function as an on-switch when regulating the glycine degradation pathway and an off-switch when regulating amino acid transporters [54,57–60]. These types of functional alterations can be artificially introduced as well, as has been seen with the iron-sensing riboswitch which can be modified to switch its binding specificity to cobalt with a few nucleotide mutations [140]. These examples highlight the importance of looking beyond the broad classifications of riboswitches and to the variation within each class to truly address the functional diversity represented.

While our ability to identify broad homology of structural RNA classes has improved dramatically [18,129,141], the field of RNA investigation is still lacking in computational methods for quickly identifying divisions within these identified classes that may have functional importance. Since the discovery of riboswitches, identification of classes has broadly been broken into two processes: *de novo* classification of homologous structured ncRNAs [141–146] and identification of new members of existing classes [17,18,87,90,119,147–150]. While this is helpful for binning sets of homologous riboswitches together, it reduces the complexity within those sets down to a binary "member or nonmember" relationship with broad models. These tools are ideal for identifying putative members of distinct classes, but are not designed to investigate the nuance within those classes. To gain a better understanding of riboswitch functionality and diversity, it is necessary to look at sets of riboswitches based on the pairwise relationships and clustering patterns within their respective class.

To address these limitations, graph clustering can be implemented to group sets of ncRNAs based on pairwise similarity. This is done by generating networks of nodes corresponding to individual instances of a ncRNA. Each node is connected by an edge to each other node, and each edge is weighted based on a chosen pairwise similarity metric comparing the two connected ncRNA. These edges can then be pruned through a thresholding process in which edges which fail to meet a minimum similarity score are removed from the network, altering the overall topology and leading to clustering of more highly similar nodes. This provides a visual representation of similarity between ncRNA instances for nuanced, hands-on investigation, as well as providing a basis for *de novo* community detection within networks across a range of thresholds. Relative conservation of groups within the network can be investigated using the change in edge-density between nodes within that group across incremental thresholding.

Graph clustering was applied to a set of ykkC riboswitches. This group of riboswitch classes was originally grouped as one class [5,137], but further experimentation later found it to be composed of five distinct subtypes each binding distinct ligands. These include the guanidine-I riboswitch [62,67] and four ykkC subtype 2 riboswitches, three of which have had their specific ligands identified. ykkC type-2A was shown to bind guanosine tetraphosphate (ppGpp), an alarmone that responds to amino acid starvation conditions, and regulate genes involved in branches chain amino acid synthesis [61]. ykkC type-2B was shown to bind phosphoribosyl pyrophosphate (PRPP), a molecule required for *de novo* purine synthesis, and regulate genes involved in purine biosynthesis [50]. ykkC type-2C was shown to bind adenosine/cytidine diphosphate (ADP/CDP) and regulate genes that hydrolyze the phosphoester bond of nucleoside diphosphate [37]. ykkC type-2D is yet to have its ligand identified and published. The high sequence and structural similarity between the ykkC subtype 2A-C riboswitches offer an ideal dataset to determine the ability of graph clustering to distinguish closely related riboswitch subtypes. In this analysis, I curated a set of riboswitches representative of a dataset predating ykkC subtype classification and performed a graph clustering analysis to determine whether the distinct riboswitches could have been detected. I then extended this to published sets of ykkC subtype 2A-C riboswitches, referred to as ppGpp, PRPP, and ADP-CDP, respectively.

3.3 Methods

3.3.1 Riboswitch Identification

The initial set of riboswitches was identified by curating a set of published riboswitches encompassing ykkC subtype 2A-C: ppGpp, PRPP, and ADP-CDP respectively. A set of 105 ppGpp riboswitch sequences [61], 253 PRPP riboswitch sequences[50], and 45 ADP-CDP riboswitch sequences [37] were gathered from existing literature within the field.

The second set of ykkC riboswitch data was generated using an RFAM covariance model from 2014. This was before the determination of the ykkC riboswitch subtype functions. The covariance model was used to query the ReSeq77 bacterial database to find instances that fit the model with a better than 10^-5 e-value. This was then further reduced to remove redundant sequences from identical species and then pruned down to one representative sequence per genus. The bedtools closest function was used to determine the nearest downstream gene and gene function for each sequence, and thus the most likely gene directly regulated by the riboswitch. These were then used to categorize riboswitches into a putative ykkC subtype based on the regulated gene function. If a ykkC subtype could not be determined based on the function, or if the gene was marked as a hypothetical protein, the sequence was left unclassified and binned into a category called “UNDETERMINED.”

3.3.2 Graph clustering

Graph clustering was performed on sets of riboswitches, with individual riboswitches representing the vertices and edges corresponding to pairwise similarity metrics relating pairs of vertices. Edges were thresholded based on pairwise similarity score, resulting in pruned networks of clustered riboswitches containing vertices with higher similarity than the threshold value. Multiple distance metrics were used to evaluate riboswitch clustering, including, partition function (RNApdist) [104,105], sequence and structure based on ensemble expected mountain height (RNAmountAlign) [85], and sequence similarity (Muscle)[76]. Visualization of riboswitch similarity networks were generated using the igraph and qgraph R-libraries [110,124]. Optimal threshold for visualization of the datasets were selected based on high separation and modularity of observed sub clusters.

3.3.3 Community Detection

Modular clusters were detected using a combination of *de novo* community detection algorithms provided in igraph: cluster_fast_greedy (hierarchical clustering starting with addition of strong edges), cluster_walktrap (random-walk), cluster_edge_betweenness (hierarchical clustering starting with removal of weak edges), and cluster_leading_eigen (spectral clustering). These were chosen to encompass a range of algorithmic approaches to community detection. I then performed 100 replicates of a parametric bootstrapping analysis on the detected communities of interest. This process introduces a 5% perturbation of the network by adding/removing edges at random in a 1:1 ratio. Upon completion, the resulting networks contain the same nodes and an equivalent number of edges, but 5% of the edges connect different nodes. After each iteration, the *de novo* community detection algorithms were rerun to detect and record community composition within the perturbed network. These were then compared to the original group that had no perturbation. Following the 100 iterations of this comparison, I used Jaccard Similarity Index to calculate compositional similarity of each post-perturbation cluster to the original cluster. Average Jaccard Similarity Index across all 100 iterations was used to determine the robustness of the clusters.

3.3.4 Edge Density Graph Generation

Generation of network density graphs was performed across thresholding ranges. This was done by determining the total number of pairwise comparisons within a cluster and calculating the percent of edges that passed the thresholding test as threshold stringency increased across the total range. These calculations were made on an in-group and out-group basis, comparing the number of possible edges within a defined group and the number of edges connecting distinct groups respectively. These graphs equate to a relative similarity measure, which can be read as a measure of conservation, and represent a flipped cumulative distribution of possible edges and observed edges for a given cluster across a range of edge-weight thresholds.

3.4 Results

3.4.1 PRPP riboswitch forms distinct cluster in ykkC data predating subtype distinctions

To evaluate clustering of ykkC riboswitches in a way that would mimic *de novo* detection of subtypes, I generated a dataset using available sequences and models predating ykkC subtyping. A covariance model was generated using 80 aligned sequences from the 2014 RFAM model of the ykkC riboswitch (**Table 3.1**). It was used to identify a set of riboswitches from RefSeq77 which fit the non-specific "ykkC riboswitch" model. This set was then curated to remove hits with low e-value scores and duplicates from closely related species. This resulted in a set of 167 "ykkC riboswitches" that were gathered independent of any subtype knowledge (**Table 3.2**). I then used bedtools to identify the most likely candidate gene that each riboswitch regulates and classified each riboswitch as a putative guanidine-I or ykkC subtype 2A-C riboswitch. The final set of 167 "ykkC" riboswitches is primarily composed of guanidine-I riboswitches, but some riboswitches were able to be sorted as ppGpp and PRPP based on their regulated genes.

To determine whether I could distinguish the distinct riboswitch classes in the dataset I performed graph clustering using three distance metrics: sequence similarity, structural similarity, and a hybrid of both. First, I implemented RNAmountAlign for pairwise comparisons because of its optimized speed and use of both sequence and ensemble structure when calculating pairwise distance. I then incorporated muscle and RNApdist for sequence and secondary structure similarity, respectively. After graph clustering using these metrics, the clustering was compared to the assignments determined based on regulated genes to determine if the clustering separated the riboswitches based on putative functionality.

The clustering pattern seen based on these putative classifications indicates that PRPP can be detected within the dataset as a distinct subgroup. While structure-based metrics show highly overlapping clustering of all types, clustering based on sequence similarity is able to distinguish a group of nodes corresponding to PRPP riboswitches, distinct from a cluster corresponding to ppGpp and guanidine-I riboswitches (**Figure 3.1A-C**). This high structural similarity fits previous findings within the field, which show that these riboswitches have highly homologous structures and it is sequence differences within the binding pocket that allow for the distinction between the subtypes [37,50,61,62,67].

To validate these findings, I implemented *de novo* community detection algorithms to verify that the PRPP cluster was a distinct subgroup of the network (**Figure 3.1D-F**). I then determined the stability of the PRPP cluster by bootstrapping the networks using a 5% perturbation of the existing edges, re-performing the *de novo* community detection, and comparing identified clusters back to the original. After 100 rounds of bootstrapping, I used the Jaccard similarity index to quantify stability of the PRPP cluster in each network based on their resistance to change after perturbation. Calculations were done across four community detection algorithms for each distance metric (**Table 3.3**). This confirmed my observation that graph clustering using sequence similarity metrics generated a clearly identifiable, stable cluster of PRPP riboswitches that was distinct from the ppGpp and guanidine-I riboswitches in the set.

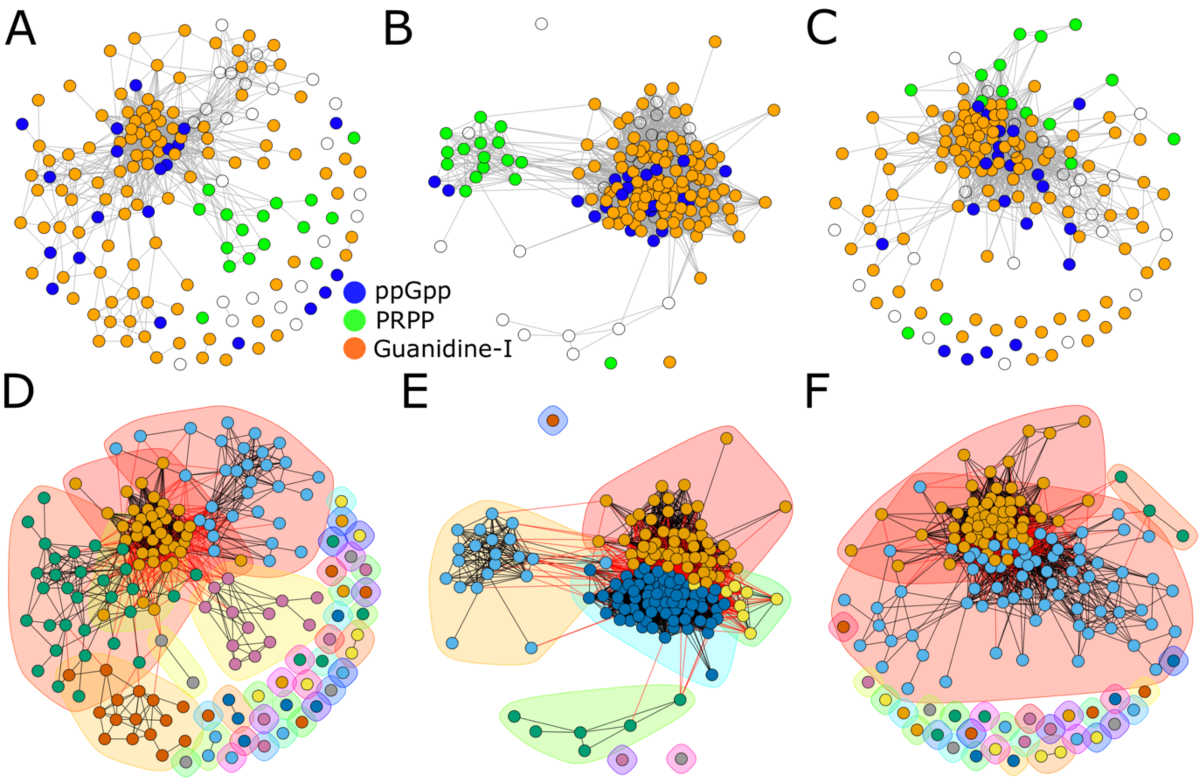

Figure 3.1: ***Clustering ykkC riboswitches gathered from data predating subtyping.*** ppGpp, PRPP, and Guanidine-I riboswitches were clustered using (A) RNAmountAlign with a threshold of 0, (B) muscle with a threshold of 63, and (C) RNApdist with a threshold of 30. Sub-clusters (communities) were identified within the (D) RNAmountAlign, (E) muscle, and (F) RNApdist networks as described previously. After 100 rounds of bootstrapping, the Jaccard similarity index was calculated to represent confidence in the PRPP cluster identification. For RNAmountAlign, muscle, and RNApdist, these values came to 0.584, 0.998, and 0.296, respectively.

3.4.2 ppGpp, PRPP, and ADP-CDP separate when clustered using a metric that combines sequence and structural similarity

A majority of the initially identified ykkC riboswitches were found to be guanidine-sensing and the remainder were classified as a rarer variant, dubbed subtype 2[62]. We now know that subtype 2 is actually comprised of four distinct riboswitches[37,50]. I chose to investigate the three subtype 2 riboswitches that have been characterized in order to gain a better understanding of their relationship and the ability of graph clustering to distinguish these highly homologous, rarer subtypes. To do this, I used published datasets for the ykkC subtype 2A (ppGpp), 2B (PRPP), and 2C (ADP-CDP) riboswitches (**Table 3.4, 3.5, 3.6**). I began by combining these datasets and calculating pairwise similarity using RNAmountAlign. Separation was observed between the three riboswitches, with the ADP-CDP instances standing apart as the most distinct cluster and the ppGpp cluster and the PRPP cluster showing much greater connectivity between each other (**Figure 3.2A**). Despite the greater connectivity of the ppGpp and PRPP clusters, *de novo* community detection methods are able to distinguish each type as distinct groups **(Figure 3.2B**). This higher connectivity can similarly be observed in the network density graphs, which depicts the edge-density within and between clusters as edges are removed through thresholding across the range of pairwise distance scores (**Figure 3.2C**). While ADP-CDP shows highest conservation within its cluster, it shows the lowest edge connectivity to the other clusters. ppGpp and PRPP show more variation within their clusters and higher edge connectivity between them, compared to ADP-CDP. Although ppGpp and PRPP show higher connectivity within the network, they can still be distinguished using *de novo* community. This begs the question: is divergence of these riboswitches occurring on a sequence or structural level or both.

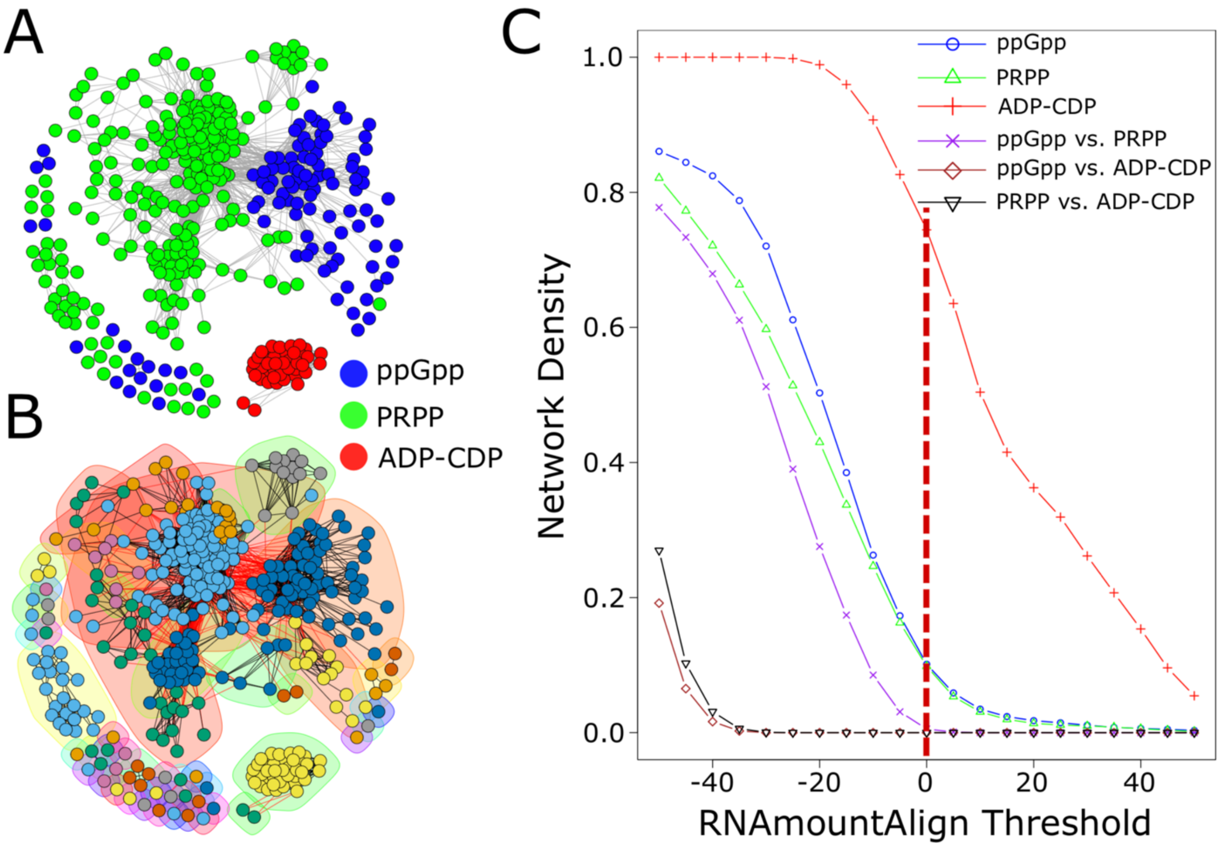

Figure 3.2: ***Clustering ykkC subtype 2A-C riboswitches using RNAmountAlign.*** A) ppGpp, PRPP, and ADP-CDP riboswitches were clustered using RNAmountAlign as a distance metric (threshold of 0) for comparing pairwise sequence and ensemble structure similarity. Aptamers are colored based on aptamer type, blue for ppGpp, green for PRPP, and red for ADP-CDP. B) Sub-clusters (communities) were identified using cluster_fast_greedy, a *de novo* community detection function within R's igraph package. Node colors correspond to distinct clusters detected. C) Network density was calculated for each riboswitch across a range of RNAmountAlign thresholds. Dotted red line indicates the RNAmountAlign threshold (0) at which the network is visualized.

3.4.3 ppGpp, PRPP, and ADP-CDP separate when clustered using sequence similarity as a metric

To address whether sequence divergence between the riboswitches was sufficient to distinguish ppGpp and PRPP as distinct clusters, I performed graph clustering using a metric relying on raw sequence similarity. I used the pairwise identity scores from muscle multiple sequence alignment as a similarity metric. This clustering, similar to RNAmountAlign, showed strong separation of ADP-CDP from both ppGpp and PRPP, while the ppGpp and PRPP clusters display higher connectivity (**Figure 3.3A**). Despite the closer connectivity of ppGpp and PRPP, the two riboswitches are clearly represented within distinct hemispheres of the network and were able to be distinguished using *de novo* community detection methods (**Figure 3.3B**). The network density graphs also showed a similar trend to RNAmountAlign; ADP-CDP is the most highly conserved and distinct from ppGpp and PRPP, while ppGpp and PRPP are less highly conserved individually, but more closely connected (**Figure 3.3C**). These results are consistent with what was observed using RNAmountAlign, demonstrating that ppGpp and PRPP riboswitch clusters can be distinguished using sequence identity for our graph clustering analysis and community detection.

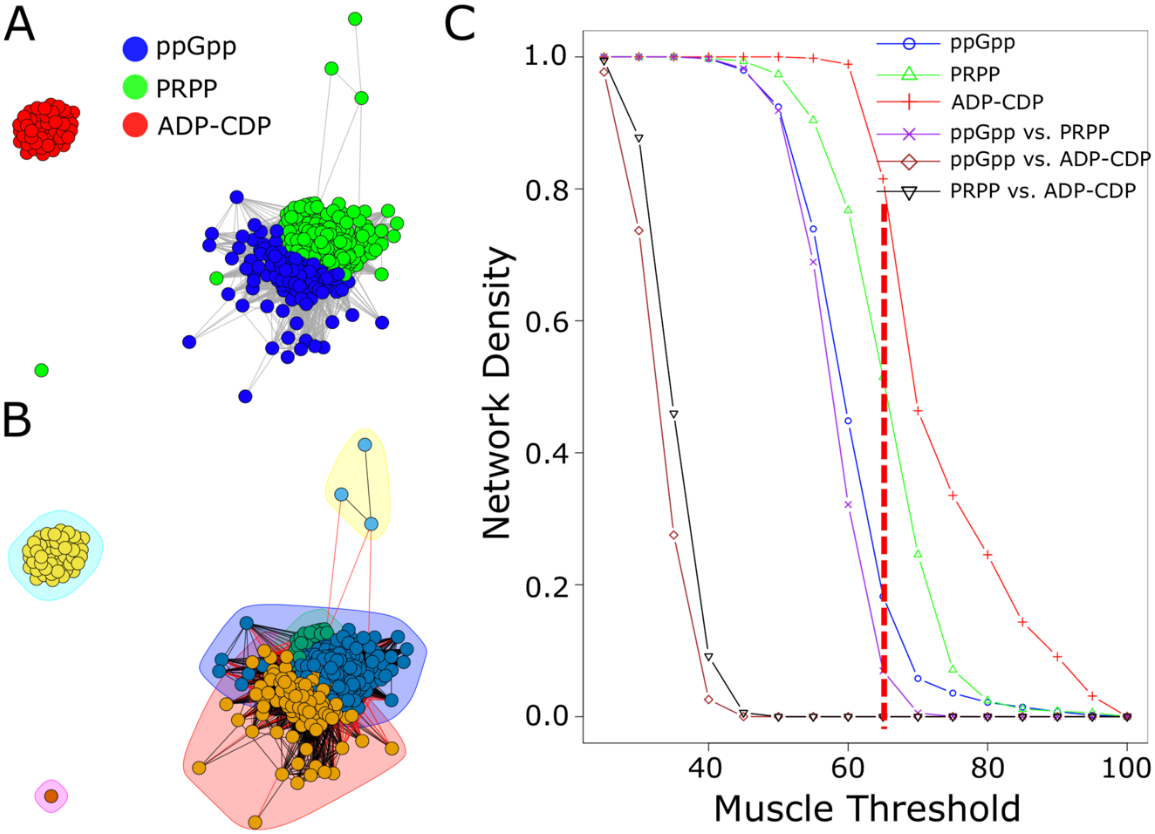

Figure 3.3: ***Clustering ykkC subtype 2A-C riboswitches using muscle.***
A) ppGpp, PRPP, and ADP-CDP riboswitches were clustered using muscle as a distance metric (threshold of 65) for comparing pairwise sequence similarity. Aptamers are colored based on aptamer type, blue for ppGpp, green for PRPP, and red for ADP-CDP. B) Sub-clusters (communities) were identified using cluster_fast_greedy, a *de novo* community detection function within R's igraph package. Node colors correspond to distinct clusters detected. C) Network density was calculated for each riboswitch across a range of RNAmountAlign thresholds. Dotted red line indicates the RNAmountAlign threshold (65) at which the network is visualized.

3.4.4 ppGpp and PRPP are indistinguishable when clustered using ensemble structure similarity as a metric

Next, I evaluated clustering patterns based on ensemble structure similarity. RNApdist was used as the graph clustering metric to quantify the pairwise ensemble structure similarity within the set of riboswitches. When this was done, I observed a different result than when using metrics which incorporate sequence similarity. While the ADP-CDP cluster still remains the most highly conserved, the difference was not as stark. Moreover, the ppGpp and PRPP nodes show high amounts of overlap when looking at the visualization of the graph clustering (**Figure 3.4A**). The inability to distinguish ppGpp and PRPP nodes is clearly seen when implementing *de novo* community detections methods, as identified clusters do not correlate with ppGpp or PRPP nodes within the network (**Figure 3.4B**). The network density graph highlights this as well, showing that the conservation within the PRPP cluster is indistinguishable from the conservation between PRPP and ppGpp (**Figure 3.4C**). These findings indicate that ppGpp and PRPP have highly conserved overall structural similarity, and it is the sequence variation that primarily distinguishes them. This supports observations from the dataset curated using a model predating ykkC subtyping, which similarly indicated that it is primarily sequence variation that separates the PRPP from ppGpp riboswitches. Taken together, this may suggest that PRPP and ppGpp share a closer evolutionary relationship than either with ADP-CDP.

These findings highlight the power of graph clustering to distinguish between these closely related riboswitches and investigate the relationships between motifs with homologous sequences and structures. This led us to ask about localized differences within the riboswitches, specifically within the P3 stem-loop, which houses the main components for ligand binding.

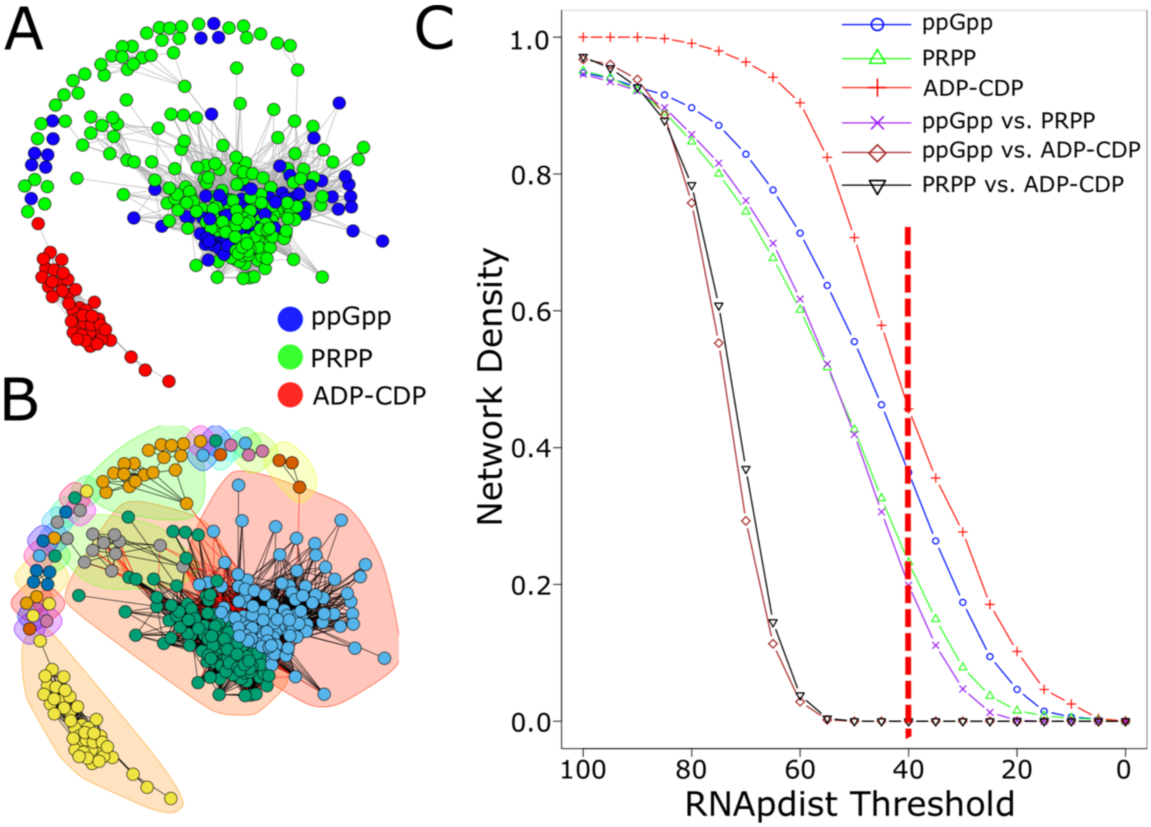

Figure 3.4: *Clustering ykkC subtype 2A-C riboswitches using RNApdist.* A) ppGpp, PRPP, and ADP-CDP riboswitches were clustered using RNAmountAlign as a distance metric (threshold of 40) for comparing pairwise ensemble structure similarity. Aptamers are colored based on aptamer type, blue for ppGpp, green for PRPP, and red for ADP-CDP. B) Sub-clusters (communities) were identified using cluster_fast_greedy, a *de novo* community detection function within R's igraph package. Node colors correspond to distinct clusters detected. C) Network density was calculated for each riboswitch across a range of RNAmountAlign thresholds. Dotted red line indicates the RNAmountAlign threshold (40) at which the network is visualized.

3.4.5 P3 stem-loop shows structural conservation between PRPP and ADP-CDP

To investigate the P3 stem-loop (which contains the major components for ligand binding) I aligned each riboswitch back to the covariance model for its type and trimmed away all of the sequence that was not contained within the third stem-loop of the secondary structure. I then performed a graph clustering analysis on these truncated sequences. Clustering based on sequence similarity of the P3 stem-loop showed similar results to clustering the whole riboswitch, with greater connectivity between the ppGpp and PRPP clusters and ADP-CDP more separated. However, the metrics that incorporated structure into the analysis cluster PRPP and ADP-CDP very closely and leaves ppGpp as farther removed (**Figure 3.5A-C**). This diverges from the analysis of the entire riboswitch, which found the ppGpp and PRPP clusters indistinguishable when using ensemble structure as the similarity metric. These findings are supported by *de novo* community detection (**Figure 3.5D-F**).

This raises the question of why higher overall conservation between ppGpp and PRPP is observed, but the P3 stem-loop shows higher structural conservation between PRPP and ADP-CDP. The specific ligands each riboswitch binds could be one possible explanation. PRPP and ADP-CDP each have a more isolated diphosphate group, compared to ppGpp's bulkier structure. This could lead to more highly similar binding pocket structure for PRPP and ADP-CDP. This is supported by the fact that the similarities between the P3 stem of PRPP and ADP-CDP led researchers to investigate ligands for the ADP-CDP riboswitch that were similar to PRPP[37].

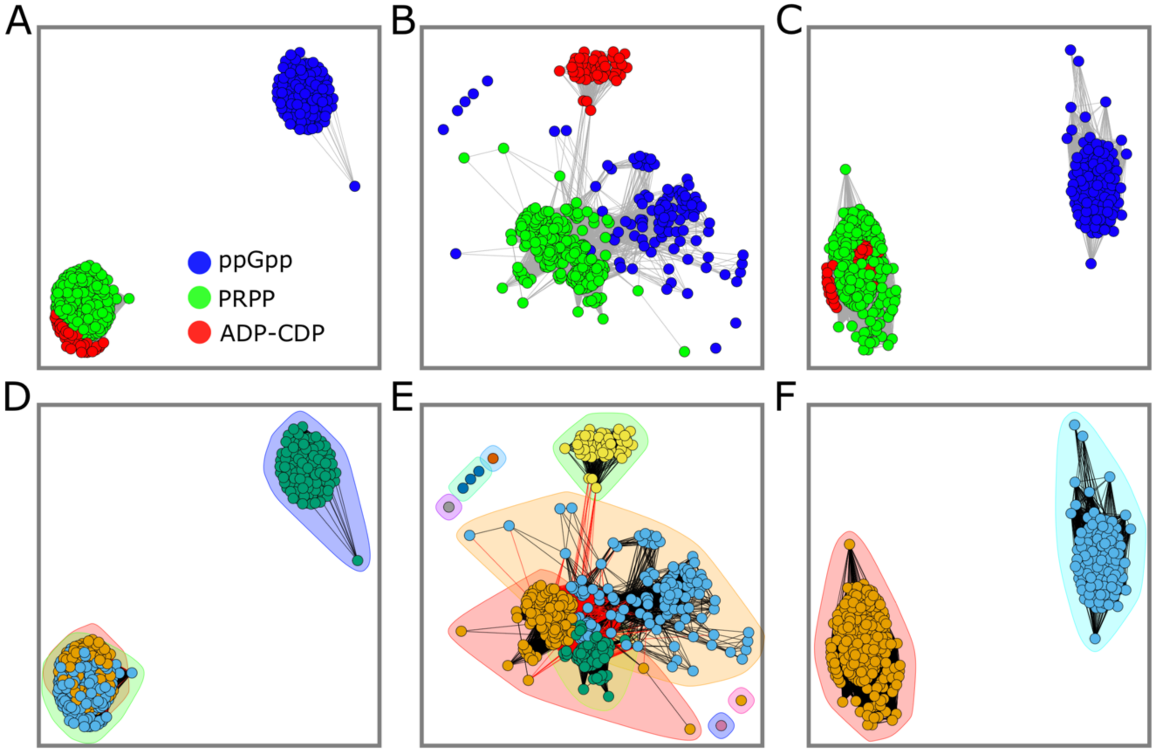

Figure 3.5: ***Clustering P3 stem of ykkC subtype 2A-C riboswitches.***
ppGpp, PRPP, and ADP-CDP riboswitches were clustered using (A) RNAmountAlign with a threshold of 0, (B) muscle with a threshold of 80, and (C) RNApdist with a threshold of 7. Sub-clusters (communities) were identified within the (D) RNAmountAlign, (E) muscle, and (F) RNApdist networks as described previously.

3.5 Discussion

Graph clustering has been demonstrated to be useful in better understanding the relationships between, and evolution of, ncRNA. It shows promise as a predictive measure for distinguishing closely related riboswitches which are functionally distinct. My studies have found that when performing graph clustering analysis, distance metric selection is crucial, as different metrics can reveal distinct clustering patterns. While this may initially appear to be a drawback, it is actually a useful feature of graph clustering, as implementation of multiple distance metrics when clustering can reveal the impact of different factors affecting ncRNA clustering, allowing for a multi-dimensional evaluation.

Moreover, while many classification methods rely on a sequence's similarity to a model, graph clustering provides a methodology for investigating pairwise similarity of ncRNA relative to other members of the defined classes. This allows researchers to identify clustering and patterns within the data based on defined distance metrics relevant to the questions they are asking, thus partitioning and parsing out subclusters within the dataset relative to specific parameters. I have previously shown that this methodology can be implemented to investigate evolutionary and functional relationships between variant riboswitches of the same class (glycine riboswitch). Now I highlight the ability of graph clustering to separate distinct riboswitches that were originally identified as members of the same class (ykkC subtypes).

In this investigation I also demonstrate the ability of graph clustering to evaluate specific components of a ncRNA. Using data reflective of what would be seen predating the ykkC subtype classifications, I was able to determine that PRPP could be distinguished when clustering based on sequence similarity. When applying graph clustering to three published datasets corresponding to ppGpp, PRPP, and ADP-CDP, I saw similar findings. The ppGpp and PRPP riboswitches clustered closely together based on structural metrics, but were able to be distinguished using sequence similarity metrics. It was also found that ADP-CDP is more highly conserved than either ppGpp or PRPP, and there was very low connectivity between ADP-CDP and either other riboswitch class. This suggests that the ppGpp and PRPP riboswitches may share a closer evolutionary relationship than either does to ADP-CDP. A closer investigation into the P3 stem-loop of these three riboswitches display a different clustering pattern than the overall riboswitch. The clustering pattern of the P3 stem-loop shows higher structural similarity between PRPP and ADP-CDP. While this pattern is different than that seen with the overall riboswitches, it is potentially due to the similarity of the PRPP and ADP-CDP ligands placing structural constraints on the binding pockets that are distinct form the ppGpp riboswitch. These findings together demonstrate the ability of graph clustering to investigate diversity within and between riboswitch classes.

Chapter 4: Discussion

In recent decades, the study of structured RNA has evolved with the advent of improved computational and experimental techniques. The characterization of over 40 riboswitch classes has deepened the understanding of riboswitches from both a phylogenetic and functional perspective[9]. The ability to identify *de novo* structured RNA motifs from sequence data, generate models for RNA family classification, and perform ligand-binding assays to identify the binding partner of riboswitches has driven these breakthroughs[18,90,141,142]. However, while these methods are helpful for classifying sets of homologous riboswitches, they simplify the variation within those datasets down to a binary "member or nonmember" relationship with broad models. To better understand riboswitch diversity and functionality, we need to study riboswitches based on the pairwise relationships within defined classes. All homologous riboswitches are not necessarily functionally the same or similar.

Graph clustering provides a framework for this type of pairwise investigation. Using graph clustering, investigators can evaluate the similarity of riboswitch instances by generating networks comprised of nodes correlating to each instance. Within the network, each node acts as a charged particle, repelling other nodes, and is connected to each other node by an edge weighted by a pairwise similarity score, which acts as a spring. By removing weaker edges, the topology of the network changes to create clusters of highly related nodes which retained their edges.

In this work I have shown the effectiveness of graph clustering in studying the relationships between, and evolution of, ncRNA. Using graph clustering, I evaluated the relationship between the homologous aptamers of the tandem glycine riboswitch, as well as the correlation between singleton glycine riboswitch structure and tandem glycine riboswitch aptamers regulating similar genes. I identified clustering patterns within the glycine riboswitch aptamers corresponding to whether the riboswitch regulates GCV or TP. This was taken further when clusters containing both singleton and tandem glycine riboswitch aptamers reveled that singleton aptamers are more highly related to a specific tandem aptamer based on the genomic context. This work culminated in the development of a model for the evolutionary path of the glycine riboswitch based on the genomic context. The model dictates that tandem riboswitches degrade into functional singleton riboswitches with a ghost aptamer which retains structures relevant to tertiary interaction. These findings fit recent studies showing evidence for glycine riboswitches that display On- or Off-switch functionality based on the regulatory context; whether the riboswitch is regulating GCV or TP[31,54,58,59].I extended this idea to investigate the ability of graph clustering to distinguish closely related riboswitches and demonstrate its viability for *de novo* identification of functionally distinct variants within defined riboswitch classes. I used the ykkC riboswitches for this demonstration, as they represent a set of highly homologous and functionally distinct riboswitches. The ability of graph clustering to distinguish highly similar riboswitches with functional distinctions is highlighted by the identification of a cluster of PRPP riboswitches from a dataset representative of sequences predating the classification of ykkC subtypes. The ability of graph clustering to identify functional differences was further demonstrated using three published datasets corresponding to the ppGpp, PRPP, and ADP-CDP riboswitches. The clustering patters of these three datasets showed that sequence variation rather than structural changes that distinguishes the ppGpp and PRPP riboswitches. This finding fits with published data which identifies key nucleotides difference within the binding pocket that dictate difference in ligand binding. This finding was further supported by investigation into the P3 stem of these riboswitches, which showed structural similarity between the PRPP and ADP-CDP riboswitches, correlating with higher similarity in their bound ligand. My findings demonstrate the ability of graph clustering to distinguish functional differences in groups of homologous riboswitches. Moreover, my research highlights the importance of distance metric selection when performing graph clustering analysis, as different metrics can reveal distinct clustering patterns. For this reason, the implementation of multiple distance metrics when performing a graph clustering analysis is encouraged to reveal the impact of different factors impacting clustering, allowing for a more thorough investigation.

These projects demonstrate graph clustering to be a powerful tool when applied to investigating motifs that exhibited different behavior in distinct contexts. This investigation could be generalized to apply to a host of other investigations regarding variation within pre-defined sets, outside of the ncRNA field. For instance, a researcher could define a set of genes representative of a cellular pathway and compare transcription profiles under different conditions to identify which stimuli elicit similar responses. This would require the researcher to utilize a vector distance metric when graph clustering, and provide vectors representative of expression levels for the set of genes they were interested in. Graph clustering has already been applied to a wide range of fields outside of biology, and there is a lot of room for application within biological research.

Chapter 4: Discussion

A major benefit of graph clustering is the applicability to any network that can be broken down to pairwise connections, and particularly the ease of adding distance metrics to investigation a range of variable that could affect variation within the dataset. Moreover, the ability to threshold the network across a range of distance scores allows researchers to see how the network topology and relationships within the data change at a scale that can be tailored to the question being asked. For instance, when comparing homologous protein sequences across a range of bacterial phyla, a lower thresholding score may be more appropriate than when comparing homologous protein sequences within one bacterial family. The ability to visualize the networks across a range of thresholds allows researchers to investigate trends within the data that could otherwise be missed.

However, thresholding also creates the largest pitfall of graph-clustering: over-clustering your data. Without an appropriate knowledge of the datasets, a researcher may over-threshold their networks and disperse relevant clusters. For this reason, when performing clustering on datasets with little *a priori* knowledge of the context, it is likely best to take one of two actions when identifying clusters: 1) select a thresholding score that generates 2-3 major clusters when performing the *de novo* community detection, or 2) choose a community detection method that will give you the best bifurcation of your data (such as non-recursive spectral clustering). This will provide the clusters which have the strongest variation within your dataset. A researcher can then investigate causes for the distinction between clusters (based on phylogenetic knowledge and the application of multiple distance metrics) and determine whether clustering has biological relevance and if more clustering need to be done. For this reason, in these instances it is recommended to perform a “top-down” clustering methodology in which multiple distance metrics are used and broad clusters are identified first and evaluated for relevance.

Taken together, my works demonstrate the ability of graph clustering to probe the underappreciated diversity within riboswitch classes and provide a methodology for identifying members of a riboswitch class which may be divergent from the other members. As I, and others, have demonstrated, these divergent riboswitch sub-clusters could be slight variants of the parent class[57,68] or have completely distinct functionality distinct from the family of original classification[37,50,61,62]. Understanding the diversity and nuance of riboswitch structure and functionality has increased drastically within the last decade, from identifying structurally distinct aptamers that bind the same ligand to distinguishing riboswitches with clear structural similarity that bind distinct ligands. Graph clustering provides a methodology for investigating the relationship between sets of structured ncRNA across a range of metrics at a pairwise level in order to reveal nuanced patterns that other tools fail to distinguish. This approach could easily be extended to incorporate additional metrics based on a researcher’s needs. Similarly, with the availability of proper distance metrics, graph clustering could be adapted to the investigation of different biological components, from protein structure to single-cell transcriptional profiles.

References:

1. Hoeppner MP, Gardner PP, Poole AM. Comparative Analysis of RNA Families Reveals Distinct Repertoires for Each Domain of Life. PLoS Comput Biol. 2012;8(11):1–8.
2. Gilbert W. The RNA world. Vol. 319, Nature. 1986. p. 618.
3. Breaker RR, Gilmore JH. The Expanding View of RNA and DNA Function Ronald. ACS Chem Biol. 2014;21(9):1059–65.
4. Kaczanowska M, Rydén-Aulin M. Ribosome Biogenesis and the Translation Process in Escherichia coli . Microbiol Mol Biol Rev. 2007;71(3):477–94.
5. Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, et al. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. Proc Natl Acad Sci. 2004;101(17):6421–6.
6. McCown PJ, Winkler WC, Breaker RR. Mechanism and Distribution of glmS Ribozymes. Methods Mol Biol. 2012;(848):113–29.
7. Adams a, Lindahl T, Fresco JR. Conformational differences between the biologically active and inactive forms of a transfer ribonucleic acid. Proc Natl Acad Sci U S A. 1967;57(6):1684–91.
8. Hammann C, Luptak A, Perreault J, De La Peña M. The ubiquitous hammerhead ribozyme. Rna. 2012;18(5):871–85.
9. McCown PJ, Corbino KA, Stav S, Sherlock ME, Breaker RR. Riboswitch diversity and distribution. RNA. 2017;23(7):995–1011.
10. Barrick JE, Breaker RR. The distributions, mechanisms, and structures of

metabolite-binding riboswitches. Genome Biol. 2007;8(11).

11. Pipas JM, McMahon JE. Method for predicting RNA secondary structure. Proc Natl Acad Sci. 1975;72(6):2017–21.

12. McCaskill JS. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. Biopolymers. 1990;29(6–7):1105–19.

13. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic paramenters improves prediction of RNA secondary structure. J Mol Biol. 1999;288:911–40.

14. Hajiaghayi M, Condon A, Hoos HH. Analysis of energy-based algorithms for RNA secondary structure prediction. BMC Bioinformatics. 2012;13(1).

15. Nussinov R, Jacobson AB. Fast algorithm for predicting the secondary structure of single-stranded RNA. PNAS. 1980;77(11):6309–13.

16. Eddy SR, Durbin R. RNA sequence analysis using covariance models. Nucleic Acids Res. 1994;22(11):2079–88.

17. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, et al. Rfam 12.0: Updates to the RNA families database. Nucleic Acids Res. 2014;43(D1):D130–7.

18. Barquist L, Burge SW, Gardner PP. Studying RNA homology and conservation with infernal: From single sequences to RNA families. Curr Protoc Bioinforma. 2016;2016:12.13.1-12.13.25.

19. Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case D a, Sampath R. RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Res.

2001;29(22):4724–35.

20. Roth A, Breaker RR. The Structural and Functional Diversity of Metabolite-binding Riboswitches. Annu Rev Biochem. 2009;78:305–34.

21. Garst AD, Edwards AL, Batey RT. Riboswitches: Structures and mechanisms. Cold Spring Harb Perspect Biol. 2011;3(6):1–13.

22. Harvey I, Garneau P, Pelletier J. Inhibition of translation by RNA-small molecule interactions. Rna. 2002;8(4):452–63.

23. Bastet L, Dubé A, Massé E, Lafontaine DA. New insights into riboswitch regulation mechanisms. Mol Microbiol. 2011;80(5):1148–54.

24. Li S, Breaker RR. Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. Nucleic Acids Res. 2013;41(5):3022–31.

25. Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, et al. Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. Cell. 2002;111(5):747–56.

26. Nudler E, Mironov AS. The riboswitch control of bacterial metabolism. Trends Biochem Sci. 2004;29(1):11–7.

27. Winkler WC, Breaker RR. Genetic control by metabolite-binding riboswitches. ChemBioChem. 2003;4(10):1024–32.

28. Sherlock ME, Sudarsan N, Stav S, Breaker RR. Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. Elife. 2018;

29. Ruff KM, Strobel SA. Ligand binding by the tandem glycine riboswitch depends on aptamer dimerization but not double ligand occupancy. Rna.

2014;20(11):1775–88.

30. Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR. An mRNA structure in bacteria that controls gene expression by binding lysine. Genes Dev. 2003;17(21):2688–97.

31. Mandal M, Lee M, Barrick JE, Weinberg Z, Mitchell G, Science S, et al. A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression Emilsson , Walter L . Ruzzo and Ronald R . Breaker. Science (80-). 2004;306:275–9.

32. Ames TD, Breaker RR. Bacterial aptamers that selectively bind glutamine. RNA Biol. 2011;8(1):82–9.

33. Batey RT, Gilbert SD, Montange RK. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. Nature. 2004;432:411–5.

34. Roth A, Winkler WC, Regulski EE, Lee BWK, Lim J, Jona I, et al. A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. Nat Struct Mol Biol. 2007;14(4):308–17.

35. Meyer MM, Roth A, Chervin SM, Garcia GA, Breaker RR. Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria. RNA. 2008;14(4):685–95.

36. McCown PJ, Liang JJ, Weinberg Z, Breaker RR. Structural, Functional, and Taxonomic Diversity of Three PreQ1 Riboswitch Classes. Chem Biol. 2014;21(7):880–9.

37. Sherlock ME, Sadeeshkumar H, Breaker RR. Variant Bacterial Riboswitches

Associated with Nucleotide Hydrolase Genes Sense Nucleoside Diphosphates. Biochemistry. 2019;58(5):401–10.

38. Green NJ, Grundy FJ, Henkin TM. The T box mechanism: tRNA as a regulatory molecule. FEBS Lett. 2010;584(2):318–24.

39. Zhang J, Ferré-D'Amaré AR. Structure and mechanism of the T-box riboswitches. WIREs RNA. 2015;6(4):419–33.

40. Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, et al. Riboswitches in eubacteria sense the second messenger cyclic Di-GMP. Science (80-). 2008;321(5887):411–3.

41. Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR. An Allosteric Self-Splicing Ribozyme Triggered by a Bacterial Second Messenger. Science (80-). 2010;329(5993):845–8.

42. Nelson JW, Sudarsan N, Furukawa K, Weinberg Z, Wang JX, Breaker RR. Riboswitches in eubacteria sense the second messenger c-di-AMP. Nat Chem Biol. 2013;9(12):834–9.

43. Nelson JW, Sudarsan N, Phillips GE, Stav S, Lünse CE, McCown PJ, et al. Control of bacterial exoelectrogenesis by c-AMP-GMP. Proc Natl Acad Sci U S A. 2015;112(17):5389–94.

44. Kim P, Nelson J. An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. Mol Cell. 2015;57(2):317–28.

45. Sundarsan N, Hammond MC, Block KF, Welz R, Barrick JE, Roth A, et al. Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions.

2006;314(October).

46. Welz R, Breaker RR. Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. Rna. 2007;13(4):573–82.

47. Rodionov DA, Dubchak I, Arkin A, Alm E, Gelfand MS. Reconstruction of regulatory and metabolic pathways in metal-reducing δ-proteobacteria. 2004;5(11):1–27.

48. Poiata E, Meyer MM, Ames TD, Breaker RR. A variant riboswitch aptamer class for S-adenosylmethionine common in marine bacteria. Rna. 2009;15(11):2046–56.

49. Nahvi A, Barrick JE, Breaker RR. Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. Nucleic Acids Res. 2004;32(1):143–50.

50. Sherlock ME, Sudarsan N, Stav S, Breaker RR. Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. Elife. 2018;7(subtype 1):1–17.

51. Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, et al. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome Biol. 2010;11(3).

52. Ren A, Xue Y, Peselis A, Serganov A, Al-Hashimi H, Patel D. Structural and Dynamic Basis for Low Affinity-High Selectivity Binding of L-glutamine by the Glutamine Riboswitch Aiming. Cell Rep. 2015;13(9):1800–13.

53. Sherman EM, Esquiaqui J, Elsayed G, Ye J-D. An energetically beneficial leader-linker interaction abolishes ligand-binding cooperativity in glycine riboswitches. Rna. 2012;18(3):496–507.

54. Khani A, Popp N, Kreikemeyer B, Patenge N. A Glycine Riboswitch in Streptococcus pyogenes Controls Expression of a Sodium:Alanine Symporter Family Protein Gene. Front Microbiol. 2018;9(February):1–10.

55. Erion T V., Strobel SA. Identification of a tertiary interaction important for cooperative ligand binding by the glycine riboswitch. Rna. 2011;17(1):74–84.

56. Ruff KM, Muhammad A, McCown PJ, Breaker RR, Strobel SA. Singlet glycine riboswitches bind ligand as well as tandem riboswitches. RNA. 2016;22(11):1728–38.

57. Crum M, Ram-Mohan N, Meyer MM. Regulatory context drives conservation of glycine riboswitch aptamers. PLoS Comput Biol. 2019;15(12):1–24.

58. Torgerson CD, Hiller DA, Stav S, Strobel SA. Gene regulation by a glycine riboswitch singlet uses a finely tuned energetic landscape for helical switching. Rna. 2019;24(12):1813–27.

59. Babina AM, Lea NE, Meyer MM. In vivo behavior of the tandem glycine riboswitch in Bacillus subtilis. MBio. 2017;8(5):1–16.

60. Torgerson CD, Hiller DA, Strobel SA. The asymmetry and cooperativity of tandem glycine riboswitch aptamers. Rna. 2020;26(5):564–80.

61. Sherlock ME, Sudarsan N, Breaker RR. Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. Proc Natl Acad Sci. 2018;115(23):6052–7.

62. Nelson JW, Atilho RM, Sherlock ME, Stockbridge RB, Breaker RR. Metabolism of Free Guanidine in Bacteria is Regulated by a Widespread Riboswitch Class.

2017;65(2):220–30.

63. Mirihana Arachchilage G, Sherlock ME, Weinberg Z, Breaker RR. SAM-VI RNAs selectively bind S-adenosylmethionine and exhibit similarities to SAM-III riboswitches. RNA Biol. 2018;15(3):371–8.

64. Price IR, Grigg JC, Ke A. Common themes and differences in SAM recognition among SAM riboswitches. Biochim Biophys Acta - Gene Regul Mech. 2014;1839(10):931–8.

65. Sun A, Gasser C, Li F, Chen H, Mair S, Krasheninina O, et al. SAM-VI riboswitch structure and signature for ligand discrimination. Nat Commun. 2019;10(1):1–13.

66. Weickhmann AK, Keller H, Wurm JP, Strebitzer E, Juen MA, Kremser J, et al. The structure of the SAM/SAH-binding riboswitch. Nucleic Acids Res. 2019;47(5):2654–65.

67. Battaglia RA, Price IR, Ke A. Structural basis for guanidine sensing by the ykkC family of riboswitches. Rna. 2017;23(4):578–85.

68. Weinberg Z, Nelson JW, Lünse CE, Sherlock ME, Breaker RR. Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. Proc Natl Acad Sci. 2017;114(11):E2077–85.

69. Nussinoc R, Pieczenik G, Griggs J, Kleitman D. Algorithms for Loop Matchings. SIAM J Appl Math. 1978;35(1):68–82.

70. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. 1981;9(1):133–48.

71. Freier SM, Kierzek R, Jaeger JA, Sugimoto N, Caruthers MH, Neilson T, et al.

Improved free-energy parameters for predictions of RNA duplex stability. Proc Natl Acad Sci U S A. 1986;83(24):9373–7.

72. Zuker M, Sankoff D. RNA secondary structures and their prediction. Bull Math Biol. 1984;46(4):591–621.

73. Ding Y, Chi YC, Lawrence CE. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. Rna. 2005;11(8):1157–66.

74. Wang L, Jiang T. On the complexity of multiple sequence alignment. J Comput Biol. 1994;1(4).

75. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7(539).

76. Edgar RC. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:1–19.

77. Mathews DH, Moss WN, Turner DH. Folding and finding RNA secondary structure. Cold Spring Harb Perspect Biol. 2010;2(12):1–15.

78. Gardner PP, Giegerich R. A comprehensive comparison of comparative RNA structure prediction approaches. BMC Bioinformatics. 2004;5:1–18.

79. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: Improved consensus structure prediction for RNA alignments. BMC Bioinformatics. 2008;9:1–13.

80. Höchsmann M, Voss B, Giegerich R. Pure multiple RNA secondary structure alignments: A progressive profile approach. IEEE/ACM Trans Comput Biol

Bioinforma. 2004;1(1):53–62.

81. Höchsmann M, Töller T, Giegerich R, Kurtz S. Local similarity in RNA secondary structures. Proc Comput Syst. 2003;159–68.

82. Sundfeld D, Havgaard JH, De Melo ACMA, Gorodkin J. Foldalign 2.5: Multithreaded implementation for pairwise structural RNA alignment. Bioinformatics. 2016;32(8):1238–40.

83. Mathews DH, Turner DH. Dynalign: An algorithm for finding the secondary structure common to two RNA sequences. J Mol Biol. 2002;317(2):191–203.

84. Will S, Joshi T, Hofacker IL, Stadler PF, Backofen R. LocARNA-P: Accurate Boundary Prediction and Improved Detection of Structured RNAs for Genome-wide Screens. Rna. 2012;18:900–14.

85. Bayegan AH, Clote P. RNAmountAlign: Efficient software for local, global, semiglobal pairwise and multiple RNA sequence/structure alignment [Internet]. Vol. 15, PLoS ONE. 2020. 1–34 p.

86. Yoon B-J. Hidden Markov Models and their Applications in Biological Sequence Analysis. Curr Genomics. 2009;10(6):402–15.

87. Singh P, Bandyopadhyay P, Bhattacharya S, Krishnamachari A, Sengupta S. Riboswitch detection using Profile Hidden Markov Models. BMC Bioinformatics. 2009;10:1–13.

88. Dowell RD, Eddy SR. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. BMC Bioinformatics. 2004;5:1–14.

89. Eddy S, Durbin R. RNA sequence analysis using covariance models. Nucleic Acids Res. 1994;22(11):2079–88.

90. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29(22):2933–5.

91. Rivas E, Clements J, Eddy SR. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. Nat Methods. 2017;14(1):45–8.

92. Rivas E, Clements J, Eddy SR. Estimating the power of sequence covariation for detecting conserved RNA structure. Bioinformatics. 2020;36(10):3072–6.

93. Weinberg Z, Breaker RR. R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. BMC Bioinformatics. 2011;12(1):3.

94. Charleston MiA, Hendy MD, Penny D. The Effects of Sequence Length, Tree Topology, and Number of Taxa on the Performance of Phylogenetic Methods. J Comput Biol. 2009;1(2):133–51.

95. Mukherjee S, Barash D, Sengupta S. Comparative genomics and phylogenomic analyses of lysine riboswitch distributions in bacteria. PLoS One. 2017;12(9):1–24.

96. Singh P, Sengupta S. Phylogenetic analysis and comparative genomics of purine riboswitch distribution in prokaryotes. Evol Bioinforma. 2012;2012(8):589–609.

97. Emmons S, Kobourov S, Gallant M, Börner K. Analysis of network clustering algorithms and cluster quality metrics at scale. PLoS One. 2016;11(7):1–18.

98. Girvan M, Newman MEJ. Community structure in social and biological networks. Proc Natl Acad Sci U S A. 2002;99(12):7821–6.

99. Schaeffer SE. Graph clustering. Comput Sci Rev. 2007;1(1):27–64.

100. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95(25):14863–8.

101. Datta S, Datta S. Evaluation of clustering algorithms for gene expression data. BMC Bioinformatics. 2006;7(SUPPL.4):1–9.

102. Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput. 2000;426:418–29.

103. Lu Z, Wahlström J, Nehorai A. Community Detection in Complex Networks via Clique Conductance. Sci Rep. 2018;8(1):1–17.

104. Lorenz R, Bernhart SH, Siederdissen CH zu, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011;1–14.

105. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M. Fast Folding and Comparison of RNA Secondary Structures. 1994;188:1–22.

106. Madeira F, Park Y mi, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019;47(W1):W636–41.

107. Harmanci AO, Sharma G, Mathews DH. Efficient pairwise RNA structure prediction using probabilistic alignment constraints in Dynalign. BMC Bioinformatics. 2007;8:1–21.

108. Reuter J, Mathews D. RNAstructure: software for RNA secondary structure

prediction and analysis. BMC Bioinformatics. 2010;11.

109. Bayegan AH, Clote P. RNAmountAlign: efficient software for local, global, semiglobal pairwise and multiple RNA sequence/structure alignment. bioRxiv. 2018;(Md):1–22.

110. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal. 2006;Complex Sy:1695.

111. Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks. Phys Rev E - Stat Physics, Plasmas, Fluids, Relat Interdiscip Top. 2004;70(6):6.

112. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys Rev E - Stat Nonlinear, Soft Matter Phys. 2004;69(2 2):1–15.

113. Pons P, Latapy M. Computing communities in large networks using random walks. J Graph Algorithms Appl. 2006;10(2):191–218.

114. Newman MEJ. Finding community structure in networks using the eigenvectors of matrices. Phys Rev E - Stat Nonlinear, Soft Matter Phys. 2006;74(3):1–19.

115. Kwon M, Strobel SA. Chemical basis of glycine riboswitch cooperativity. Rna. 2008;14(1):25–34.

116. Butler EB, Xiong Y, Wang J, Strobel SA. Structural Basis of Cooperative Ligand Binding by the Glycine Riboswitch. ACS Chem Biol. 2011;18(3):293–8.

117. Kladwang W, Chou FC, Das R. Automated RNA structure prediction uncovers a kink-turn linker in double glycine riboswitches. J Am Chem Soc. 2012;134(3):1404–7.

118. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. 2016;44(D1):D733–45.

119. Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, et al. Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 2017;46(D1):D335–42.

120. Quinlan AR. BEDTools: The Swiss-Army tool for genome feature analysis. Vol. 2014, Current Protocols in Bioinformatics. 2014. 11.12.1-11.12.34.

121. Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. PLoS Comput Biol. 2007;3(4):680–91.

122. Raden M, Ali SM, Alkhnbashi OS, Busch A, Costa F, Davis JA, et al. Freiburg RNA tools: A central online resource for RNA-focused research and teaching. Nucleic Acids Res. 2018;46(W1):W25–9.

123. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.

124. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. **qgraph** : Network Visualizations of Relationships in Psychometric Data. J Stat Softw. 2012;48(4).

125. Jäger ML. MCL: Markov Cluster Algorithm. 2015.

126. Hennig C. fpc: Flexible Procedures for Clustering [Internet]. 2019.

127. Griffiths-Jones S. RALEE - RNA alignment editor in Emacs. Bioinformatics.

2005;21(2):257–9.

128. Darty K, Denise A, Ponty Y. VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics. 2009;25(15):1974–5.

129. Kalvari I, Nawrocki EP, Argasinska J, Quinones-Olvera N, Finn RD, Bateman A, et al. Non-Coding RNA Analysis Using the Rfam Database. Curr Protoc Bioinforma. 2018;62(1):1–27.

130. Snell E, Guirard M. Some Interrelationships of pyrixodine, alanine, and glycine in their effect on certain lactic acid bacteria. Proc Natl Acad Sci. 1943;874(1942):66–73.

131. Maculla ES, Cowles PB. The Use of Glycine in the Disruption of Bacterial Cells. Science (80-). 1948;107.

132. Hishinuma F, Izaki K, Takahashi H. Effects of glycine and d-amino acids on growth of various microorganisms. Agric Biol Chem. 1969;33(11):1577–86.

133. Hammes W, Schleifer KH, Kandler O. Mode of action of glycine on the biosynthesis of peptidoglycan. J Bacteriol. 1973;116(2):1029–53.

134. Tezuka T, Ohnishi Y. Two glycine riboswitches activate the glycine cleavage system essential for glycine detoxification in Streptomyces griseus. J Bacteriol. 2014;196(7):1369–76.

135. Atilho RM, Perkins KR, Breaker RR. Rare variants of the FMN riboswitch class in Clostridium difficile and other bacteria exhibit altered ligand specificity. RNA. 2019;25(1):23–34.

136. Kim JN, Roth A, Breaker RR. Guanine riboswitch variants from Mesoplasma

florum selectively recognize 2'-deoxyguanosine. PNAS. 2007;104(41).

137. Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N, Breaker RR. Challenges of ligand identification for riboswitch candidates. RNA Biol. 2011;8(1):5–10.

138. Wang J, Breaker R. Riboswitches that sense S-adenosylmethionine and S-adenosylhomocysteine. Biochem Cell Biol. 2008;

139. Breaker RR. Prospects for Riboswitch Discovery and Analysis. Mol Cell. 2011;43(6):867–79.

140. Bandyopadhyay S, Chaudhury S, Mehta D, Ramesh A. Discovery of iron-sensing bacterial riboswitches. Nat Chem Biol. 2020;

141. Gruber AR, Findeiß S, Washietl S, Hofacker IL, Stadler PF. RNAz 2.0: improved noncoding RNA detection. Pac Symp Biocomput. 2010;69–79.

142. Washietl S, Hofacker IL, Stadler PF. Fast and reliable prediction of noncoding RNAs. Proc Natl Acad Sci USA. 2005;102(7):2454–9.

143. Heyne S, Costa F, Rose D, Backofen R. Graphclust: Alignment-free structural clustering of local RNA secondary structures. Bioinformatics. 2012;28(12):224–32.

144. Miladi M, Junge A, Costa F, Seemann SE, Havgaard JH, Gorodkin J, et al. RNAscClust: Clustering RNA sequences using structure conservation and graph based motifs. Bioinformatics. 2017;33(14):2089–96.

145. Havill JT, Bhatiya C, Johnson SM, Sheets JD, Thompson JS. A new approach for detecting riboswitches in DNA sequences. Bioinformatics. 2014;30(21):3012–9.

146. Miladi M, Sokhoyan E, Houwaart T, Heyne S, Costa F, Grüning B, et al. GraphClust2: Annotation and discovery of structured RNAs with scalable and accessible integrative clustering. Gigascience. 2019;8(12):1–16.

147. Mukherjee S, Das Mandal S, Gupta N, Drory-Retwitzer M, Barash D, Sengupta S. RiboD: a comprehensive database for prokaryotic riboswitches. Bioinformatics. 2019;35(18):3541–3.

148. Abreu-Goodger C, Merino E. RibEx: A web server for locating riboswitches and other conserved bacterial regulatory elements. Nucleic Acids Res. 2005;33(SUPPL. 2):690–2.

149. Chang TH, Huang H-D, Wu L-C, Yeh C-T, Liu B-J, Horng J-T. Computational identification of riboswitches based on RNA conserved functional sequences and conformations. RNA. 2009;15:1426–30.

150. Mukherjee S, Sengupta S. Riboswitch Scanner: An efficient pHMM-based web-server to detect riboswitches in genomic sequences. Bioinformatics. 2016;32(5):776–8.

Appendix

Table 2.1: 48 Bacillaceae and 37 Vibrionaceae tandem riboswitch sequences for used phylogenetic analysis.

NZ_CM001792/1912588-1912785(-)/Bacillaceae/Oceanobacillus kimchii/tandem/GCV
NC_004193/1936128-1936325(-)/Bacillaceae/Oceanobacillus iheyensis/tandem/GCV
NZ_CP011974/3995995-3996179(-)/Bacillaceae/Bacillus endophyticus/tandem/GCV
NZ_CP012024/2219785-2219957(-)/Bacillaceae/Bacillus smithii/tandem/GCV
NC_002570/2941415-2941604(-)/Bacillaceae/Bacillus halodurans/tandem/GCV
NC_013791/2602919-2603097(-)/Bacillaceae/Bacillus pseudofirmus/tandem/GCV
NC_014171/2223922-2224136(-)/Bacillaceae/Bacillus thuringiensis/tandem/TP
NC_022781/4886033-4886247(-)/Bacillaceae/Bacillus toyonensis/tandem/TP
NC_010184/2253226-2253439(-)/Bacillaceae/Bacillus weihenstephanensis/tandem/TP
NC_003997/2221504-2221716(-)/Bacillaceae/Bacillus anthracis/tandem/TP
NC_003909/2304702-2304914(-)/Bacillaceae/Bacillus cereus/tandem/TP
NZ_CP007512/2078497-2078711(-)/Bacillaceae/Bacillus bombysepticus/tandem/TP
NZ_LN554846/1395685-1395898(-)/Vibrionaceae/Aliivibrio wodanis/tandem/TP
NZ_CP016307/1666806-1667024(-)/Vibrionaceae/Vibrio scophthalmi/tandem/TP
NC_015633/1730293-1730506(-)/Vibrionaceae/Vibrio anguillarum/tandem/TP
NZ_CP011406/1684498-1684708(-)/Vibrionaceae/Vibrio parahaemolyticus/tandem/TP
NC_021848/1063075-1063285(-)/Vibrionaceae/Vibrio parahaemolyticus/tandem/TP
NC_022269/245694-245904(-)/Vibrionaceae/Vibrio campbellii/tandem/TP
NZ_CP013484/549960-550168(-)/Vibrionaceae/Vibrio alginolyticus/tandem/TP
NC_022349/470880-471088(-)/Vibrionaceae/Vibrio alginolyticus/tandem/TP
NZ_CP016347/1636768-1636977(-)/Vibrionaceae/Vibrio natriegens/tandem/TP
NZ_CP009354/1442154-1442357(-)/Vibrionaceae/Vibrio tubiashii/tandem/TP
NZ_CP009264/1183330-1183543(-)/Vibrionaceae/Vibrio coralliilyticus/tandem/TP
NZ_CP016556/3409979-3410192(-)/Vibrionaceae/Vibrio coralliilyticus/tandem/TP
NZ_CP009261/1622641-1622870(-)/Vibrionaceae/Vibrio vulnificus/tandem/TP
NC_005139/1639249-1639478(-)/Vibrionaceae/Vibrio vulnificus/tandem/TP
NZ_CP011775/1463508-1463737(-)/Vibrionaceae/Vibrio vulnificus/tandem/TP
NZ_CP015506/4039196-4039369(-)/Bacillaceae/Bacillus oceanisediminis/tandem/GCV
NC_021362/2708987-2709200(-)/Bacillaceae/Bacillus paralicheniformis/tandem/GCV
NC_000964/2549408-2549606(-)/Bacillaceae/Bacillus subtilis/tandem/GCV
NC_014639/2070204-2070404(-)/Bacillaceae/Bacillus atrophaeus/tandem/GCV

Appendix

NC_014551/2417262-2417468(-)/Bacillaceae/Bacillus amyloliquefaciens/tandem/GCV
NZ_CP015911/2409495-2409700(-)/Bacillaceae/Bacillus velezensis/tandem/GCV
NC_014019/4324630-4324819(-)/Bacillaceae/Bacillus megaterium/tandem/GCV
NC_006270/2547358-2547569(-)/Bacillaceae/Bacillus licheniformis/tandem/GCV
NZ_LT603683/2822575-2822779(-)/Bacillaceae/Bacillus glycinifermentans/tandem/GCV
NC_022080/2350097-2350306(-)/Bacillaceae/Geobacillus sp/tandem/GCV
NC_006510/2456137-2456344(-)/Bacillaceae/Geobacillus kaustophilus/tandem/GCV
NC_009328/2482842-2483052(-)/Bacillaceae/Geobacillus thermodenitrificans/tandem/GCV
NC_017668/2468838-2469025(-)/Bacillaceae/Halobacillus halophilus/tandem/GCV
NZ_CP012152/1897558-1897726(-)/Bacillaceae/Anoxybacillus gonensis/tandem/GCV
NC_012793/2406299-2406478(-)/Bacillaceae/Geobacillus sp/tandem/GCV
NZ_CP012712/3379286-3379464(-)/Bacillaceae/Geobacillus thermoglucosidasius/tandem/GCV
NZ_CM002692/2459035-2459205(-)/Bacillaceae/Geobacillus stearothermophilus/tandem/GCV
NC_015634/1722666-1722849(-)/Bacillaceae/Bacillus coagulans/tandem/GCV
NC_022524/3285938-3286106(-)/Bacillaceae/Bacillus infantis/tandem/GCV
NC_021171/3397531-3397700(-)/Bacillaceae/Bacillus sp/tandem/GCV
NZ_CP012600/1674290-1674489(-)/Bacillaceae/Bacillus gobiensis/tandem/GCV
NZ_CP003923/2412887-2413083(-)/Bacillaceae/Bacillus lehensis/tandem/GCV
NZ_CP007739/2294995-2295177(-)/Bacillaceae/Bacillus methanolicus/tandem/GCV
NC_006582/2619769-2619955(-)/Bacillaceae/Bacillus clausii/tandem/GCV
NC_009848/2216054-2216268(-)/Bacillaceae/Bacillus pumilus/tandem/GCV
NZ_CP006837/2926928-2927133(-)/Bacillaceae/Lysinibacillus varians/tandem/GCV
NZ_CP011008/3501658-3501839(-)/Bacillaceae/Bacillus simplex/tandem/GCV
NZ_CP008876/1734313-1734490(-)/Bacillaceae/Terribacillus aidingensis/tandem/GCV
NC_014829/1874202-1874428(+)/Bacillaceae/Bacillus cellulosilyticus/tandem/GCV
NZ_CM000745/4807363-4807579(+)/Bacillaceae/Bacillus pseudomycoides/tandem/TP
NC_016602/2046029-2046257(+)/Vibrionaceae/Vibrio furnissii/tandem/TP
NC_011312/1917427-1917640(+)/Vibrionaceae/Aliivibrio salmonicida/tandem/TP
NZ_LK021130/1565347-1565560(+)/Vibrionaceae/Vibrio anguillarum/tandem/TP
NZ_CP011460/1509763-1509976(+)/Vibrionaceae/Vibrio anguillarum/tandem/TP
NZ_CP016177/1449571-1449779(+)/Vibrionaceae/Vibrio breoganii/tandem/TP
NC_006840/1607738-1607953(+)/Vibrionaceae/Vibrio fischeri/tandem/TP
NC_011184/1633222-1633437(+)/Vibrionaceae/Vibrio fischeri/tandem/TP
NC_004603/1855938-1856148(+)/Vibrionaceae/Vibrio parahaemolyticus/tandem/TP

NZ_CP009847/1884813-1885023(+)/Vibrionaceae/Vibrio parahaemolyticus/tandem/TP
NZ_CP014038/335935-336145(+)/Vibrionaceae/Vibrio harveyi/tandem/TP
NC_009783/2485129-2485339(+)/Vibrionaceae/Vibrio campbellii/tandem/TP
NC_016613/1821145-1821354(+)/Vibrionaceae/Vibrio sp/tandem/TP
NC_013456/2228325-2228533(+)/Vibrionaceae/Vibrio antiquarius/tandem/TP
NZ_CP016324/1619624-1619854(+)/Vibrionaceae/Vibrio cholerae/tandem/TP
NC_012668/2123415-2123645(+)/Vibrionaceae/Vibrio cholerae/tandem/TP
NZ_CP010811/1520067-1520297(+)/Vibrionaceae/Vibrio cholerae/tandem/TP
NC_022528/1969582-1969756(+)/Vibrionaceae/Vibrio nigripulchritudo/tandem/TP
NC_004459/2743752-2743981(+)/Vibrionaceae/Vibrio vulnificus/tandem/TP
NZ_CP016351/1896934-1897143(+)/Vibrionaceae/Vibrio natriegens/tandem/TP
NZ_CP008934/2794157-2794364(+)/Bacillaceae/Geobacillus stearothermophilus/tandem/GCV
NZ_CP014342/825424-825634(+)/Bacillaceae/Geobacillus subterraneus/tandem/GCV
NC_011567/927747-927918(+)/Bacillaceae/Anoxybacillus flavithermus/tandem/GCV
NC_015660/1196873-1197051(+)/Bacillaceae/Geobacillus thermoglucosidasius/tandem/GCV
NZ_CP007161/1859758-1859984(+)/Bacillaceae/Virgibacillus sp/tandem/GCV
NZ_CP010820/1718510-1718723(+)/Bacillaceae/Lysinibacillus fusiformis/tandem/GCV
NC_002505/1520426-1520656(-)/Vibrionaceae/Vibrio cholerae/tandem/TP
NC_017270/1416320-1416550(-)/Vibrionaceae/Vibrio cholerae/tandem/TP
NC_012582/1610260-1610490(-)/Vibrionaceae/Vibrio cholerae/tandem/TP

Table 2.2: 48 Bacillaceae tandem riboswitch aptamer-1 sequences used for phylogenetic analysis.

NC_000964/2549502-2549606(-)/Bacillaceae/Bacillus_subtilis/GCV
NC_002570/2941507-2941604(-)/Bacillaceae/Bacillus_halodurans/GCV
NC_004193/1936231-1936325(-)/Bacillaceae/Oceanobacillus_iheyensis/GCV
NC_006270/2547462-2547569(-)/Bacillaceae/Bacillus_licheniformis/GCV
NC_006510/2456253-2456344(-)/Bacillaceae/Geobacillus_kaustophilus/GCV
NC_006582/2619859-2619955(-)/Bacillaceae/Bacillus_clausii/GCV
NC_009328/2482961-2483052(-)/Bacillaceae/Geobacillus_thermodenitrificans/GCV
NC_009848/2216158-2216268(-)/Bacillaceae/Bacillus_pumilus/GCV
NC_011567/927748-927838(+)/Bacillaceae/Anoxybacillus_flavithermus/GCV
NC_012793/2406387-2406478(-)/Bacillaceae/Geobacillus_sp/GCV
NC_013791/2602998-2603097(-)/Bacillaceae/Bacillus_pseudofirmus/GCV
NC_014019/4324724-4324819(-)/Bacillaceae/Bacillus_megaterium/GCV

Appendix

NC_014551/2417365-2417468(-)/Bacillaceae/Bacillus_amyloliquefaciens/GCV
NC_014639/2070301-2070404(-)/Bacillaceae/Bacillus_atrophaeus/GCV
NC_014829/1874203-1874297(+)/Bacillaceae/Bacillus_cellulosilyticus/GCV
NC_015634/1722755-1722849(-)/Bacillaceae/Bacillus_coagulans/GCV
NC_015660/1196874-1196964(+)/Bacillaceae/Geobacillus_thermoglucosidasius/GCV
NC_016593/2515901-2515992(-)/Bacillaceae/Geobacillus_thermoleovorans/GCV
NC_017668/2468936-2469025(-)/Bacillaceae/Halobacillus_halophilus/GCV
NC_018520/2529006-2529110(-)/Bacillaceae/Bacillus_subtilis/GCV
NC_021171/3397609-3397700(-)/Bacillaceae/Bacillus_sp/GCV
NC_021362/2709093-2709200(-)/Bacillaceae/Bacillus_paralicheniformis/GCV
NC_022080/2350215-2350306(-)/Bacillaceae/Geobacillus_sp/GCV
NC_022524/3286016-3286106(-)/Bacillaceae/Bacillus_infantis/GCV
NZ_CM000742/2048642-2048751(-)/Bacillaceae/Bacillus_mycoides/GCV
NZ_CM001792/1912691-1912785(-)/Bacillaceae/Oceanobacillus_kimchii/GCV
NZ_CM002692/2459114-2459205(-)/Bacillaceae/Geobacillus_stearothermophilus/GCV
NZ_CP003923/2412985-2413083(-)/Bacillaceae/Bacillus_lehensis/GCV
NZ_CP006837/2927038-2927133(-)/Bacillaceae/Lysinibacillus_varians/GCV
NZ_CP007161/1859759-1859870(+)/Bacillaceae/Virgibacillus_sp/GCV
NZ_CP007739/2295082-2295177(-)/Bacillaceae/Bacillus_methanolicus/GCV
NZ_CP008876/1734400-1734490(-)/Bacillaceae/Terribacillus_aidingensis/GCV
NZ_CP008934/2794158-2794248(+)/Bacillaceae/Geobacillus_stearothermophilus/GCV
NZ_CP010820/1718511-1718606(+)/Bacillaceae/Lysinibacillus_fusiformis/GCV
NZ_CP011008/3501749-3501839(-)/Bacillaceae/Bacillus_simplex/GCV
NZ_CP011974/3996087-3996179(-)/Bacillaceae/Bacillus_endophyticus/GCV
NZ_CP012024/2219866-2219957(-)/Bacillaceae/Bacillus_smithii/GCV
NZ_CP012152/1897635-1897726(-)/Bacillaceae/Anoxybacillus_gonensis/GCV
NZ_CP012600/1674390-1674489(-)/Bacillaceae/Bacillus_gobiensis/GCV
NZ_CP012712/3379373-3379464(-)/Bacillaceae/Geobacillus_thermoglucosidasius/GCV
NZ_CP014342/825425-825515(+)/Bacillaceae/Geobacillus_subterraneus/GCV
NZ_CP015506/4039279-4039369(-)/Bacillaceae/Bacillus_oceanisediminis/GCV
NZ_CP015911/2409598-2409700(-)/Bacillaceae/Bacillus_velezensis/GCV
NZ_LT603683/2822672-2822779(-)/Bacillaceae/Bacillus_glycinifermentans/GCV
NC_010184/2253330-2253439(-)/Bacillaceae/Bacillus_weihenstephanensis/TP
NZ_CM000745/4807364-4807472(+)/Bacillaceae/Bacillus_pseudomycoides/TP
NC_003909/2304806-2304914(-)/Bacillaceae/Bacillus_cereus/TP
NC_003997/2221608-2221716(-)/Bacillaceae/Bacillus_anthracis/TP

NC_022781/4886138-4886247(-)/Bacillaceae/Bacillus_toyonensis/TP
NC_014171/2224027-2224136(-)/Bacillaceae/Bacillus_thuringiensis/TP
NZ_CP007512/2078602-2078711(-)/Bacillaceae/Bacillus_bombysepticus/TP

Table 2.3: 48 Bacillaceae tandem riboswitch aptamer-2 sequences used for phylogenetic analysis.

NC_000964/2549409-2549501(-)/Bacillaceae/Bacillus_subtilis/GCV
NC_002570/2941416-2941506(-)/Bacillaceae/Bacillus_halodurans/GCV
NC_004193/1936129-1936230(-)/Bacillaceae/Oceanobacillus_iheyensis/GCV
NC_006270/2547359-2547461(-)/Bacillaceae/Bacillus_licheniformis/GCV
NC_006510/2456138-2456252(-)/Bacillaceae/Geobacillus_kaustophilus/GCV
NC_006582/2619770-2619858(-)/Bacillaceae/Bacillus_clausii/GCV
NC_009328/2482843-2482960(-)/Bacillaceae/Geobacillus_thermodenitrificans/GCV
NC_009848/2216055-2216157(-)/Bacillaceae/Bacillus_pumilus/GCV
NC_011567/927839-927918(+)/Bacillaceae/Anoxybacillus_flavithermus/GCV
NC_012793/2406300-2406386(-)/Bacillaceae/Geobacillus_sp/GCV
NC_013791/2602920-2602997(-)/Bacillaceae/Bacillus_pseudofirmus/GCV
NC_014019/4324631-4324723(-)/Bacillaceae/Bacillus_megaterium/GCV
NC_014551/2417263-2417364(-)/Bacillaceae/Bacillus_amyloliquefaciens/GCV
NC_014639/2070205-2070300(-)/Bacillaceae/Bacillus_atrophaeus/GCV
NC_014829/1874298-1874428(+)/Bacillaceae/Bacillus_cellulosilyticus/GCV
NC_015634/1722667-1722754(-)/Bacillaceae/Bacillus_coagulans/GCV
NC_015660/1196965-1197051(+)/Bacillaceae/Geobacillus_thermoglucosidasius/GCV
NC_016593/2515786-2515900(-)/Bacillaceae/Geobacillus_thermoleovorans/GCV
NC_017668/2468839-2468935(-)/Bacillaceae/Halobacillus_halophilus/GCV
NC_018520/2528913-2529005(-)/Bacillaceae/Bacillus_subtilis/GCV
NC_021171/3397532-3397608(-)/Bacillaceae/Bacillus_sp/GCV
NC_021362/2708988-2709092(-)/Bacillaceae/Bacillus_paralicheniformis/GCV
NC_022080/2350098-2350214(-)/Bacillaceae/Geobacillus_sp/GCV
NC_022524/3285939-3286015(-)/Bacillaceae/Bacillus_infantis/GCV
NZ_CM000742/2048539-2048641(-)/Bacillaceae/Bacillus_mycoides/GCV
NZ_CM001792/1912589-1912690(-)/Bacillaceae/Oceanobacillus_kimchii/GCV
NZ_CM002692/2459036-2459113(-)/Bacillaceae/Geobacillus_stearothermophilus/GCV
NZ_CP003923/2412888-2412984(-)/Bacillaceae/Bacillus_lehensis/GCV
NZ_CP006837/2926929-2927037(-)/Bacillaceae/Lysinibacillus_varians/GCV
NZ_CP007161/1859871-1859984(+)/Bacillaceae/Virgibacillus_sp/GCV
NZ_CP007739/2294996-2295081(-)/Bacillaceae/Bacillus_methanolicus/GCV

NZ_CP008876/1734314-1734399(-)/Bacillaceae/Terribacillus_aidingensis/GCV
NZ_CP008934/2794249-2794364(+)/Bacillaceae/Geobacillus_stearothermophilus/GCV
NZ_CP010820/1718607-1718723(+)/Bacillaceae/Lysinibacillus_fusiformis/GCV
NZ_CP011008/3501659-3501748(-)/Bacillaceae/Bacillus_simplex/GCV
NZ_CP011974/3995996-3996086(-)/Bacillaceae/Bacillus_endophyticus/GCV
NZ_CP012024/2219786-2219865(-)/Bacillaceae/Bacillus_smithii/GCV
NZ_CP012152/1897559-1897634(-)/Bacillaceae/Anoxybacillus_gonensis/GCV
NZ_CP012600/1674291-1674389(-)/Bacillaceae/Bacillus_gobiensis/GCV
NZ_CP012712/3379287-3379372(-)/Bacillaceae/Geobacillus_thermoglucosidasius/GCV
NZ_CP014342/825516-825634(+)/Bacillaceae/Geobacillus_subterraneus/GCV
NZ_CP015506/4039197-4039278(-)/Bacillaceae/Bacillus_oceanisediminis/GCV
NZ_CP015911/2409496-2409597(-)/Bacillaceae/Bacillus_velezensis/GCV
NZ_LT603683/2822576-2822671(-)/Bacillaceae/Bacillus_glycinifermentans/GCV
NZ_CM000745/4807473-4807579(+)/Bacillaceae/Bacillus_pseudomycoides/TP
NC_010184/2253227-2253329(-)/Bacillaceae/Bacillus_weihenstephanensis/TP
NC_003909/2304703-2304805(-)/Bacillaceae/Bacillus_cereus/TP
NC_003997/2221505-2221607(-)/Bacillaceae/Bacillus_anthracis/TP
NC_022781/4886034-4886137(-)/Bacillaceae/Bacillus_toyonensis/TP
NC_014171/2223923-2224026(-)/Bacillaceae/Bacillus_thuringiensis/TP
NZ_CP007512/2078498-2078601(-)/Bacillaceae/Bacillus_bombysepticus/TP

Table 2.4: 168 Bacillaceae tandem riboswitch aptamer sequences used for graph clustering analysis.

NC_000964/2549510-2549601(-)/Bacillaceae/Bacillus_subtilis/apt_A/GCV
NC_000964/2549415-2549501(-)/Bacillaceae/Bacillus_subtilis/apt_B/GCV
NC_002570/2941515-2941599(-)/Bacillaceae/Bacillus_halodurans/apt_A/GCV
NC_002570/2941417-2941508(-)/Bacillaceae/Bacillus_halodurans/apt_B/GCV
NC_004193/1936239-1936320(-)/Bacillaceae/Oceanobacillus_iheyensis/apt_A/GCV
NC_004193/1936134-1936232(-)/Bacillaceae/Oceanobacillus_iheyensis/apt_B/GCV
NC_006270/2547470-2547564(-)/Bacillaceae/Bacillus_licheniformis/apt_A/GCV
NC_006270/2547365-2547461(-)/Bacillaceae/Bacillus_licheniformis/apt_B/GCV
NC_006510/2456261-2456339(-)/Bacillaceae/Geobacillus_kaustophilus/apt_A/GCV
NC_006510/2456138-2456251(-)/Bacillaceae/Geobacillus_kaustophilus/apt_B/GCV
NC_006582/2619867-2619950(-)/Bacillaceae/Bacillus_clausii/apt_A/GCV
NC_006582/2619777-2619858(-)/Bacillaceae/Bacillus_clausii/apt_B/GCV
NC_009328/2482969-2483047(-)/Bacillaceae/Geobacillus_thermodenitrificans/apt_A/GCV

NC_009328/2482843-2482959(-)/Bacillaceae/Geobacillus_thermodenitrificans/apt_B/GCV
NC_009725/2422935-2423025(-)/Bacillaceae/Bacillus_velezensis/apt_A/GCV
NC_009725/2422832-2422926(-)/Bacillaceae/Bacillus_velezensis/apt_B/GCV
NC_009848/2216166-2216263(-)/Bacillaceae/Bacillus_pumilus/apt_A/GCV
NC_009848/2216062-2216157(-)/Bacillaceae/Bacillus_pumilus/apt_B/GCV
NC_011567/927752-927830(+)/Bacillaceae/Anoxybacillus_flavithermus/apt_A/GCV
NC_011567/927839-927918(+)/Bacillaceae/Anoxybacillus_flavithermus/apt_B/GCV
NC_012793/2406395-2406473(-)/Bacillaceae/Geobacillus_sp/apt_A/GCV
NC_012793/2406302-2406387(-)/Bacillaceae/Geobacillus_sp/apt_B/GCV
NC_013411/218783-218861(+)/Bacillaceae/Geobacillus_sp/apt_A/GCV
NC_013411/218870-218985(+)/Bacillaceae/Geobacillus_sp/apt_B/GCV
NC_013791/2603007-2603091(-)/Bacillaceae/Bacillus_pseudofirmus/apt_A/GCV
NC_013791/2602920-2602996(-)/Bacillaceae/Bacillus_pseudofirmus/apt_B/GCV
NC_014019/4324732-4324814(-)/Bacillaceae/Bacillus_megaterium/apt_A/GCV
NC_014019/4324638-4324723(-)/Bacillaceae/Bacillus_megaterium/apt_B/GCV
NC_014479/2406250-2406341(-)/Bacillaceae/Bacillus_subtilis/apt_A/GCV
NC_014479/2406155-2406241(-)/Bacillaceae/Bacillus_subtilis/apt_B/GCV
NC_014551/2417373-2417463(-)/Bacillaceae/Bacillus_amyloliquefaciens/apt_A/GCV
NC_014551/2417270-2417364(-)/Bacillaceae/Bacillus_amyloliquefaciens/apt_B/GCV
NC_014639/2070309-2070399(-)/Bacillaceae/Bacillus_atrophaeus/apt_A/GCV
NC_014639/2070211-2070300(-)/Bacillaceae/Bacillus_atrophaeus/apt_B/GCV
NC_014650/1132999-1133077(+)/Bacillaceae/Geobacillus_sp/apt_A/GCV
NC_014650/1133085-1133169(+)/Bacillaceae/Geobacillus_sp/apt_B/GCV
NC_014829/1874207-1874289(+)/Bacillaceae/Bacillus_cellulosilyticus/apt_A/GCV
NC_014829/1874295-1874423(+)/Bacillaceae/Bacillus_cellulosilyticus/apt_B/GCV
NC_014915/2520626-2520704(-)/Bacillaceae/Geobacillus_sp/apt_A/GCV
NC_014915/2520503-2520616(-)/Bacillaceae/Geobacillus_sp/apt_B/GCV
NC_015634/1722764-1722844(-)/Bacillaceae/Bacillus_coagulans/apt_A/GCV
NC_015634/1722673-1722755(-)/Bacillaceae/Bacillus_coagulans/apt_B/GCV
NC_015660/1196878-1196956(+)/Bacillaceae/Geobacillus_thermoglucosidasius/apt_A/GCV
NC_015660/1196964-1197048(+)/Bacillaceae/Geobacillus_thermoglucosidasius/apt_B/GCV
NC_016023/3047686-3047766(+)/Bacillaceae/Bacillus_coagulans/apt_A/GCV
NC_016023/3047775-3047855(+)/Bacillaceae/Bacillus_coagulans/apt_B/GCV
NC_016047/2530173-2530264(-)/Bacillaceae/Bacillus_subtilis/apt_A/GCV
NC_016047/2530078-2530164(-)/Bacillaceae/Bacillus_subtilis/apt_B/GCV
NC_016593/2515909-2515987(-)/Bacillaceae/Geobacillus_thermoleovorans/apt_A/GCV

Appendix

NC_016593/2515786-2515899(-)/Bacillaceae/Geobacillus_thermoleovorans/apt_B/GCV
NC_016784/2514621-2514711(-)/Bacillaceae/Bacillus_velezensis/apt_A/GCV
NC_016784/2514518-2514612(-)/Bacillaceae/Bacillus_velezensis/apt_B/GCV
NC_017061/2662425-2662515(-)/Bacillaceae/Bacillus_velezensis/apt_A/GCV
NC_017061/2662322-2662416(-)/Bacillaceae/Bacillus_velezensis/apt_B/GCV
NC_017195/2398174-2398265(-)/Bacillaceae/Bacillus_subtilis/apt_A/GCV
NC_017195/2398079-2398165(-)/Bacillaceae/Bacillus_subtilis/apt_B/GCV
NC_017668/2468944-2469020(-)/Bacillaceae/Halobacillus_halophilus/apt_A/GCV
NC_017668/2468839-2468934(-)/Bacillaceae/Halobacillus_halophilus/apt_B/GCV
NC_017743/2456064-2456155(-)/Bacillaceae/Bacillus_sp/apt_A/GCV
NC_017743/2455969-2456055(-)/Bacillaceae/Bacillus_sp/apt_B/GCV
NC_017912/2658551-2658641(-)/Bacillaceae/Bacillus_amyloliquefaciens/apt_A/GCV
NC_017912/2658448-2658542(-)/Bacillaceae/Bacillus_amyloliquefaciens/apt_B/GCV
NC_019842/2480102-2480192(-)/Bacillaceae/Bacillus_velezensis/apt_A/GCV
NC_019842/2479999-2480093(-)/Bacillaceae/Bacillus_velezensis/apt_B/GCV
NC_020272/1498766-1498856(+)/Bacillaceae/Bacillus_amyloliquefaciens/apt_A/GCV
NC_020272/1498865-1498959(+)/Bacillaceae/Bacillus_amyloliquefaciens/apt_B/GCV
NC_021171/3397617-3397695(-)/Bacillaceae/Bacillus_sp/apt_A/GCV
NC_021171/3397532-3397607(-)/Bacillaceae/Bacillus_sp/apt_B/GCV
NC_021362/2709101-2709195(-)/Bacillaceae/Bacillus_paralicheniformis/apt_A/GCV
NC_021362/2708995-2709092(-)/Bacillaceae/Bacillus_paralicheniformis/apt_B/GCV
NC_022080/2350223-2350301(-)/Bacillaceae/Geobacillus_sp/apt_A/GCV
NC_022080/2350098-2350213(-)/Bacillaceae/Geobacillus_sp/apt_B/GCV
NC_022524/3286024-3286101(-)/Bacillaceae/Bacillus_infantis/apt_A/GCV
NC_022524/3285939-3286014(-)/Bacillaceae/Bacillus_infantis/apt_B/GCV
NC_022653/2420010-2420100(-)/Bacillaceae/Bacillus_amyloliquefaciens/apt_A/GCV
NC_022653/2419907-2420001(-)/Bacillaceae/Bacillus_amyloliquefaciens/apt_B/GCV
NZ_AP013294/2865878-2865959(-)/Bacillaceae/Bacillus_sp/apt_A/GCV
NZ_AP013294/2865793-2865871(-)/Bacillaceae/Bacillus_sp/apt_B/GCV
NZ_AP014928/2205267-2205364(-)/Bacillaceae/Bacillus_pumilus/apt_A/GCV
NZ_AP014928/2205163-2205258(-)/Bacillaceae/Bacillus_pumilus/apt_B/GCV
NZ_CM001483/1078319-1078397(+)/Bacillaceae/Geobacillus_thermoglucosidans/apt_A/GCV
NZ_CM001483/1078405-1078489(+)/Bacillaceae/Geobacillus_thermoglucosidans/apt_B/GCV
NZ_CM001792/1912699-1912780(-)/Bacillaceae/Oceanobacillus_kimchii/apt_A/GCV
NZ_CM001792/1912594-1912692(-)/Bacillaceae/Oceanobacillus_kimchii/apt_B/GCV

NZ_CM002692/2459122-2459200(-)/Bacillaceae/Geobacillus_stearothermophilus/apt_A/GCV
NZ_CM002692/2459036-2459112(-)/Bacillaceae/Geobacillus_stearothermophilus/apt_B/GCV
NZ_CP003923/2412993-2413078(-)/Bacillaceae/Bacillus_lehensis/apt_A/GCV
NZ_CP003923/2412895-2412984(-)/Bacillaceae/Bacillus_lehensis/apt_B/GCV
NZ_CP006837/2927046-2927128(-)/Bacillaceae/Lysinibacillus_varians/apt_A/GCV
NZ_CP006837/2926932-2927037(-)/Bacillaceae/Lysinibacillus_varians/apt_B/GCV
NZ_CP006890/2652127-2652217(-)/Bacillaceae/Bacillus_velezensis/apt_A/GCV
NZ_CP006890/2652024-2652118(-)/Bacillaceae/Bacillus_velezensis/apt_B/GCV
NZ_CP007161/1859763-1859862(+)/Bacillaceae/Virgibacillus_sp/apt_A/GCV
NZ_CP007161/1859869-1859978(+)/Bacillaceae/Virgibacillus_sp/apt_B/GCV
NZ_CP007739/2295090-2295172(-)/Bacillaceae/Bacillus_methanolicus/apt_A/GCV
NZ_CP007739/2295003-2295081(-)/Bacillaceae/Bacillus_methanolicus/apt_B/GCV
NZ_CP008855/778913-778991(-)/Bacillaceae/Bacillus_sp/apt_A/GCV
NZ_CP008855/778828-778903(-)/Bacillaceae/Bacillus_sp/apt_B/GCV
NZ_CP008876/1734408-1734485(-)/Bacillaceae/Terribacillus_aidingensis/apt_A/GCV
NZ_CP008876/1734314-1734398(-)/Bacillaceae/Terribacillus_aidingensis/apt_B/GCV
NZ_CP008903/733183-733261(-)/Bacillaceae/Geobacillus_sp/apt_A/GCV
NZ_CP008903/733060-733175(-)/Bacillaceae/Geobacillus_sp/apt_B/GCV
NZ_CP008934/2794162-2794240(+)/Bacillaceae/Geobacillus_stearothermophilus/apt_A/GCV
NZ_CP008934/2794249-2794364(+)/Bacillaceae/Geobacillus_stearothermophilus/apt_B/GCV
NZ_CP009611/2444930-2445020(-)/Bacillaceae/Bacillus_subtilis/apt_A/GCV
NZ_CP009611/2444827-2444921(-)/Bacillaceae/Bacillus_subtilis/apt_B/GCV
NZ_CP009709/721275-721355(-)/Bacillaceae/Bacillus_coagulans/apt_A/GCV
NZ_CP009709/721184-721266(-)/Bacillaceae/Bacillus_coagulans/apt_B/GCV
NZ_CP009748/2420320-2420410(-)/Bacillaceae/Bacillus_subtilis/apt_A/GCV
NZ_CP009748/2420217-2420311(-)/Bacillaceae/Bacillus_subtilis/apt_B/GCV
NZ_CP009749/2480838-2480928(-)/Bacillaceae/Bacillus_subtilis/apt_A/GCV
NZ_CP009749/2480735-2480829(-)/Bacillaceae/Bacillus_subtilis/apt_B/GCV
NZ_CP009920/1299846-1299928(-)/Bacillaceae/Bacillus_megaterium/apt_A/GCV
NZ_CP009920/1299752-1299837(-)/Bacillaceae/Bacillus_megaterium/apt_B/GCV
NZ_CP009938/3068648-3068738(-)/Bacillaceae/Bacillus_sp/apt_A/GCV
NZ_CP009938/3068545-3068639(-)/Bacillaceae/Bacillus_sp/apt_B/GCV
NZ_CP010014/2156144-2156235(-)/Bacillaceae/Bacillus_sp/apt_A/GCV
NZ_CP010014/2156049-2156135(-)/Bacillaceae/Bacillus_sp/apt_B/GCV
NZ_CP010075/2217406-2217502(-)/Bacillaceae/Bacillus_sp/apt_A/GCV
NZ_CP010075/2217304-2217397(-)/Bacillaceae/Bacillus_sp/apt_B/GCV

Appendix

NZ_CP010406/3568525-3568615(+)/Bacillaceae/Bacillus_sp/apt_A/GCV
NZ_CP010406/3568624-3568718(+)/Bacillaceae/Bacillus_sp/apt_B/GCV
NZ_CP010525/1603444-1603524(+)/Bacillaceae/Bacillus_coagulans/apt_A/GCV
NZ_CP010525/1603533-1603615(+)/Bacillaceae/Bacillus_coagulans/apt_B/GCV
NZ_CP010820/1718515-1718598(+)/Bacillaceae/Lysinibacillus_fusiformis/apt_A/GCV
NZ_CP010820/1718607-1718719(+)/Bacillaceae/Lysinibacillus_fusiformis/apt_B/GCV
NZ_CP011007/2233640-2233737(-)/Bacillaceae/Bacillus_pumilus/apt_A/GCV
NZ_CP011007/2233536-2233631(-)/Bacillaceae/Bacillus_pumilus/apt_B/GCV
NZ_CP011008/3501757-3501834(-)/Bacillaceae/Bacillus_simplex/apt_A/GCV
NZ_CP011008/3501666-3501748(-)/Bacillaceae/Bacillus_simplex/apt_B/GCV
NZ_CP011109/2082927-2083024(-)/Bacillaceae/Bacillus_pumilus/apt_A/GCV
NZ_CP011109/2082823-2082918(-)/Bacillaceae/Bacillus_pumilus/apt_B/GCV
NZ_CP011346/2597939-2598029(-)/Bacillaceae/Bacillus_velezensis/apt_A/GCV
NZ_CP011346/2597833-2597933(-)/Bacillaceae/Bacillus_velezensis/apt_B/GCV
NZ_CP011361/1138533-1138627(+)/Bacillaceae/Salimicrobium_jeotgali/apt_A/GCV
NZ_CP011361/1138636-1138713(+)/Bacillaceae/Salimicrobium_jeotgali/apt_B/GCV
NZ_CP011832/2370825-2370903(-)/Bacillaceae/Geobacillus_sp/apt_A/GCV
NZ_CP011832/2370702-2370817(-)/Bacillaceae/Geobacillus_sp/apt_B/GCV
NZ_CP011974/3996095-3996174(-)/Bacillaceae/Bacillus_endophyticus/apt_A/GCV
NZ_CP011974/3996002-3996087(-)/Bacillaceae/Bacillus_endophyticus/apt_B/GCV
NZ_CP012024/2219874-2219952(-)/Bacillaceae/Bacillus_smithii/apt_A/GCV
NZ_CP012024/2219793-2219865(-)/Bacillaceae/Bacillus_smithii/apt_B/GCV
NZ_CP012152/1897643-1897721(-)/Bacillaceae/Anoxybacillus_gonensis/apt_A/GCV
NZ_CP012152/1897559-1897633(-)/Bacillaceae/Anoxybacillus_gonensis/apt_B/GCV
NZ_CP012600/1674398-1674484(-)/Bacillaceae/Bacillus_gobiensis/apt_A/GCV
NZ_CP012600/1674298-1674389(-)/Bacillaceae/Bacillus_gobiensis/apt_B/GCV
NZ_CP012601/681422-681504(+)/Bacillaceae/Bacillus_sp/apt_A/GCV
NZ_CP012601/681514-681599(+)/Bacillaceae/Bacillus_sp/apt_B/GCV
NZ_CP012602/2037426-2037505(+)/Bacillaceae/Bacillus_sp/apt_A/GCV
NZ_CP012602/2037514-2037595(+)/Bacillaceae/Bacillus_sp/apt_B/GCV
NZ_CP013950/3735085-3735175(+)/Bacillaceae/Bacillus_sp/apt_A/GCV
NZ_CP013950/3735184-3735278(+)/Bacillaceae/Bacillus_sp/apt_B/GCV
NZ_CP013984/2409261-2409352(-)/Bacillaceae/Bacillus_subtilis/apt_A/GCV
NZ_CP013984/2409165-2409252(-)/Bacillaceae/Bacillus_subtilis/apt_B/GCV
NZ_CP014342/825429-825507(+)/Bacillaceae/Geobacillus_subterraneus/apt_A/GCV
NZ_CP014342/825516-825634(+)/Bacillaceae/Geobacillus_subterraneus/apt_B/GCV

NZ_CP014643/2310694-2310777(+)/Bacillaceae/Lysinibacillus_sphaericus/apt_A/GCV
NZ_CP014643/2310786-2310898(+)/Bacillaceae/Lysinibacillus_sphaericus/apt_B/GCV
NZ_CP014840/1408676-1408766(+)/Bacillaceae/Bacillus_subtilis/apt_A/GCV
NZ_CP014840/1408775-1408864(+)/Bacillaceae/Bacillus_subtilis/apt_B/GCV
NZ_CP015226/4221431-4221513(-)/Bacillaceae/Bacillus_sp/apt_A/GCV
NZ_CP015226/4221337-4221422(-)/Bacillaceae/Bacillus_sp/apt_B/GCV
NZ_CP015506/4039287-4039364(-)/Bacillaceae/Bacillus_oceanisediminis/apt_A/GCV
NZ_CP015506/4039202-4039280(-)/Bacillaceae/Bacillus_oceanisediminis/apt_B/GCV
NZ_CP015911/2409606-2409695(-)/Bacillaceae/Bacillus_velezensis/apt_A/GCV
NZ_CP015911/2409503-2409597(-)/Bacillaceae/Bacillus_velezensis/apt_B/GCV
NZ_LT603683/2822680-2822774(-)/Bacillaceae/Bacillus_glycinifermentans/apt_A/GCV
NZ_LT603683/2822576-2822670(-)/Bacillaceae/Bacillus_glycinifermentans/apt_B/GCV

Table 2.5: 72 Vibrionaceae tandem riboswitch aptamer sequences used for graph clustering analysis.

NC_002505/1520526-1520651(-)/Vibrionaceae/Vibrio_cholerae/apt_A/TP
NC_002505/1520433-1520518(-)/Vibrionaceae/Vibrio_cholerae/apt_B/TP
NC_004459/2743757-2743881(+)/Vibrionaceae/Vibrio_vulnificus/apt_A/TP
NC_004459/2743889-2743974(+)/Vibrionaceae/Vibrio_vulnificus/apt_B/TP
NC_004603/1855943-1856046(+)/Vibrionaceae/Vibrio_parahaemolyticus/apt_A/TP
NC_004603/1856054-1856141(+)/Vibrionaceae/Vibrio_parahaemolyticus/apt_B/TP
NC_005139/1639349-1639473(-)/Vibrionaceae/Vibrio_vulnificus/apt_A/TP
NC_005139/1639256-1639341(-)/Vibrionaceae/Vibrio_vulnificus/apt_B/TP
NC_006840/1607743-1607847(+)/Vibrionaceae/Vibrio_fischeri/apt_A/TP
NC_006840/1607855-1607946(+)/Vibrionaceae/Vibrio_fischeri/apt_B/TP
NC_009457/1078941-1079066(-)/Vibrionaceae/Vibrio_cholerae/apt_A/TP
NC_009457/1078848-1078933(-)/Vibrionaceae/Vibrio_cholerae/apt_B/TP
NC_009783/2485134-2485237(+)/Vibrionaceae/Vibrio_campbellii/apt_A/TP
NC_009783/2485245-2485332(+)/Vibrionaceae/Vibrio_campbellii/apt_B/TP
NC_011184/1633227-1633331(+)/Vibrionaceae/Vibrio_fischeri/apt_A/TP
NC_011184/1633339-1633430(+)/Vibrionaceae/Vibrio_fischeri/apt_B/TP
NC_011312/1917432-1917534(+)/Vibrionaceae/Aliivibrio_salmonicida/apt_A/TP
NC_011312/1917542-1917633(+)/Vibrionaceae/Aliivibrio_salmonicida/apt_B/TP
NC_013456/2228330-2228433(+)/Vibrionaceae/Vibrio_antiquarius/apt_A/TP

Appendix

NC_013456/2228441-2228526(+)/Vibrionaceae/Vibrio_antiquarius/apt_B/TP
NC_015633/1730387-1730501(-)/Vibrionaceae/Vibrio_anguillarum/apt_A/TP
NC_015633/1730300-1730379(-)/Vibrionaceae/Vibrio_anguillarum/apt_B/TP
NC_016602/2046034-2046152(+)/Vibrionaceae/Vibrio_furnissii/apt_A/TP
NC_016602/2046160-2046250(+)/Vibrionaceae/Vibrio_furnissii/apt_B/TP
NC_016613/1821150-1821253(+)/Vibrionaceae/Vibrio_sp/apt_A/TP
NC_016613/1821261-1821347(+)/Vibrionaceae/Vibrio_sp/apt_B/TP
NC_017270/1416420-1416545(-)/Vibrionaceae/Vibrio_cholerae/apt_A/TP
NC_017270/1416327-1416412(-)/Vibrionaceae/Vibrio_cholerae/apt_B/TP
NC_022349/470980-471083(-)/Vibrionaceae/Vibrio_alginolyticus/apt_A/TP
NC_022349/470887-470972(-)/Vibrionaceae/Vibrio_alginolyticus/apt_B/TP
NC_022528/1969588-1969662(+)/Vibrionaceae/Vibrio_nigripulchritudo/apt_A/TP
NC_022528/1969670-1969749(+)/Vibrionaceae/Vibrio_nigripulchritudo/apt_B/TP
NZ_AOCM01000113/12937-13061(-)/Vibrionaceae/Vibrio_vulnificus/apt_A/TP
NZ_AOCM01000113/12844-12929(-)/Vibrionaceae/Vibrio_vulnificus/apt_B/TP
NZ_CP009261/1622741-1622865(-)/Vibrionaceae/Vibrio_vulnificus/apt_A/TP
NZ_CP009261/1622648-1622733(-)/Vibrionaceae/Vibrio_vulnificus/apt_B/TP
NZ_CP009264/1183435-1183538(-)/Vibrionaceae/Vibrio_coralliilyticus/apt_A/TP
NZ_CP009264/1183337-1183427(-)/Vibrionaceae/Vibrio_coralliilyticus/apt_B/TP
NZ_CP009354/1442249-1442352(-)/Vibrionaceae/Vibrio_tubiashii/apt_A/TP
NZ_CP009354/1442161-1442241(-)/Vibrionaceae/Vibrio_tubiashii/apt_B/TP
NZ_CP009847/1884818-1884921(+)/Vibrionaceae/Vibrio_parahaemolyticus/apt_A/TP
NZ_CP009847/1884929-1885016(+)/Vibrionaceae/Vibrio_parahaemolyticus/apt_B/TP
NZ_CP011406/1684600-1684703(-)/Vibrionaceae/Vibrio_parahaemolyticus/apt_A/TP
NZ_CP011406/1684505-1684592(-)/Vibrionaceae/Vibrio_parahaemolyticus/apt_B/TP
NZ_CP011460/1509768-1509882(+)/Vibrionaceae/Vibrio_anguillarum/apt_A/TP
NZ_CP011460/1509890-1509969(+)/Vibrionaceae/Vibrio_anguillarum/apt_B/TP
NZ_CP013484/550060-550163(-)/Vibrionaceae/Vibrio_alginolyticus/apt_A/TP
NZ_CP013484/549967-550052(-)/Vibrionaceae/Vibrio_alginolyticus/apt_B/TP
NZ_CP014035/1314775-1314893(+)/Vibrionaceae/Vibrio_fluvialis/apt_A/TP
NZ_CP014035/1314901-1314991(+)/Vibrionaceae/Vibrio_fluvialis/apt_B/TP
NZ_CP014038/335940-336043(+)/Vibrionaceae/Vibrio_harveyi/apt_A/TP
NZ_CP014038/336051-336138(+)/Vibrionaceae/Vibrio_harveyi/apt_B/TP
NZ_CP014043/2211708-2211832(-)/Vibrionaceae/Vibrio_mimicus/apt_A/TP
NZ_CP014043/2211615-2211700(-)/Vibrionaceae/Vibrio_mimicus/apt_B/TP
NZ_CP016177/1449576-1449672(+)/Vibrionaceae/Vibrio_breoganii/apt_A/TP

NZ_CP016177/1449680-1449772(+)/Vibrionaceae/Vibrio_breoganii/apt_B/TP
NZ_CP016307/1666914-1667019(-)/Vibrionaceae/Vibrio_scophthalmi/apt_A/TP
NZ_CP016307/1666813-1666906(-)/Vibrionaceae/Vibrio_scophthalmi/apt_B/TP
NZ_CP016324/1619629-1619754(+)/Vibrionaceae/Vibrio_cholerae/apt_A/TP
NZ_CP016324/1619762-1619847(+)/Vibrionaceae/Vibrio_cholerae/apt_B/TP
NZ_CP016345/1812579-1812682(+)/Vibrionaceae/Vibrio_natriegens/apt_A/TP
NZ_CP016345/1812690-1812776(+)/Vibrionaceae/Vibrio_natriegens/apt_B/TP
NZ_CP016347/1636869-1636972(-)/Vibrionaceae/Vibrio_natriegens/apt_A/TP
NZ_CP016347/1636775-1636861(-)/Vibrionaceae/Vibrio_natriegens/apt_B/TP
NZ_CP016351/1896939-1897042(+)/Vibrionaceae/Vibrio_natriegens/apt_A/TP
NZ_CP016351/1897050-1897136(+)/Vibrionaceae/Vibrio_natriegens/apt_B/TP
NZ_CP016556/3410084-3410187(-)/Vibrionaceae/Vibrio_coralliilyticus/apt_A/TP
NZ_CP016556/3409986-3410076(-)/Vibrionaceae/Vibrio_coralliilyticus/apt_B/TP
NZ_LGLS01000001/580810-580913(+)/Vibrionaceae/Vibrio_coralliilyticus/apt_A/TP
NZ_LGLS01000001/580921-581011(+)/Vibrionaceae/Vibrio_coralliilyticus/apt_B/TP
NZ_LN554846/1395791-1395893(-)/Vibrionaceae/Aliivibrio_wodanis/apt_A/TP
NZ_LN554846/1395692-1395783(-)/Vibrionaceae/Aliivibrio_wodanis/apt_B/TP

Table 2.6: 782 glycine riboswitch singleton aptamer sequences labeled by type.

NC_002745/1576871-1576958(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_002758/1653283-1653370(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_002944/1693368-1693455(+)/Actinobacteria/Mycobacterium_avium/singlet_0/GCV
NC_002951/1629949-1630036(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_002952/1687639-1687726(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_002953/1602173-1602260(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_003923/1622418-1622505(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_004116/1147549-1147671(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NC_004350/1115970-1116039(+)/Firmicutes/Streptococcus_mutans/singlet_0/TP
NC_004368/1250645-1250767(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NC_006177/3096082-3096178(-)/Firmicutes/Symbiobacterium_thermophilum/singlet_0/OTHER
NC_007347/3638456-3638534(+)/Proteobacteria/Ralstonia_eutropha/singlet_0/GCV
NC_007432/1203256-1203378(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NC_007614/228143-228238(-)/Proteobacteria/Nitrosospira_multiformis/singlet_0/GCV
NC_007622/1532050-1532137(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV

NC_007793/1652785-1652872(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_007795/1555988-1556075(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_007907/4798232-4798312(-)/Firmicutes/Desulfitobacterium_hafniense/singlet_0/OTHER
NC_007973/3761303-3761380(+)/Proteobacteria/Cupriavidus_metallidurans/singlet_0/GCV
NC_007974/2332733-2332820(+)/Proteobacteria/Cupriavidus_metallidurans/singlet_0/GLY_MET
NC_008146/3007263-3007351(-)/Actinobacteria/Mycobacterium_sp/singlet_0/GCV
NC_008268/963522-963635(-)/Actinobacteria/Rhodococcus_jostii/singlet_0/GCV
NC_008344/2070207-2070293(-)/Proteobacteria/Nitrosomonas_eutropha/singlet_0/GCV
NC_008705/3025151-3025239(-)/Actinobacteria/Mycobacterium_sp/singlet_0/GCV
NC_008752/3470675-3470795(+)/Proteobacteria/Acidovorax_citrulli/singlet_0/GCV
NC_008752/5005963-5006060(+)/Proteobacteria/Acidovorax_citrulli/singlet_0/GLY_MET
NC_009077/2998451-2998539(-)/Actinobacteria/Mycobacterium_sp/singlet_0/GCV
NC_009380/2673652-2673741(-)/Actinobacteria/Salinispora_tropica/singlet_0/GCV
NC_009487/1701651-1701738(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_009632/1701525-1701612(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_009633/3462240-3462320(+)/Firmicutes/Alkaliphilus_metalliredigens/singlet_0/OTHER
NC_009641/1612787-1612874(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_009706/769790-769871(-)/Firmicutes/Clostridium_kluyveri/singlet_0/OTHER
NC_009782/1654683-1654770(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_009831/4748515-4748664(-)/Proteobacteria/Shewanella_sediminis/singlet_0/TP
NC_009953/2875689-2875778(-)/Actinobacteria/Salinispora_arenicola/singlet_0/GCV
NC_010002/4732127-4732235(+)/Proteobacteria/Delftia_acidovorans/singlet_0/GCV
NC_010079/1653375-1653462(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_010506/4619903-4620018(-)/Proteobacteria/Shewanella_woodyi/singlet_0/TP
NC_010524/3723425-3723507(+)/Proteobacteria/Leptothrix_cholodnii/singlet_0/TP
NC_011663/1002919-1003055(+)/Proteobacteria/Shewanella_baltica/singlet_0/TP
NC_011753/1389794-1389914(-)/Proteobacteria/Vibrio_tasmaniensis/singlet_0/TP
NC_011830/1219406-1219486(+)/Firmicutes/Desulfitobacterium_hafniense/singlet_0/OTHER
NC_011837/769765-769846(-)/Firmicutes/Clostridium_kluyveri/singlet_0/OTHER
NC_011999/1288490-1288565(-)/Firmicutes/Macrococcus_caseolyticus/singlet_0/GCV
NC_012522/664679-664786(-)/Actinobacteria/Rhodococcus_opacus/singlet_0/GCV
NC_012791/822769-822863(-)/Proteobacteria/Variovorax_paradoxus/singlet_0/GLY_MET

Appendix

NC_012803/318030-318136(+)/Actinobacteria/Micrococcus_luteus/singlet_0/GCV
NC_012943/2347220-2347307(-)/Actinobacteria/Mycobacterium_tuberculosis/singlet_0/GCV
NC_013093/6369333-6369415(-)/Actinobacteria/Actinosynnema_mirum/singlet_0/GCV
NC_013165/2795479-2795572(-)/Actinobacteria/Slackia_heliotrinireducens/singlet_0/TP
NC_013441/2957852-2957947(-)/Actinobacteria/Gordonia_bronchialis/singlet_0/GCV
NC_013450/1591119-1591206(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_013521/46303-46404(+)/Actinobacteria/Sanguibacter_keddieii/singlet_0/GCV
NC_013893/1413864-1413953(+)/Firmicutes/Staphylococcus_lugdunensis/singlet_0/GCV
NC_013928/910302-910371(-)/Firmicutes/Streptococcus_mutans/singlet_0/TP
NC_014011/1428648-1428755(-)/Synergistetes/Aminobacterium_colombiense/singlet_0/OTHER
NC_014165/1085978-1086065(-)/Actinobacteria/Thermobispora_bispora/singlet_0/GCV
NC_014328/3228387-3228466(+)/Firmicutes/Clostridium_ljungdahlii/singlet_0/OTHER
NC_014483/4053479-4053557(-)/Firmicutes/Paenibacillus_polymyxa/singlet_0/TP
NC_014539/3950709-3950800(+)/Proteobacteria/Burkholderia_sp/singlet_0/GCV
NC_014624/1317373-1317451(+)/Firmicutes/Eubacterium_limosum/singlet_0/TP
NC_014640/6045891-6046001(-)/Proteobacteria/Achromobacter_xylosoxidans/singlet_0/GCV
NC_014925/1304501-1304586(-)/Firmicutes/Staphylococcus_pseudintermedius/singlet_0/GCV
NC_015138/5036675-5036772(+)/Proteobacteria/Acidovorax_avenae/singlet_0/GLY_MET
NC_015291/1570417-1570493(-)/Firmicutes/Streptococcus_oralis/singlet_0/TP
NC_015385/888585-888673(+)/Spirochaetes/Treponema_succinifaciens/singlet_0/TP
NC_015433/243709-243803(+)/Firmicutes/Streptococcus_suis/singlet_0/TP
NC_015434/3603550-3603638(-)/Actinobacteria/Verrucosispora_maris/singlet_0/GCV
NC_015514/4173243-4173324(-)/Actinobacteria/Cellulomonas_fimi/singlet_0/GCV
NC_015563/2861620-2861728(-)/Proteobacteria/Delftia_sp/singlet_0/GCV
NC_015576/2096336-2096419(+)/Actinobacteria/Mycobacterium_sinense/singlet_0/GCV
NC_015760/1112053-1112149(-)/Firmicutes/Streptococcus_salivarius/singlet_0/TP
NC_016002/3438442-3438527(-)/Proteobacteria/Pseudogulbenkiania_sp/singlet_0/GCV
NC_016011/1352611-1352692(+)/Firmicutes/Listeria_ivanovii/singlet_0/GCV
NC_016048/120153-120225(+)/Firmicutes/Oscillibacter_valericigenes/singlet_0/TP

NC_016078/498931-499014(+)/Proteobacteria/Pelagibacterium_halotolerans/singlet_0/GCV
NC_016148/1361420-1361524(-)/Synergistetes/Thermovirga_lienii/singlet_0/OTHER
NC_016768/2343951-2344038(-)/Actinobacteria/Mycobacterium_tuberculosis/singlet_0/GCV
NC_016907/69039-69145(+)/Actinobacteria/Gordonia_polyisoprenivorans/singlet_0/GCV
NC_016912/1513616-1513703(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_016928/1623108-1623195(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_016941/1539556-1539642(-)/Firmicutes/Staphylococcus_argenteus/singlet_0/GCV
NC_017093/4414241-4414322(-)/Actinobacteria/Actinoplanes_missouriensis/singlet_0/GCV
NC_017331/1666967-1667054(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017333/1690674-1690761(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017337/1612311-1612397(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017338/1613148-1613235(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017340/1611422-1611509(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017341/1624788-1624875(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017342/1697490-1697577(+)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017343/1531289-1531376(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017347/1625084-1625171(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017349/1601107-1601194(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017351/1651885-1651972(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017353/1443456-1443545(+)/Firmicutes/Staphylococcus_lugdunensis/singlet_0/GCV
NC_017673/1511921-1512008(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017763/1615934-1616021(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_017768/922860-922929(-)/Firmicutes/Streptococcus_mutans/singlet_0/TP
NC_018017/1041279-1041359(+)/Firmicutes/Desulfitobacterium_dehalogenans/singlet_0/OTHER
NC_018027/2851539-2851625(+)/Actinobacteria/Mycobacterium_chubuense/singlet_0/GCV
NC_018078/2347102-2347189(-)/Actinobacteria/Mycobacterium_tuberculosis/singlet_0/GCV
NC_018089/1098316-1098385(+)/Firmicutes/Streptococcus_mutans/singlet_0/TP
NC_018524/4297874-4297954(+)/Actinobacteria/Nocardiopsis_alba/singlet_0/GCV
NC_018525/2294193-2294276(+)/Proteobacteria/Pectobacterium_carotovorum/singlet_0/TP
NC_018581/3026835-3026926(-)/Actinobacteria/Gordonia_sp/singlet_0/GCV
NC_018608/1623681-1623768(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_018646/1157511-1157633(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP

Appendix

NC_018697/2049640-2049718(-)/Proteobacteria/Cycloclasticus_sp/singlet_0/GCV
NC_020529/1555408-1555495(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_020532/1555708-1555795(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_020533/1555344-1555431(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_020536/1555390-1555477(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_020537/1555161-1555248(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_020561/288142-288247(-)/Proteobacteria/Sphingomonas_sp/singlet_0/GCV
NC_020564/1555397-1555484(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_020566/1555349-1555436(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_020568/1555213-1555300(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_021059/1636473-1636560(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_021182/4174192-4174272(-)/Firmicutes/Clostridium_pasteurianum/singlet_0/OTHER
NC_021200/2515791-2515878(-)/Actinobacteria/Mycobacterium_avium/singlet_0/GCV
NC_021285/3116831-3116941(+)/Proteobacteria/Achromobacter_xylosoxidans/singlet_0/GCV
NC_021485/1173029-1173151(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NC_021486/1150859-1150981(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NC_021554/1626386-1626473(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_021670/1729549-1729636(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_021917/478326-478404(+)/Proteobacteria/Cycloclasticus_zancles/singlet_0/GCV
NC_022113/1554452-1554539(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_022222/1526428-1526515(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_022226/1543673-1543760(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_022442/1623196-1623283(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_022443/1562605-1562692(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_022516/250292-250386(+)/Firmicutes/Streptococcus_suis/singlet_0/TP
NC_022567/2147388-2147481(+)/Actinobacteria/Adlercreutzia_equolifaciens/singlet_0/TP
NC_022571/3840542-3840615(-)/Firmicutes/Clostridium_saccharobutylicum/singlet_0/TP
NC_022592/986597-986676(+)/Firmicutes/Clostridium_autoethanogenum/singlet_0/OTHER
NC_022604/1668147-1668234(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_022665/214056-214150(-)/Firmicutes/Streptococcus_suis/singlet_0/TP
NZ_AP014652/1559495-1559582(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_AP014653/1559495-1559582(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV

Appendix

NZ_AP014942/1571963-1572050(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_AP017320/1654444-1654531(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CCSE01000001/796366-796446(+)/Firmicutes/Jeotgalicoccus_sp/singlet_0/GCV
NZ_CM000952/2219376-2219463(+)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CM001149/756331-756427(+)/Actinobacteria/Rhodococcus_equi/singlet_0/GCV
NZ_CM001438/3989728-3989832(+)/Proteobacteria/Burkholderiales_bacterium/singlet_0/GCV
NZ_CM001487/679472-679559(-)/Firmicutes/Eubacterium_cellulosolvens/singlet_0/TP
NZ_CM002132/1194545-1194641(-)/Firmicutes/Streptococcus_sp/singlet_0/TP
NZ_CM003594/2250613-2250740(+)/Proteobacteria/Psychrobacter_sp/singlet_0/GCV
NZ_CP002007/243475-243569(+)/Firmicutes/Streptococcus_suis/singlet_0/TP
NZ_CP003949/5172988-5173101(-)/Actinobacteria/Rhodococcus_opacus/singlet_0/GCV
NZ_CP004078/5201331-5201397(-)/Firmicutes/Paenibacillus_sabinae/singlet_0/TP
NZ_CP005974/1990877-1990956(-)/Proteobacteria/Photobacterium_gaetbulicola/singlet_0/TP
NZ_CP006630/1610565-1610652(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP006910/1152980-1153102(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP007016/1115853-1115922(+)/Firmicutes/Streptococcus_mutans/singlet_0/TP
NZ_CP007032/2758647-2758728(-)/Firmicutes/Desulfitobacterium_metallireducens/singlet_0/OTHER
NZ_CP007172/1399240-1399321(+)/Firmicutes/Listeria_ivanovii/singlet_0/GCV
NZ_CP007176/1659141-1659228(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP007208/1035762-1035858(-)/Firmicutes/Staphylococcus_xylosus/singlet_0/GCV
NZ_CP007220/2402285-2402368(+)/Actinobacteria/Mycobacterium_chelonae/singlet_0/GCV
NZ_CP007437/318053-318159(+)/Actinobacteria/Micrococcus_luteus/singlet_0/GCV
NZ_CP007454/2313901-2313988(+)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP007499/1729687-1729774(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP007539/1642331-1642418(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP007570/1204598-1204720(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP

Appendix

NZ_CP007571/1234769-1234891(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP007572/1230979-1231101(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP007631/1201993-1202115(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP007632/1123273-1123395(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP007657/1635805-1635892(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP007659/1630121-1630208(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP007670/1634465-1634552(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP007672/1622092-1622179(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP007674/1618964-1619051(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP007676/1623708-1623795(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP007690/1653372-1653459(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP007739/592729-592792(-)/Firmicutes/Bacillus_methanolicus/singlet_0/OTHER
NZ_CP008724/1368020-1368116(+)/Firmicutes/Staphylococcus_xylosus/singlet_0/GCV
NZ_CP009267/3593879-3593958(-)/Firmicutes/Clostridium_pasteurianum/singlet_0/OTHER
NZ_CP009268/3594087-3594166(-)/Firmicutes/Clostridium_pasteurianum/singlet_0/OTHER
NZ_CP009361/1609070-1609157(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP009448/5405172-5405282(-)/Proteobacteria/Achromobacter_xylosoxidans/singlet_0/GCV
NZ_CP009554/1657826-1657913(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP009575/1447806-1447887(+)/Firmicutes/Listeria_ivanovii/singlet_0/GCV
NZ_CP009576/1526057-1526138(+)/Firmicutes/Listeria_ivanovii/singlet_0/GCV
NZ_CP009577/1399236-1399317(+)/Firmicutes/Listeria_ivanovii/singlet_0/GCV
NZ_CP009623/728092-728173(-)/Firmicutes/Staphylococcus_agnetis/singlet_0/GCV
NZ_CP009681/866776-866863(+)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP009687/4000376-4000454(+)/Firmicutes/Clostridium_aceticum/singlet_0/OTHER

NZ_CP009828/1670652-1670739(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP010113/2488504-2488591(-)/Actinobacteria/Mycobacterium_avium/singlet_0/GCV
NZ_CP010114/2491211-2491298(-)/Actinobacteria/Mycobacterium_avium/singlet_0/GCV
NZ_CP010295/1652782-1652869(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP010296/1652782-1652869(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP010297/1652782-1652869(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP010298/1652781-1652868(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP010299/1652782-1652869(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP010300/1652782-1652869(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP010319/1086894-1087016(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP010402/1597648-1597735(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP010526/1691522-1691609(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP010867/1172824-1172946(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP010890/1583286-1583373(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP010946/2337694-2337777(+)/Actinobacteria/Mycobacterium_chelonae/singlet_0/GCV
NZ_CP010989/2365768-2365853(+)/Actinobacteria/Pseudonocardia_sp/singlet_0/GCV
NZ_CP010998/781764-781851(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP011129/187548-187644(+)/Proteobacteria/Lysobacter_antibioticus/singlet_0/OTHER
NZ_CP011147/1558608-1558695(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP011325/1158749-1158871(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP011326/1152893-1153015(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP011327/1158545-1158667(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP011329/1145363-1145485(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP

NZ_CP011366/1420311-1420389(+)/Firmicutes/Salinicoccus_halodurans/singlet_0/GCV
NZ_CP011419/245186-245280(+)/Firmicutes/Streptococcus_suis/singlet_0/TP
NZ_CP011512/4592275-4592353(-)/Firmicutes/Paenibacillus_peoriae/singlet_0/TP
NZ_CP011526/1575300-1575387(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP011528/1527333-1527420(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP011685/1630065-1630152(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP011773/3545398-3545483(-)/Actinobacteria/Mycobacterium_sp/singlet_0/GCV
NZ_CP011774/1340505-1340586(-)/Proteobacteria/Limnohabitans_sp/singlet_0/GLY_MET
NZ_CP011834/2347795-2347915(+)/Proteobacteria/Limnohabitans_sp/singlet_0/GCV
NZ_CP011883/2241095-2241182(+)/Actinobacteria/Mycobacterium_haemophilum/singlet_0/GCV
NZ_CP011914/2725783-2725861(-)/Firmicutes/Eubacterium_limosum/singlet_0/TP
NZ_CP012011/1781179-1781266(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012012/1726570-1726657(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012013/1736933-1737020(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012015/1732437-1732524(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012018/1731700-1731787(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012076/701238-701337(+)/Proteobacteria/Bordetella_hinzii/singlet_0/GCV
NZ_CP012077/645121-645220(+)/Proteobacteria/Bordetella_hinzii/singlet_0/GCV
NZ_CP012119/270000-270087(+)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012120/2168577-2168664(+)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012181/4322133-4322218(-)/Actinobacteria/Pseudonocardia_sp/singlet_0/GCV
NZ_CP012184/6038170-6038255(-)/Actinobacteria/Pseudonocardia_sp/singlet_0/GCV
NZ_CP012192/3970406-3970496(+)/Proteobacteria/Burkholderia_sp/singlet_0/GCV
NZ_CP012371/1581942-1582038(+)/Proteobacteria/Nitrosospira_briensis/singlet_0/GCV
NZ_CP012395/986662-986741(+)/Firmicutes/Clostridium_autoethanogenum/singlet_0/OTHER
NZ_CP012409/2776173-2776260(+)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV

NZ_CP012480/1169820-1169942(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP012593/1581491-1581578(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012678/2425320-2425441(-)/Proteobacteria/Psychrobacter_urativorans/singlet_0/GCV
NZ_CP012731/815353-815447(-)/Firmicutes/Streptococcus_suis/singlet_0/TP
NZ_CP012756/103385-103472(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012911/1961589-1961683(-)/Firmicutes/Streptococcus_suis/singlet_0/TP
NZ_CP012968/415792-415888(+)/Firmicutes/Staphylococcus_sp/singlet_0/GCV
NZ_CP012970/1574816-1574903(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012972/1574309-1574396(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012974/1533466-1533553(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012976/1533479-1533566(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012978/1525541-1525628(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP012979/1526101-1526188(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP013114/1423770-1423866(-)/Firmicutes/Staphylococcus_equorum/singlet_0/GCV
NZ_CP013137/1092897-1092984(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP013140/161181-161275(+)/Proteobacteria/Lysobacter_enzymogenes/singlet_0/OTHER
NZ_CP013141/3630613-3630709(-)/Proteobacteria/Lysobacter_antibioticus/singlet_0/OTHER
NZ_CP013202/1197748-1197870(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_CP013216/1043399-1043495(+)/Firmicutes/Streptococcus_salivarius/singlet_0/TP
NZ_CP013231/682582-682669(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP013237/508672-508741(-)/Firmicutes/Streptococcus_mutans/singlet_0/TP
NZ_CP013254/3165306-3165401(-)/Actinobacteria/Kocuria_flava/singlet_0/GCV
NZ_CP013341/2305779-2305868(+)/Proteobacteria/Nitrosomonas_ureae/singlet_0/GCV
NZ_CP013616/1623447-1623534(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP013619/1609543-1609630(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP013621/1686365-1686452(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV

NZ_CP013714/1387545-1387641(-)/Firmicutes/Staphylococcus_equorum/singlet_0/GCV
NZ_CP013729/3489767-3489895(+)/Proteobacteria/Roseateles_depolymerans/singlet_0/GCV
NZ_CP013923/5543579-5543689(-)/Proteobacteria/Achromobacter_denitrificans/singlet_0/GCV
NZ_CP013953/1714121-1714208(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP013955/1718779-1718866(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP013957/1668765-1668852(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP013959/1711045-1711132(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP013980/1355646-1355742(-)/Firmicutes/Staphylococcus_equorum/singlet_0/GCV
NZ_CP014016/809402-809511(+)/Firmicutes/Staphylococcus_simulans/singlet_0/GCV
NZ_CP014022/252064-252153(+)/Firmicutes/Staphylococcus_lugdunensis/singlet_0/GCV
NZ_CP014023/568595-568684(-)/Firmicutes/Staphylococcus_lugdunensis/singlet_0/GCV
NZ_CP014060/5033118-5033228(+)/Proteobacteria/Achromobacter_xylosoxidans/singlet_0/GCV
NZ_CP014064/778987-779074(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP014170/1125000-1125078(-)/Firmicutes/Clostridium_tyrobutyricum/singlet_0/OTHER
NZ_CP014480/1270197-1270294(-)/Actinobacteria/Kocuria_turfanensis/singlet_0/GCV
NZ_CP014672/2148803-2148894(-)/Firmicutes/Clostridium_stercorarium/singlet_0/TP
NZ_CP014791/1516370-1516457(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP014951/2428835-2428920(+)/Actinobacteria/Mycobacterium_abscessus/singlet_0/GCV
NZ_CP014953/2295564-2295649(+)/Actinobacteria/Mycobacterium_abscessus/singlet_0/GCV
NZ_CP014954/2396053-2396138(+)/Actinobacteria/Mycobacterium_abscessus/singlet_0/GCV
NZ_CP014958/2423994-2424079(+)/Actinobacteria/Mycobacterium_abscessus/singlet_0/GCV
NZ_CP015173/2233592-2233679(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP015405/4901775-4901855(+)/Firmicutes/Blautia_sp/singlet_0/TP

NZ_CP015427/1347666-1347766(-)/Proteobacteria/Haemophilus_ducreyi/singlet_0/TP
NZ_CP015428/643654-643754(-)/Proteobacteria/Haemophilus_ducreyi/singlet_0/TP
NZ_CP015432/1449905-1450005(-)/Proteobacteria/Haemophilus_ducreyi/singlet_0/TP
NZ_CP015495/1693368-1693455(+)/Actinobacteria/Mycobacterium_avium/singlet_0/GCV
NZ_CP015645/550638-550725(+)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_CP015646/1688267-1688354(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_HG939456/1104375-1104497(+)/Firmicutes/Streptococcus_agalactiae/singlet_0/TP
NZ_LN554852/4482985-4483070(-)/Proteobacteria/Moritella_viscosa/singlet_0/TP
NZ_LN554884/1362844-1362940(+)/Firmicutes/Staphylococcus_xylosus/singlet_0/GCV
NZ_LN626917/1634208-1634295(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_LN831036/1602201-1602288(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_LN831049/1879212-1879299(+)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_LN854556/1576730-1576817(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_LT009690/1621930-1622017(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NZ_LT546645/1868974-1869075(+)/Proteobacteria/Bordetella_trematum/singlet_0/GCV
NZ_LT575468/1507980-1508061(+)/Proteobacteria/Plesiomonas_shigelloides/singlet_0/TP
NZ_LT598688/1612788-1612875(-)/Firmicutes/Staphylococcus_aureus/singlet_0/GCV
NC_000962/2075620-2075703(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_002755/2072845-2072928(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_002945/2065972-2066055(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NC_002973/1352550-1352631(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_002976/1146875-1147007(-)/Firmicutes/Staphylococcus_epidermidis/singlet_1/GCV
NC_003210/1372845-1372926(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_003212/1379631-1379712(+)/Firmicutes/Listeria_innocua/singlet_1/GCV
NC_004461/1258820-1258952(-)/Firmicutes/Staphylococcus_epidermidis/singlet_1/GCV

Appendix

NC_004557/1925417-1925514(+)/Firmicutes/Clostridium_tetani/singlet_1/TP
NC_007168/1408552-1408657(+)/Firmicutes/Staphylococcus_haemolyticus/singlet_1/GCV
NC_007333/1619925-1620006(+)/Actinobacteria/Thermobifida_fusca/singlet_1/GCV
NC_007350/1259149-1259250(+)/Firmicutes/Staphylococcus_saprophyticus/singlet_1/GCV
NC_008313/3890153-3890230(+)/Proteobacteria/Ralstonia_eutropha/singlet_1/GCV
NC_008555/1351694-1351775(+)/Firmicutes/Listeria_welshimeri/singlet_1/GCV
NC_008699/2884572-2884656(-)/Actinobacteria/Nocardioides_sp/singlet_1/GCV
NC_008769/2076575-2076658(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NC_009524/2129190-2129317(+)/Proteobacteria/Psychrobacter_sp/singlet_1/GCV
NC_009525/2085552-2085635(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_009565/2083816-2083899(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_009664/3908899-3908989(+)/Actinobacteria/Kineococcus_radiotolerans/singlet_1/GCV
NC_010528/3242798-3242875(+)/Proteobacteria/Cupriavidus_taiwanensis/singlet_1/GCV
NC_010617/650982-651087(+)/Actinobacteria/Kocuria_rhizophila/singlet_1/GCV
NC_010622/3367073-3367173(+)/Proteobacteria/Paraburkholderia_phymatum/singlet_1/GCV
NC_011566/4572393-4572508(-)/Proteobacteria/Shewanella_piezotolerans/singlet_1/TP
NC_011660/1267236-1267317(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_011899/2563206-2563288(-)/Firmicutes/Halothermothrix_orenii/singlet_1/GCV
NC_012121/1177746-1177867(-)/Firmicutes/Staphylococcus_carnosus/singlet_1/GCV
NC_012207/2058566-2058649(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NC_012488/1356540-1356621(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_013169/1537225-1537300(-)/Actinobacteria/Kytococcus_sedentarius/singlet_1/GCV
NC_013235/4424245-4424338(-)/Actinobacteria/Nakamurella_multipartita/singlet_1/GCV
NC_013522/1198865-1198950(-)/Synergistetes/Thermanaerovibrio_acidaminovorans/singlet_1/OTHER
NC_013729/3388461-3388546(+)/Actinobacteria/Kribbella_flavida/singlet_1/GCV
NC_013729/7009633-7009726(-)/Actinobacteria/Kribbella_flavida/singlet_1/GLY_MET
NC_013766/1441837-1441918(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_013768/1408537-1408618(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_013891/1282081-1282162(+)/Firmicutes/Listeria_seeligeri/singlet_1/GCV

NC_014147/1271717-1271825(+)/Proteobacteria/Moraxella_catarrhalis/singlet_1/GCV
NC_014158/2430914-2431005(-)/Actinobacteria/Tsukamurella_paurometabola/singlet_1/GCV
NC_014210/1803416-1803495(+)/Actinobacteria/Nocardiopsis_dassonvillei/singlet_1/GCV
NC_014328/919839-919911(+)/Firmicutes/Clostridium_ljungdahlii/singlet_1/OTHER
NC_014391/3621234-3621314(-)/Actinobacteria/Micromonospora_aurantiaca/singlet_1/GCV
NC_014541/1102873-1102953(+)/Proteobacteria/Ferrimonas_balearica/singlet_1/TP
NC_014614/1261555-1261643(+)/Firmicutes/Clostridium_sticklandii/singlet_1/GCV
NC_014654/2603079-2603154(-)/Firmicutes/Halanaerobium_hydrogeniformans/singlet_1/GCV
NC_014659/2606010-2606106(-)/Actinobacteria/Rhodococcus_equi/singlet_1/GCV
NC_014815/5602963-5603043(+)/Actinobacteria/Micromonospora_sp/singlet_1/GCV
NC_014830/2048165-2048269(-)/Actinobacteria/Intrasporangium_calvum/singlet_1/GCV
NC_015145/836223-836311(+)/Actinobacteria/Pseudarthrobacter_phenanthrenivorans/singlet_1/GCV
NC_015588/56386-56480(+)/Actinobacteria/Isoptericola_variabilis/singlet_1/GCV
NC_015635/2553274-2553364(+)/Actinobacteria/Microlunatus_phosphovorus/singlet_1/GCV
NC_015726/3709013-3709090(+)/Proteobacteria/Cupriavidus_necator/singlet_1/GCV
NC_015731/1342704-1342794(+)/Proteobacteria/Nitrosomonas_sp/singlet_1/GCV
NC_015758/2074938-2075021(+)/Actinobacteria/Mycobacterium_africanum/singlet_1/GCV
NC_015848/2114061-2114144(+)/Actinobacteria/Mycobacterium_canettii/singlet_1/GCV
NC_016109/1692732-1692828(-)/Actinobacteria/Kitasatospora_setae/singlet_1/GCV
NC_016513/230264-230398(+)/Proteobacteria/Aggregatibacter_actinomycetemcomitans/singlet_1/TP
NC_016804/2055940-2056023(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NC_016934/2078881-2078964(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_017522/2056354-2056437(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_017524/2069329-2069412(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_017529/1388681-1388762(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_017537/1386640-1386721(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_017544/1331767-1331848(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_017545/1373235-1373316(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV

NC_017546/1393231-1393312(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_017547/1352970-1353051(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_017568/1287237-1287322(+)/Firmicutes/Staphylococcus_pseudintermedius/singlet_1/GCV
NC_017728/1349378-1349459(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_017803/4829529-4829605(-)/Actinobacteria/Actinoplanes_sp/singlet_1/GCV
NC_018143/2075627-2075710(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_018150/2505536-2505623(-)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NC_018584/1364860-1364941(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_018585/1353552-1353633(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_018586/1382084-1382165(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_018587/1356771-1356852(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_018588/1393227-1393308(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_018589/1392887-1392968(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_018590/1320833-1320914(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_018591/1390987-1391068(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_018592/1374860-1374941(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_018593/1319489-1319570(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_018642/1356355-1356436(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_019048/1005779-1005901(+)/Firmicutes/Streptococcus_agalactiae/singlet_1/TP
NC_019556/1352548-1352629(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_019950/2086421-2086504(+)/Actinobacteria/Mycobacterium_canettii/singlet_1/GCV
NC_019951/2159215-2159298(+)/Actinobacteria/Mycobacterium_canettii/singlet_1/GCV
NC_019965/2093962-2094045(+)/Actinobacteria/Mycobacterium_canettii/singlet_1/GCV
NC_019966/3204883-3204969(+)/Actinobacteria/Mycobacterium_sp/singlet_1/GCV
NC_020089/2079393-2079476(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_020164/1270725-1270813(+)/Firmicutes/Staphylococcus_warneri/singlet_1/GCV
NC_020245/2046730-2046813(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NC_020559/2066439-2066522(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_021054/2075427-2075510(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_021191/4712561-4712720(-)/Actinobacteria/Actinoplanes_sp/singlet_1/GCV
NC_021194/2069847-2069930(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV

NC_021251/2060903-2060986(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_021282/2389520-2389605(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NC_021740/2070201-2070284(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_021823/867912-867993(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_021824/1907893-1907974(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_021825/1820128-1820209(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_021826/222153-222234(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_021827/1828645-1828726(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_021829/1160078-1160159(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_021830/2722801-2722882(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_021837/2101751-2101832(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_021838/1329340-1329421(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_021839/1117419-1117500(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_021840/1225423-1225504(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_022350/2074869-2074952(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_022513/4333303-4333379(+)/Proteobacteria/Ralstonia_pickettii/singlet_1/GCV
NC_022568/1375351-1375432(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_022592/3168961-3169033(+)/Firmicutes/Clostridium_autoethanogenum/singlet_1/OTHER
NC_022657/4614050-4614129(+)/Actinobacteria/Actinoplanes_friuliensis/singlet_1/GCV
NC_022737/146880-146968(-)/Firmicutes/Staphylococcus_pasteuri/singlet_1/GCV
NC_022777/1891808-1891905(+)/Firmicutes/Clostridium_tetani/singlet_1/TP
NZ_AM412059/2058492-2058575(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NZ_ANOI01000001/1493333-1493418(+)/Firmicutes/Staphylococcus_pseudintermedius/singlet_1/GCV
NZ_AP012555/2644236-2644323(+)/Actinobacteria/Mycobacterium_avium/singlet_1/GCV
NZ_AP014520/483704-483838(-)/Proteobacteria/Aggregatibacter_actinomycetemcomitans/singlet_1/TP
NZ_AP014547/2368726-2368811(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_AP014573/2081080-2081163(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_AP014944/1354206-1354289(+)/Firmicutes/Staphylococcus_schleiferi/singlet_1/GCV
NZ_CM000787/2065508-2065591(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV

Appendix

NZ_CM000788/2065507-2065590(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM000789/2065507-2065590(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM001024/1121044-1121126(+)/Actinobacteria/Aeromicrobium_marinum/singlet_1/GCV
NZ_CM001043/2056331-2056414(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM001044/2056292-2056375(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM001045/2056514-2056597(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM001049/1322031-1322112(+)/Firmicutes/Listeria_innocua/singlet_1/GCV
NZ_CM001051/1453677-1453758(+)/Firmicutes/Listeria_seeligeri/singlet_1/GCV
NZ_CM001159/1427889-1427970(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CM001225/2045484-2045567(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM001226/2060543-2060626(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM001227/2046648-2046731(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM001469/1032568-1032649(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CM001515/2076985-2077068(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM001852/1977937-1978047(-)/Actinobacteria/Nocardioides_sp/singlet_1/GCV
NZ_CM002022/2071084-2071167(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002048/2075480-2075563(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002049/2072509-2072592(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002050/2078811-2078894(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002051/2078804-2078887(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002052/2071162-2071245(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002053/2077363-2077446(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002054/2078844-2078927(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002055/2071440-2071523(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV

NZ_CM002056/2072460-2072543(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002057/2065917-2066000(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002058/2068025-2068108(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002059/2076094-2076177(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002060/2075959-2076042(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002061/2072479-2072562(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002062/2076208-2076291(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002063/2075408-2075491(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002064/2072468-2072551(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002065/2074201-2074284(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002066/2079726-2079809(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002067/2076117-2076200(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002068/2073710-2073793(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002069/2072059-2072142(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002070/2078757-2078840(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002071/2070789-2070872(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002072/2072505-2072588(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002073/2060215-2060298(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002074/2070071-2070154(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002075/2072072-2072155(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002076/2064020-2064103(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002077/2076290-2076373(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV

NZ_CM002078/2062882-2062965(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002079/2076889-2076972(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002080/2075574-2075657(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002097/2078768-2078851(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002098/2075512-2075595(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002099/2069972-2070055(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002100/2077146-2077229(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002101/2070268-2070351(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002102/2077718-2077801(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002103/2078133-2078216(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002104/2076043-2076126(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002105/2072477-2072560(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002106/2069500-2069583(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002107/2073463-2073546(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002108/2076034-2076117(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002109/2076913-2076996(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002110/2075800-2075883(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002111/2075610-2075693(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002112/2076048-2076131(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002113/2075590-2075673(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002114/2076081-2076164(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002115/2075438-2075521(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV

NZ_CM002116/2075960-2076043(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002117/2076092-2076175(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002118/2072439-2072522(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002119/2075122-2075205(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002120/2065529-2065612(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002121/2076102-2076185(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002122/2079585-2079668(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002123/2075114-2075197(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002124/2058255-2058338(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002125/2076072-2076155(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002126/2062896-2062979(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002127/2078762-2078845(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002882/2075024-2075107(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002883/2075630-2075713(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002884/2075010-2075093(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM003382/1369615-1369696(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CM003596/2175578-2175707(+)/Proteobacteria/Psychrobacter_sp/singlet_1/GCV
NZ_CP002871/2058144-2058227(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP002882/2062539-2062622(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP002883/2058860-2058943(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP002885/2059355-2059438(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP003494/1776220-1776303(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NZ_CP006858/1441837-1441918(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV

NZ_CP006859/1441837-1441918(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP006860/1441837-1441918(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP006861/1407420-1407501(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP006862/1403598-1403679(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP006874/1386193-1386274(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP007027/2075008-2075091(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP007160/142678-142759(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP007210/762293-762374(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP007299/2056719-2056802(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP007452/1705578-1705676(-)/Firmicutes/Eubacterium_acidaminophilum/singlet_1/OTHER
NZ_CP007452/1705900-1706000(-)/Firmicutes/Eubacterium_acidaminophilum/singlet_1/OTHER
NZ_CP007452/860145-860238(-)/Firmicutes/Eubacterium_acidaminophilum/singlet_1/OTHER
NZ_CP007492/1352548-1352629(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP007502/118131-118265(+)/Proteobacteria/Aggregatibacter_actinomycetemcomitans/singlet_1/TP
NZ_CP007600/1793559-1793640(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP007601/876997-877124(-)/Firmicutes/Staphylococcus_capitis/singlet_1/GCV
NZ_CP007669/1351979-1352087(+)/Proteobacteria/Moraxella_catarrhalis/singlet_1/GCV
NZ_CP007684/606488-606569(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP007685/1374036-1374117(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP007686/1390387-1390468(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP007687/1364940-1365021(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP007688/1375107-1375188(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP007689/609323-609404(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP007803/2046750-2046833(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV

NZ_CP007809/2053349-2053432(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP008744/2057565-2057648(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NZ_CP008747/1398221-1398305(+)/Firmicutes/Staphylococcus_hyicus/singlet_1/GCV
NZ_CP008889/1506750-1506835(+)/Actinobacteria/Dermacoccus_nishinomiyaensis/singlet_1/GCV
NZ_CP008984/118132-118266(+)/Proteobacteria/Aggregatibacter_actinomycetemcomitans/singlet_1/TP
NZ_CP009046/212645-212777(-)/Firmicutes/Staphylococcus_epidermidis/singlet_1/GCV
NZ_CP009100/2075612-2075695(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP009101/2075642-2075725(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP009111/4794790-4794903(-)/Actinobacteria/Rhodococcus_opacus/singlet_1/GCV
NZ_CP009242/1306699-1306780(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP009243/2058569-2058652(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NZ_CP009258/664631-664712(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP009426/2055073-2055156(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP009427/2082869-2082952(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP009447/2389487-2389574(-)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP009449/2035320-2035403(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NZ_CP009470/1286978-1287061(+)/Firmicutes/Staphylococcus_schleiferi/singlet_1/GCV
NZ_CP009480/2070921-2071004(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP009615/2505353-2505440(-)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP009616/2505500-2505587(-)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP009617/1903460-1903550(+)/Proteobacteria/Vibrio_coralliilyticus/singlet_1/TP
NZ_CP009676/1312231-1312314(+)/Firmicutes/Staphylococcus_schleiferi/singlet_1/GCV
NZ_CP009762/1287067-1287150(+)/Firmicutes/Staphylococcus_schleiferi/singlet_1/GCV

Appendix

NZ_CP009896/2646886-2646973(+)/Actinobacteria/Pimelobacter_simplex/singlet_1/GCV
NZ_CP009897/1351953-1352034(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP010309/1350950-1351033(+)/Firmicutes/Staphylococcus_schleiferi/singlet_1/GCV
NZ_CP010330/2082579-2082662(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP010337/2078366-2078449(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP010338/2070645-2070728(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP010339/2073521-2073604(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP010340/2085864-2085947(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP010346/1362199-1362280(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP010516/3342694-3342792(+)/Proteobacteria/Cupriavidus_gilardii/singlet_1/GCV
NZ_CP010517/2007900-2008004(+)/Proteobacteria/Cupriavidus_gilardii/singlet_1/GLY_MET
NZ_CP010536/4333944-4334019(+)/Proteobacteria/Cupriavidus_basilensis/singlet_1/GCV
NZ_CP010797/953856-953941(+)/Actinobacteria/Rhodococcus_sp/singlet_1/GCV
NZ_CP010968/2545486-2545569(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP011004/1476271-1476352(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP011112/3231756-3231842(+)/Actinobacteria/Luteipulveratus_mongoliensis/singlet_1/GCV
NZ_CP011345/1657289-1657370(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP011397/1380839-1380920(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP011398/1455659-1455740(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP011451/3146724-3146819(+)/Proteobacteria/Nitrosomonas_communis/singlet_1/GCV
NZ_CP011455/2054460-2054543(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NZ_CP011530/2625418-2625503(+)/Actinobacteria/Mycobacterium_immunogenum/singlet_1/GCV
NZ_CP012021/1386257-1386338(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV

NZ_CP012090/1839980-1840063(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP012095/2070581-2070664(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NZ_CP012395/3169155-3169227(+)/Firmicutes/Clostridium_autoethanogenum/singlet_1/OTHER
NZ_CP012410/2112371-2112460(-)/Fusobacteria/Leptotrichia_sp/singlet_1/GCV
NZ_CP012506/2049329-2049412(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP012605/3630228-3630315(+)/Proteobacteria/Ralstonia_insidiosa/singlet_1/GCV
NZ_CP012958/1496661-1496795(+)/Proteobacteria/Aggregatibacter_actinomycetemcomitans/singlet_1/TP
NZ_CP012959/1470628-1470762(-)/Proteobacteria/Aggregatibacter_actinomycetemcomitans/singlet_1/TP
NZ_CP013285/2290566-2290647(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP013286/1089769-1089850(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP013287/2738112-2738193(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP013288/1850119-1850200(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP013289/591831-591912(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP013722/179213-179294(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP013723/692045-692126(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP013724/2891529-2891610(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP013741/2058865-2058948(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NZ_CP013911/1204865-1204970(-)/Firmicutes/Staphylococcus_haemolyticus/singlet_1/GCV
NZ_CP013919/1407271-1407352(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP014057/1296854-1296955(+)/Firmicutes/Staphylococcus_saprophyticus/singlet_1/GCV
NZ_CP014113/1815724-1815825(+)/Firmicutes/Staphylococcus_saprophyticus/singlet_1/GCV
NZ_CP014229/3020056-3020160(-)/Proteobacteria/Desulfovibrio_fairfieldensis/singlet_1/OTHER
NZ_CP014234/711783-711889(+)/Proteobacteria/Moraxella_osloensis/singlet_1/GCV
NZ_CP014250/1434742-1434823(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV

Appendix

NZ_CP014252/1400056-1400137(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP014261/1331721-1331802(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP014566/2058564-2058647(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NZ_CP014617/2174218-2174301(+)/Actinobacteria/Mycobacterium_africanum/singlet_1/GCV
NZ_CP014790/1362756-1362837(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP014844/108133-108210(+)/Proteobacteria/Cupriavidus_nantongensis/singlet_1/GCV
NZ_CP014945/2182067-2182194(+)/Proteobacteria/Psychrobacter_alimentarius/singlet_1/GCV
NZ_CP014950/2273716-2273801(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP014952/2295059-2295144(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP014959/2386831-2386916(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP014961/2428681-2428766(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP015114/316352-316474(+)/Firmicutes/Staphylococcus_condimenti/singlet_1/GCV
NZ_CP015220/4301108-4301195(-)/Actinobacteria/Rhodococcus_sp/singlet_1/GCV
NZ_CP015235/4373127-4373212(+)/Actinobacteria/Rhodococcus_fascians/singlet_1/GCV
NZ_CP015626/1198799-1198884(+)/Firmicutes/Staphylococcus_pseudintermedius/singlet_1/GCV
NZ_CP016022/3585998-3586085(+)/Proteobacteria/Ralstonia_insidiosa/singlet_1/GCV
NZ_CP016072/1370469-1370554(+)/Firmicutes/Staphylococcus_pseudintermedius/singlet_1/GCV
NZ_CP016073/1370828-1370913(+)/Firmicutes/Staphylococcus_pseudintermedius/singlet_1/GCV
NZ_CP016188/2317520-2317605(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP016189/2652353-2652438(+)/Actinobacteria/Mycobacterium_immunogenum/singlet_1/GCV
NZ_CP016191/2398496-2398581(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP016193/2428870-2428955(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP016553/746597-746731(+)/Proteobacteria/Aggregatibacter_actinomycetemcomitans/singlet_1/TP

Appendix

NZ_CP016760/1101477-1101598(+)/Firmicutes/Staphylococcus_carnosus/singlet_1/GCV
NZ_CP016793/3926557-3926638(+)/Actinobacteria/Lentzea_sp/singlet_1/GCV
NZ_CP016794/2059652-2059735(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_HG813240/2059375-2059458(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_HG813242/1276788-1276920(-)/Firmicutes/Staphylococcus_epidermidis/singlet_1/GCV
NZ_HG813247/1381564-1381645(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_HG813249/1987610-1987691(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_HG969191/1506922-1507011(-)/Proteobacteria/Bartonella_henselae/singlet_1/GCV
NZ_LT571449/1319102-1319234(-)/Firmicutes/Staphylococcus_epidermidis/singlet_1/GCV
NC_002940/363037-363137(+)/Proteobacteria/Haemophilus_ducreyi/singlet_2/TP
NC_003028/387500-387576(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_003098/368916-368992(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_006370/2453503-2453604(+)/Proteobacteria/Photobacterium_profundum/singlet_2/TP
NC_006448/885873-885969(-)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NC_006449/887995-888091(-)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NC_007204/958065-958186(-)/Proteobacteria/Psychrobacter_arcticus/singlet_2/GCV
NC_007969/951652-951779(-)/Proteobacteria/Psychrobacter_cryohalolentis/singlet_2/GCV
NC_008532/914895-914991(-)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NC_008533/376445-376521(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_009009/353828-353904(+)/Firmicutes/Streptococcus_sanguinis/singlet_2/TP
NC_009338/3615662-3615750(-)/Actinobacteria/Mycobacterium_gilvum/singlet_2/GCV
NC_009655/806471-806559(+)/Proteobacteria/Actinobacillus_succinogenes/singlet_2/TP
NC_010380/469833-469909(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_010582/402689-402765(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_011072/367310-367386(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_011900/371192-371268(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_012466/381221-381297(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_012467/393372-393448(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_012468/431788-431864(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_012469/412662-412738(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP

NC_012924/236094-236188(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_012925/236286-236380(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_012926/238360-238454(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_013165/2020904-2020977(-)/Actinobacteria/Slackia_heliotrinireducens/singlet_2/TP
NC_013204/146216-146311(-)/Actinobacteria/Eggerthella_lenta/singlet_2/TP
NC_013798/1261523-1261704(+)/Firmicutes/Streptococcus_gallolyticus/singlet_2/TP
NC_013853/1774989-1775065(-)/Firmicutes/Streptococcus_mitis/singlet_2/TP
NC_014251/637935-638011(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_014376/4317589-4317664(+)/Firmicutes/Clostridium_saccharolyticum/singlet_2/TP
NC_014393/2492842-2492927(-)/Firmicutes/Clostridium_cellulovorans/singlet_2/OTHER
NC_014494/397456-397532(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_014498/439569-439645(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_014814/2829309-2829397(-)/Actinobacteria/Mycobacterium_gilvum/singlet_2/GCV
NC_015215/1240953-1241134(+)/Firmicutes/Streptococcus_gallolyticus/singlet_2/TP
NC_015600/1091462-1091642(+)/Firmicutes/Streptococcus_pasteurianus/singlet_2/TP
NC_015678/1265775-1265853(-)/Firmicutes/Streptococcus_parasanguinis/singlet_2/TP
NC_015738/2850373-2850463(+)/Actinobacteria/Eggerthella_sp/singlet_2/TP
NC_015875/463623-463699(+)/Firmicutes/Streptococcus_pseudopneumoniae/singlet_2/TP
NC_016641/70766-70844(-)/Firmicutes/Paenibacillus_terrae/singlet_2/TP
NC_016749/1083579-1083760(+)/Firmicutes/Streptococcus_macedonicus/singlet_2/TP
NC_016791/2529809-2529896(+)/Firmicutes/Clostridium_sp/singlet_2/TP
NC_016826/1050065-1050234(+)/Firmicutes/Streptococcus_infantarius/singlet_2/TP
NC_017563/895286-895382(-)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NC_017576/1230862-1231043(+)/Firmicutes/Streptococcus_gallolyticus/singlet_2/TP
NC_017581/990874-990970(-)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NC_017591/393078-393154(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_017592/394531-394607(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_017593/375181-375257(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_017595/1133091-1133187(+)/Firmicutes/Streptococcus_salivarius/singlet_2/TP
NC_017617/237674-237768(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_017618/236178-236272(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_017619/236865-236959(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_017620/251074-251168(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_017621/235293-235387(+)/Firmicutes/Streptococcus_suis/singlet_2/TP

NC_017622/236983-237077(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_017769/451657-451733(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_017905/1829191-1829269(-)/Firmicutes/Streptococcus_parasanguinis/singlet_2/TP
NC_017927/884389-884485(+)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NC_017950/237728-237822(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_018526/233220-233314(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_018594/1670349-1670425(-)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_018630/401914-401990(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_019903/936898-936976(+)/Firmicutes/Desulfitobacterium_dichloroeliminans/singlet_2/OTHER
NC_020134/2092567-2092658(-)/Firmicutes/Clostridium_stercorarium/singlet_2/TP
NC_020526/239243-239337(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_020833/915411-915492(-)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NC_020834/1664100-1664181(+)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NC_020887/2088319-2088410(-)/Firmicutes/Clostridium_stercorarium/singlet_2/TP
NC_021003/1552337-1552413(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_021004/1553653-1553729(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_021005/383812-383888(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_021006/1494416-1494492(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_021021/1431657-1431745(-)/Actinobacteria/Gordonibacter_pamelaeae/singlet_2/TP
NC_021026/383812-383888(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_021028/394235-394311(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_021082/2384103-2384184(+)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NC_021175/1794922-1794998(-)/Firmicutes/Streptococcus_oligofermentans/singlet_2/TP
NC_021213/262477-262571(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_021661/984190-984317(-)/Proteobacteria/Psychrobacter_sp/singlet_2/GCV
NC_021738/575514-575595(+)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NC_021739/701888-701969(+)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NC_021743/661759-661840(+)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NC_021883/1803842-1803923(+)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NC_021900/1053821-1053986(+)/Firmicutes/Streptococcus_lutetiensis/singlet_2/TP
NC_022247/892054-892148(-)/Proteobacteria/Variovorax_paradoxus/singlet_2/GLY_MET
NC_022655/372914-372990(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_023037/4576573-4576651(-)/Firmicutes/Paenibacillus_polymyxa/singlet_2/TP

Appendix

NZ_AKVY01000001/387955-388031(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NZ_CM001240/4441757-4441898(-)/Firmicutes/Clostridium_sp/singlet_2/TP
NZ_CM001835/128888-128964(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NZ_CM002130/952340-952436(-)/Firmicutes/Streptococcus_sp/singlet_2/TP
NZ_CP006819/1093234-1093330(-)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NZ_CP006872/4502101-4502179(-)/Firmicutes/Paenibacillus_polymyxa/singlet_2/TP
NZ_CP006957/1729354-1729435(+)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NZ_CP007155/4089027-4089118(+)/Actinobacteria/Kutzneria_albida/singlet_2/GCV
NZ_CP007453/560134-560218(+)/Firmicutes/Eubacterium_acidaminophilum/singlet_2/OTHER
NZ_CP007497/236095-236189(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NZ_CP007593/366649-366725(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NZ_CP007628/1580158-1580234(-)/Firmicutes/Streptococcus_sp/singlet_2/TP
NZ_CP008876/2534631-2534695(+)/Firmicutes/Terribacillus_aidingensis/singlet_2/TP
NZ_CP008921/220395-220489(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NZ_CP009909/1628526-1628604(-)/Firmicutes/Paenibacillus_polymyxa/singlet_2/TP
NZ_CP009913/1145058-1145154(+)/Firmicutes/Streptococcus_salivarius/singlet_2/TP
NZ_CP010996/248126-248209(+)/Actinobacteria/Mycobacterium_simiae/singlet_2/GCV
NZ_CP010999/902982-903078(-)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NZ_CP011098/915293-915374(-)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NZ_CP011099/915311-915392(-)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NZ_CP011217/915668-915764(-)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NZ_CP012588/698044-698140(+)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NZ_CP012646/743371-743447(+)/Firmicutes/Streptococcus_mitis/singlet_2/TP
NZ_CP013111/5911287-5911379(+)/Proteobacteria/Bordetella_sp/singlet_2/GCV
NZ_CP013651/1185245-1185319(-)/Firmicutes/Streptococcus_sp/singlet_2/TP
NZ_CP013688/855630-855811(-)/Firmicutes/Streptococcus_gallolyticus/singlet_2/TP
NZ_CP013689/911642-911807(+)/Firmicutes/Streptococcus_infantarius/singlet_2/TP
NZ_CP013939/48918-49014(+)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NZ_CP014056/2747787-2747881(-)/Proteobacteria/Grimontia_hollisae/singlet_2/TP

NZ_CP014144/1074159-1074255(+)/Firmicutes/Streptococcus_salivarius/singlet_2/TP
NZ_CP014264/635631-635703(+)/Firmicutes/Streptococcus_sp/singlet_2/TP
NZ_CP014326/1775437-1775513(-)/Firmicutes/Streptococcus_mitis/singlet_2/TP
NZ_CP014673/2167283-2167374(-)/Firmicutes/Clostridium_stercorarium/singlet_2/TP
NZ_CP015424/47158-47258(+)/Proteobacteria/Haemophilus_ducreyi/singlet_2/TP
NZ_CP015426/551362-551462(+)/Proteobacteria/Haemophilus_ducreyi/singlet_2/TP
NZ_CP015431/385026-385126(+)/Proteobacteria/Haemophilus_ducreyi/singlet_2/TP
NZ_CP015433/566502-566602(+)/Proteobacteria/Haemophilus_ducreyi/singlet_2/TP
NZ_CP015434/960828-960928(+)/Proteobacteria/Haemophilus_ducreyi/singlet_2/TP
NZ_CP015483/1670373-1670450(+)/Proteobacteria/Acinetobacter_baumannii/singlet_2/TP
NZ_CP015557/224967-225061(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NZ_CP016026/218197-218293(-)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NZ_CP016207/1048293-1048369(+)/Firmicutes/Streptococcus_sp/singlet_2/TP
NZ_CP016439/27550-27646(+)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NZ_HG322949/3092595-3092669(+)/Proteobacteria/Janthinobacterium_agaricidamnosum/singlet_2/GCV
NZ_LN831029/5941333-5941443(-)/Proteobacteria/Achromobacter_xylosoxidans/singlet_2/GCV
NZ_LN831051/96684-96760(-)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NZ_LN847353/372397-372473(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP

Table 2.7: 436 Bacilli riboswitch aptamer sequences for graph clustering analysis.

NZ_CP014140/166646-166760(-)/Bacilli/Aneurinibacillus_sp/apt_B/TP
NZ_CP011974/3996095-3996174(-)/Bacilli/Bacillus_endophyticus/apt_A/GCV
NC_011567/927752-927830(+)/Bacilli/Anoxybacillus_flavithermus/apt_A/GCV
NZ_CP008855/778913-778991(-)/Bacilli/Bacillus_sp/apt_A/GCV
NZ_CP012152/1897643-1897721(-)/Bacilli/Anoxybacillus_gonensis/apt_A/GCV
NC_012793/2406395-2406473(-)/Bacilli/Geobacillus_sp/apt_A/GCV
NC_014650/1132999-1133077(+)/Bacilli/Geobacillus_sp/apt_A/GCV
NC_015660/1196878-1196956(+)/Bacilli/Geobacillus_thermoglucosidasius/apt_A/GCV
NZ_CM001483/1078319-1078397(+)/Bacilli/Geobacillus_thermoglucosidans/apt_A/GCV
NZ_CM002692/2459122-2459200(-)/Bacilli/Geobacillus_stearothermophilus/apt_A/GCV
NC_021171/3397617-3397695(-)/Bacilli/Bacillus_sp/apt_A/GCV

NZ_CP015506/4039287-4039364(-)/Bacilli/Bacillus_oceanisediminis/apt_A/GCV
NC_014829/1874207-1874289(+)/Bacilli/Bacillus_cellulosilyticus/apt_A/GCV
NZ_CP012024/2219874-2219952(-)/Bacilli/Bacillus_smithii/apt_A/GCV
NZ_CP010976/2514229-2514316(-)/Bacilli/Paenibacillus_sp/apt_B/TP
NC_003028/387500-387576(+)/Bacilli/Streptococcus_pneumoniae/singlet_2/TP
NC_015875/463623-463699(+)/Bacilli/Streptococcus_pseudopneumoniae/singlet_2/TP
NZ_CP012646/743371-743447(+)/Bacilli/Streptococcus_mitis/singlet_2/TP
NZ_CP007628/1580158-1580234(-)/Bacilli/Streptococcus_sp/singlet_2/TP
NC_013853/1774989-1775065(-)/Bacilli/Streptococcus_mitis/singlet_2/TP
NC_015291/1570417-1570493(-)/Bacilli/Streptococcus_oralis/singlet_0/TP
NC_004193/1936134-1936232(-)/Bacilli/Oceanobacillus_iheyensis/apt_B/GCV
NC_016023/3047686-3047766(+)/Bacilli/Bacillus_coagulans/apt_A/GCV
NC_015634/1722764-1722844(-)/Bacilli/Bacillus_coagulans/apt_A/GCV
NZ_CP009709/721275-721355(-)/Bacilli/Bacillus_coagulans/apt_A/GCV
NC_022524/3286024-3286101(-)/Bacilli/Bacillus_infantis/apt_A/GCV
NZ_CP010783/915018-915093(-)/Bacilli/Streptococcus_iniae/apt_A/TP
NZ_CP007161/1859869-1859978(+)/Bacilli/Virgibacillus_sp/apt_B/GCV
NZ_CM001792/1912594-1912692(-)/Bacilli/Oceanobacillus_kimchii/apt_B/GCV
NZ_CP011974/3996002-3996087(-)/Bacilli/Bacillus_endophyticus/apt_B/GCV
NZ_CP012600/1674398-1674484(-)/Bacilli/Bacillus_gobiensis/apt_A/GCV
NC_021362/2709101-2709195(-)/Bacilli/Bacillus_paralicheniformis/apt_A/GCV
NC_000964/2549510-2549601(-)/Bacilli/Bacillus_subtilis/apt_A/GCV
NZ_CP010014/2156144-2156235(-)/Bacilli/Bacillus_sp/apt_A/GCV
NC_017743/2456064-2456155(-)/Bacilli/Bacillus_sp/apt_A/GCV
NC_016047/2530173-2530264(-)/Bacilli/Bacillus_subtilis/apt_A/GCV
NC_014479/2406250-2406341(-)/Bacilli/Bacillus_subtilis/apt_A/GCV
NC_014639/2070309-2070399(-)/Bacilli/Bacillus_atrophaeus/apt_A/GCV
NZ_CP014840/1408676-1408766(+)/Bacilli/Bacillus_subtilis/apt_A/GCV
NZ_CP012602/2037426-2037505(+)/Bacilli/Bacillus_sp/apt_A/GCV
NC_014551/2417373-2417463(-)/Bacilli/Bacillus_amyloliquefaciens/apt_A/GCV
NZ_CP009748/2420320-2420410(-)/Bacilli/Bacillus_subtilis/apt_A/GCV
NZ_CP015911/2409606-2409695(-)/Bacilli/Bacillus_velezensis/apt_A/GCV
NC_009725/2422935-2423025(-)/Bacilli/Bacillus_velezensis/apt_A/GCV
NC_020272/1498766-1498856(+)/Bacilli/Bacillus_amyloliquefaciens/apt_A/GCV
NZ_CP009611/2444930-2445020(-)/Bacilli/Bacillus_subtilis/apt_A/GCV
NZ_CP010406/3568525-3568615(+)/Bacilli/Bacillus_sp/apt_A/GCV
NC_017061/2662425-2662515(-)/Bacilli/Bacillus_velezensis/apt_A/GCV
NC_017912/2658551-2658641(-)/Bacilli/Bacillus_amyloliquefaciens/apt_A/GCV

Appendix

NZ_CP009938/3068648-3068738(-)/Bacilli/Bacillus_sp/apt_A/GCV
NZ_CP013950/3735085-3735175(+)/Bacilli/Bacillus_sp/apt_A/GCV
NC_019842/2480102-2480192(-)/Bacilli/Bacillus_velezensis/apt_A/GCV
NZ_CP009749/2480838-2480928(-)/Bacilli/Bacillus_subtilis/apt_A/GCV
NC_002570/2941515-2941599(-)/Bacilli/Bacillus_halodurans/apt_A/GCV
NC_002570/2941417-2941508(-)/Bacilli/Bacillus_halodurans/apt_B/GCV
NZ_CP015506/4039202-4039280(-)/Bacilli/Bacillus_oceanisediminis/apt_B/GCV
NC_012891/1062484-1062547(+)/Bacilli/Streptococcus_dysgalactiae/apt_A/TP
NC_015558/862065-862130(+)/Bacilli/Streptococcus_parauberis/apt_A/TP
NC_003485/1022153-1022216(+)/Bacilli/Streptococcus_pyogenes/apt_A/TP
NZ_CP007023/920524-920586(+)/Bacilli/Streptococcus_pyogenes/apt_A/TP
NC_002737/1046368-1046430(+)/Bacilli/Streptococcus_pyogenes/apt_A/TP
NC_011134/935489-935551(-)/Bacilli/Streptococcus_equi/apt_A/TP
NC_017040/957614-957675(+)/Bacilli/Streptococcus_pyogenes/apt_A/TP
NZ_CP003923/2412993-2413078(-)/Bacilli/Bacillus_lehensis/apt_A/GCV
NZ_CP012600/4551035-4551127(-)/Bacilli/Bacillus_gobiensis/apt_A/GLY_MET
NZ_CP011008/2539251-2539343(+)/Bacilli/Bacillus_simplex/apt_A/GLY_MET
NZ_CP012601/681422-681504(+)/Bacilli/Bacillus_sp/apt_A/GCV
NC_014219/2513191-2513277(-)/Bacilli/Bacillus_selenitireducens/apt_A/GCV
NZ_CP016537/642144-642223(+)/Bacilli/Planococcus_halocryophilus/apt_B/GCV
NZ_CP016540/3072813-3072889(+)/Bacilli/Planococcus_sp/apt_B/GCV
NZ_CP013661/214554-214630(-)/Bacilli/Planococcus_kocurii/apt_B/GCV
NZ_CP016534/2665454-2665530(-)/Bacilli/Planococcus_antarcticus/apt_B/GCV
NZ_CP016543/1860764-1860843(-)/Bacilli/Planococcus_donghaensis/apt_B/GCV
NZ_CP009129/1365473-1365552(+)/Bacilli/Planococcus_sp/apt_B/GCV
NZ_CP011008/2537188-2537275(+)/Bacilli/Bacillus_simplex/apt_A/TP
NC_012491/2385825-2385939(+)/Bacilli/Brevibacillus_brevis/apt_B/GCV
NZ_CP012600/4550933-4551029(-)/Bacilli/Bacillus_gobiensis/apt_B/GLY_MET
NZ_CP011346/2597833-2597933(-)/Bacilli/Bacillus_velezensis/apt_B/GCV
NZ_CP008747/1398221-1398305(+)/Bacilli/Staphylococcus_hyicus/singlet_1/GCV
NZ_CP009623/728092-728173(-)/Bacilli/Staphylococcus_agnetis/singlet_0/GCV
NZ_AP014944/1354206-1354289(+)/Bacilli/Staphylococcus_schleiferi/singlet_1/GCV
NC_014925/1304501-1304586(-)/Bacilli/Staphylococcus_pseudintermedius/singlet_0/GCV
NZ_CP011366/1420311-1420389(+)/Bacilli/Salinicoccus_halodurans/singlet_0/GCV
NC_014829/1874295-1874423(+)/Bacilli/Bacillus_cellulosilyticus/apt_B/GCV
NZ_CP015286/5044257-5044368(-)/Bacilli/Paenibacillus_glucanolyticus/apt_B/GCV
NC_013406/4198371-4198482(+)/Bacilli/Paenibacillus_sp/apt_B/GCV
NZ_CP011008/2539349-2539442(+)/Bacilli/Bacillus_simplex/apt_B/GLY_MET

NC_012004/961950-962014(+)/Bacilli/Streptococcus_uberis/apt_A/TP
NC_007168/1408552-1408657(+)/Bacilli/Staphylococcus_haemolyticus/singlet_1/GCV
NZ_CP007601/876997-877124(-)/Bacilli/Staphylococcus_capitis/singlet_1/GCV
NC_002976/1146875-1147007(-)/Bacilli/Staphylococcus_epidermidis/singlet_1/GCV
NC_016941/1539556-1539642(-)/Bacilli/Staphylococcus_argenteus/singlet_0/GCV
NC_017337/1612311-1612397(-)/Bacilli/Staphylococcus_aureus/singlet_0/GCV
NC_002745/1576871-1576958(-)/Bacilli/Staphylococcus_aureus/singlet_0/GCV
NC_007622/1532050-1532137(-)/Bacilli/Staphylococcus_aureus/singlet_0/GCV
NZ_CP011528/1527333-1527420(-)/Bacilli/Staphylococcus_aureus/singlet_0/GCV
NC_017351/1651885-1651972(-)/Bacilli/Staphylococcus_aureus/singlet_0/GCV
NC_007350/1259149-1259250(+)/Bacilli/Staphylococcus_saprophyticus/singlet_1/GCV
NZ_CP012968/415792-415888(+)/Bacilli/Staphylococcus_sp/singlet_0/GCV
NZ_CP013114/1423770-1423866(-)/Bacilli/Staphylococcus_equorum/singlet_0/GCV
NZ_CP013714/1387545-1387641(-)/Bacilli/Staphylococcus_equorum/singlet_0/GCV
NC_022737/146880-146968(-)/Bacilli/Staphylococcus_pasteuri/singlet_1/GCV
NC_020164/1270725-1270813(+)/Bacilli/Staphylococcus_warneri/singlet_1/GCV
NC_013893/1413864-1413953(+)/Bacilli/Staphylococcus_lugdunensis/singlet_0/GCV
NZ_CP014016/809402-809511(+)/Bacilli/Staphylococcus_simulans/singlet_0/GCV
NZ_CP007208/1035762-1035858(-)/Bacilli/Staphylococcus_xylosus/singlet_0/GCV
NZ_CP015114/316352-316474(+)/Bacilli/Staphylococcus_condimenti/singlet_1/GCV
NC_012121/1177746-1177867(-)/Bacilli/Staphylococcus_carnosus/singlet_1/GCV
NZ_CP009241/7045113-7045218(+)/Bacilli/Paenibacillus_sp/apt_B/GCV
NZ_CP009283/6704576-6704684(+)/Bacilli/Paenibacillus_sp/apt_B/GCV
NZ_CP009284/6366570-6366679(+)/Bacilli/Paenibacillus_sp/apt_B/GCV
NZ_CP015506/431959-432029(+)/Bacilli/Bacillus_oceanisediminis/apt_B/OTHER
NZ_CP014140/166766-166870(-)/Bacilli/Aneurinibacillus_sp/apt_A/TP
NC_018491/4489912-4490008(+)/Bacilli/Bacillus_cereus/apt_A/TP
NZ_CM000743/2013396-2013492(-)/Bacilli/Bacillus_mycoides/apt_A/TP
NZ_CM000745/4807368-4807464(+)/Bacilli/Bacillus_pseudomycoides/apt_A/TP
NZ_CP009651/5217675-5217771(+)/Bacilli/Bacillus_thuringiensis/apt_A/TP
NZ_CM000739/2124803-2124899(+)/Bacilli/Bacillus_cereus/apt_A/TP
NZ_CM000733/1896277-1896372(+)/Bacilli/Bacillus_cereus/apt_A/TP
NZ_CM000737/2142917-2143013(-)/Bacilli/Bacillus_cereus/apt_A/TP
NZ_CP013274/2245274-2245370(-)/Bacilli/Bacillus_thuringiensis/apt_A/TP
NC_014171/2224035-2224131(-)/Bacilli/Bacillus_thuringiensis/apt_A/TP
NC_018693/1544707-1544803(-)/Bacilli/Bacillus_thuringiensis/apt_A/TP
NC_022781/4886146-4886242(-)/Bacilli/Bacillus_toyonensis/apt_A/TP

NZ_CM000728/2134133-2134229(-)/Bacilli/Bacillus_cereus/apt_A/TP
NZ_CM000740/2179273-2179369(-)/Bacilli/Bacillus_cereus/apt_A/TP
NC_010184/2253338-2253434(-)/Bacilli/Bacillus_weihenstephanensis/apt_A/TP
NZ_CM000719/2129380-2129476(-)/Bacilli/Bacillus_cereus/apt_A/TP
NZ_CM000742/2048650-2048746(-)/Bacilli/Bacillus_mycoides/apt_A/TP
NZ_CM000718/2219437-2219533(+)/Bacilli/Bacillus_cereus/apt_A/TP
NC_011772/2210581-2210677(-)/Bacilli/Bacillus_cereus/apt_A/TP
NC_018508/1772597-1772693(-)/Bacilli/Bacillus_thuringiensis/apt_A/TP
NZ_CM000732/2139914-2140008(+)/Bacilli/Bacillus_cereus/apt_A/TP
NC_003909/2304814-2304909(-)/Bacilli/Bacillus_cereus/apt_A/TP
NC_003997/2221616-2221711(-)/Bacilli/Bacillus_anthracis/apt_A/TP
NC_005957/2242862-2242957(-)/Bacilli/Bacillus_thuringiensis/apt_A/TP
NZ_CP001974/2221112-2221207(-)/Bacilli/Bacillus_anthracis/apt_A/TP
NZ_CP009369/2743627-2743722(+)/Bacilli/Bacillus_cereus/apt_A/TP
NC_011725/2251884-2251980(-)/Bacilli/Bacillus_cereus/apt_A/TP
NC_017208/2301767-2301863(-)/Bacilli/Bacillus_thuringiensis/apt_A/TP
NC_004722/2266452-2266548(-)/Bacilli/Bacillus_cereus/apt_A/TP
NC_020238/2489373-2489469(-)/Bacilli/Bacillus_thuringiensis/apt_A/TP
NZ_CP007512/2078610-2078706(-)/Bacilli/Bacillus_bombysepticus/apt_A/TP
NC_006274/2269910-2270005(-)/Bacilli/Bacillus_cereus/apt_A/TP
NZ_CP011058/2165864-2165955(-)/Bacilli/Paenibacillus_beijingensis/apt_B/TP
NZ_CP014140/1265837-1265932(+)/Bacilli/Aneurinibacillus_sp/apt_B/GCV
NZ_CP013023/4575610-4575795(+)/Bacilli/Paenibacillus_bovis/apt_B/GCV
NZ_CP011058/1147936-1148024(+)/Bacilli/Paenibacillus_beijingensis/apt_A/GCV
NZ_CP011361/1138533-1138627(+)/Bacilli/Salimicrobium_jeotgali/apt_A/GCV
NC_014219/2513091-2513182(-)/Bacilli/Bacillus_selenitireducens/apt_B/GCV
NZ_CP013652/3047205-3047287(+)/Bacilli/Paenibacillus_naphthalenovorans/apt_B/GCV
NZ_CP011058/1148033-1148142(+)/Bacilli/Paenibacillus_beijingensis/apt_B/GCV
NZ_CP011058/1650021-1650127(+)/Bacilli/Paenibacillus_beijingensis/apt_B/GCV
NC_017905/1829191-1829269(-)/Bacilli/Streptococcus_parasanguinis/singlet_2/TP
NC_015678/1265775-1265853(-)/Bacilli/Streptococcus_parasanguinis/singlet_2/TP
NC_009009/353828-353904(+)/Bacilli/Streptococcus_sanguinis/singlet_2/TP
NC_021175/1794922-1794998(-)/Bacilli/Streptococcus_oligofermentans/singlet_2/TP
NC_014483/4053479-4053557(-)/Bacilli/Paenibacillus_polymyxa/singlet_0/TP
NZ_CP011512/4592275-4592353(-)/Bacilli/Paenibacillus_peoriae/singlet_0/TP
NC_016641/70766-70844(-)/Bacilli/Paenibacillus_terrae/singlet_2/TP
NZ_CP006872/4502101-4502179(-)/Bacilli/Paenibacillus_polymyxa/singlet_2/TP
NC_014622/4416707-4416786(-)/Bacilli/Paenibacillus_polymyxa/apt_B/TP

Appendix

NZ_CP014264/635631-635703(+)/Bacilli/Streptococcus_sp/singlet_2/TP
NZ_CP013651/1185245-1185319(-)/Bacilli/Streptococcus_sp/singlet_2/TP
NZ_CM000733/1896381-1896486(+)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CM000745/4807473-4807579(+)/Bacilli/Bacillus_pseudomycoides/apt_B/TP
NZ_CP009651/5217780-5217886(+)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_CP009351/1643071-1643174(+)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_CP007607/3491242-3491346(+)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_AP014864/3419358-3419462(+)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_CM000715/2054459-2054563(+)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CM000714/2084988-2085091(+)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CM000725/2160468-2160571(+)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CP009692/3547951-3548054(+)/Bacilli/Bacillus_mycoides/apt_B/TP
NZ_CP009746/2489207-2489310(+)/Bacilli/Bacillus_weihenstephanensis/apt_B/TP
NZ_CM000732/2140017-2140120(+)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CM000752/2067335-2067438(+)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NC_012581/2017479-2017582(+)/Bacilli/Bacillus_anthracis/apt_B/TP
NZ_CM000754/2085906-2086009(+)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_CP009300/2151102-2151205(+)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CM000739/2124908-2125011(+)/Bacilli/Bacillus_cereus/apt_B/TP
NC_018491/4490017-4490120(+)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CP009369/2743731-2743834(+)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CM000730/2143408-2143512(+)/Bacilli/Bacillus_cereus/apt_B/TP
NC_018500/1611447-1611551(+)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_CP012483/3228891-3228995(+)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CM000718/2219542-2219646(+)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CP006837/2926932-2927037(-)/Bacilli/Lysinibacillus_varians/apt_B/GCV
NZ_CP010820/1718607-1718719(+)/Bacilli/Lysinibacillus_fusiformis/apt_B/GCV
NZ_CP014643/2310786-2310898(+)/Bacilli/Lysinibacillus_sphaericus/apt_B/GCV
NZ_CP011058/1147591-1147671(-)/Bacilli/Paenibacillus_beijingensis/apt_B/GCV
NZ_CP012602/2037514-2037595(+)/Bacilli/Bacillus_sp/apt_B/GCV
NZ_CP007806/1670143-1670225(+)/Bacilli/Brevibacillus_laterosporus/apt_B/GCV
NZ_CP010268/2227050-2227127(+)/Bacilli/Paenibacillus_polymyxa/apt_B/GLY_MET
NZ_CP013023/4371872-4371972(-)/Bacilli/Paenibacillus_bovis/apt_B/GLY_MET
NZ_CP012600/1674298-1674389(-)/Bacilli/Bacillus_gobiensis/apt_B/GCV
NC_006270/2547365-2547461(-)/Bacilli/Bacillus_licheniformis/apt_B/GCV
NC_021362/2708995-2709092(-)/Bacilli/Bacillus_paralicheniformis/apt_B/GCV
NZ_CP009286/5152320-5152418(+)/Bacilli/Paenibacillus_stellifer/apt_B/GCV
NZ_CP009288/5440520-5440675(+)/Bacilli/Paenibacillus_durus/apt_B/GCV

NZ_CP011114/4266609-4266764(+)/Bacilli/Paenibacillus_durus/apt_B/GCV
NZ_CP004078/4789751-4789908(+)/Bacilli/Paenibacillus_sabinae/apt_B/GCV
NZ_CP009285/7501003-7501102(+)/Bacilli/Paenibacillus_borealis/apt_B/GCV
NZ_CP009282/7083774-7083873(+)/Bacilli/Paenibacillus_sp/apt_B/GCV
NZ_CP009280/7773780-7773879(+)/Bacilli/Paenibacillus_sp/apt_B/GCV
NZ_CP011008/3501666-3501748(-)/Bacilli/Bacillus_simplex/apt_B/GCV
NC_013411/218870-218985(+)/Bacilli/Geobacillus_sp/apt_B/GCV
NZ_CP008934/2794249-2794364(+)/Bacilli/Geobacillus_stearothermophilus/apt_B/GCV
NZ_CP003923/2412895-2412984(-)/Bacilli/Bacillus_lehensis/apt_B/GCV
NC_006582/2619777-2619858(-)/Bacilli/Bacillus_clausii/apt_B/GCV
NZ_CP011361/1138636-1138713(+)/Bacilli/Salimicrobium_jeotgali/apt_B/GCV
NZ_CP014342/825516-825634(+)/Bacilli/Geobacillus_subterraneus/apt_B/GCV
NC_011567/927839-927918(+)/Bacilli/Anoxybacillus_flavithermus/apt_B/GCV
NZ_LN831776/7239766-7239866(+)/Bacilli/Paenibacillus_riograndensis/apt_B/GCV
NZ_CP009287/6537365-6537465(+)/Bacilli/Paenibacillus_graminis/apt_B/GCV
NC_014098/704461-704567(+)/Bacilli/Kyrpidia_tusciae/apt_B/GCV
NZ_CP011058/1642625-1642724(-)/Bacilli/Paenibacillus_beijingensis/apt_B/OTHER
NZ_CP007739/2295003-2295081(-)/Bacilli/Bacillus_methanolicus/apt_B/GCV
NC_018065/1720440-1720548(+)/Bacilli/Solibacillus_silvestris/apt_B/GCV
NZ_CP012601/681514-681599(+)/Bacilli/Bacillus_sp/apt_B/GCV
NZ_CP015286/5044784-5044872(+)/Bacilli/Paenibacillus_glucanolyticus/apt_B/GCV
NC_013406/4197872-4197958(-)/Bacilli/Paenibacillus_sp/apt_B/GCV
NZ_CP009920/1299752-1299837(-)/Bacilli/Bacillus_megaterium/apt_B/GCV
NC_014019/4324638-4324723(-)/Bacilli/Bacillus_megaterium/apt_B/GCV
NZ_CP015226/4221337-4221422(-)/Bacilli/Bacillus_sp/apt_B/GCV
NC_009725/2422832-2422926(-)/Bacilli/Bacillus_velezensis/apt_B/GCV
NC_022653/2419907-2420001(-)/Bacilli/Bacillus_amyloliquefaciens/apt_B/GCV
NC_016784/2514518-2514612(-)/Bacilli/Bacillus_velezensis/apt_B/GCV
NC_017912/2658448-2658542(-)/Bacilli/Bacillus_amyloliquefaciens/apt_B/GCV
NZ_CP009611/2444827-2444921(-)/Bacilli/Bacillus_subtilis/apt_B/GCV
NZ_CP009938/3068545-3068639(-)/Bacilli/Bacillus_sp/apt_B/GCV
NC_014551/2417270-2417364(-)/Bacilli/Bacillus_amyloliquefaciens/apt_B/GCV
NZ_CP009748/2420217-2420311(-)/Bacilli/Bacillus_subtilis/apt_B/GCV
NZ_CP013950/3735184-3735278(+)/Bacilli/Bacillus_sp/apt_B/GCV
NZ_CP006890/2652024-2652118(-)/Bacilli/Bacillus_velezensis/apt_B/GCV
NZ_CP009749/2480735-2480829(-)/Bacilli/Bacillus_subtilis/apt_B/GCV
NZ_CP012024/2219793-2219865(-)/Bacilli/Bacillus_smithii/apt_B/GCV
NZ_AP014928/2205163-2205258(-)/Bacilli/Bacillus_pumilus/apt_B/GCV

NC_017743/2455969-2456055(-)/Bacilli/Bacillus_sp/apt_B/GCV
NC_014479/2406155-2406241(-)/Bacilli/Bacillus_subtilis/apt_B/GCV
NC_016047/2530078-2530164(-)/Bacilli/Bacillus_subtilis/apt_B/GCV
NZ_CP013984/2409165-2409252(-)/Bacilli/Bacillus_subtilis/apt_B/GCV
NC_009848/2216062-2216157(-)/Bacilli/Bacillus_pumilus/apt_B/GCV
NZ_CP010075/2217304-2217397(-)/Bacilli/Bacillus_sp/apt_B/GCV
NC_014639/2070211-2070300(-)/Bacilli/Bacillus_atrophaeus/apt_B/GCV
NZ_CP014840/1408775-1408864(+)/Bacilli/Bacillus_subtilis/apt_B/GCV
NC_000964/2549415-2549501(-)/Bacilli/Bacillus_subtilis/apt_B/GCV
NZ_CP010014/2156049-2156135(-)/Bacilli/Bacillus_sp/apt_B/GCV
NC_017195/2398079-2398165(-)/Bacilli/Bacillus_subtilis/apt_B/GCV
NC_023134/3590089-3590260(+)/Bacilli/Paenibacillus_larvae/apt_B/GCV
NZ_CP007739/592729-592792(-)/Bacilli/Bacillus_methanolicus/singlet_0/OTHER
NZ_CP004078/5201331-5201397(-)/Bacilli/Paenibacillus_sabinae/singlet_0/TP
NZ_CP008876/2534631-2534695(+)/Bacilli/Terribacillus_aidingensis/singlet_2/TP
NZ_CP014835/478795-478869(-)/Bacilli/Streptococcus_sp/apt_B/TP
NC_015433/243709-243803(+)/Bacilli/Streptococcus_suis/singlet_0/TP
NC_012924/236094-236188(+)/Bacilli/Streptococcus_suis/singlet_2/TP
NC_017950/237728-237822(+)/Bacilli/Streptococcus_suis/singlet_2/TP
NZ_CP011058/1147680-1147784(-)/Bacilli/Paenibacillus_beijingensis/apt_A/GCV
NZ_CP013652/3047112-3047196(+)/Bacilli/Paenibacillus_naphthalenovorans/apt_A/GCV
NZ_CP007739/2295090-2295172(-)/Bacilli/Bacillus_methanolicus/apt_A/GCV
NC_014019/4324732-4324814(-)/Bacilli/Bacillus_megaterium/apt_A/GCV
NZ_CP015226/4221431-4221513(-)/Bacilli/Bacillus_sp/apt_A/GCV
NC_016826/1050065-1050234(+)/Bacilli/Streptococcus_infantarius/singlet_2/TP
NC_021900/1053821-1053986(+)/Bacilli/Streptococcus_lutetiensis/singlet_2/TP
NZ_CP013689/911642-911807(+)/Bacilli/Streptococcus_infantarius/singlet_2/TP
NC_021485/1173029-1173151(+)/Bacilli/Streptococcus_agalactiae/singlet_0/TP
NZ_CP011329/1145363-1145485(+)/Bacilli/Streptococcus_agalactiae/singlet_0/TP
NZ_CP010874/1121792-1121914(+)/Bacilli/Streptococcus_agalactiae/apt_B/TP
NC_004116/1147549-1147671(+)/Bacilli/Streptococcus_agalactiae/singlet_0/TP
NC_015558/862143-862220(+)/Bacilli/Streptococcus_parauberis/apt_B/TP
NZ_CP010783/914935-915012(-)/Bacilli/Streptococcus_iniae/apt_B/TP
NC_012891/1062560-1062636(+)/Bacilli/Streptococcus_dysgalactiae/apt_B/TP
NC_002737/1046443-1046519(+)/Bacilli/Streptococcus_pyogenes/apt_B/TP
NC_017040/957688-957764(+)/Bacilli/Streptococcus_pyogenes/apt_B/TP
NC_016749/1083579-1083760(+)/Bacilli/Streptococcus_macedonicus/singlet_2/TP
NZ_CP013688/855630-855811(-)/Bacilli/Streptococcus_gallolyticus/singlet_2/TP

Appendix

NC_013798/1261523-1261704(+)/Bacilli/Streptococcus_gallolyticus/singlet_2/TP
NC_015600/1091462-1091642(+)/Bacilli/Streptococcus_pasteurianus/singlet_2/TP
NC_012004/962025-962102(+)/Bacilli/Streptococcus_uberis/apt_B/TP
NZ_CM002130/952340-952436(-)/Bacilli/Streptococcus_sp/singlet_2/TP
NC_006448/885873-885969(-)/Bacilli/Streptococcus_thermophilus/singlet_2/TP
NZ_CM002132/1194545-1194641(-)/Bacilli/Streptococcus_sp/singlet_0/TP
NZ_CP013216/1043399-1043495(+)/Bacilli/Streptococcus_salivarius/singlet_0/TP
NC_015760/1112053-1112149(-)/Bacilli/Streptococcus_salivarius/singlet_0/TP
NZ_CP011008/3501757-3501834(-)/Bacilli/Bacillus_simplex/apt_A/GCV
NZ_CP011058/1642733-1642820(-)/Bacilli/Paenibacillus_beijingensis/apt_A/OTHER
NC_018065/1720349-1720431(+)/Bacilli/Solibacillus_silvestris/apt_A/GCV
NZ_CP013023/4371981-4372083(-)/Bacilli/Paenibacillus_bovis/apt_A/GLY_MET
NC_016641/2086069-2086172(-)/Bacilli/Paenibacillus_terrae/apt_A/GLY_MET
NC_014483/429899-430004(-)/Bacilli/Paenibacillus_polymyxa/apt_A/GLY_MET
NC_023037/371062-371167(-)/Bacilli/Paenibacillus_polymyxa/apt_A/GLY_MET
NZ_CP011512/463227-463332(-)/Bacilli/Paenibacillus_peoriae/apt_A/GLY_MET
NC_014622/488126-488230(-)/Bacilli/Paenibacillus_polymyxa/apt_A/GLY_MET
NZ_CP006872/442540-442644(-)/Bacilli/Paenibacillus_polymyxa/apt_A/GLY_MET
NZ_CP009909/3434951-3435055(-)/Bacilli/Paenibacillus_polymyxa/apt_A/GLY_MET
NZ_CP011058/1649650-1649759(-)/Bacilli/Paenibacillus_beijingensis/apt_A/GLY_MET
NZ_CP015286/5044664-5044775(+)/Bacilli/Paenibacillus_glucanolyticus/apt_A/GCV
NC_013406/4197967-4198079(-)/Bacilli/Paenibacillus_sp/apt_A/GCV
NZ_CP014140/1265751-1265830(+)/Bacilli/Aneurinibacillus_sp/apt_A/GCV
NZ_CP009416/1967796-1967874(-)/Bacilli/Jeotgalibacillus_malaysiensis/apt_A/GCV
NZ_CP011832/2370702-2370817(-)/Bacilli/Geobacillus_sp/apt_B/GCV
NZ_CP008903/733060-733175(-)/Bacilli/Geobacillus_sp/apt_B/GCV
NC_014650/1133085-1133169(+)/Bacilli/Geobacillus_sp/apt_B/GCV
NC_015660/1196964-1197048(+)/Bacilli/Geobacillus_thermoglucosidasius/apt_B/GCV
NZ_CM001483/1078405-1078489(+)/Bacilli/Geobacillus_thermoglucosidans/apt_B/GCV
NC_012793/2406302-2406387(-)/Bacilli/Geobacillus_sp/apt_B/GCV
NC_011999/1288490-1288565(-)/Bacilli/Macrococcus_caseolyticus/singlet_0/GCV
NC_006270/2547470-2547564(-)/Bacilli/Bacillus_licheniformis/apt_A/GCV
NZ_LT603683/2822680-2822774(-)/Bacilli/Bacillus_glycinifermentans/apt_A/GCV
NC_006510/2456261-2456339(-)/Bacilli/Geobacillus_kaustophilus/apt_A/GCV
NC_009328/2482969-2483047(-)/Bacilli/Geobacillus_thermodenitrificans/apt_A/GCV
NC_013411/218783-218861(+)/Bacilli/Geobacillus_sp/apt_A/GCV

NC_016593/2515909-2515987(-)/Bacilli/Geobacillus_thermoleovorans/apt_A/GCV
NZ_CP008934/2794162-2794240(+)/Bacilli/Geobacillus_stearothermophilus/apt_A/GCV
NZ_CP014342/825429-825507(+)/Bacilli/Geobacillus_subterraneus/apt_A/GCV
NZ_CP009288/5440424-5440511(+)/Bacilli/Paenibacillus_durus/apt_A/GCV
NZ_CP009241/7045021-7045107(+)/Bacilli/Paenibacillus_sp/apt_A/GCV
NZ_CP009280/7773686-7773771(+)/Bacilli/Paenibacillus_sp/apt_A/GCV
NZ_CP009285/7500908-7500994(+)/Bacilli/Paenibacillus_borealis/apt_A/GCV
NZ_CP009283/6704485-6704570(+)/Bacilli/Paenibacillus_sp/apt_A/GCV
NZ_CP009287/6537270-6537356(+)/Bacilli/Paenibacillus_graminis/apt_A/GCV
NZ_CP011114/4266512-4266600(+)/Bacilli/Paenibacillus_durus/apt_A/GCV
NZ_CP009284/6366478-6366564(+)/Bacilli/Paenibacillus_sp/apt_A/GCV
NZ_LN831776/7239671-7239757(+)/Bacilli/Paenibacillus_riograndensis/apt_A/GCV
NZ_CP009281/6216261-6216348(+)/Bacilli/Paenibacillus_sp/apt_A/GCV
NZ_CP009279/6201358-6201446(+)/Bacilli/Paenibacillus_sp/apt_A/GCV
NZ_CP009428/6255438-6255526(+)/Bacilli/Paenibacillus_odorifer/apt_A/GCV
NZ_CP015506/431863-431953(+)/Bacilli/Bacillus_oceanisediminis/apt_A/OTHER
NC_012470/1186136-1186212(+)/Bacilli/Streptococcus_equi/apt_B/TP
NC_011134/935400-935476(-)/Bacilli/Streptococcus_equi/apt_B/TP
NC_014098/704361-704452(+)/Bacilli/Kyrpidia_tusciae/apt_A/GCV
NC_006582/2619867-2619950(-)/Bacilli/Bacillus_clausii/apt_A/GCV
NZ_CP011058/1649919-1650012(+)/Bacilli/Paenibacillus_beijingensis/apt_A/GCV
NZ_CP015286/5044374-5044465(-)/Bacilli/Paenibacillus_glucanolyticus/apt_A/GCV
NZ_CP007806/1670045-1670134(+)/Bacilli/Brevibacillus_laterosporus/apt_A/GCV
NC_012491/2385732-2385819(+)/Bacilli/Brevibacillus_brevis/apt_A/GCV
NZ_CP004078/4789655-4789742(+)/Bacilli/Paenibacillus_sabinae/apt_A/GCV
NZ_CP011007/2233640-2233737(-)/Bacilli/Bacillus_pumilus/apt_A/GCV
NZ_AP014928/2205267-2205364(-)/Bacilli/Bacillus_pumilus/apt_A/GCV
NZ_CP011109/2082927-2083024(-)/Bacilli/Bacillus_pumilus/apt_A/GCV
NZ_CP010075/2217406-2217502(-)/Bacilli/Bacillus_sp/apt_A/GCV
NC_009848/2216166-2216263(-)/Bacilli/Bacillus_pumilus/apt_A/GCV
NC_013406/4198275-4198365(+)/Bacilli/Paenibacillus_sp/apt_A/GCV
NC_023134/3589997-3590080(+)/Bacilli/Paenibacillus_larvae/apt_A/GCV
NZ_CP009286/5152216-5152311(+)/Bacilli/Paenibacillus_stellifer/apt_A/GCV
NZ_CP011058/2165963-2166050(-)/Bacilli/Paenibacillus_beijingensis/apt_A/TP
NZ_CP013023/4575518-4575603(+)/Bacilli/Paenibacillus_bovis/apt_A/GCV
NC_013791/2603007-2603091(-)/Bacilli/Bacillus_pseudofirmus/apt_A/GCV
NZ_AP013294/2865793-2865871(-)/Bacilli/Bacillus_sp/apt_B/GCV
NZ_CP010976/2514323-2514407(-)/Bacilli/Paenibacillus_sp/apt_A/TP

NC_014622/4416795-4416885(-)/Bacilli/Paenibacillus_polymyxa/apt_A/TP
NZ_CCSE01000001/796366-796446(+)/Bacilli/Jeotgalicoccus_sp/singlet_0/GCV
NZ_CP007161/1859763-1859862(+)/Bacilli/Virgibacillus_sp/apt_A/GCV
NZ_CM001792/1912699-1912780(-)/Bacilli/Oceanobacillus_kimchii/apt_A/GCV
NC_004193/1936239-1936320(-)/Bacilli/Oceanobacillus_iheyensis/apt_A/GCV
NZ_CP009279/6201452-6201557(+)/Bacilli/Paenibacillus_sp/apt_B/GCV
NZ_CP010820/1718515-1718598(+)/Bacilli/Lysinibacillus_fusiformis/apt_A/GCV
NZ_CP009428/6255532-6255638(+)/Bacilli/Paenibacillus_odorifer/apt_B/GCV
NZ_CP009281/6216354-6216460(+)/Bacilli/Paenibacillus_sp/apt_B/GCV
NZ_CM001076/1119386-1119461(+)/Bacilli/Streptococcus_dysgalactiae/apt_A/TP
NZ_CP010874/1121712-1121785(+)/Bacilli/Streptococcus_agalactiae/apt_A/TP
NZ_CP011008/2537284-2537342(+)/Bacilli/Bacillus_simplex/apt_B/TP
NC_016023/3047775-3047855(+)/Bacilli/Bacillus_coagulans/apt_B/GCV
NC_015634/1722673-1722755(-)/Bacilli/Bacillus_coagulans/apt_B/GCV
NZ_CM000743/2013283-2013386(-)/Bacilli/Bacillus_mycoides/apt_B/TP
NZ_CP007626/4849364-4849468(-)/Bacilli/Bacillus_mycoides/apt_B/TP
NZ_CM000758/2267189-2267291(-)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_CP013274/2245162-2245264(-)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NC_011772/2210470-2210571(-)/Bacilli/Bacillus_cereus/apt_B/TP
NC_018508/1772486-1772587(-)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_CM000719/2129269-2129370(-)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CM000734/2197540-2197642(-)/Bacilli/Bacillus_cereus/apt_B/TP
NC_020238/2489261-2489363(-)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_CM000717/2070797-2070899(-)/Bacilli/Bacillus_cereus/apt_B/TP
NC_004722/2266340-2266442(-)/Bacilli/Bacillus_cereus/apt_B/TP
NC_014171/2223923-2224025(-)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_CP007512/2078498-2078600(-)/Bacilli/Bacillus_bombysepticus/apt_B/TP
NC_010184/2253227-2253328(-)/Bacilli/Bacillus_weihenstephanensis/apt_B/TP
NZ_CM000742/2048539-2048640(-)/Bacilli/Bacillus_mycoides/apt_B/TP
NC_003997/2221505-2221606(-)/Bacilli/Bacillus_anthracis/apt_B/TP
NC_005957/2242751-2242852(-)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NC_011773/2285457-2285558(-)/Bacilli/Bacillus_cereus/apt_B/TP
NC_006274/2269799-2269900(-)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CM000720/2138064-2138165(-)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CM000740/2179162-2179263(-)/Bacilli/Bacillus_cereus/apt_B/TP
NC_003909/2304703-2304804(-)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CM000746/2123594-2123695(-)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NC_017200/2393420-2393522(-)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NC_018693/1544595-1544697(-)/Bacilli/Bacillus_thuringiensis/apt_B/TP

NC_022781/4886034-4886136(-)/Bacilli/Bacillus_toyonensis/apt_B/TP
NZ_CM000728/2134021-2134123(-)/Bacilli/Bacillus_cereus/apt_B/TP
NC_017208/2301655-2301757(-)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_CP011151/2295690-2295792(-)/Bacilli/Bacillus_cereus/apt_B/TP
NZ_CP012099/2234053-2234155(-)/Bacilli/Bacillus_thuringiensis/apt_B/TP
NZ_CP011058/1649565-1649640(-)/Bacilli/Paenibacillus_beijingensis/apt_B/GLY_MET
NC_014483/429814-429889(-)/Bacilli/Paenibacillus_polymyxa/apt_B/GLY_MET
NZ_CP011512/463142-463217(-)/Bacilli/Paenibacillus_peoriae/apt_B/GLY_MET
NZ_CP006872/442455-442530(-)/Bacilli/Paenibacillus_polymyxa/apt_B/GLY_MET
NC_014622/488041-488116(-)/Bacilli/Paenibacillus_polymyxa/apt_B/GLY_MET
NC_016641/2085986-2086059(-)/Bacilli/Paenibacillus_terrae/apt_B/GLY_MET
NZ_LT603683/2822576-2822670(-)/Bacilli/Bacillus_glycinifermentans/apt_B/GCV
NC_022080/2350098-2350213(-)/Bacilli/Geobacillus_sp/apt_B/GCV
NC_006510/2456138-2456251(-)/Bacilli/Geobacillus_kaustophilus/apt_B/GCV
NC_014915/2520503-2520616(-)/Bacilli/Geobacillus_sp/apt_B/GCV
NC_016593/2515786-2515899(-)/Bacilli/Geobacillus_thermoleovorans/apt_B/GCV
NC_021171/3397532-3397607(-)/Bacilli/Bacillus_sp/apt_B/GCV
NC_009328/2482843-2482959(-)/Bacilli/Geobacillus_thermodenitrificans/apt_B/GCV
NZ_CM002692/2459036-2459112(-)/Bacilli/Geobacillus_stearothermophilus/apt_B/GCV
NZ_CP012152/1897559-1897633(-)/Bacilli/Anoxybacillus_gonensis/apt_B/GCV
NC_017668/2468839-2468934(-)/Bacilli/Halobacillus_halophilus/apt_B/GCV
NC_013791/2602920-2602996(-)/Bacilli/Bacillus_pseudofirmus/apt_B/GCV
NZ_CP008855/778828-778903(-)/Bacilli/Bacillus_sp/apt_B/GCV
NC_022524/3285939-3286014(-)/Bacilli/Bacillus_infantis/apt_B/GCV
NZ_CP008876/1734314-1734398(-)/Bacilli/Terribacillus_aidingensis/apt_B/GCV
NZ_CP009416/1967708-1967785(-)/Bacilli/Jeotgalibacillus_malaysiensis/apt_B/GCV
NC_017668/2468944-2469020(-)/Bacilli/Halobacillus_halophilus/apt_A/GCV
NC_016011/1352611-1352692(+)/Bacilli/Listeria_ivanovii/singlet_0/GCV
NC_011660/1267236-1267317(-)/Bacilli/Listeria_monocytogenes/singlet_1/GCV
NC_003212/1379631-1379712(+)/Bacilli/Listeria_innocua/singlet_1/GCV
NC_008555/1351694-1351775(+)/Bacilli/Listeria_welshimeri/singlet_1/GCV
NC_003210/1372845-1372926(+)/Bacilli/Listeria_monocytogenes/singlet_1/GCV
NC_017545/1373235-1373316(+)/Bacilli/Listeria_monocytogenes/singlet_1/GCV
NC_002973/1352550-1352631(+)/Bacilli/Listeria_monocytogenes/singlet_1/GCV
NC_013891/1282081-1282162(+)/Bacilli/Listeria_seeligeri/singlet_1/GCV
NZ_CP016540/3072725-3072806(+)/Bacilli/Planococcus_sp/apt_A/GCV
NZ_CP013661/214637-214718(-)/Bacilli/Planococcus_kocurii/apt_A/GCV

NZ_CP009129/1365385-1365466(+)/Bacilli/Planococcus_sp/apt_A/GCV
NZ_CP016537/642056-642137(+)/Bacilli/Planococcus_halocryophilus/apt_A/GCV
NZ_CP016543/1860850-1860931(-)/Bacilli/Planococcus_donghaensis/apt_A/GCV
NZ_CP014643/2310694-2310777(+)/Bacilli/Lysinibacillus_sphaericus/apt_A/GCV
NZ_CP006837/2927046-2927128(-)/Bacilli/Lysinibacillus_varians/apt_A/GCV
NZ_CP016534/2665537-2665618(-)/Bacilli/Planococcus_antarcticus/apt_A/GCV
NZ_AP013294/2865878-2865959(-)/Bacilli/Bacillus_sp/apt_A/GCV
NZ_CP008876/1734408-1734485(-)/Bacilli/Terribacillus_aidingensis/apt_A/GCV
NZ_CP014835/478878-478958(-)/Bacilli/Streptococcus_sp/apt_A/TP
NC_004350/1115970-1116039(+)/Bacilli/Streptococcus_mutans/singlet_0/TP

Table 2.8: Cluster stability after 100 bootstrap replicates for Bacilli and Actinobacteria clustering.

	Bacilli GCV Cluster	*Bacilli* TP Cluster	*Actinobacteria* GCV Cluster
cluster_walktrap	.91	.92	.89
cluster_leading_eigen	.79	.86	.90
cluster_edge_betweenness	.81	.61	.78
cluster_fast_greedy	.64	.40	.72

Average Jaccard Similarity Index after 100 bootstrap replicates for *Bacilli* cluster regulating GCV, *Bacilli* cluster regulating TP, and *Actinobacteria* cluster regulating GCV. The first row indicates the cluster and the first column indicates the community detection method used. The methods tend to show good cluster stability, particularly cluster_walktrap.

Table 2.9: 124 Bacilli riboswitch aptamer sequences from aptamer-1 and singleton type-1 sub-cluster with paired aptamer-2 supplemented in.

NC_002570/2941515-2941599(-)/Bacilli/Bacillus_halodurans/apt_A/GCV
NC_002570/2941417-2941508(-)/Bacilli/Bacillus_halodurans/apt_B/GCV
NC_002745/1576871-1576958(-)/Bacilli/Staphylococcus_aureus/singlet_0/GCV
NC_002973/1352550-1352631(+)/Bacilli/Listeria_monocytogenes/singlet_1/GCV
NC_003210/1372845-1372926(+)/Bacilli/Listeria_monocytogenes/singlet_1/GCV
NC_003212/1379631-1379712(+)/Bacilli/Listeria_innocua/singlet_1/GCV
NC_004193/1936239-1936320(-)/Bacilli/Oceanobacillus_iheyensis/apt_A/GCV
NC_004193/1936134-1936232(-)/Bacilli/Oceanobacillus_iheyensis/apt_B/GCV
NC_006510/2456261-2456339(-)/Bacilli/Geobacillus_kaustophilus/apt_A/GCV
NC_006510/2456138-2456251(-)/Bacilli/Geobacillus_kaustophilus/apt_B/GCV
NC_006582/2619867-2619950(-)/Bacilli/Bacillus_clausii/apt_A/GCV
NC_006582/2619777-2619858(-)/Bacilli/Bacillus_clausii/apt_B/GCV

Appendix

NC_007622/1532050-1532137(-)/Bacilli/Staphylococcus_aureus/singlet_0/GCV
NC_008555/1351694-1351775(+)/Bacilli/Listeria_welshimeri/singlet_1/GCV
NC_009328/2482969-2483047(-)/Bacilli/Geobacillus_thermodenitrificans/apt_A/GCV
NC_009328/2482843-2482959(-)/Bacilli/Geobacillus_thermodenitrificans/apt_B/GCV
NC_011567/927752-927830(+)/Bacilli/Anoxybacillus_flavithermus/apt_A/GCV
NC_011567/927839-927918(+)/Bacilli/Anoxybacillus_flavithermus/apt_B/GCV
NC_011660/1267236-1267317(-)/Bacilli/Listeria_monocytogenes/singlet_1/GCV
NC_011999/1288490-1288565(-)/Bacilli/Macrococcus_caseolyticus/singlet_0/GCV
NC_012793/2406395-2406473(-)/Bacilli/Geobacillus_sp/apt_A/GCV
NC_012793/2406302-2406387(-)/Bacilli/Geobacillus_sp/apt_B/GCV
NC_013411/218783-218861(+)/Bacilli/Geobacillus_sp/apt_A/GCV
NC_013411/218870-218985(+)/Bacilli/Geobacillus_sp/apt_B/GCV
NC_013791/2603007-2603091(-)/Bacilli/Bacillus_pseudofirmus/apt_A/GCV
NC_013791/2602920-2602996(-)/Bacilli/Bacillus_pseudofirmus/apt_B/GCV
NC_013891/1282081-1282162(+)/Bacilli/Listeria_seeligeri/singlet_1/GCV
NC_013893/1413864-1413953(+)/Bacilli/Staphylococcus_lugdunensis/singlet_0/GCV
NC_014019/4324732-4324814(-)/Bacilli/Bacillus_megaterium/apt_A/GCV
NC_014019/4324638-4324723(-)/Bacilli/Bacillus_megaterium/apt_B/GCV
NC_014219/2513191-2513277(-)/Bacilli/Bacillus_selenitireducens/apt_A/GCV
NC_014219/2513091-2513182(-)/Bacilli/Bacillus_selenitireducens/apt_B/GCV
NC_014650/1132999-1133077(+)/Bacilli/Geobacillus_sp/apt_A/GCV
NC_014650/1133085-1133169(+)/Bacilli/Geobacillus_sp/apt_B/GCV
NC_014829/1874207-1874289(+)/Bacilli/Bacillus_cellulosilyticus/apt_A/GCV
NC_014829/1874295-1874423(+)/Bacilli/Bacillus_cellulosilyticus/apt_B/GCV
NC_014925/1304501-1304586(-)/Bacilli/Staphylococcus_pseudintermedius/singlet_0/GCV
NC_015634/1722764-1722844(-)/Bacilli/Bacillus_coagulans/apt_A/GCV
NC_015634/1722673-1722755(-)/Bacilli/Bacillus_coagulans/apt_B/GCV
NC_015660/1196878-1196956(+)/Bacilli/Geobacillus_thermoglucosidasius/apt_A/GCV
NC_015660/1196964-1197048(+)/Bacilli/Geobacillus_thermoglucosidasius/apt_B/GCV
NC_016011/1352611-1352692(+)/Bacilli/Listeria_ivanovii/singlet_0/GCV
NC_016023/3047686-3047766(+)/Bacilli/Bacillus_coagulans/apt_A/GCV
NC_016023/3047775-3047855(+)/Bacilli/Bacillus_coagulans/apt_B/GCV
NC_016593/2515909-2515987(-)/Bacilli/Geobacillus_thermoleovorans/apt_A/GCV
NC_016593/2515786-2515899(-)/Bacilli/Geobacillus_thermoleovorans/apt_B/GCV
NC_016941/1539556-1539642(-)/Bacilli/Staphylococcus_argenteus/singlet_0/GCV
NC_017337/1612311-1612397(-)/Bacilli/Staphylococcus_aureus/singlet_0/GCV

NC_017351/1651885-1651972(-)/Bacilli/Staphylococcus_aureus/singlet_0/GCV
NC_017545/1373235-1373316(+)/Bacilli/Listeria_monocytogenes/singlet_1/GCV
NC_017668/2468944-2469020(-)/Bacilli/Halobacillus_halophilus/apt_A/GCV
NC_017668/2468839-2468934(-)/Bacilli/Halobacillus_halophilus/apt_B/GCV
NC_018065/1720349-1720431(+)/Bacilli/Solibacillus_silvestris/apt_A/GCV
NC_018065/1720440-1720548(+)/Bacilli/Solibacillus_silvestris/apt_B/GCV
NC_020164/1270725-1270813(+)/Bacilli/Staphylococcus_warneri/singlet_1/GCV
NC_021171/3397617-3397695(-)/Bacilli/Bacillus_sp/apt_A/GCV
NC_021171/3397532-3397607(-)/Bacilli/Bacillus_sp/apt_B/GCV
NC_022524/3286024-3286101(-)/Bacilli/Bacillus_infantis/apt_A/GCV
NC_022524/3285939-3286014(-)/Bacilli/Bacillus_infantis/apt_B/GCV
NC_022737/146880-146968(-)/Bacilli/Staphylococcus_pasteuri/singlet_1/GCV
NZ_AP013294/2865878-2865959(-)/Bacilli/Bacillus_sp/apt_A/GCV
NZ_AP013294/2865793-2865871(-)/Bacilli/Bacillus_sp/apt_B/GCV
NZ_AP014944/1354206-1354289(+)/Bacilli/Staphylococcus_schleiferi/singlet_1/GCV
NZ_CCSE01000001/796366-796446(+)/Bacilli/Jeotgalicoccus_sp/singlet_0/GCV
NZ_CM001483/1078319-1078397(+)/Bacilli/Geobacillus_thermoglucosidans/apt_A/GCV
NZ_CM001483/1078405-1078489(+)/Bacilli/Geobacillus_thermoglucosidans/apt_B/GCV
NZ_CM001792/1912699-1912780(-)/Bacilli/Oceanobacillus_kimchii/apt_A/GCV
NZ_CM001792/1912594-1912692(-)/Bacilli/Oceanobacillus_kimchii/apt_B/GCV
NZ_CM002692/2459122-2459200(-)/Bacilli/Geobacillus_stearothermophilus/apt_A/GCV
NZ_CM002692/2459036-2459112(-)/Bacilli/Geobacillus_stearothermophilus/apt_B/GCV
NZ_CP003923/2412993-2413078(-)/Bacilli/Bacillus_lehensis/apt_A/GCV
NZ_CP003923/2412895-2412984(-)/Bacilli/Bacillus_lehensis/apt_B/GCV
NZ_CP006837/2927046-2927128(-)/Bacilli/Lysinibacillus_varians/apt_A/GCV
NZ_CP006837/2926932-2927037(-)/Bacilli/Lysinibacillus_varians/apt_B/GCV
NZ_CP007739/2295090-2295172(-)/Bacilli/Bacillus_methanolicus/apt_A/GCV
NZ_CP007739/2295003-2295081(-)/Bacilli/Bacillus_methanolicus/apt_B/GCV
NZ_CP008747/1398221-1398305(+)/Bacilli/Staphylococcus_hyicus/singlet_1/GCV
NZ_CP008855/778913-778991(-)/Bacilli/Bacillus_sp/apt_A/GCV
NZ_CP008855/778828-778903(-)/Bacilli/Bacillus_sp/apt_B/GCV
NZ_CP008876/1734408-1734485(-)/Bacilli/Terribacillus_aidingensis/apt_A/GCV
NZ_CP008876/1734314-1734398(-)/Bacilli/Terribacillus_aidingensis/apt_B/GCV
NZ_CP008934/2794162-2794240(+)/Bacilli/Geobacillus_stearothermophilus/apt_A/GCV

Appendix

NZ_CP008934/2794249-2794364(+)/Bacilli/Geobacillus_stearothermophilus/apt_B/GCV
NZ_CP009129/1365385-1365466(+)/Bacilli/Planococcus_sp/apt_A/GCV
NZ_CP009129/1365473-1365552(+)/Bacilli/Planococcus_sp/apt_B/GCV
NZ_CP009416/1967796-1967874(-)/Bacilli/Jeotgalibacillus_malaysiensis/apt_A/GCV
NZ_CP009416/1967708-1967785(-)/Bacilli/Jeotgalibacillus_malaysiensis/apt_B/GCV
NZ_CP009623/728092-728173(-)/Bacilli/Staphylococcus_agnetis/singlet_0/GCV
NZ_CP009709/721275-721355(-)/Bacilli/Bacillus_coagulans/apt_A/GCV
NZ_CP009709/721184-721266(-)/Bacilli/Bacillus_coagulans/apt_B/GCV
NZ_CP010820/1718515-1718598(+)/Bacilli/Lysinibacillus_fusiformis/apt_A/GCV
NZ_CP010820/1718607-1718719(+)/Bacilli/Lysinibacillus_fusiformis/apt_B/GCV
NZ_CP011008/3501757-3501834(-)/Bacilli/Bacillus_simplex/apt_A/GCV
NZ_CP011008/3501666-3501748(-)/Bacilli/Bacillus_simplex/apt_B/GCV
NZ_CP011366/1420311-1420389(+)/Bacilli/Salinicoccus_halodurans/singlet_0/GCV
NZ_CP011528/1527333-1527420(-)/Bacilli/Staphylococcus_aureus/singlet_0/GCV
NZ_CP011974/3996095-3996174(-)/Bacilli/Bacillus_endophyticus/apt_A/GCV
NZ_CP011974/3996002-3996087(-)/Bacilli/Bacillus_endophyticus/apt_B/GCV
NZ_CP012024/2219874-2219952(-)/Bacilli/Bacillus_smithii/apt_A/GCV
NZ_CP012024/2219793-2219865(-)/Bacilli/Bacillus_smithii/apt_B/GCV
NZ_CP012152/1897643-1897721(-)/Bacilli/Anoxybacillus_gonensis/apt_A/GCV
NZ_CP012152/1897559-1897633(-)/Bacilli/Anoxybacillus_gonensis/apt_B/GCV
NZ_CP012601/681422-681504(+)/Bacilli/Bacillus_sp/apt_A/GCV
NZ_CP012601/681514-681599(+)/Bacilli/Bacillus_sp/apt_B/GCV
NZ_CP012602/2037426-2037505(+)/Bacilli/Bacillus_sp/apt_A/GCV
NZ_CP012602/2037514-2037595(+)/Bacilli/Bacillus_sp/apt_B/GCV
NZ_CP013661/214637-214718(-)/Bacilli/Planococcus_kocurii/apt_A/GCV
NZ_CP013661/214554-214630(-)/Bacilli/Planococcus_kocurii/apt_B/GCV
NZ_CP014342/825429-825507(+)/Bacilli/Geobacillus_subterraneus/apt_A/GCV
NZ_CP014342/825516-825634(+)/Bacilli/Geobacillus_subterraneus/apt_B/GCV
NZ_CP014643/2310694-2310777(+)/Bacilli/Lysinibacillus_sphaericus/apt_A/GCV
NZ_CP014643/2310786-2310898(+)/Bacilli/Lysinibacillus_sphaericus/apt_B/GCV
NZ_CP015226/4221431-4221513(-)/Bacilli/Bacillus_sp/apt_A/GCV
NZ_CP015226/4221337-4221422(-)/Bacilli/Bacillus_sp/apt_B/GCV
NZ_CP015506/4039287-4039364(-)/Bacilli/Bacillus_oceanisediminis/apt_A/GCV
NZ_CP015506/4039202-4039280(-)/Bacilli/Bacillus_oceanisediminis/apt_B/GCV
NZ_CP016534/2665537-2665618(-)/Bacilli/Planococcus_antarcticus/apt_A/GCV
NZ_CP016534/2665454-2665530(-)/Bacilli/Planococcus_antarcticus/apt_B/GCV
NZ_CP016537/642056-642137(+)/Bacilli/Planococcus_halocryophilus/apt_A/GCV
NZ_CP016537/642144-642223(+)/Bacilli/Planococcus_halocryophilus/apt_B/GCV

NZ_CP016540/3072725-3072806(+)/Bacilli/Planococcus_sp/apt_A/GCV
NZ_CP016540/3072813-3072889(+)/Bacilli/Planococcus_sp/apt_B/GCV
NZ_CP016543/1860850-1860931(-)/Bacilli/Planococcus_donghaensis/apt_A/GCV
NZ_CP016543/1860764-1860843(-)/Bacilli/Planococcus_donghaensis/apt_B/GCV

Table 2.10: 35 Bacilli riboswitch aptamer sequences from aptamer-2 and singleton type-2 sub-cluster with paired aptamer-1 supplemented in.

NC_002737/1046368-1046430(+)/Bacilli/Streptococcus_pyogenes/apt_A/TP
NC_002737/1046443-1046519(+)/Bacilli/Streptococcus_pyogenes/apt_B/TP
NC_003028/387500-387576(+)/Bacilli/Streptococcus_pneumoniae/singlet_2/TP
NC_004350/1115970-1116039(+)/Bacilli/Streptococcus_mutans/singlet_0/TP
NC_009009/353828-353904(+)/Bacilli/Streptococcus_sanguinis/singlet_2/TP
NC_011134/935489-935551(-)/Bacilli/Streptococcus_equi/apt_A/TP
NC_011134/935400-935476(-)/Bacilli/Streptococcus_equi/apt_B/TP
NC_012004/961950-962014(+)/Bacilli/Streptococcus_uberis/apt_A/TP
NC_012004/962025-962102(+)/Bacilli/Streptococcus_uberis/apt_B/TP
NC_012470/1186061-1186123(+)/Bacilli/Streptococcus_equi/apt_A/TP
NC_012470/1186136-1186212(+)/Bacilli/Streptococcus_equi/apt_B/TP
NC_012891/1062484-1062547(+)/Bacilli/Streptococcus_dysgalactiae/apt_A/TP
NC_012891/1062560-1062636(+)/Bacilli/Streptococcus_dysgalactiae/apt_B/TP
NC_013853/1774989-1775065(-)/Bacilli/Streptococcus_mitis/singlet_2/TP
NC_014483/4053479-4053557(-)/Bacilli/Paenibacillus_polymyxa/singlet_0/TP
NC_014622/4416795-4416885(-)/Bacilli/Paenibacillus_polymyxa/apt_A/TP
NC_014622/4416707-4416786(-)/Bacilli/Paenibacillus_polymyxa/apt_B/TP
NC_015291/1570417-1570493(-)/Bacilli/Streptococcus_oralis/singlet_0/TP
NC_015558/862065-862130(+)/Bacilli/Streptococcus_parauberis/apt_A/TP
NC_015558/862143-862220(+)/Bacilli/Streptococcus_parauberis/apt_B/TP
NC_015678/1265775-1265853(-)/Bacilli/Streptococcus_parasanguinis/singlet_2/TP
NC_015875/463623-463699(+)/Bacilli/Streptococcus_pseudopneumoniae/singlet_2/TP
NC_016641/70766-70844(-)/Bacilli/Paenibacillus_terrae/singlet_2/TP
NC_017040/957614-957675(+)/Bacilli/Streptococcus_pyogenes/apt_A/TP
NC_017040/957688-957764(+)/Bacilli/Streptococcus_pyogenes/apt_B/TP
NC_017905/1829191-1829269(-)/Bacilli/Streptococcus_parasanguinis/singlet_2/TP
NC_021175/1794922-1794998(-)/Bacilli/Streptococcus_oligofermentans/singlet_2/TP
NZ_CP006872/4502101-4502179(-)/Bacilli/Paenibacillus_polymyxa/singlet_2/TP
NZ_CP007628/1580158-1580234(-)/Bacilli/Streptococcus_sp/singlet_2/TP
NZ_CP010783/915018-915093(-)/Bacilli/Streptococcus_iniae/apt_A/TP

NZ_CP010783/914935-915012(-)/Bacilli/Streptococcus_iniae/apt_B/TP
NZ_CP011512/4592275-4592353(-)/Bacilli/Paenibacillus_peoriae/singlet_0/TP
NZ_CP012646/743371-743447(+)/Bacilli/Streptococcus_mitis/singlet_2/TP
NZ_CP013651/1185245-1185319(-)/Bacilli/Streptococcus_sp/singlet_2/TP
NZ_CP014264/635631-635703(+)/Bacilli/Streptococcus_sp/singlet_2/TP

Table 2.11: 40 randomly selected aptamers used as out-group for inter-edge network density.

NZ_CP016328/835794-835874(+)/Proteobacteria/Gluconobacter_oxydans/apt_A/GCV
NC_015589/3340289-3340382(-)/Firmicutes/Desulfotomaculum_ruminis/apt_A/GCV
NZ_CM002097/80499-80601(-)/Actinobacteria/Mycobacterium_tuberculosis/apt_B/GLY_MET
NZ_CP010537/1157197-1157283(+)/Proteobacteria/Cupriavidus_basilensis/apt_A/GLY_MET
NZ_CM000728/2134021-2134123(-)/Firmicutes/Bacillus_cereus/apt_B/TP
NZ_CP006992/657797-657890(-)/Proteobacteria/Methylobacterium_sp/apt_B/GCV
NZ_CP013297/963598-963685(+)/Actinobacteria/Arthrobacter_sp/apt_A/GCV
NZ_CP012573/2298658-2298734(-)/Actinobacteria/Clavibacter_michiganensis/apt_A/GCV
NC_008322/3918318-3918426(-)/Proteobacteria/Shewanella_sp/apt_A/TP
NZ_HG316453/958105-958167(+)/Firmicutes/Streptococcus_pyogenes/apt_A/TP
NZ_LN868200/1597609-1597738(+)/Proteobacteria/Acinetobacter_baumannii/apt_B/GLY_MET
NZ_CP013200/1443775-1443860(-)/Actinobacteria/Arthrobacter_alpinus/apt_A/TP
NC_012442/692328-692420(-)/Proteobacteria/Brucella_melitensis/apt_B/GCV
NC_003155/3402036-3402122(+)/Actinobacteria/Streptomyces_avermitilis/apt_A/GCV
NC_016010/1329013-1329117(-)/Proteobacteria/Xanthomonas_axonopodis/apt_B/GCV
NC_014923/4320568-4320644(+)/Proteobacteria/Mesorhizobium_ciceri/apt_A/GCV
NC_010674/1000474-1000563(+)/Firmicutes/Clostridium_botulinum/apt_B/TP
NZ_CM000736/2015117-2015221(+)/Firmicutes/Bacillus_cereus/apt_B/TP
NC_008781/2825251-2825342(+)/Proteobacteria/Polaromonas_naphthalenivorans/apt_B/GCV
NZ_CP013652/3047205-3047287(+)/Firmicutes/Paenibacillus_naphthalenovorans/apt_B/GCV
NZ_CP009447/2389487-2389574(-)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CM002073/2060215-2060298(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CP008921/220395-220489(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NZ_CM002130/952340-952436(-)/Firmicutes/Streptococcus_sp/singlet_2/TP

NZ_CP016026/218197-218293(-)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NZ_HG969191/1506922-1507011(-)/Proteobacteria/Bartonella_henselae/singlet_1/GCV
NC_021005/383812-383888(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_017576/1230862-1231043(+)/Firmicutes/Streptococcus_gallolyticus/singlet_2/TP
NC_015145/836223-836311(+)/Actinobacteria/Pseudarthrobacter_phenanthrenivorans/singlet_1/GCV
NC_014147/1271717-1271825(+)/Proteobacteria/Moraxella_catarrhalis/singlet_1/GCV
NC_019966/3204883-3204969(+)/Actinobacteria/Mycobacterium_sp/singlet_1/GCV
NZ_CP014945/2182067-2182194(+)/Proteobacteria/Psychrobacter_alimentarius/singlet_1/GCV
NZ_CP008744/2057565-2057648(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NZ_LN831029/5941333-5941443(-)/Proteobacteria/Achromobacter_xylosoxidans/singlet_2/GCV
NC_022592/3168961-3169033(+)/Firmicutes/Clostridium_autoethanogenum/singlet_1/OTHER
NZ_CP014234/711783-711889(+)/Proteobacteria/Moraxella_osloensis/singlet_1/GCV
NZ_CP015483/1670373-1670450(+)/Proteobacteria/Acinetobacter_baumannii/singlet_2/TP
NC_021883/1803842-1803923(+)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NC_006370/2453503-2453604(+)/Proteobacteria/Photobacterium_profundum/singlet_2/TP
NC_013204/146216-146311(-)/Actinobacteria/Eggerthella_lenta/singlet_2/TP

Table 2.12: 606 Actinobacteria riboswitch aptamer sequences used for graph clustering analysis.

NC_014550/3470645-3470748(+)/Actinobacteria/Glutamicibacter_arilaitensis/apt_B/TP
NC_018681/4124587-4124673(+)/Actinobacteria/Nocardia_brasiliensis/apt_B/GLY_MET
NC_014165/1966405-1966506(+)/Actinobacteria/Thermobispora_bispora/apt_B/GCV
NC_014550/2942762-2942862(-)/Actinobacteria/Glutamicibacter_arilaitensis/apt_B/GCV
NZ_CP014196/778614-778711(+)/Actinobacteria/Arthrobacter_sp/apt_B/TP
NZ_CP012752/3350647-3350734(+)/Actinobacteria/Kibdelosporangium_phytohabitans/apt_B/GLY_MET
NC_007164/259392-259491(+)/Actinobacteria/Corynebacterium_jeikeium/apt_B/GCV
NC_015673/350132-350236(-)/Actinobacteria/Corynebacterium_resistens/apt_B/GCV

Appendix

NC_015859/462532-462648(+)/Actinobacteria/Corynebacterium_variabile/apt_B/GCV
NC_021663/365015-365131(+)/Actinobacteria/Corynebacterium_terpenotabidum/apt_B/GCV
NC_011886/583465-583553(+)/Actinobacteria/Pseudarthrobacter_chlorophenolicus/apt_B/TP
NZ_CP014196/911934-912051(+)/Actinobacteria/Arthrobacter_sp/apt_B/GCV
NC_008711/1042888-1043007(+)/Actinobacteria/Paenarthrobacter_aurescens/apt_B/GCV
NC_018531/998323-998442(+)/Actinobacteria/Arthrobacter_sp/apt_B/GCV
NC_008541/825754-825895(+)/Actinobacteria/Arthrobacter_sp/apt_B/GCV
NZ_CP012479/3051810-3051905(-)/Actinobacteria/Arthrobacter_sp/apt_B/GCV
NZ_CP007595/1926337-1926436(+)/Actinobacteria/Arthrobacter_sp/apt_B/TP
NZ_CP007595/1925837-1925928(-)/Actinobacteria/Arthrobacter_sp/apt_B/GCV
NC_019673/7527172-7527260(-)/Actinobacteria/Saccharothrix_espanaensis/apt_B/GCV
NZ_CP016793/4776199-4776288(+)/Actinobacteria/Lentzea_sp/apt_B/GCV
NC_013131/5445899-5446016(-)/Actinobacteria/Catenulispora_acidiphila/apt_B/GCV
NZ_CP013200/1444196-1444288(+)/Actinobacteria/Arthrobacter_alpinus/apt_B/GCV
NZ_CP013745/2764458-2764550(-)/Actinobacteria/Arthrobacter_sp/apt_B/GCV
NZ_CP013745/2764966-2765066(+)/Actinobacteria/Arthrobacter_sp/apt_B/TP
NC_008711/883526-883622(+)/Actinobacteria/Paenarthrobacter_aurescens/apt_B/TP
NC_018531/838982-839078(+)/Actinobacteria/Arthrobacter_sp/apt_B/TP
NZ_CP012479/3052245-3052344(+)/Actinobacteria/Arthrobacter_sp/apt_B/TP
NC_008711/4380596-4380687(-)/Actinobacteria/Paenarthrobacter_aurescens/apt_B/GCV
NC_015145/3634462-3634564(-)/Actinobacteria/Pseudarthrobacter_phenanthrenivorans/apt_B/GCV
NZ_CP014196/4139088-4139179(-)/Actinobacteria/Arthrobacter_sp/apt_B/GCV
NC_018531/4540932-4541023(-)/Actinobacteria/Arthrobacter_sp/apt_B/GCV
NC_015656/806884-806993(+)/Actinobacteria/Frankia_symbiont/apt_B/GCV
NZ_CP012677/2738662-2738760(+)/Actinobacteria/Arthrobacter_alpinus/apt_B/TP
NC_013530/47394-47508(+)/Actinobacteria/Xylanimonas_cellulosilytica/apt_B/GCV
NC_008578/2089950-2090035(-)/Actinobacteria/Acidothermus_cellulolyticus/apt_B/GCV
NC_015312/2700577-2700674(-)/Actinobacteria/Pseudonocardia_dioxanivorans/apt_B/GCV
NC_014830/3963733-3963822(+)/Actinobacteria/Intrasporangium_calvum/apt_B/TP
NC_011886/582911-583010(-)/Actinobacteria/Pseudarthrobacter_chlorophenolicus/apt_B/GCV
NZ_CP012171/2917122-2917222(-)/Actinobacteria/Arthrobacter_sp/apt_B/GCV

NZ_CP012750/2923497-2923597(-)/Actinobacteria/Glutamicibacter_arilaitensis/apt_B/GCV
NZ_CP013747/3362652-3362752(+)/Actinobacteria/Pseudarthrobacter_sulfonivorans/apt_B/GCV
NZ_CP013200/1443670-1443767(-)/Actinobacteria/Arthrobacter_alpinus/apt_B/TP
NC_008541/66063-66162(+)/Actinobacteria/Arthrobacter_sp/apt_B/GLY_MET
NZ_CP012677/2738147-2738252(-)/Actinobacteria/Arthrobacter_alpinus/apt_B/GCV
NC_010168/1481681-1481777(+)/Actinobacteria/Renibacterium_salmoninarum/apt_B/GCV
NZ_CP011005/2008781-2008878(+)/Actinobacteria/Arthrobacter_sp/apt_B/GCV
NC_008278/6999578-6999703(-)/Actinobacteria/Frankia_alni/apt_B/GCV
NZ_CM001489/6119637-6119762(+)/Actinobacteria/Frankia_sp/apt_B/GCV
NC_007777/4868066-4868187(-)/Actinobacteria/Frankia_sp/apt_B/GLY_MET
NZ_CP013747/1435877-1436017(-)/Actinobacteria/Pseudarthrobacter_sulfonivorans/apt_B/GCV
NC_012522/2853627-2853729(+)/Actinobacteria/Rhodococcus_opacus/apt_B/GCV
NC_008268/3094413-3094515(+)/Actinobacteria/Rhodococcus_jostii/apt_B/GCV
NZ_CP009111/6794851-6794953(+)/Actinobacteria/Rhodococcus_opacus/apt_B/GCV
NZ_CP003949/7432755-7432857(+)/Actinobacteria/Rhodococcus_opacus/apt_B/GCV
NZ_CP009914/781784-781874(+)/Actinobacteria/Mycobacterium_sp/apt_B/GLY_MET
NZ_CP011269/696540-696630(+)/Actinobacteria/Mycobacterium_fortuitum/apt_B/GLY_MET
NC_000962/79100-79202(-)/Actinobacteria/Mycobacterium_tuberculosis/apt_B/GLY_MET
NC_002945/79127-79229(-)/Actinobacteria/Mycobacterium_bovis/apt_B/GLY_MET
NC_015758/79188-79290(-)/Actinobacteria/Mycobacterium_africanum/apt_B/GLY_MET
NC_015848/81161-81263(-)/Actinobacteria/Mycobacterium_canettii/apt_B/GLY_MET
NC_019952/80807-80909(-)/Actinobacteria/Mycobacterium_canettii/apt_B/GLY_MET
NZ_CM001225/79284-79386(-)/Actinobacteria/Mycobacterium_tuberculosis/apt_B/GLY_MET
NC_018581/1356080-1356169(+)/Actinobacteria/Gordonia_sp/apt_B/GLY_MET
NZ_CP012697/1788415-1788503(+)/Actinobacteria/Microbacterium_sp/apt_B/OTHER
NC_020520/2864792-2864879(-)/Actinobacteria/Ilumatobacter_coccineus/apt_B/GCV
NC_015312/2700681-2700759(-)/Actinobacteria/Pseudonocardia_dioxanivorans/apt_A/GCV
NC_020504/3338148-3338250(+)/Actinobacteria/Streptomyces_davawensis/apt_B/GCV
NZ_CP009754/2774026-2774134(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV

NC_003888/5958891-5959000(-)/Actinobacteria/Streptomyces_coelicolor/apt_B/GCV
NZ_CM001889/5976509-5976618(-)/Actinobacteria/Streptomyces_lividans/apt_B/GCV
NZ_CP010519/2282352-2282471(+)/Actinobacteria/Streptomyces_albus/apt_B/GCV
NC_016111/4607580-4607703(-)/Actinobacteria/Streptomyces_cattleya/apt_B/GCV
NZ_CP015849/5110228-5110343(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP009438/5330898-5331005(-)/Actinobacteria/Streptomyces_glaucescens/apt_B/GCV
NZ_CM001165/2152879-2152980(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CM002271/1754253-1754366(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_020990/1617785-1617898(+)/Actinobacteria/Streptomyces_albus/apt_B/GCV
NZ_CP014485/1894621-1894734(+)/Actinobacteria/Streptomyces_albus/apt_B/GCV
NZ_CM000951/6448744-6448841(-)/Actinobacteria/Streptomyces_sviceus/apt_B/GCV
NZ_CP015866/5090218-5090329(-)/Actinobacteria/Streptomyces_parvulus/apt_B/GCV
NZ_CP012382/5531552-5531662(-)/Actinobacteria/Streptomyces_ambofaciens/apt_B/GCV
NZ_LN997842/6693476-6693582(-)/Actinobacteria/Streptomyces_reticuli/apt_B/GCV
NZ_CP013743/3659350-3659451(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP013142/5507328-5507444(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_013929/3128007-3128141(+)/Actinobacteria/Streptomyces_scabiei/apt_B/GCV
NZ_CP009313/5122156-5122261(-)/Actinobacteria/Streptomyces_nodosus/apt_B/GCV
NC_003155/3402129-3402230(+)/Actinobacteria/Streptomyces_avermitilis/apt_B/GCV
NZ_CP010849/5479859-5479969(-)/Actinobacteria/Streptomyces_cyaneogriseus/apt_B/GCV
NZ_LN831790/2680853-2680963(+)/Actinobacteria/Streptomyces_leeuwenhoekii/apt_B/GCV
NC_017765/7469079-7469185(-)/Actinobacteria/Streptomyces_hygroscopicus/apt_B/GCV
NZ_CM002285/2815896-2816002(+)/Actinobacteria/Streptomyces_roseochromogenus/apt_B/GCV
NZ_CP015098/6272882-6272996(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_021985/5904714-5904821(-)/Actinobacteria/Streptomyces_collinus/apt_B/GCV
NZ_CP012171/3268663-3268765(+)/Actinobacteria/Arthrobacter_sp/apt_B/TP
NC_012704/371667-371823(+)/Actinobacteria/Corynebacterium_kroppenstedtii/apt_B/GCV
NC_013441/2957852-2957947(-)/Actinobacteria/Gordonia_bronchialis/singlet_0/GCV
NZ_CP007156/301966-302064(+)/Actinobacteria/Corynebacterium_falsenii/apt_B/GCV

NC_013947/3988940-3989043(-)/Actinobacteria/Stackebrandtia_nassauensis/apt_B/GCV
NC_011886/967316-967444(+)/Actinobacteria/Pseudarthrobacter_chlorophenolicus/apt_B/GCV
NZ_CP012507/138337-138435(+)/Actinobacteria/Kocuria_palustris/apt_B/GCV
NZ_CP014480/2726068-2726167(-)/Actinobacteria/Kocuria_turfanensis/apt_B/GLY_MET
NZ_CM000950/2417694-2417778(+)/Actinobacteria/Streptomyces_pristinaespiralis/apt_A/GCV
NC_009142/6403700-6403792(+)/Actinobacteria/Saccharopolyspora_erythraea/apt_B/GCV
NC_018142/2071062-2071130(+)/Actinobacteria/Propionibacterium_propionicum/apt_B/GCV
NZ_CP006842/3107043-3107127(-)/Actinobacteria/Corynebacterium_glyciniphilum/apt_A/GCV
NZ_CM002280/2499972-2500090(+)/Actinobacteria/Streptomyces_niveus/apt_B/GCV
NZ_CP007699/2827920-2828031(+)/Actinobacteria/Streptomyces_lydicus/apt_B/GCV
NZ_CP003987/3509108-3509219(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_021177/5913230-5913365(-)/Actinobacteria/Streptomyces_fulvissimus/apt_B/GCV
NC_016582/4674555-4674658(+)/Actinobacteria/Streptomyces_bingchenggensis/apt_B/GCV
NC_010572/2416231-2416362(+)/Actinobacteria/Streptomyces_griseus/apt_B/GCV
NZ_CM003601/6173944-6174074(-)/Actinobacteria/Streptomyces_anulatus/apt_B/GCV
NZ_CM000950/2417787-2417893(+)/Actinobacteria/Streptomyces_pristinaespiralis/apt_B/GCV
NC_015957/1718555-1718655(-)/Actinobacteria/Streptomyces_violaceusniger/apt_B/GCV
NZ_CP011664/5500255-5500389(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP013738/5714203-5714334(-)/Actinobacteria/Streptomyces_globisporus/apt_B/GCV
NZ_CP006871/6732983-6733095(-)/Actinobacteria/Streptomyces_albulus/apt_B/GCV
NZ_CP013129/6152730-6152849(-)/Actinobacteria/Streptomyces_venezuelae/apt_B/GCV
NC_018750/5558377-5558470(-)/Actinobacteria/Streptomyces_venezuelae/apt_B/GCV
NZ_CP010407/5872050-5872143(-)/Actinobacteria/Streptomyces_vietnamensis/apt_B/GCV
NC_016114/2273576-2273675(+)/Actinobacteria/Streptomyces_pratensis/apt_B/GCV
NC_021055/5158479-5158578(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP011522/2528505-2528624(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV

NC_015953/5168614-5168715(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CM000913/5109390-5109505(-)/Actinobacteria/Streptomyces_clavuligerus/apt_B/GCV
NZ_CP011492/5868408-5868519(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP015726/7543774-7543887(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_015635/2553274-2553364(+)/Actinobacteria/Microlunatus_phosphovorus/singlet_1/GCV
NZ_CP006842/3106954-3107036(-)/Actinobacteria/Corynebacterium_glyciniphilum/apt_B/GCV
NC_013510/2826589-2826702(-)/Actinobacteria/Thermomonospora_curvata/apt_B/GCV
NC_016907/69039-69145(+)/Actinobacteria/Gordonia_polyisoprenivorans/singlet_0/GCV
NC_009480/2494566-2494642(-)/Actinobacteria/Clavibacter_michiganensis/apt_A/GCV
NC_021663/364904-365008(+)/Actinobacteria/Corynebacterium_terpenotabidum/apt_A/GCV
NZ_CP011269/696453-696533(+)/Actinobacteria/Mycobacterium_fortuitum/apt_A/GLY_MET
NZ_CP009914/781695-781777(+)/Actinobacteria/Mycobacterium_sp/apt_A/GLY_MET
NC_007333/1619925-1620006(+)/Actinobacteria/Thermobifida_fusca/singlet_1/GCV
NC_014830/2048165-2048269(-)/Actinobacteria/Intrasporangium_calvum/singlet_1/GCV
NC_021191/4712561-4712720(-)/Actinobacteria/Actinoplanes_sp/singlet_1/GCV
NZ_CP012299/212293-212377(+)/Actinobacteria/Microbacterium_sp/apt_A/GCV
NC_008578/1364130-1364214(-)/Actinobacteria/Acidothermus_cellulolyticus/apt_A/GCV
NZ_CM000913/796880-796967(-)/Actinobacteria/Streptomyces_clavuligerus/apt_A/GCV
NZ_CM001852/1977937-1978047(-)/Actinobacteria/Nocardioides_sp/singlet_1/GCV
NC_021177/5913372-5913459(-)/Actinobacteria/Streptomyces_fulvissimus/apt_A/GCV
NC_015434/3603550-3603638(-)/Actinobacteria/Verrucosispora_maris/singlet_0/GCV
NC_014391/3621234-3621314(-)/Actinobacteria/Micromonospora_aurantiaca/singlet_1/GCV
NC_014815/5602963-5603043(+)/Actinobacteria/Micromonospora_sp/singlet_1/GCV
NZ_CP012573/2298560-2298651(-)/Actinobacteria/Clavibacter_michiganensis/apt_B/GCV
NC_010407/2364506-2364597(-)/Actinobacteria/Clavibacter_michiganensis/apt_B/GCV
NC_009480/2494468-2494559(-)/Actinobacteria/Clavibacter_michiganensis/apt_B/GCV

NZ_CP014761/1026452-1026547(-)/Actinobacteria/Leifsonia_xyli/apt_B/GCV
NC_006087/1714297-1714418(-)/Actinobacteria/Leifsonia_xyli/apt_B/GCV
NC_022438/1801132-1801253(-)/Actinobacteria/Leifsonia_xyli/apt_B/GCV
NZ_CP011043/2169872-2169963(-)/Actinobacteria/Clavibacter_michiganensis/apt_B/GCV
NZ_CP012299/212384-212484(+)/Actinobacteria/Microbacterium_sp/apt_B/GCV
NZ_CM001466/4359260-4359351(+)/Actinobacteria/Saccharomonospora_azurea/apt_B/GCV
NZ_CM001484/1432500-1432593(+)/Actinobacteria/Saccharomonospora_glauca/apt_B/GCV
NC_013159/1369717-1369805(+)/Actinobacteria/Saccharomonospora_viridis/apt_B/GCV
NC_014318/1944254-1944356(+)/Actinobacteria/Amycolatopsis_mediterranei/apt_B/GCV
NC_021252/6835908-6835999(-)/Actinobacteria/Amycolatopsis_orientalis/apt_B/GCV
NZ_CP008953/1835137-1835228(+)/Actinobacteria/Amycolatopsis_japonica/apt_B/GCV
NZ_CP016174/6864381-6864472(-)/Actinobacteria/Amycolatopsis_orientalis/apt_B/GCV
NZ_CP009110/5533167-5533257(-)/Actinobacteria/Amycolatopsis_methanolica/apt_B/GCV
NZ_CM001439/1845491-1845575(+)/Actinobacteria/Saccharomonospora_marina/apt_B/GCV
NC_014210/516684-516775(-)/Actinobacteria/Nocardiopsis_dassonvillei/apt_B/GCV
NC_018524/3432814-3432905(-)/Actinobacteria/Nocardiopsis_alba/apt_B/GCV
NC_012669/2179460-2179556(+)/Actinobacteria/Beutenbergia_cavernae/apt_B/GCV
NC_015953/905931-906058(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP011664/1732212-1732342(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP009922/340177-340303(-)/Actinobacteria/Streptomyces_xiamenensis/apt_B/GCV
NZ_CP010848/1806603-1806701(-)/Actinobacteria/Rathayibacter_toxicus/apt_B/GCV
NC_015125/3325745-3325854(+)/Actinobacteria/Microbacterium_testaceum/apt_B/GCV
NZ_CM001440/1580527-1580618(+)/Actinobacteria/Saccharomonospora_cyanea/apt_B/GCV
NZ_CP011530/2625418-2625503(+)/Actinobacteria/Mycobacterium_immunogenum/singlet_1/GCV
NZ_CP007699/7788291-7788378(+)/Actinobacteria/Streptomyces_lydicus/apt_A/GCV
NZ_CP003987/8094436-8094523(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CP006871/2046804-2046891(-)/Actinobacteria/Streptomyces_albulus/apt_A/GCV

NZ_CP009896/3317636-3317722(+)/Actinobacteria/Pimelobacter_simplex/apt_A/GCV
NC_013235/4424245-4424338(-)/Actinobacteria/Nakamurella_multipartita/singlet_1/GCV
NC_009664/2839982-2840065(+)/Actinobacteria/Kineococcus_radiotolerans/apt_A/GCV
NC_015656/806787-806878(+)/Actinobacteria/Frankia_symbiont/apt_A/GCV
NC_008278/6999709-6999794(-)/Actinobacteria/Frankia_alni/apt_A/GCV
NC_007777/4868193-4868280(-)/Actinobacteria/Frankia_sp/apt_A/GLY_MET
NC_019673/8312390-8312471(-)/Actinobacteria/Saccharothrix_espanaensis/apt_A/GCV
NZ_CP016793/3926557-3926638(+)/Actinobacteria/Lentzea_sp/singlet_1/GCV
NC_008699/2884572-2884656(-)/Actinobacteria/Nocardioides_sp/singlet_1/GCV
NC_014210/1803416-1803495(+)/Actinobacteria/Nocardiopsis_dassonvillei/singlet_1/GCV
NZ_CP008953/4347870-4347961(-)/Actinobacteria/Amycolatopsis_japonica/apt_A/GCV
NC_021252/4103669-4103760(+)/Actinobacteria/Amycolatopsis_orientalis/apt_A/GCV
NZ_CP009110/3119884-3119974(+)/Actinobacteria/Amycolatopsis_methanolica/singlet_0/OTHER
NC_018524/4297874-4297954(+)/Actinobacteria/Nocardiopsis_alba/singlet_0/GCV
NC_022657/4614050-4614129(+)/Actinobacteria/Actinoplanes_friuliensis/singlet_1/GCV
NZ_CP010519/7543934-7544019(+)/Actinobacteria/Streptomyces_albus/apt_A/GCV
NZ_CP011862/564805-564889(+)/Actinobacteria/Pseudonocardia_sp/apt_A/GCV
NZ_CP011868/2333238-2333322(+)/Actinobacteria/Pseudonocardia_sp/apt_A/GCV
NC_015312/3379007-3379087(+)/Actinobacteria/Pseudonocardia_dioxanivorans/apt_A/GCV
NZ_CP010519/2282259-2282345(+)/Actinobacteria/Streptomyces_albus/apt_A/GCV
NC_014165/1085978-1086065(-)/Actinobacteria/Thermobispora_bispora/singlet_0/GCV
NC_012803/318030-318136(+)/Actinobacteria/Micrococcus_luteus/singlet_0/GCV
NC_013929/3127914-3128000(+)/Actinobacteria/Streptomyces_scabiei/apt_A/GCV
NZ_CP013743/3659252-3659343(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CP015098/6273003-6273092(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CM002285/2815802-2815889(+)/Actinobacteria/Streptomyces_roseochromogenus/apt_A/GCV
NZ_CP009438/5331012-5331101(-)/Actinobacteria/Streptomyces_glaucescens/apt_A/GCV
NZ_CP015866/5090336-5090420(-)/Actinobacteria/Streptomyces_parvulus/apt_A/GCV
NC_003888/5959007-5959090(-)/Actinobacteria/Streptomyces_coelicolor/apt_A/GCV

NZ_CM001889/5976625-5976708(-)/Actinobacteria/Streptomyces_lividans/apt_A/GCV
NZ_CP009754/2773936-2774019(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CP013142/5507451-5507541(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CP015849/5110350-5110440(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CP012382/5531669-5531753(-)/Actinobacteria/Streptomyces_ambofaciens/apt_A/GCV
NZ_CP003987/3509016-3509101(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CP006871/6733102-6733187(-)/Actinobacteria/Streptomyces_albulus/apt_A/GCV
NZ_CP007699/2827828-2827913(+)/Actinobacteria/Streptomyces_lydicus/apt_A/GCV
NZ_CP011664/5500396-5500489(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CM001165/2152779-2152872(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CM000951/6448848-6448935(-)/Actinobacteria/Streptomyces_sviceus/apt_A/GCV
NC_015953/5168722-5168808(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NC_016114/2273482-2273569(+)/Actinobacteria/Streptomyces_pratensis/apt_A/GCV
NC_021055/5158585-5158672(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CP011492/5868526-5868614(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NC_019673/7216349-7216429(-)/Actinobacteria/Saccharothrix_espanaensis/apt_A/GCV
NC_013093/6369333-6369415(-)/Actinobacteria/Actinosynnema_mirum/singlet_0/GCV
NC_014318/5512696-5512784(-)/Actinobacteria/Amycolatopsis_mediterranei/apt_A/GCV
NC_009142/4206459-4206551(-)/Actinobacteria/Saccharopolyspora_erythraea/apt_A/GCV
NZ_CP009754/7278332-7278417(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CM000951/1681006-1681091(-)/Actinobacteria/Streptomyces_sviceus/apt_A/GCV
NC_003888/1457849-1457934(-)/Actinobacteria/Streptomyces_coelicolor/apt_A/GCV
NZ_CM001889/1681098-1681183(-)/Actinobacteria/Streptomyces_lividans/apt_A/GCV
NZ_CP015866/1091868-1091953(-)/Actinobacteria/Streptomyces_parvulus/apt_A/GCV
NC_016582/4674448-4674548(+)/Actinobacteria/Streptomyces_bingchenggensis/apt_A/GCV
NZ_CP015726/7543894-7543992(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CM002280/2499878-2499965(+)/Actinobacteria/Streptomyces_niveus/apt_A/GCV
NC_015957/1718662-1718762(-)/Actinobacteria/Streptomyces_violaceusniger/apt_A/GCV

NZ_CP010849/5479976-5480059(-)/Actinobacteria/Streptomyces_cyaneogriseus/apt_A/GCV
NZ_LN831790/2680763-2680846(+)/Actinobacteria/Streptomyces_leeuwenhoekii/apt_A/GCV
NC_017765/7469192-7469277(-)/Actinobacteria/Streptomyces_hygroscopicus/apt_A/GCV
NC_020504/3338056-3338141(+)/Actinobacteria/Streptomyces_davawensis/apt_A/GCV
NC_021985/5904828-5904915(-)/Actinobacteria/Streptomyces_collinus/apt_A/GCV
NZ_LN997842/6693589-6693683(-)/Actinobacteria/Streptomyces_reticuli/apt_A/GCV
NC_018750/5558477-5558566(-)/Actinobacteria/Streptomyces_venezuelae/apt_A/GCV
NZ_CP010407/5872150-5872239(-)/Actinobacteria/Streptomyces_vietnamensis/apt_A/GCV
NZ_CP013129/6152856-6152944(-)/Actinobacteria/Streptomyces_venezuelae/apt_A/GCV
NC_010572/2416137-2416224(+)/Actinobacteria/Streptomyces_griseus/apt_A/GCV
NZ_CM003601/6174081-6174168(-)/Actinobacteria/Streptomyces_anulatus/apt_A/GCV
NZ_CP011522/2528411-2528498(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CP013738/5714341-5714428(-)/Actinobacteria/Streptomyces_globisporus/apt_A/GCV
NZ_CP009313/5122268-5122358(-)/Actinobacteria/Streptomyces_nodosus/apt_A/GCV
NC_003155/3402036-3402122(+)/Actinobacteria/Streptomyces_avermitilis/apt_A/GCV
NC_020990/1617683-1617778(+)/Actinobacteria/Streptomyces_albus/apt_A/GCV
NZ_CM002271/1754151-1754246(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NC_013169/1537225-1537300(-)/Actinobacteria/Kytococcus_sedentarius/singlet_1/GCV
NC_017803/4829529-4829605(-)/Actinobacteria/Actinoplanes_sp/singlet_1/GCV
NC_014158/2430914-2431005(-)/Actinobacteria/Tsukamurella_paurometabola/singlet_1/GCV
NC_016109/1692732-1692828(-)/Actinobacteria/Kitasatospora_setae/singlet_1/GCV
NZ_CP014313/2537717-2537811(+)/Actinobacteria/Microbacterium_sp/apt_B/GCV
NC_009664/2840072-2840160(+)/Actinobacteria/Kineococcus_radiotolerans/apt_B/GCV
NC_013093/7326537-7326626(-)/Actinobacteria/Actinosynnema_mirum/apt_B/GCV
NZ_CP015810/2574960-2575041(-)/Actinobacteria/Microbacterium_chocolatum/apt_A/GCV
NZ_CP011112/3231756-3231842(+)/Actinobacteria/Luteipulveratus_mongoliensis/singlet_1/GCV
NC_013093/7326634-7326714(-)/Actinobacteria/Actinosynnema_mirum/apt_A/GCV

NC_009142/7144784-7144881(-)/Actinobacteria/Saccharopolyspora_erythraea/apt_A/GCV
NZ_CP007155/8798971-8799054(-)/Actinobacteria/Kutzneria_albida/apt_A/GLY_MET
NC_013947/3989049-3989144(-)/Actinobacteria/Stackebrandtia_nassauensis/apt_A/GCV
NC_014210/516782-516873(-)/Actinobacteria/Nocardiopsis_dassonvillei/apt_A/GCV
NC_018524/3432912-3433003(-)/Actinobacteria/Nocardiopsis_alba/apt_A/GCV
NC_013510/2826709-2826794(-)/Actinobacteria/Thermomonospora_curvata/apt_A/GCV
NC_014165/1966315-1966398(+)/Actinobacteria/Thermobispora_bispora/apt_A/GCV
NC_013729/7009633-7009726(-)/Actinobacteria/Kribbella_flavida/singlet_1/GLY_MET
NZ_CP010989/2365768-2365853(+)/Actinobacteria/Pseudonocardia_sp/singlet_0/GCV
NC_016111/4607710-4607798(-)/Actinobacteria/Streptomyces_cattleya/apt_A/GCV
NZ_CM002280/7611562-7611649(+)/Actinobacteria/Streptomyces_niveus/apt_A/GCV
NC_013595/1874341-1874429(-)/Actinobacteria/Streptosporangium_roseum/apt_A/GCV
NC_016582/10368443-10368530(+)/Actinobacteria/Streptomyces_bingchenggensis/apt_A/GCV
NC_018750/1091574-1091663(-)/Actinobacteria/Streptomyces_venezuelae/apt_A/GCV
NZ_CP010407/1437023-1437112(-)/Actinobacteria/Streptomyces_vietnamensis/apt_A/GCV
NC_006361/2650094-2650182(+)/Actinobacteria/Nocardia_farcinica/apt_A/GCV
NZ_LN868938/1621027-1621121(-)/Actinobacteria/Nocardia_farcinica/apt_A/GCV
NZ_CM001489/6119546-6119631(+)/Actinobacteria/Frankia_sp/apt_A/GCV
NC_016111/788066-788160(-)/Actinobacteria/Streptomyces_cattleya/apt_A/GCV
NC_020504/8089952-8090037(+)/Actinobacteria/Streptomyces_davawensis/apt_A/GCV
NZ_CP010849/1475461-1475546(-)/Actinobacteria/Streptomyces_cyaneogriseus/apt_A/GCV
NZ_LN831790/6698479-6698564(+)/Actinobacteria/Streptomyces_leeuwenhoekii/apt_A/GCV
NZ_CP015849/1057259-1057343(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CP013142/1478628-1478713(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NC_021985/1691298-1691382(-)/Actinobacteria/Streptomyces_collinus/apt_A/GCV
NC_017765/3307343-3307427(-)/Actinobacteria/Streptomyces_hygroscopicus/apt_A/GCV
NZ_LN997842/2143029-2143113(-)/Actinobacteria/Streptomyces_reticuli/apt_A/GCV

NZ_CP009313/1365801-1365886(-)/Actinobacteria/Streptomyces_nodosus/apt_A/GCV
NC_003155/8344964-8345049(+)/Actinobacteria/Streptomyces_avermitilis/apt_A/GCV
NZ_CP015098/1567551-1567636(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NC_013929/8442527-8442613(+)/Actinobacteria/Streptomyces_scabiei/apt_A/GCV
NZ_CP012382/1458787-1458872(-)/Actinobacteria/Streptomyces_ambofaciens/apt_A/GCV
NC_015953/906065-906151(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CM002285/8110928-8111012(+)/Actinobacteria/Streptomyces_roseochromogenus/apt_A/GCV
NZ_CP011492/425023-425109(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NC_015957/6466040-6466127(-)/Actinobacteria/Streptomyces_violaceusniger/apt_A/GCV
NZ_CP009438/1255388-1255473(-)/Actinobacteria/Streptomyces_glaucescens/apt_A/GCV
NZ_CP013743/8253683-8253768(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NC_020990/6182033-6182120(+)/Actinobacteria/Streptomyces_albus/apt_A/GCV
NZ_CM002271/6184297-6184384(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NC_021177/1028304-1028391(-)/Actinobacteria/Streptomyces_fulvissimus/apt_A/GCV
NZ_CP011522/7388394-7388481(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CM003601/1085076-1085163(-)/Actinobacteria/Streptomyces_anulatus/apt_A/GCV
NC_010572/7259079-7259166(+)/Actinobacteria/Streptomyces_griseus/apt_A/GCV
NZ_CP013738/1090499-1090586(-)/Actinobacteria/Streptomyces_globisporus/apt_A/GCV
NZ_CP007155/9112607-9112691(+)/Actinobacteria/Kutzneria_albida/apt_A/GCV
NZ_CP015726/9704223-9704310(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NC_013510/4395172-4395258(+)/Actinobacteria/Thermomonospora_curvata/apt_A/GCV
NZ_CM002177/2385725-2385809(-)/Actinobacteria/Rhodococcus_aetherivorans/apt_A/GCV
NZ_CP015529/2403172-2403256(+)/Actinobacteria/Rhodococcus_sp/apt_A/GCV
NZ_CP006850/3470162-3470246(+)/Actinobacteria/Nocardia_nova/apt_A/GCV
NC_018681/4014559-4014641(+)/Actinobacteria/Nocardia_brasiliensis/apt_A/GCV
NZ_CP003949/5172988-5173101(-)/Actinobacteria/Rhodococcus_opacus/singlet_0/GCV
NC_008268/963522-963635(-)/Actinobacteria/Rhodococcus_jostii/singlet_0/GCV
NZ_CP009111/4794790-4794903(-)/Actinobacteria/Rhodococcus_opacus/singlet_1/GCV
NC_012522/664679-664786(-)/Actinobacteria/Rhodococcus_opacus/singlet_0/GCV

NC_016887/2839773-2839859(+)/Actinobacteria/Nocardia_cyriacigeorgica/apt_A/GCV
NZ_CP007155/4089027-4089118(+)/Actinobacteria/Kutzneria_albida/singlet_2/GCV
NC_009380/2673652-2673741(-)/Actinobacteria/Salinispora_tropica/singlet_0/GCV
NC_009953/2875689-2875778(-)/Actinobacteria/Salinispora_arenicola/singlet_0/GCV
NC_013510/4395268-4395356(+)/Actinobacteria/Thermomonospora_curvata/apt_B/GCV
NC_013595/1874248-1874332(-)/Actinobacteria/Streptosporangium_roseum/apt_B/GCV
NC_008578/1364035-1364121(-)/Actinobacteria/Acidothermus_cellulolyticus/apt_B/GCV
NZ_CP007514/2846984-2847048(-)/Actinobacteria/Rubrobacter_radiotolerans/apt_B/GCV
NZ_CP008889/1506750-1506835(+)/Actinobacteria/Dermacoccus_nishinomiyaensis/singlet_1/GCV
NZ_CP003949/7429666-7429747(-)/Actinobacteria/Rhodococcus_opacus/apt_A/OTHER
NC_008268/3091409-3091490(-)/Actinobacteria/Rhodococcus_jostii/apt_A/OTHER
NC_012522/2849234-2849315(-)/Actinobacteria/Rhodococcus_opacus/apt_A/OTHER
NZ_CP009922/4588460-4588544(-)/Actinobacteria/Streptomyces_xiamenensis/apt_B/GCV
NC_010612/243691-243772(-)/Actinobacteria/Mycobacterium_marinum/apt_A/GCV
NC_020133/197234-197315(-)/Actinobacteria/Mycobacterium_liflandii/apt_A/GCV
NZ_CP014313/2537626-2537710(+)/Actinobacteria/Microbacterium_sp/apt_A/GCV
NZ_CP010848/1806708-1806785(-)/Actinobacteria/Rathayibacter_toxicus/apt_A/GCV
NC_015514/4173243-4173324(-)/Actinobacteria/Cellulomonas_fimi/singlet_0/GCV
NZ_CP007155/9112698-9112780(+)/Actinobacteria/Kutzneria_albida/apt_B/GCV
NZ_CP015810/2574853-2574953(-)/Actinobacteria/Microbacterium_chocolatum/apt_B/GCV
NZ_CP007155/8798868-8798964(-)/Actinobacteria/Kutzneria_albida/apt_B/GLY_MET
NZ_CP012752/6518126-6518213(+)/Actinobacteria/Kibdelosporangium_phytohabitans/apt_B/OTHER
NC_013595/4461504-4461604(+)/Actinobacteria/Streptosporangium_roseum/apt_B/GCV
NC_009142/7144690-7144777(-)/Actinobacteria/Saccharopolyspora_erythraea/apt_B/GCV
NC_019673/8312299-8312383(-)/Actinobacteria/Saccharothrix_espanaensis/apt_B/GCV
NC_010612/243580-243684(-)/Actinobacteria/Mycobacterium_marinum/apt_B/GCV
NC_020133/197123-197227(-)/Actinobacteria/Mycobacterium_liflandii/apt_B/GCV
NC_015671/98062-98139(+)/Actinobacteria/Cellulomonas_gilvus/apt_A/GCV

NC_014318/1944163-1944246(+)/Actinobacteria/Amycolatopsis_mediterranei/apt_A/GCV
NC_020520/2864887-2864962(-)/Actinobacteria/Ilumatobacter_coccineus/apt_A/GCV
NC_015125/3325654-3325738(+)/Actinobacteria/Microbacterium_testaceum/apt_A/GCV
NC_006087/1714425-1714513(-)/Actinobacteria/Leifsonia_xyli/apt_A/GCV
NZ_CP014761/1026554-1026636(-)/Actinobacteria/Leifsonia_xyli/apt_A/GCV
NC_000962/79209-79289(-)/Actinobacteria/Mycobacterium_tuberculosis/apt_A/GLY_MET
NC_002945/79236-79316(-)/Actinobacteria/Mycobacterium_bovis/apt_A/GLY_MET
NC_015758/79297-79377(-)/Actinobacteria/Mycobacterium_africanum/apt_A/GLY_MET
NC_015848/81270-81350(-)/Actinobacteria/Mycobacterium_canettii/apt_A/GLY_MET
NC_019952/80916-80996(-)/Actinobacteria/Mycobacterium_canettii/apt_A/GLY_MET
NC_022438/1801260-1801348(-)/Actinobacteria/Leifsonia_xyli/apt_A/GCV
NC_008268/3094325-3094406(+)/Actinobacteria/Rhodococcus_jostii/apt_A/GCV
NZ_CP003949/7432667-7432748(+)/Actinobacteria/Rhodococcus_opacus/apt_A/GCV
NC_012522/2853539-2853620(+)/Actinobacteria/Rhodococcus_opacus/apt_A/GCV
NZ_CP014480/1270197-1270294(-)/Actinobacteria/Kocuria_turfanensis/singlet_0/GCV
NC_010168/1481582-1481674(+)/Actinobacteria/Renibacterium_salmoninarum/apt_A/GCV
NC_014830/3963358-3963444(-)/Actinobacteria/Intrasporangium_calvum/apt_A/GCV
NZ_CP013254/77235-77336(+)/Actinobacteria/Kocuria_flava/apt_B/GCV
NZ_CP012697/1791729-1791809(+)/Actinobacteria/Microbacterium_sp/apt_A/GLY_MET
NC_017803/4926329-4926416(-)/Actinobacteria/Actinoplanes_sp/apt_A/GCV
NC_015312/3379094-3379190(+)/Actinobacteria/Pseudonocardia_dioxanivorans/apt_B/GCV
NZ_CP011868/2333329-2333444(+)/Actinobacteria/Pseudonocardia_sp/apt_B/GCV
NZ_CP011862/564896-565012(+)/Actinobacteria/Pseudonocardia_sp/apt_B/GCV
NC_014318/5512576-5512689(-)/Actinobacteria/Amycolatopsis_mediterranei/apt_B/GCV
NC_019673/7216234-7216342(-)/Actinobacteria/Saccharothrix_espanaensis/apt_B/GCV
NC_016887/2839867-2839965(+)/Actinobacteria/Nocardia_cyriacigeorgica/apt_B/GCV
NZ_CP006850/3470254-3470351(+)/Actinobacteria/Nocardia_nova/apt_B/GCV
NC_006361/2650190-2650300(+)/Actinobacteria/Nocardia_farcinica/apt_B/GCV

NZ_CP016793/5149482-5149593(+)/Actinobacteria/Lentzea_sp/apt_B/GCV
NZ_CP012752/7309666-7309764(-)/Actinobacteria/Kibdelosporangium_phytohabitans/apt_B/GCV
NC_015564/3208553-3208662(-)/Actinobacteria/Hoyosella_subflava/apt_B/GCV
NC_021252/4103767-4103882(+)/Actinobacteria/Amycolatopsis_orientalis/apt_B/GCV
NZ_CP008953/4347748-4347863(-)/Actinobacteria/Amycolatopsis_japonica/apt_B/GCV
NZ_CP016174/4400771-4400886(+)/Actinobacteria/Amycolatopsis_orientalis/apt_B/GCV
NC_013530/47295-47387(+)/Actinobacteria/Xylanimonas_cellulosilytica/apt_A/GCV
NC_015588/56386-56480(+)/Actinobacteria/Isoptericola_variabilis/singlet_1/GCV
NZ_CP011773/3545398-3545483(-)/Actinobacteria/Mycobacterium_sp/singlet_0/GCV
NZ_CP009922/4588552-4588634(-)/Actinobacteria/Streptomyces_xiamenensis/apt_A/GCV
NZ_CP010797/953856-953941(+)/Actinobacteria/Rhodococcus_sp/singlet_1/GCV
NZ_CM000913/5109514-5109596(-)/Actinobacteria/Streptomyces_clavuligerus/apt_A/GCV
NZ_CP014480/2726174-2726263(-)/Actinobacteria/Kocuria_turfanensis/apt_A/GLY_MET
NZ_CP011005/2008682-2008774(+)/Actinobacteria/Arthrobacter_sp/apt_A/GCV
NC_012669/2179372-2179453(+)/Actinobacteria/Beutenbergia_cavernae/apt_A/GCV
NC_023150/584439-584518(-)/Actinobacteria/Rhodococcus_pyridinivorans/apt_A/TP
NZ_CP013254/3165306-3165401(-)/Actinobacteria/Kocuria_flava/singlet_0/GCV
NZ_CP012479/3052151-3052238(+)/Actinobacteria/Arthrobacter_sp/apt_A/TP
NZ_CP014196/778518-778607(+)/Actinobacteria/Arthrobacter_sp/apt_A/TP
NZ_CP012752/3350552-3350640(+)/Actinobacteria/Kibdelosporangium_phytohabitans/apt_A/GLY_MET
NZ_CP013745/2764853-2764959(+)/Actinobacteria/Arthrobacter_sp/apt_A/TP
NZ_CP007595/1926243-1926330(+)/Actinobacteria/Arthrobacter_sp/apt_A/TP
NC_018681/4124497-4124579(+)/Actinobacteria/Nocardia_brasiliensis/apt_A/GLY_MET
NZ_CP007156/301876-301959(+)/Actinobacteria/Corynebacterium_falsenii/apt_A/GCV
NZ_CM001024/1121044-1121126(+)/Actinobacteria/Aeromicrobium_marinum/singlet_1/GCV
NC_018581/3026835-3026926(-)/Actinobacteria/Gordonia_sp/singlet_0/GCV
NZ_CM001440/1580438-1580520(+)/Actinobacteria/Saccharomonospora_cyanea/apt_A/GCV
NZ_CP009110/5533265-5533349(-)/Actinobacteria/Amycolatopsis_methanolica/apt_A/GCV

NC_008148/3199887-3199952(-)/Actinobacteria/Rubrobacter_xylanophilus/apt_B/GCV
NZ_CM002177/2385610-2385717(-)/Actinobacteria/Rhodococcus_aetherivorans/apt_B/GCV
NZ_CP015529/2403264-2403371(+)/Actinobacteria/Rhodococcus_sp/apt_B/GCV
NC_023150/953945-954044(-)/Actinobacteria/Rhodococcus_pyridinivorans/apt_B/GCV
NZ_CP011853/2420253-2420359(-)/Actinobacteria/Gordonia_sp/apt_B/GCV
NC_015564/3208671-3208754(-)/Actinobacteria/Hoyosella_subflava/apt_A/GCV
NC_014550/3470549-3470637(+)/Actinobacteria/Glutamicibacter_arilaitensis/apt_A/TP
NZ_CP012171/3268568-3268656(+)/Actinobacteria/Arthrobacter_sp/apt_A/TP
NZ_CP012677/2738568-2738655(+)/Actinobacteria/Arthrobacter_alpinus/apt_A/TP
NZ_CP012479/3051912-3051996(-)/Actinobacteria/Arthrobacter_sp/apt_A/GCV
NZ_CP007595/1925935-1926020(-)/Actinobacteria/Arthrobacter_sp/apt_A/GCV
NZ_CP013745/2764557-2764642(-)/Actinobacteria/Arthrobacter_sp/apt_A/GCV
NZ_CP013200/1444103-1444189(+)/Actinobacteria/Arthrobacter_alpinus/apt_A/GCV
NZ_CP012677/2738259-2738345(-)/Actinobacteria/Arthrobacter_alpinus/apt_A/GCV
NC_015859/462419-462525(+)/Actinobacteria/Corynebacterium_variabile/apt_A/GCV
NC_015673/350243-350327(-)/Actinobacteria/Corynebacterium_resistens/apt_A/GCV
NC_023150/584348-584432(-)/Actinobacteria/Rhodococcus_pyridinivorans/apt_B/TP
NC_013521/46303-46404(+)/Actinobacteria/Sanguibacter_keddieii/singlet_0/GCV
NC_014830/3963266-3963347(-)/Actinobacteria/Intrasporangium_calvum/apt_B/GCV
NZ_CP009896/3317733-3317822(+)/Actinobacteria/Pimelobacter_simplex/apt_B/GCV
NC_022567/2147388-2147481(+)/Actinobacteria/Adlercreutzia_equolifaciens/singlet_0/TP
NC_013165/2020904-2020977(-)/Actinobacteria/Slackia_heliotrinireducens/singlet_2/TP
NC_021021/1431657-1431745(-)/Actinobacteria/Gordonibacter_pamelaeae/singlet_2/TP
NC_013204/146216-146311(-)/Actinobacteria/Eggerthella_lenta/singlet_2/TP
NC_015738/2850373-2850463(+)/Actinobacteria/Eggerthella_sp/singlet_2/TP
NC_012522/2849125-2849225(-)/Actinobacteria/Rhodococcus_opacus/apt_B/OTHER
NZ_CP003949/7429557-7429657(-)/Actinobacteria/Rhodococcus_opacus/apt_B/OTHER
NC_008268/3091300-3091400(-)/Actinobacteria/Rhodococcus_jostii/apt_B/OTHER
NC_014815/4642819-4642898(+)/Actinobacteria/Micromonospora_sp/apt_A/GCV
NZ_CP009896/2646886-2646973(+)/Actinobacteria/Pimelobacter_simplex/singlet_1/GCV
NZ_CP015220/4301108-4301195(-)/Actinobacteria/Rhodococcus_sp/singlet_1/GCV

Appendix

NC_019952/2153469-2153574(+)/Actinobacteria/Mycobacterium_canettii/apt_B/GCV
NC_017093/4414241-4414322(-)/Actinobacteria/Actinoplanes_missouriensis/singlet_0/GCV
NC_013595/4461416-4461497(+)/Actinobacteria/Streptosporangium_roseum/apt_A/GCV
NZ_CP013129/1502753-1502842(-)/Actinobacteria/Streptomyces_venezuelae/apt_A/GCV
NC_016114/6306038-6306123(+)/Actinobacteria/Streptomyces_pratensis/apt_A/GCV
NC_021055/1118529-1118614(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_CM000950/6764090-6764177(+)/Actinobacteria/Streptomyces_pristinaespiralis/apt_A/GCV
NZ_CM001165/6833233-6833321(+)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NC_018681/4014649-4014755(+)/Actinobacteria/Nocardia_brasiliensis/apt_B/GCV
NC_013159/1369628-1369708(+)/Actinobacteria/Saccharomonospora_viridis/apt_A/GCV
NZ_CM001466/4359170-4359250(+)/Actinobacteria/Saccharomonospora_azurea/apt_A/GCV
NC_014550/2942869-2942956(-)/Actinobacteria/Glutamicibacter_arilaitensis/apt_A/GCV
NZ_CP012171/2917229-2917316(-)/Actinobacteria/Arthrobacter_sp/apt_A/GCV
NZ_CP012750/2923604-2923691(-)/Actinobacteria/Glutamicibacter_arilaitensis/apt_A/GCV
NZ_CP009427/79249-79327(-)/Actinobacteria/Mycobacterium_tuberculosis/apt_A/GLY_MET
NC_013729/3388461-3388546(+)/Actinobacteria/Kribbella_flavida/singlet_1/GCV
NZ_CP011853/2420369-2420456(-)/Actinobacteria/Gordonia_sp/apt_A/GCV
NC_015576/2096336-2096419(+)/Actinobacteria/Mycobacterium_sinense/singlet_0/GCV
NZ_CP010996/248126-248209(+)/Actinobacteria/Mycobacterium_simiae/singlet_2/GCV
NC_010617/650982-651087(+)/Actinobacteria/Kocuria_rhizophila/singlet_1/GCV
NZ_CP013200/1443775-1443860(-)/Actinobacteria/Arthrobacter_alpinus/apt_A/TP
NZ_CP012697/1788332-1788411(+)/Actinobacteria/Microbacterium_sp/apt_A/OTHER
NC_022663/88846-88947(+)/Actinobacteria/Mycobacterium_kansasii/apt_B/GCV
NZ_CM001439/1845402-1845482(+)/Actinobacteria/Saccharomonospora_marina/apt_A/GCV
NZ_CP016174/6864480-6864561(-)/Actinobacteria/Amycolatopsis_orientalis/apt_A/GCV
NZ_CM001484/1432412-1432491(+)/Actinobacteria/Saccharomonospora_glauca/apt_A/GCV
NC_008148/3199959-3200048(-)/Actinobacteria/Rubrobacter_xylanophilus/apt_A/GCV

NZ_CP007514/2847058-2847173(-)/Actinobacteria/Rubrobacter_radiotolerans/apt_A/GCV
NZ_CP007220/2402285-2402368(+)/Actinobacteria/Mycobacterium_chelonae/singlet_0/GCV
NZ_CP012507/138242-138330(+)/Actinobacteria/Kocuria_palustris/apt_A/GCV
NC_018150/2505536-2505623(-)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NC_008596/3710883-3710967(-)/Actinobacteria/Mycobacterium_smegmatis/apt_A/GCV
NZ_CP012150/4632892-4632976(+)/Actinobacteria/Mycobacterium_goodii/apt_A/GCV
NC_008726/3330509-3330593(-)/Actinobacteria/Mycobacterium_vanbaalenii/apt_A/GCV
NC_000962/2075620-2075703(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NC_002945/2065972-2066055(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NC_015758/2074938-2075021(+)/Actinobacteria/Mycobacterium_africanum/singlet_1/GCV
NC_015848/2114061-2114144(+)/Actinobacteria/Mycobacterium_canettii/singlet_1/GCV
NC_014830/3963632-3963724(+)/Actinobacteria/Intrasporangium_calvum/apt_A/TP
NZ_CP010071/2241405-2241492(+)/Actinobacteria/Mycobacterium_sp/apt_A/GCV
NC_013165/2795479-2795572(-)/Actinobacteria/Slackia_heliotrinireducens/singlet_0/TP
NC_008146/3007263-3007351(-)/Actinobacteria/Mycobacterium_sp/singlet_0/GCV
NC_018581/1355991-1356074(+)/Actinobacteria/Gordonia_sp/apt_A/GLY_MET
NC_019952/2153368-2153455(+)/Actinobacteria/Mycobacterium_canettii/apt_A/GCV
NC_012943/2347220-2347307(-)/Actinobacteria/Mycobacterium_tuberculosis/singlet_0/GCV
NC_023150/954052-954137(-)/Actinobacteria/Rhodococcus_pyridinivorans/apt_A/GCV
NZ_CP011883/2241095-2241182(+)/Actinobacteria/Mycobacterium_haemophilum/singlet_0/GCV
NZ_CP012697/1791815-1791912(+)/Actinobacteria/Microbacterium_sp/apt_B/GLY_MET
NZ_CM002280/7611656-7611770(+)/Actinobacteria/Streptomyces_niveus/apt_B/GCV
NZ_CP010407/1436904-1437016(-)/Actinobacteria/Streptomyces_vietnamensis/apt_B/GCV
NZ_CM000913/796758-796873(-)/Actinobacteria/Streptomyces_clavuligerus/apt_B/GCV
NZ_CP013129/1502637-1502746(-)/Actinobacteria/Streptomyces_venezuelae/apt_B/GCV

NZ_CM000950/6764184-6764296(+)/Actinobacteria/Streptomyces_pristinaespiralis/apt_B/GCV
NC_021985/1691185-1691291(-)/Actinobacteria/Streptomyces_collinus/apt_B/GCV
NZ_CP015849/1057144-1057252(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP012382/1458668-1458780(-)/Actinobacteria/Streptomyces_ambofaciens/apt_B/GCV
NZ_LN997842/2142916-2143022(-)/Actinobacteria/Streptomyces_reticuli/apt_B/GCV
NZ_CP009438/1255273-1255381(-)/Actinobacteria/Streptomyces_glaucescens/apt_B/GCV
NZ_CP015866/1091753-1091861(-)/Actinobacteria/Streptomyces_parvulus/apt_B/GCV
NC_018750/1091455-1091567(-)/Actinobacteria/Streptomyces_venezuelae/apt_B/GCV
NZ_CM002271/6184391-6184502(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CM001165/6833328-6833450(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_016111/787937-788059(-)/Actinobacteria/Streptomyces_cattleya/apt_B/GCV
NZ_CM003601/1084944-1085069(-)/Actinobacteria/Streptomyces_anulatus/apt_B/GCV
NZ_CP013738/1090365-1090492(-)/Actinobacteria/Streptomyces_globisporus/apt_B/GCV
NC_017765/3307230-3307336(-)/Actinobacteria/Streptomyces_hygroscopicus/apt_B/GCV
NC_003155/8345056-8345165(+)/Actinobacteria/Streptomyces_avermitilis/apt_B/GCV
NZ_CP006871/2046677-2046797(-)/Actinobacteria/Streptomyces_albulus/apt_B/GCV
NZ_CP003987/8094530-8094651(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_015957/6465911-6466033(-)/Actinobacteria/Streptomyces_violaceusniger/apt_B/GCV
NC_016582/10368537-10368659(+)/Actinobacteria/Streptomyces_bingchenggensis/apt_B/GCV
NC_010572/7259173-7259298(+)/Actinobacteria/Streptomyces_griseus/apt_B/GCV
NZ_CP010519/7544026-7544132(+)/Actinobacteria/Streptomyces_albus/apt_B/GCV
NZ_CP011522/7388488-7388613(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_020990/6182127-6182238(+)/Actinobacteria/Streptomyces_albus/apt_B/GCV
NZ_CP007699/7788385-7788506(+)/Actinobacteria/Streptomyces_lydicus/apt_B/GCV
NZ_CP011492/424902-425016(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_013929/8442620-8442724(+)/Actinobacteria/Streptomyces_scabiei/apt_B/GCV
NZ_LN831790/6698571-6698679(+)/Actinobacteria/Streptomyces_leeuwenhoekii/apt_B/GCV
NZ_CP010849/1475346-1475454(-)/Actinobacteria/Streptomyces_cyaneogriseus/apt_B/GCV

NZ_CM000951/1680890-1680999(-)/Actinobacteria/Streptomyces_sviceus/apt_B/GCV
NZ_CP013142/1478513-1478621(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP013743/8253775-8253883(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP015098/1567436-1567544(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP009754/7278424-7278535(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_003888/1457730-1457842(-)/Actinobacteria/Streptomyces_coelicolor/apt_B/GCV
NZ_CM001889/1680979-1681091(-)/Actinobacteria/Streptomyces_lividans/apt_B/GCV
NC_021177/1028171-1028297(-)/Actinobacteria/Streptomyces_fulvissimus/apt_B/GCV
NZ_CP009313/1365685-1365794(-)/Actinobacteria/Streptomyces_nodosus/apt_B/GCV
NC_020504/8090044-8090151(+)/Actinobacteria/Streptomyces_davawensis/apt_B/GCV
NZ_CM002285/8111019-8111120(+)/Actinobacteria/Streptomyces_roseochromogenus/apt_B/GCV
NZ_CP015726/9704317-9704438(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_021055/1118397-1118522(-)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_016114/6306130-6306255(+)/Actinobacteria/Streptomyces_pratensis/apt_B/GCV
NZ_CP016793/5149389-5149475(+)/Actinobacteria/Lentzea_sp/apt_A/GCV
NC_013131/5446022-5446097(-)/Actinobacteria/Catenulispora_acidiphila/apt_A/GCV
NZ_AP012555/2644236-2644323(+)/Actinobacteria/Mycobacterium_avium/singlet_1/GCV
NC_019673/7527267-7527344(-)/Actinobacteria/Saccharothrix_espanaensis/apt_A/GCV
NC_008595/2916244-2916331(-)/Actinobacteria/Mycobacterium_avium/apt_A/GCV
NC_002944/1693368-1693455(+)/Actinobacteria/Mycobacterium_avium/singlet_0/GCV
NC_016946/2833102-2833189(-)/Actinobacteria/Mycobacterium_intracellulare/apt_A/GCV
NC_017904/2863396-2863483(-)/Actinobacteria/Mycobacterium_sp/apt_A/GCV
NC_018612/2880619-2880706(+)/Actinobacteria/Mycobacterium_indicus/apt_A/GCV
NC_021715/2685464-2685551(-)/Actinobacteria/Mycobacterium_yongonense/apt_A/GCV
NZ_CP016793/4776112-4776192(+)/Actinobacteria/Lentzea_sp/apt_A/GCV
NC_014659/2606010-2606106(-)/Actinobacteria/Rhodococcus_equi/singlet_1/GCV
NC_022663/88743-88830(+)/Actinobacteria/Mycobacterium_kansasii/apt_A/GCV
NC_008578/2090042-2090125(-)/Actinobacteria/Acidothermus_cellulolyticus/apt_A/GCV
NZ_CP012752/7309771-7309859(-)/Actinobacteria/Kibdelosporangium_phytohabitans/apt_A/GCV

NZ_CP012752/6518030-6518119(+)/Actinobacteria/Kibdelosporangium_phytohabitans/apt_A/OTHER
NZ_CP009922/340310-340397(-)/Actinobacteria/Streptomyces_xiamenensis/apt_A/GCV
NC_009142/6403604-6403687(+)/Actinobacteria/Saccharopolyspora_erythraea/apt_A/GCV
NC_010612/3302478-3302565(+)/Actinobacteria/Mycobacterium_marinum/apt_A/GCV
NC_020133/2648630-2648717(+)/Actinobacteria/Mycobacterium_liflandii/apt_A/GCV
NZ_CP011269/3459660-3459749(-)/Actinobacteria/Mycobacterium_fortuitum/apt_A/GCV
NZ_CP009914/3460267-3460356(-)/Actinobacteria/Mycobacterium_sp/apt_A/GCV
NC_009664/3908899-3908989(+)/Actinobacteria/Kineococcus_radiotolerans/singlet_1/GCV
NC_015671/98150-98247(+)/Actinobacteria/Cellulomonas_gilvus/apt_B/GCV
NC_017803/4926238-4926318(-)/Actinobacteria/Actinoplanes_sp/apt_B/GCV
NC_014815/4642905-4642991(+)/Actinobacteria/Micromonospora_sp/apt_B/GCV
NC_012704/371550-371660(+)/Actinobacteria/Corynebacterium_kroppenstedtii/apt_A/GCV
NZ_CP010071/2241501-2241631(+)/Actinobacteria/Mycobacterium_sp/apt_B/GCV
NZ_CP012150/4632987-4633087(+)/Actinobacteria/Mycobacterium_goodii/apt_B/GCV
NC_008596/3710770-3710872(-)/Actinobacteria/Mycobacterium_smegmatis/apt_B/GCV
NZ_CM001762/3823411-3823513(-)/Actinobacteria/Mycobacterium_smegmatis/apt_B/GCV
NC_021715/2685348-2685455(-)/Actinobacteria/Mycobacterium_yongonense/apt_B/GCV
NC_017904/2863280-2863387(-)/Actinobacteria/Mycobacterium_sp/apt_B/GCV
NC_016947/2839751-2839858(-)/Actinobacteria/Mycobacterium_intracellulare/apt_B/GCV
NC_016946/2832986-2833093(-)/Actinobacteria/Mycobacterium_intracellulare/apt_B/GCV
NZ_CP014475/2420294-2420399(+)/Actinobacteria/Mycobacterium_phlei/apt_B/GCV
NC_023036/3313278-3313364(-)/Actinobacteria/Mycobacterium_neoaurum/apt_B/GCV
NZ_CP011022/3313277-3313363(-)/Actinobacteria/Mycobacterium_sp/apt_B/GCV
NC_008595/2916129-2916236(-)/Actinobacteria/Mycobacterium_avium/apt_B/GCV
NC_010612/3302574-3302681(+)/Actinobacteria/Mycobacterium_marinum/apt_B/GCV
NC_020133/2648726-2648833(+)/Actinobacteria/Mycobacterium_liflandii/apt_B/GCV

Appendix

NC_016604/4876258-4876345(+)/Actinobacteria/Mycobacterium_rhodesiae/apt_B/GCV
NZ_CP009914/3460170-3460258(-)/Actinobacteria/Mycobacterium_sp/apt_B/GCV
NZ_CP011269/3459563-3459651(-)/Actinobacteria/Mycobacterium_fortuitum/apt_B/GCV
NC_008726/3330415-3330498(-)/Actinobacteria/Mycobacterium_vanbaalenii/apt_B/GCV
NC_018027/2851539-2851625(+)/Actinobacteria/Mycobacterium_chubuense/singlet_0/GCV
NZ_CP014950/2273716-2273801(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NC_021282/2389520-2389605(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP015235/4373127-4373212(+)/Actinobacteria/Rhodococcus_fascians/singlet_1/GCV
NC_023036/3313374-3313459(-)/Actinobacteria/Mycobacterium_neoaurum/apt_A/GCV
NZ_CP011022/3313373-3313458(-)/Actinobacteria/Mycobacterium_sp/apt_A/GCV
NC_008711/4380694-4380782(-)/Actinobacteria/Paenarthrobacter_aurescens/apt_A/GCV
NC_018531/4541030-4541118(-)/Actinobacteria/Arthrobacter_sp/apt_A/GCV
NZ_CP013254/77138-77228(+)/Actinobacteria/Kocuria_flava/apt_A/GCV
NC_008711/1042794-1042881(+)/Actinobacteria/Paenarthrobacter_aurescens/apt_A/GCV
NC_018531/998229-998316(+)/Actinobacteria/Arthrobacter_sp/apt_A/GCV
NZ_CP014196/911839-911927(+)/Actinobacteria/Arthrobacter_sp/apt_A/GCV
NC_011886/967223-967309(+)/Actinobacteria/Pseudarthrobacter_chlorophenolicus/apt_A/GCV
NC_008541/65969-66056(+)/Actinobacteria/Arthrobacter_sp/apt_A/GLY_MET
NZ_CP013747/3362558-3362645(+)/Actinobacteria/Pseudarthrobacter_sulfonivorans/apt_A/GCV
NC_015145/3634571-3634658(-)/Actinobacteria/Pseudarthrobacter_phenanthrenivorans/apt_A/GCV
NZ_CP014196/4139186-4139275(-)/Actinobacteria/Arthrobacter_sp/apt_A/GCV
NC_011886/583370-583458(+)/Actinobacteria/Pseudarthrobacter_chlorophenolicus/apt_A/TP
NC_008711/883431-883519(+)/Actinobacteria/Paenarthrobacter_aurescens/apt_A/TP
NC_018531/838887-838975(+)/Actinobacteria/Arthrobacter_sp/apt_A/TP
NC_015145/836223-836311(+)/Actinobacteria/Pseudarthrobacter_phenanthrenivorans/singlet_1/GCV
NC_008541/825659-825747(+)/Actinobacteria/Arthrobacter_sp/apt_A/GCV
NZ_CP013747/1436024-1436111(-)/Actinobacteria/Pseudarthrobacter_sulfonivorans/apt_A/GCV

NC_018612/2880721-2880827(+)/Actinobacteria/Mycobacterium_indicus/apt_B/GCV
NC_021252/6836007-6836090(-)/Actinobacteria/Amycolatopsis_orientalis/apt_A/GCV
NZ_CP008953/1835046-1835129(+)/Actinobacteria/Amycolatopsis_japonica/apt_A/GCV
NC_018142/2070979-2071054(+)/Actinobacteria/Propionibacterium_propionicum/apt_A/GCV
NC_011886/583017-583106(-)/Actinobacteria/Pseudarthrobacter_chlorophenolicus/apt_A/GCV
NZ_CP011664/1732351-1732437(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NC_016604/4876162-4876248(+)/Actinobacteria/Mycobacterium_rhodesiae/apt_A/GCV
NC_007164/259295-259385(+)/Actinobacteria/Corynebacterium_jeikeium/apt_A/GCV
NC_019966/3204883-3204969(+)/Actinobacteria/Mycobacterium_sp/singlet_1/GCV
NC_014814/2829309-2829397(-)/Actinobacteria/Mycobacterium_gilvum/singlet_2/GCV
NC_009338/3615662-3615750(-)/Actinobacteria/Mycobacterium_gilvum/singlet_2/GCV
NZ_CP014475/2420198-2420285(+)/Actinobacteria/Mycobacterium_phlei/apt_A/GCV
NC_009142/4206355-4206452(-)/Actinobacteria/Saccharopolyspora_erythraea/apt_B/GCV

Table 2.13: 50 Actinobacteria riboswitch aptamer sequences from aptamer-1 and singleton type-1 sub-cluster with paired aptamer-2 supplemented.

NC_014814/2829309-2829397(-)/Actinobacteria/Mycobacterium_gilvum/singlet_2/GCV
NC_015564/3208671-3208754(-)/Actinobacteria/Hoyosella_subflava/apt_A/GCV
NC_015564/3208553-3208662(-)/Actinobacteria/Hoyosella_subflava/apt_B/GCV
NC_015576/2096336-2096419(+)/Actinobacteria/Mycobacterium_sinense/singlet_0/GCV
NC_015758/2074938-2075021(+)/Actinobacteria/Mycobacterium_africanum/singlet_1/GCV
NC_015848/2114061-2114144(+)/Actinobacteria/Mycobacterium_canettii/singlet_1/GCV
NC_016604/4876162-4876248(+)/Actinobacteria/Mycobacterium_rhodesiae/apt_A/GCV
NC_016604/4876258-4876345(+)/Actinobacteria/Mycobacterium_rhodesiae/apt_B/GCV
NC_016946/2833102-2833189(-)/Actinobacteria/Mycobacterium_intracellulare/apt_A/GCV
NC_016946/2832986-2833093(-)/Actinobacteria/Mycobacterium_intracellulare/apt_B/GCV

NC_017904/2863396-2863483(-)/Actinobacteria/Mycobacterium_sp/apt_A/GCV
NC_017904/2863280-2863387(-)/Actinobacteria/Mycobacterium_sp/apt_B/GCV
NC_018027/2851539-2851625(+)/Actinobacteria/Mycobacterium_chubuense/singlet_0/GCV
NC_018150/2505536-2505623(-)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NC_018612/2880619-2880706(+)/Actinobacteria/Mycobacterium_indicus/apt_A/GCV
NC_018612/2880721-2880827(+)/Actinobacteria/Mycobacterium_indicus/apt_B/GCV
NC_019952/2153368-2153455(+)/Actinobacteria/Mycobacterium_canettii/apt_A/GCV
NC_019952/2153469-2153574(+)/Actinobacteria/Mycobacterium_canettii/apt_B/GCV
NC_019966/3204883-3204969(+)/Actinobacteria/Mycobacterium_sp/singlet_1/GCV
NC_020133/2648630-2648717(+)/Actinobacteria/Mycobacterium_liflandii/apt_A/GCV
NC_020133/2648726-2648833(+)/Actinobacteria/Mycobacterium_liflandii/apt_B/GCV
NC_021282/2389520-2389605(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NC_021715/2685464-2685551(-)/Actinobacteria/Mycobacterium_yongonense/apt_A/GCV
NC_021715/2685348-2685455(-)/Actinobacteria/Mycobacterium_yongonense/apt_B/GCV
NC_022663/88743-88830(+)/Actinobacteria/Mycobacterium_kansasii/apt_A/GCV
NC_022663/88846-88947(+)/Actinobacteria/Mycobacterium_kansasii/apt_B/GCV
NC_023036/3313374-3313459(-)/Actinobacteria/Mycobacterium_neoaurum/apt_A/GCV
NC_023036/3313278-3313364(-)/Actinobacteria/Mycobacterium_neoaurum/apt_B/GCV
NZ_AP012555/2644236-2644323(+)/Actinobacteria/Mycobacterium_avium/singlet_1/GCV
NZ_CP007220/2402285-2402368(+)/Actinobacteria/Mycobacterium_chelonae/singlet_0/GCV
NZ_CP009914/3460267-3460356(-)/Actinobacteria/Mycobacterium_sp/apt_A/GCV
NZ_CP009914/3460170-3460258(-)/Actinobacteria/Mycobacterium_sp/apt_B/GCV
NZ_CP010071/2241405-2241492(+)/Actinobacteria/Mycobacterium_sp/apt_A/GCV
NZ_CP010071/2241501-2241631(+)/Actinobacteria/Mycobacterium_sp/apt_B/GCV
NZ_CP010797/953856-953941(+)/Actinobacteria/Rhodococcus_sp/singlet_1/GCV
NZ_CP010996/248126-248209(+)/Actinobacteria/Mycobacterium_simiae/singlet_2/GCV
NZ_CP011022/3313373-3313458(-)/Actinobacteria/Mycobacterium_sp/apt_A/GCV
NZ_CP011022/3313277-3313363(-)/Actinobacteria/Mycobacterium_sp/apt_B/GCV
NZ_CP011269/3459660-3459749(-)/Actinobacteria/Mycobacterium_fortuitum/apt_A/GCV

NZ_CP011269/3459563-3459651(-)/Actinobacteria/Mycobacterium_fortuitum/apt_B/GCV
NZ_CP011530/2625418-2625503(+)/Actinobacteria/Mycobacterium_immunogenum/singlet_1/GCV
NZ_CP011773/3545398-3545483(-)/Actinobacteria/Mycobacterium_sp/singlet_0/GCV
NZ_CP011883/2241095-2241182(+)/Actinobacteria/Mycobacterium_haemophilum/singlet_0/GCV
NZ_CP012150/4632892-4632976(+)/Actinobacteria/Mycobacterium_goodii/apt_A/GCV
NZ_CP012150/4632987-4633087(+)/Actinobacteria/Mycobacterium_goodii/apt_B/GCV
NZ_CP014475/2420198-2420285(+)/Actinobacteria/Mycobacterium_phlei/apt_A/GCV
NZ_CP014475/2420294-2420399(+)/Actinobacteria/Mycobacterium_phlei/apt_B/GCV
NZ_CP014950/2273716-2273801(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP015220/4301108-4301195(-)/Actinobacteria/Rhodococcus_sp/singlet_1/GCV
NZ_CP015235/4373127-4373212(+)/Actinobacteria/Rhodococcus_fascians/singlet_1/GCV

Table 2.14: 150 riboswitch aptamer sequences regulating GCV for graph clustering analysis.

NC_009720/2461124-2461239(-)/Proteobacteria/Xanthobacter_autotrophicus/apt_B/GCV
NZ_CP010052/2549426-2549512(-)/Firmicutes/Bacillus_subtilis/apt_B/GCV
NZ_CP009416/1967796-1967874(-)/Firmicutes/Jeotgalibacillus_malaysiensis/apt_A/GCV
NC_019382/2803252-2803359(-)/Proteobacteria/Bordetella_bronchiseptica/apt_B/GCV
NZ_CP016886/1636786-1636931(+)/Proteobacteria/Neisseria_meningitidis/apt_B/GCV
NC_014098/704461-704567(+)/Firmicutes/Kyrpidia_tusciae/apt_B/GCV
NZ_CP010406/3568624-3568718(+)/Firmicutes/Bacillus_sp/apt_B/GCV
NZ_CP015285/1237718-1237828(-)/Proteobacteria/Azospirillum_humicireducens/apt_B/GCV
NZ_CP008953/1835137-1835228(+)/Actinobacteria/Amycolatopsis_japonica/apt_B/GCV
NC_013929/3127914-3128000(+)/Actinobacteria/Streptomyces_scabiei/apt_A/GCV
NZ_CP009626/654252-654326(+)/Proteobacteria/Brucella_abortus/apt_A/GCV
NC_014034/1212818-1212885(+)/Proteobacteria/Rhodobacter_capsulatus/apt_B/GCV
NZ_CM000951/1681006-1681091(-)/Actinobacteria/Streptomyces_sviceus/apt_A/GCV

Appendix

NZ_CP010434/2406241-2406332(-)/Firmicutes/Bacillus_subtilis/apt_A/GCV
NZ_CP007688/1375107-1375188(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP009031/3505718-3505787(+)/Proteobacteria/Xanthomonas_citri/apt_A/GCV
NZ_CP010524/2728023-2728117(-)/Firmicutes/Bacillus_paralicheniformis/apt_A/GCV
NC_015953/906065-906151(-)/Actinobacteria/Streptomyces_sp/apt_A/GCV
NZ_HG916765/3297637-3297723(-)/Proteobacteria/Castellaniella_defragrans/apt_A/GCV
NZ_CM001043/2056331-2056414(+)/Actinobacteria/Mycobacterium_tuberculosis/singlet_1/GCV
NZ_CM002273/1795610-1795723(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CP007683/1050191-1050283(+)/Proteobacteria/Brucella_abortus/apt_B/GCV
NC_008048/1967802-1967896(-)/Proteobacteria/Sphingopyxis_alaskensis/apt_B/GCV
NZ_CP013129/1502753-1502842(-)/Actinobacteria/Streptomyces_venezuelae/apt_A/GCV
NC_010170/4308874-4308960(-)/Proteobacteria/Bordetella_petrii/apt_A/GCV
NC_018142/2070979-2071054(+)/Actinobacteria/Propionibacterium_propionicum/apt_A/GCV
NZ_CP003987/3509108-3509219(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NC_003919/1392470-1392539(-)/Proteobacteria/Xanthomonas_axonopodis/apt_A/GCV
NZ_CP007597/1273116-1273201(-)/Proteobacteria/Stenotrophomonas_rhizophila/apt_B/GCV
NZ_CP012152/1897559-1897633(-)/Firmicutes/Anoxybacillus_gonensis/apt_B/GCV
NC_007404/191454-191563(-)/Proteobacteria/Thiobacillus_denitrificans/apt_B/GCV
NZ_CP006058/663355-663445(-)/Firmicutes/Bacillus_amyloliquefaciens/apt_A/GCV
NC_020832/2365515-2365606(-)/Firmicutes/Bacillus_subtilis/apt_A/GCV
NC_018584/1364860-1364941(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_010104/712604-712678(-)/Proteobacteria/Brucella_canis/apt_A/GCV
NC_017186/1944184-1944267(+)/Actinobacteria/Amycolatopsis_mediterranei/apt_A/GCV
NZ_CP007244/2461803-2461893(-)/Firmicutes/Bacillus_velezensis/apt_A/GCV
NZ_CP003915/3757195-3757302(-)/Proteobacteria/Advenella_mimigardefordensis/apt_B/GCV
NZ_CP011112/3231756-3231842(+)/Actinobacteria/Luteipulveratus_mongoliensis/singlet_1/GCV
NC_014752/1777817-1777962(-)/Proteobacteria/Neisseria_lactamica/apt_B/GCV
NZ_CP014057/1296854-1296955(+)/Firmicutes/Staphylococcus_saprophyticus/singlet_1/GCV
NC_021663/364904-365008(+)/Actinobacteria/Corynebacterium_terpenotabidum/apt_A/GCV

NZ_CP011058/1147680-1147784(-)/Firmicutes/Paenibacillus_beijingensis/apt_A/GCV
NZ_CP008995/1392361-1392465(-)/Proteobacteria/Xanthomonas_citri/apt_B/GCV
NZ_CP010961/961851-961931(-)/Proteobacteria/Bordetella_pertussis/apt_A/GCV
NZ_CP011043/2169970-2170046(-)/Actinobacteria/Clavibacter_michiganensis/apt_A/GCV
NZ_CP016328/835794-835874(+)/Proteobacteria/Gluconobacter_oxydans/apt_A/GCV
NZ_CP011269/3459563-3459651(-)/Actinobacteria/Mycobacterium_fortuitum/apt_B/GCV
NZ_CP009679/2454867-2454961(-)/Firmicutes/Bacillus_velezensis/apt_B/GCV
NZ_CP010841/197164-197271(+)/Proteobacteria/Bordetella_pertussis/apt_B/GCV
NC_021985/1691298-1691382(-)/Actinobacteria/Streptomyces_collinus/apt_A/GCV
NC_015684/2708281-2708360(+)/Proteobacteria/Oligotropha_carboxidovorans/apt_B/GCV
NZ_CP012023/2806113-2806196(-)/Proteobacteria/Celeribacter_marinus/apt_A/GCV
NZ_CP013552/2308949-2309014(+)/Proteobacteria/Rhizobium_phaseoli/apt_B/GCV
NC_014410/1941720-1941788(+)/Firmicutes/Thermoanaerobacterium_thermosaccharolyticum/apt_B/GCV
NC_013791/2603007-2603091(-)/Firmicutes/Bacillus_pseudofirmus/apt_A/GCV
NZ_CP012150/4632987-4633087(+)/Actinobacteria/Mycobacterium_goodii/apt_B/GCV
NC_022657/4614050-4614129(+)/Actinobacteria/Actinoplanes_friuliensis/singlet_1/GCV
NC_009253/773769-773864(+)/Firmicutes/Desulfotomaculum_reducens/apt_A/GCV
NZ_CP011144/631183-631290(+)/Proteobacteria/Pseudoxanthomonas_suwonensis/apt_B/GCV
NZ_CM001792/1912594-1912692(-)/Firmicutes/Oceanobacillus_kimchii/apt_B/GCV
NZ_CP015210/804599-804680(+)/Proteobacteria/Rhodobacter_sphaeroides/apt_A/GCV
NC_019973/3898492-3898586(-)/Proteobacteria/Mesorhizobium_australicum/apt_B/GCV
NZ_CP010407/1437023-1437112(-)/Actinobacteria/Streptomyces_vietnamensis/apt_A/GCV
NZ_CP014196/4139088-4139179(-)/Actinobacteria/Arthrobacter_sp/apt_B/GCV
NZ_LN879547/2149499-2149593(-)/Proteobacteria/Comamonas_testosteroni/apt_A/GCV
NZ_CP011300/285385-285462(+)/Proteobacteria/Devosia_sp/apt_A/GCV
NZ_CM001439/1845402-1845482(+)/Actinobacteria/Saccharomonospora_marina/apt_A/GCV
NZ_CP011501/150177-150263(-)/Proteobacteria/Pandoraea_apista/apt_A/GCV
NC_012808/659288-659379(-)/Proteobacteria/Methylobacterium_extorquens/apt_A/GCV

NZ_CP012251/1224417-1224486(+)/Proteobacteria/Xanthomonas_arboricola/apt_A/GCV
NZ_CP016885/1637350-1637437(+)/Proteobacteria/Neisseria_meningitidis/apt_A/GCV
NZ_LN907826/3856660-3856740(-)/Proteobacteria/Bradyrhizobium_sp/apt_A/GCV
NZ_CP006850/3470254-3470351(+)/Actinobacteria/Nocardia_nova/apt_B/GCV
NZ_CP009283/6704576-6704684(+)/Firmicutes/Paenibacillus_sp/apt_B/GCV
NZ_CP011961/1297239-1297308(-)/Proteobacteria/Xanthomonas_oryzae/apt_A/GCV
NZ_CP012939/1067109-1067196(-)/Proteobacteria/Ralstonia_solanacearum/apt_A/GCV
NZ_CP011451/3146724-3146819(+)/Proteobacteria/Nitrosomonas_communis/singlet_1/GCV
NZ_CP013552/2308854-2308936(+)/Proteobacteria/Rhizobium_phaseoli/apt_A/GCV
NC_016894/3761482-3761556(-)/Firmicutes/Acetobacterium_woodii/apt_B/GCV
NC_022535/1449512-1449589(+)/Proteobacteria/Rhizobium_sp/apt_A/GCV
NZ_CP011367/859410-859509(+)/Proteobacteria/Thioalkalivibrio_versutus/apt_B/GCV
NC_015684/2708171-2708272(+)/Proteobacteria/Oligotropha_carboxidovorans/apt_A/GCV
NZ_CM002280/7611562-7611649(+)/Actinobacteria/Streptomyces_niveus/apt_A/GCV
NC_014960/1877759-1877835(+)/Chloroflexi/Anaerolinea_thermophila/apt_B/GCV
NZ_CP013068/206449-206547(-)/Proteobacteria/Pannonibacter_phragmitetus/apt_A/GCV
NC_015312/3379007-3379087(+)/Actinobacteria/Pseudonocardia_dioxanivorans/apt_A/GCV
NC_022116/1952893-1952995(+)/Actinobacteria/Amycolatopsis_mediterranei/apt_B/GCV
NZ_CP013580/2399966-2400031(+)/Proteobacteria/Rhizobium_phaseoli/apt_B/GCV
NZ_CP011832/2370825-2370903(-)/Firmicutes/Geobacillus_sp/apt_A/GCV
NZ_LT571436/1334859-1334985(-)/Proteobacteria/Neisseria_weaveri/apt_B/GCV
NC_012522/2853539-2853620(+)/Actinobacteria/Rhodococcus_opacus/apt_A/GCV
NC_007205/649875-649939(+)/Proteobacteria/Candidatus_Pelagibacter/apt_B/GCV
NZ_CM001466/4359260-4359351(+)/Actinobacteria/Saccharomonospora_azurea/apt_B/GCV
NZ_CP013595/2299920-2299984(+)/Proteobacteria/Rhizobium_sp/apt_B/GCV
NZ_CP008876/1734314-1734398(-)/Firmicutes/Terribacillus_aidingensis/apt_B/GCV
NZ_CP012382/5531552-5531662(-)/Actinobacteria/Streptomyces_ambofaciens/apt_B/GCV
NZ_CP011853/2420253-2420359(-)/Actinobacteria/Gordonia_sp/apt_B/GCV
NZ_CP007514/2846984-2847048(-)/Actinobacteria/Rubrobacter_radiotolerans/apt_B/GCV

NZ_CM002271/6184391-6184502(+)/Actinobacteria/Streptomyces_sp/apt_B/GCV
NZ_CM002258/1343611-1343682(-)/Proteobacteria/Agrobacterium_tumefaciens/apt_B/GCV
NZ_CP010536/4333944-4334019(+)/Proteobacteria/Cupriavidus_basilensis/singlet_1/GCV
NC_019965/2093962-2094045(+)/Actinobacteria/Mycobacterium_canettii/singlet_1/GCV
NC_014158/2430914-2431005(-)/Actinobacteria/Tsukamurella_paurometabola/singlet_1/GCV
NZ_CP009762/1287067-1287150(+)/Firmicutes/Staphylococcus_schleiferi/singlet_1/GCV
NZ_CP014234/711783-711889(+)/Proteobacteria/Moraxella_osloensis/singlet_1/GCV
NC_022513/4333303-4333379(+)/Proteobacteria/Ralstonia_pickettii/singlet_1/GCV
NZ_CP014844/108133-108210(+)/Proteobacteria/Cupriavidus_nantongensis/singlet_1/GCV
NZ_AP014547/2368726-2368811(+)/Actinobacteria/Mycobacterium_abscessus/singlet_1/GCV
NZ_CP007689/609323-609404(-)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NC_010528/3242798-3242875(+)/Proteobacteria/Cupriavidus_taiwanensis/singlet_1/GCV
NZ_AM412059/2058492-2058575(+)/Actinobacteria/Mycobacterium_bovis/singlet_1/GCV
NZ_CP009111/4794790-4794903(-)/Actinobacteria/Rhodococcus_opacus/singlet_1/GCV
NZ_CM001049/1322031-1322112(+)/Firmicutes/Listeria_innocua/singlet_1/GCV
NC_014210/1803416-1803495(+)/Actinobacteria/Nocardiopsis_dassonvillei/singlet_1/GCV
NC_015588/56386-56480(+)/Actinobacteria/Isoptericola_variabilis/singlet_1/GCV
NZ_CP009896/2646886-2646973(+)/Actinobacteria/Pimelobacter_simplex/singlet_1/GCV
NZ_CM001051/1453677-1453758(+)/Firmicutes/Listeria_seeligeri/singlet_1/GCV
NZ_CP016022/3585998-3586085(+)/Proteobacteria/Ralstonia_insidiosa/singlet_1/GCV
NC_015726/3709013-3709090(+)/Proteobacteria/Cupriavidus_necator/singlet_1/GCV
NC_019966/3204883-3204969(+)/Actinobacteria/Mycobacterium_sp/singlet_1/GCV
NZ_CM001024/1121044-1121126(+)/Actinobacteria/Aeromicrobium_marinum/singlet_1/GCV
NZ_CP016072/1370469-1370554(+)/Firmicutes/Staphylococcus_pseudintermedius/singlet_1/GCV
NZ_CP014617/2174218-2174301(+)/Actinobacteria/Mycobacterium_africanum/singlet_1/GCV
NZ_CP009046/212645-212777(-)/Firmicutes/Staphylococcus_epidermidis/singlet_1/GCV
NC_008313/3890153-3890230(+)/Proteobacteria/Ralstonia_eutropha/singlet_1/GCV

NZ_CP008889/1506750-1506835(+)/Actinobacteria/Dermacoccus_nishinomiyaensis/singlet_1/GCV
NC_012121/1177746-1177867(-)/Firmicutes/Staphylococcus_carnosus/singlet_1/GCV
NC_011899/2563206-2563288(-)/Firmicutes/Halothermothrix_orenii/singlet_1/GCV
NZ_CP009242/1306699-1306780(+)/Firmicutes/Listeria_monocytogenes/singlet_1/GCV
NZ_CP008747/1398221-1398305(+)/Firmicutes/Staphylococcus_hyicus/singlet_1/GCV
NZ_CP010797/953856-953941(+)/Actinobacteria/Rhodococcus_sp/singlet_1/GCV
NC_010617/650982-651087(+)/Actinobacteria/Kocuria_rhizophila/singlet_1/GCV
NZ_CP015220/4301108-4301195(-)/Actinobacteria/Rhodococcus_sp/singlet_1/GCV
NZ_CP014945/2182067-2182194(+)/Proteobacteria/Psychrobacter_alimentarius/singlet_1/GCV
NC_010622/3367073-3367173(+)/Proteobacteria/Paraburkholderia_phymatum/singlet_1/GCV
NC_014815/5602963-5603043(+)/Actinobacteria/Micromonospora_sp/singlet_1/GCV
NC_014614/1261555-1261643(+)/Firmicutes/Clostridium_sticklandii/singlet_1/GCV
NZ_CP010516/3342694-3342792(+)/Proteobacteria/Cupriavidus_gilardii/singlet_1/GCV
NZ_CP011530/2625418-2625503(+)/Actinobacteria/Mycobacterium_immunogenum/singlet_1/GCV
NZ_CP015114/316352-316474(+)/Firmicutes/Staphylococcus_condimenti/singlet_1/GCV
NC_007204/958065-958186(-)/Proteobacteria/Psychrobacter_arcticus/singlet_2/GCV
NC_007969/951652-951779(-)/Proteobacteria/Psychrobacter_cryohalolentis/singlet_2/GCV
NC_021661/984190-984317(-)/Proteobacteria/Psychrobacter_sp/singlet_2/GCV
NZ_CP007155/4089027-4089118(+)/Actinobacteria/Kutzneria_albida/singlet_2/GCV
NZ_CP010996/248126-248209(+)/Actinobacteria/Mycobacterium_simiae/singlet_2/GCV
NZ_CP013111/5911287-5911379(+)/Proteobacteria/Bordetella_sp/singlet_2/GCV
NZ_LN831029/5941333-5941443(-)/Proteobacteria/Achromobacter_xylosoxidans/singlet_2/GCV
NC_020504/3338148-3338250(+)/Actinobacteria/Streptomyces_davawensis/apt_B/GCV
NZ_CP011008/3501757-3501834(-)/Firmicutes/Bacillus_simplex/apt_A/GCV

Table 2.15: 150 riboswitch aptamer sequences regulating TP for graph clustering analysis.

NZ_CP014704/1005120-1005205(+)/Firmicutes/Clostridium_butyricum/apt_A/TP
NZ_CM000720/2138064-2138165(-)/Firmicutes/Bacillus_cereus/apt_B/TP
NZ_CM001804/2460249-2460350(-)/Firmicutes/Bacillus_thuringiensis/apt_B/TP
NZ_CM000733/1896381-1896486(+)/Firmicutes/Bacillus_cereus/apt_B/TP

NZ_CP012483/3228786-3228882(+)/Firmicutes/Bacillus_cereus/apt_A/TP
NZ_CP016210/1286070-1286161(-)/Proteobacteria/Azoarcus_olearius/apt_A/TP
NC_015565/1335281-1335359(+)/Firmicutes/Desulfotomaculum_nigrificans/apt_B/TP
NZ_CP006883/3602944-3603023(-)/Proteobacteria/Aeromonas_hydrophila/apt_A/TP
NZ_CP009610/120744-120840(+)/Proteobacteria/Haemophilus_influenzae/apt_B/TP
NC_013192/1471300-1471387(-)/Fusobacteria/Leptotrichia_buccalis/apt_B/TP
NZ_CP007512/2078610-2078706(-)/Firmicutes/Bacillus_bombysepticus/apt_A/TP
NZ_CP007566/1474907-1475005(+)/Proteobacteria/Aeromonas_hydrophila/apt_B/TP
NZ_CP009267/2916505-2916575(-)/Firmicutes/Clostridium_pasteurianum/apt_B/TP
NZ_CP012171/3268568-3268656(+)/Actinobacteria/Arthrobacter_sp/apt_A/TP
NC_017566/3692121-3692258(-)/Proteobacteria/Shewanella_putrefaciens/apt_A/TP
NC_015558/862065-862130(+)/Firmicutes/Streptococcus_parauberis/apt_A/TP
NZ_CP014140/166766-166870(-)/Firmicutes/Aneurinibacillus_sp/apt_A/TP
NC_010519/876458-876559(+)/Proteobacteria/Haemophilus_somnus/apt_B/TP
NZ_CP007618/1800269-1800364(+)/Firmicutes/Bacillus_anthracis/apt_A/TP
NZ_CM000743/2013283-2013386(-)/Firmicutes/Bacillus_mycoides/apt_B/TP
NZ_CM000734/2197652-2197748(-)/Firmicutes/Bacillus_cereus/apt_A/TP
NC_002737/1046443-1046519(+)/Firmicutes/Streptococcus_pyogenes/apt_B/TP
NZ_CP014046/2298614-2298701(-)/Proteobacteria/Vibrio_parahaemolyticus/apt_B/TP
NC_014335/2196828-2196929(-)/Firmicutes/Bacillus_cereus/apt_B/TP
NC_022781/4886146-4886242(-)/Firmicutes/Bacillus_toyonensis/apt_A/TP
NZ_CP012725/2221609-2221710(-)/Firmicutes/Bacillus_anthracis/apt_B/TP
NC_017571/1031063-1031189(+)/Proteobacteria/Shewanella_baltica/apt_B/TP
NC_014922/269342-269417(+)/Proteobacteria/Haemophilus_influenzae/apt_A/TP
NZ_CP007634/992865-992950(-)/Proteobacteria/Vibrio_cholerae/apt_B/TP
NZ_CP015150/2460634-2460730(-)/Firmicutes/Bacillus_thuringiensis/apt_A/TP
NZ_CP009240/105773-105850(-)/Firmicutes/Megasphaera_elsdenii/apt_B/TP
NZ_CP009686/2294642-2294744(-)/Firmicutes/Bacillus_cereus/apt_B/TP
NC_022196/1670335-1670405(+)/Fusobacteria/Fusobacterium_nucleatum/apt_A/TP
NZ_CM003111/1520739-1520864(-)/Proteobacteria/Vibrio_cholerae/apt_A/TP
NC_007298/3464338-3464438(-)/Proteobacteria/Dechloromonas_aromatica/apt_B/TP
NC_017040/957614-957675(+)/Firmicutes/Streptococcus_pyogenes/apt_A/TP
NC_009665/986098-986234(+)/Proteobacteria/Shewanella_baltica/apt_A/TP
NC_017567/1160788-1160851(+)/Firmicutes/Streptococcus_dysgalactiae/apt_A/TP
NZ_CM000728/2134133-2134229(-)/Firmicutes/Bacillus_cereus/apt_A/TP
NC_016779/2245358-2245453(-)/Firmicutes/Bacillus_cereus/apt_A/TP
NC_008321/827895-828003(+)/Proteobacteria/Shewanella_sp/apt_A/TP

NZ_CP011414/860123-860185(-)/Firmicutes/Streptococcus_pyogenes/apt_A/TP
NZ_CP006944/1482291-1482373(+)/Proteobacteria/Mannheimia_varigena/apt_B/TP
NZ_CP012483/3228891-3228995(+)/Firmicutes/Bacillus_cereus/apt_B/TP
NC_023150/584348-584432(-)/Actinobacteria/Rhodococcus_pyridinivorans/apt_B/TP
NC_016771/2279760-2279861(-)/Firmicutes/Bacillus_cereus/apt_B/TP
NZ_CM000758/2267301-2267397(-)/Firmicutes/Bacillus_thuringiensis/apt_A/TP
NZ_CM001977/2260761-2260838(-)/Proteobacteria/Dickeya_zeae/apt_A/TP
NC_005139/1639349-1639473(-)/Proteobacteria/Vibrio_vulnificus/apt_A/TP
NZ_CM000748/2067374-2067478(+)/Firmicutes/Bacillus_thuringiensis/apt_B/TP
NZ_CP006955/1544525-1544642(+)/Proteobacteria/Bibersteinia_trehalosi/apt_B/TP
NZ_CP009981/2269838-2269941(+)/Firmicutes/Bacillus_anthracis/apt_B/TP
NC_005139/1639256-1639341(-)/Proteobacteria/Vibrio_vulnificus/apt_B/TP
NZ_CM001435/4005734-4005869(-)/Proteobacteria/Shewanella_baltica/apt_A/TP
NC_012691/2830373-2830455(+)/Proteobacteria/Tolumonas_auensis/apt_B/TP
NC_015460/913658-913737(-)/Proteobacteria/Gallibacterium_anatis/apt_A/TP
NC_016602/2046034-2046152(+)/Proteobacteria/Vibrio_furnissii/apt_A/TP
NC_015425/1886020-1886101(-)/Firmicutes/Clostridium_botulinum/apt_A/TP
NZ_CP014196/778614-778711(+)/Actinobacteria/Arthrobacter_sp/apt_B/TP
NC_009901/879646-879777(+)/Proteobacteria/Shewanella_pealeana/apt_B/TP
NC_017295/1621303-1621378(-)/Firmicutes/Clostridium_acetobutylicum/apt_B/TP
NZ_CP011058/2165963-2166050(-)/Firmicutes/Paenibacillus_beijingensis/apt_A/TP
NC_009052/4112520-4112656(-)/Proteobacteria/Shewanella_baltica/apt_A/TP
NZ_CP012419/1131855-1131977(+)/Firmicutes/Streptococcus_agalactiae/apt_B/TP
NC_022781/4886034-4886136(-)/Firmicutes/Bacillus_toyonensis/apt_B/TP
NC_016808/1310178-1310258(-)/Proteobacteria/Pasteurella_multocida/apt_A/TP
NC_019903/3578310-3578393(+)/Firmicutes/Desulfitobacterium_dichloroeliminans/apt_B/TP
NZ_CP007062/1573209-1573278(+)/Fusobacteria/Fusobacterium_nucleatum/apt_A/TP
NZ_CP008918/2074656-2074763(-)/Proteobacteria/Pasteurella_multocida/apt_B/TP
NZ_CP008695/637205-637268(-)/Firmicutes/Streptococcus_pyogenes/apt_A/TP
NZ_CP009651/5217675-5217771(+)/Firmicutes/Bacillus_thuringiensis/apt_A/TP
NC_007493/804683-804743(+)/Proteobacteria/Rhodobacter_sphaeroides/apt_B/TP
NC_005085/2193506-2193583(-)/Proteobacteria/Chromobacterium_violaceum/apt_A/TP
NZ_CP009237/2129299-2129386(+)/Proteobacteria/Haemophilus_parasuis/apt_B/TP
NZ_CP012067/8149-8251(+)/Proteobacteria/Aggregatibacter_aphrophilus/apt_B/TP
NC_014500/2325218-2325298(+)/Proteobacteria/Dickeya_dadantii/apt_A/TP
NZ_CP016347/1636869-1636972(-)/Proteobacteria/Vibrio_natriegens/apt_A/TP

NC_009253/1747759-1747844(+)/Firmicutes/Desulfotomaculum_reducens/apt_B/TP
NZ_CP011349/279895-279999(+)/Firmicutes/Bacillus_thuringiensis/apt_B/TP
NZ_CM001974/2165992-2166076(+)/Proteobacteria/Dickeya_chrysanthemi/apt_A/TP
NZ_CM000718/2219437-2219533(+)/Firmicutes/Bacillus_cereus/apt_A/TP
NC_011772/2210470-2210571(-)/Firmicutes/Bacillus_cereus/apt_B/TP
NZ_CP014035/1314775-1314893(+)/Proteobacteria/Vibrio_fluvialis/apt_A/TP
NZ_CM000752/2067231-2067326(+)/Firmicutes/Bacillus_thuringiensis/apt_A/TP
NZ_CP009159/1342046-1342135(-)/Proteobacteria/Actinobacillus_suis/apt_B/TP
NC_021281/597658-597737(-)/Fusobacteria/Fusobacterium_nucleatum/apt_B/TP
NC_011830/5207024-5207099(+)/Firmicutes/Desulfitobacterium_hafniense/apt_A/TP
NC_014622/4416795-4416885(-)/Firmicutes/Paenibacillus_polymyxa/apt_A/TP
NZ_AP014864/3419358-3419462(+)/Firmicutes/Bacillus_thuringiensis/apt_B/TP
NC_009052/4112396-4112512(-)/Proteobacteria/Shewanella_baltica/apt_B/TP
NZ_CP015429/80829-80896(+)/Proteobacteria/Haemophilus_ducreyi/apt_A/TP
NC_012913/1474908-1474983(+)/Proteobacteria/Aggregatibacter_aphrophilus/apt_A/TP
NZ_CP015176/2240858-2240954(-)/Firmicutes/Bacillus_thuringiensis/apt_A/TP
NZ_CP006829/1768129-1768221(+)/Proteobacteria/Thalassolituus_oleivorans/apt_B/TP
NC_012917/2268393-2268484(+)/Proteobacteria/Pectobacterium_carotovorum/apt_A/TP
NZ_CM001841/2113304-2113385(+)/Proteobacteria/Dickeya_dianthicola/apt_B/TP
NZ_CP010783/914935-915012(-)/Firmicutes/Streptococcus_iniae/apt_B/TP
NC_008261/1892279-1892367(-)/Firmicutes/Clostridium_perfringens/apt_B/TP
NZ_CP006943/1569293-1569375(+)/Proteobacteria/Mannheimia_varigena/apt_B/TP
NC_003028/387500-387576(+)/Firmicutes/Streptococcus_pneumoniae/singlet_2/TP
NC_015875/463623-463699(+)/Firmicutes/Streptococcus_pseudopneumoniae/singlet_2/TP
NZ_CP012646/743371-743447(+)/Firmicutes/Streptococcus_mitis/singlet_2/TP
NZ_CP007628/1580158-1580234(-)/Firmicutes/Streptococcus_sp/singlet_2/TP
NC_013853/1774989-1775065(-)/Firmicutes/Streptococcus_mitis/singlet_2/TP
NZ_CM001240/4441757-4441898(-)/Firmicutes/Clostridium_sp/singlet_2/TP
NC_009655/806471-806559(+)/Proteobacteria/Actinobacillus_succinogenes/singlet_2/TP
NC_020833/915411-915492(-)/Proteobacteria/Mannheimia_haemolytica/singlet_2/TP
NC_002940/363037-363137(+)/Proteobacteria/Haemophilus_ducreyi/singlet_2/TP
NC_006370/2453503-2453604(+)/Proteobacteria/Photobacterium_profundum/singlet_2/TP
NC_013165/2020904-2020977(-)/Actinobacteria/Slackia_heliotrinireducens/singlet_2/TP

NC_021021/1431657-1431745(-)/Actinobacteria/Gordonibacter_pamelaeae/singlet_2/TP
NC_013204/146216-146311(-)/Actinobacteria/Eggerthella_lenta/singlet_2/TP
NC_015738/2850373-2850463(+)/Actinobacteria/Eggerthella_sp/singlet_2/TP
NC_014376/4317589-4317664(+)/Firmicutes/Clostridium_saccharolyticum/singlet_2/TP
NC_017905/1829191-1829269(-)/Firmicutes/Streptococcus_parasanguinis/singlet_2/TP
NC_015678/1265775-1265853(-)/Firmicutes/Streptococcus_parasanguinis/singlet_2/TP
NC_009009/353828-353904(+)/Firmicutes/Streptococcus_sanguinis/singlet_2/TP
NC_021175/1794922-1794998(-)/Firmicutes/Streptococcus_oligofermentans/singlet_2/TP
NC_023037/4576573-4576651(-)/Firmicutes/Paenibacillus_polymyxa/singlet_2/TP
NC_016641/70766-70844(-)/Firmicutes/Paenibacillus_terrae/singlet_2/TP
NZ_CP006872/4502101-4502179(-)/Firmicutes/Paenibacillus_polymyxa/singlet_2/TP
NZ_CP014264/635631-635703(+)/Firmicutes/Streptococcus_sp/singlet_2/TP
NZ_CP013651/1185245-1185319(-)/Firmicutes/Streptococcus_sp/singlet_2/TP
NZ_CP014056/2747787-2747881(-)/Proteobacteria/Grimontia_hollisae/singlet_2/TP
NZ_CP008876/2534631-2534695(+)/Firmicutes/Terribacillus_aidingensis/singlet_2/TP
NC_012924/236094-236188(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_017950/237728-237822(+)/Firmicutes/Streptococcus_suis/singlet_2/TP
NC_016826/1050065-1050234(+)/Firmicutes/Streptococcus_infantarius/singlet_2/TP
NC_021900/1053821-1053986(+)/Firmicutes/Streptococcus_lutetiensis/singlet_2/TP
NZ_CP013689/911642-911807(+)/Firmicutes/Streptococcus_infantarius/singlet_2/TP
NC_016749/1083579-1083760(+)/Firmicutes/Streptococcus_macedonicus/singlet_2/TP
NZ_CP013688/855630-855811(-)/Firmicutes/Streptococcus_gallolyticus/singlet_2/TP
NC_013798/1261523-1261704(+)/Firmicutes/Streptococcus_gallolyticus/singlet_2/TP
NC_015600/1091462-1091642(+)/Firmicutes/Streptococcus_pasteurianus/singlet_2/TP
NZ_CM002130/952340-952436(-)/Firmicutes/Streptococcus_sp/singlet_2/TP
NC_006448/885873-885969(-)/Firmicutes/Streptococcus_thermophilus/singlet_2/TP
NC_017595/1133091-1133187(+)/Firmicutes/Streptococcus_salivarius/singlet_2/TP
NC_016791/2529809-2529896(+)/Firmicutes/Clostridium_sp/singlet_2/TP
NC_020134/2092567-2092658(-)/Firmicutes/Clostridium_stercorarium/singlet_2/TP
NZ_CP015483/1670373-1670450(+)/Proteobacteria/Acinetobacter_baumannii/singlet_2/TP
NC_004557/1925417-1925514(+)/Firmicutes/Clostridium_tetani/singlet_1/TP
NC_011566/4572393-4572508(-)/Proteobacteria/Shewanella_piezotolerans/singlet_1/TP

NC_014541/1102873-1102953(+)/Proteobacteria/Ferrimonas_balearica/singlet_1/TP
NC_016513/230264-230398(+)/Proteobacteria/Aggregatibacter_actinomycetemcomitans/singlet_1/TP
NC_019048/1005779-1005901(+)/Firmicutes/Streptococcus_agalactiae/singlet_1/TP
NC_022777/1891808-1891905(+)/Firmicutes/Clostridium_tetani/singlet_1/TP
NZ_AP014520/483704-483838(-)/Proteobacteria/Aggregatibacter_actinomycetemcomitans/singlet_1/TP
NZ_CP010993/1965968-1966036(-)/Firmicutes/Clostridium_perfringens/apt_A/TP
NZ_LN614756/2597828-2597906(+)/Firmicutes/Peptoclostridium_difficile/apt_B/TP
NC_014171/2224035-2224131(-)/Firmicutes/Bacillus_thuringiensis/apt_A/TP

Table 2.16: 53 gene sequences used for generating gcvT (aminomethyltransferase) gene tree.

NZ_LN609302/2660179-2661315(-)/Proteobacteria/Acetobacter_ghanensis/Tandem/GCV
NZ_CP012111/546068-547204(-)/Proteobacteria/Acetobacter_pasteurianus/Tandem/GCV
NC_011985/1993994-1995130(+)/Proteobacteria/Agrobacterium_radiobacter/Tandem/GCV
NZ_CM002024/1393670-1394809(-)/Proteobacteria/Agrobacterium_tumefaciens/Tandem/GCV
NZ_CP007793/966601-967713(+)/Proteobacteria/Azospirillum_brasilense/Tandem/GCV
NZ_CP012401/2758875-2759987(-)/Proteobacteria/Azospirillum_thiophilum/Tandem/GCV
NC_019896/1651856-1652944(+)/Firmicutes/Bacillus_subtilis/Tandem/GCV
NC_010645/500603-501703(+)/Proteobacteria/Bordetella_avium/Tandem/GCV
NZ_CM002881/3256959-3258059(+)/Proteobacteria/Bordetella_bronchiseptica/Tandem/GCV
NZ_CP011446/197379-198479(+)/Proteobacteria/Bordetella_pertussis/Tandem/GCV
NZ_LN907826/3855362-3856510(-)/Proteobacteria/Bradyrhizobium_sp/Tandem/GCV
NZ_CP009099/468235-469338(+)/Proteobacteria/Brucella_abortus/Tandem/GCV
NZ_CP007630/709330-710433(-)/Proteobacteria/Brucella_canis/Tandem/GCV
NC_022906/568298-569401(+)/Proteobacteria/Brucella_ceti/Tandem/GCV
NC_003318/584622-585725(+)/Proteobacteria/Brucella_melitensis/Tandem/GCV
NC_009504/713299-714402(-)/Proteobacteria/Brucella_ovis/Tandem/GCV
NZ_CP007721/671749-672852(-)/Proteobacteria/Brucella_suis/Tandem/GCV
NZ_CP012899/287496-288641(-)/Proteobacteria/Burkholderia_sp/Tandem/GCV
NZ_CM001438/3989899-3991020(+)/Proteobacteria/Burkholderiales_bacterium/Singleton/GCV
NZ_HG966617/655501-656634(+)/Proteobacteria/Candidatus_Phaeomarinobacter/Tandem/GCV

NC_007973/3761634-3762761(+)/Proteobacteria/Cupriavidus_metallidurans/Singleton/GCV
NZ_CP007236/2786127-2787266(+)/Proteobacteria/Ensifer_adhaerens/Tandem/GCV
NC_012793/2405139-2406233(-)/Firmicutes/Geobacillus_sp/Tandem/GCV
NC_016593/2514623-2515717(-)/Firmicutes/Geobacillus_thermoleovorans/Tandem/GCV
NC_019396/1040462-1041595(-)/Proteobacteria/Gluconobacter_oxydans/Tandem/GCV
NZ_CP014327/303264-304382(-)/Proteobacteria/Halocynthiibacter_arcticus/Tandem/GCV
NC_003212/1379801-1380889(+)/Firmicutes/Listeria_innocua/Singleton/GCV
NC_013766/1442007-1443095(+)/Firmicutes/Listeria_monocytogenes/Singleton/GCV
NC_008555/1351863-1352951(+)/Firmicutes/Listeria_welshimeri/Singleton/GCV
NZ_CP014856/3736892-3737995(-)/Firmicutes/Lysinibacillus_sphaericus/Tandem/GCV
NZ_FO538765/768253-769368(-)/Proteobacteria/Magnetospira_sp/Tandem/GCV
NZ_CP015318/4397560-4398660(+)/Proteobacteria/Mesorhizobium_amorphae/Tandem/GCV
NZ_CP015064/2728610-2729710(-)/Proteobacteria/Mesorhizobium_ciceri/Tandem/GCV
NZ_CP012369/1914895-1915995(-)/Firmicutes/Moorella_thermoacetica/Tandem/GCV
NC_002946/1371197-1372297(-)/Proteobacteria/Neisseria_gonorrhoeae/Tandem/GCV
NZ_CP006869/785066-786166(-)/Proteobacteria/Neisseria_meningitidis/Tandem/GCV
NC_008344/2069016-2070107(-)/Proteobacteria/Nitrosomonas_eutropha/Singleton/GCV
NC_007948/1929372-1930538(-)/Proteobacteria/Polaromonas_sp/Tandem/GCV
NZ_CM002757/3448238-3449365(+)/Proteobacteria/Ralstonia_solanacearum/Tandem/GCV
NZ_CP007641/2286802-2287938(+)/Proteobacteria/Rhizobium_etli/Tandem/GCV
NZ_FO082820/1877688-1878845(+)/Proteobacteria/Rhizobium_sp/Tandem/GCV
NC_011420/3781673-3782788(+)/Proteobacteria/Rhodospirillum_centenum/Tandem/GCV
NZ_AP014800/1195891-1197021(+)/Proteobacteria/Rhodovulum_sulfidophilum/Tandem/GCV
NC_017075/1643624-1644763(-)/Proteobacteria/Rubrivivax_gelatinosus/Tandem/GCV
NC_009636/1264127-1265266(-)/Proteobacteria/Sinorhizobium_medicae/Tandem/GCV
NZ_CP013264/1909747-1910811(+)/Proteobacteria/Sphingobium_baderi/Tandem/GCV
NZ_CP012700/2646549-2647685(-)/Proteobacteria/Sphingopyxis_macrogoltabida/Tandem/GCV

NC_020564/1554226-1555317(-)/Firmicutes/Staphylococcus_aureus/Singleton/GCV
NZ_CM001486/279297-280403(+)/Firmicutes/Thermoanaerobacter_siderophilus/Tandem/GCV
NC_010320/243418-244524(-)/Firmicutes/Thermoanaerobacter_sp/Tandem/GCV
NC_015958/318841-319947(-)/Firmicutes/Thermoanaerobacter_wiegelii/Tandem/GCV
NC_014153/2743928-2745088(-)/Proteobacteria/Thiomonas_intermedia/Tandem/GCV
NC_014931/2534656-2535834(-)/Proteobacteria/Variovorax_paradoxus/Tandem/GCV

Table 2.17: 80 gene sequences used for generating sodium:amino-acid symporter gene tree.

NZ_CP007715/1329597-1330964(-)/Proteobacteria/Actinobacillus_equuli/Tandem/TP
NZ_CP006579/3252525-3253901(-)/Proteobacteria/Aeromonas_hydrophila/Tandem/TP
NZ_CP013067/1263393-1264769(+)/Proteobacteria/Aeromonas_schubertii/Tandem/TP
NZ_CP012504/3822993-3824369(-)/Proteobacteria/Aeromonas_veronii/Tandem/TP
NZ_AP014520/482217-483590(-)/Proteobacteria/Aggregatibacter_actinomycetemcomitans/Singleton/TP
NZ_CP012067/8362-9735(+)/Proteobacteria/Aggregatibacter_aphrophilus/Tandem/TP
NZ_LN554846/1394204-1395586(-)/Proteobacteria/Aliivibrio_wodanis/Tandem/TP
NC_003997/2220021-2221388(-)/Firmicutes/Bacillus_anthracis/Tandem/TP
NZ_CP006956/928591-929958(-)/Proteobacteria/Bibersteinia_trehalosi/Tandem/TP
NZ_CP015405/4901990-4903405(+)/Firmicutes/Blautia_sp/Singleton/TP
NZ_CP010888/3051927-3053294(+)/Firmicutes/Clostridioides_difficile/Tandem/TP
NZ_CP006905/3150801-3152129(-)/Firmicutes/Clostridium_baratii/Tandem/TP
NZ_HG917868/1302364-1303743(+)/Firmicutes/Clostridium_bornimense/Tandem/TP
NZ_CP006903/3543636-3545063(+)/Firmicutes/Clostridium_botulinum/Tandem/TP
NZ_CP013252/1333239-1334666(+)/Firmicutes/Clostridium_butyricum/Tandem/TP
NC_015275/4694284-4695744(-)/Firmicutes/Clostridium_lentocellum/Tandem/TP
NZ_CP009557/1935440-1936828(-)/Firmicutes/Clostridium_perfringens/Tandem/TP
NC_004557/2098988-2100382(-)/Firmicutes/Clostridium_tetani/Tandem/TP
NC_015565/1335500-1336897(+)/Firmicutes/Desulfotomaculum_nigrificans/Tandem/TP
NC_009253/1747963-1749333(+)/Firmicutes/Desulfotomaculum_reducens/Tandem/TP
NC_002937/806523-807929(-)/Proteobacteria/Desulfovibrio_vulgaris/Tandem/TP
NZ_CP007452/2055773-2057167(-)/Firmicutes/Eubacterium_acidaminophilum/Tandem/TP
NZ_CP011914/2724288-2725730(-)/Firmicutes/Eubacterium_limosum/Singleton/TP

NC_021021/1430006-1431538(-)/Actinobacteria/Gordonibacter_pamelaeae/Singleton/TP
NZ_CP015425/1104447-1105814(+)/Proteobacteria/Haemophilus_ducreyi/Tandem/TP
NC_000907/934743-936113(-)/Proteobacteria/Haemophilus_influenzae/Tandem/TP
NZ_CP009158/327476-328864(+)/Proteobacteria/Haemophilus_parasuis/Tandem/TP
NC_010337/1555566-1556966(+)/Firmicutes/Heliobacterium_modesticaldum/Tandem/TP
NZ_CP006942/1656766-1658133(+)/Proteobacteria/Mannheimia_varigena/Tandem/TP
NC_015873/19445-20830(-)/Firmicutes/Megasphaera_elsdenii/Tandem/TP
NZ_LN554852/4481457-4482854(-)/Proteobacteria/Moritella_viscosa/Singleton/TP
NZ_CP011512/4590689-4592107(-)/Firmicutes/Paenibacillus_peoriae/Singleton/TP
NC_014483/4051893-4053311(-)/Firmicutes/Paenibacillus_polymyxa/Singleton/TP
NZ_CP010976/2512692-2514077(-)/Firmicutes/Paenibacillus_sp/Tandem/TP
NC_016641/69181-70695(-)/Firmicutes/Paenibacillus_terrae/Singleton/TP
NZ_CP004391/692024-693397(-)/Proteobacteria/Pasteurella_multocida/Tandem/TP
NC_009089/2640783-2642150(+)/Firmicutes/Peptoclostridium_difficile/Tandem/TP
NZ_CP005974/1989396-1990769(-)/Proteobacteria/Photobacterium_gaetbulicola/Singleton/TP
NZ_LT575468/1508203-1509558(+)/Proteobacteria/Plesiomonas_shigelloides/Singleton/TP
NC_011566/4570806-4572179(-)/Proteobacteria/Shewanella_piezotolerans/Singleton/TP
NC_013165/2019317-2020807(-)/Actinobacteria/Slackia_heliotrinireducens/Singleton/TP
NC_004116/1147833-1149179(+)/Firmicutes/Streptococcus_agalactiae/Singleton/TP
NC_012891/1062762-1064099(+)/Firmicutes/Streptococcus_dysgalactiae/Tandem/TP
NC_011134/933932-935266(-)/Firmicutes/Streptococcus_equi/Tandem/TP
NC_013798/1261838-1263166(+)/Firmicutes/Streptococcus_gallolyticus/Singleton/TP
NC_016826/1050374-1051705(+)/Firmicutes/Streptococcus_infantarius/Singleton/TP
NZ_CP010783/913477-914805(-)/Firmicutes/Streptococcus_iniae/Tandem/TP
NC_021900/1054125-1055456(+)/Firmicutes/Streptococcus_lutetiensis/Singleton/TP
NC_016749/1083892-1085220(+)/Firmicutes/Streptococcus_macedonicus/Singleton/TP
NC_013853/1773570-1774892(-)/Firmicutes/Streptococcus_mitis/Singleton/TP
NC_004350/1116158-1117522(+)/Firmicutes/Streptococcus_mutans/Singleton/TP
NC_015291/1569000-1570322(-)/Firmicutes/Streptococcus_oralis/Singleton/TP
NC_015678/1264334-1265665(-)/Firmicutes/Streptococcus_parasanguinis/Singleton/TP
NC_015558/862342-863673(+)/Firmicutes/Streptococcus_parauberis/Tandem/TP
NC_015600/1091783-1093111(+)/Firmicutes/Streptococcus_pasteurianus/Singleton/TP

NC_003028/387673-388995(+)/Firmicutes/Streptococcus_pneumoniae/Singleton/TP
NC_015875/463796-465118(+)/Firmicutes/Streptococcus_pseudopneumoniae/Singleton/TP
NC_002737/1046652-1047974(+)/Firmicutes/Streptococcus_pyogenes/Tandem/TP
NC_015760/1110605-1111951(-)/Firmicutes/Streptococcus_salivarius/Singleton/TP
NZ_CM002130/950896-952239(-)/Firmicutes/Streptococcus_sp/Singleton/TP
NZ_CP007628/1578740-1580062(-)/Firmicutes/Streptococcus_sp/Singleton/TP
NZ_CP013651/1183740-1185080(-)/Firmicutes/Streptococcus_sp/Singleton/TP
NZ_CP014264/635802-637124(+)/Firmicutes/Streptococcus_sp/Singleton/TP
NZ_CP014835/477311-478636(-)/Firmicutes/Streptococcus_sp/Tandem/TP
NZ_CP016207/1048465-1049787(+)/Firmicutes/Streptococcus_sp/Singleton/TP
NC_012924/236300-237634(+)/Firmicutes/Streptococcus_suis/Singleton/TP
NC_006448/884430-885773(-)/Firmicutes/Streptococcus_thermophilus/Singleton/TP
NC_012004/962232-963557(+)/Firmicutes/Streptococcus_uberis/Tandem/TP
NC_015385/888775-890181(+)/Spirochaetes/Treponema_succinifaciens/Singleton/TP
NZ_CP016224/1401954-1403330(-)/Proteobacteria/Vibrio_alginolyticus/Tandem/TP
NZ_CP011460/1510100-1511479(+)/Proteobacteria/Vibrio_anguillarum/Tandem/TP
NZ_CP016177/1449865-1451250(+)/Proteobacteria/Vibrio_breoganii/Tandem/TP
NC_002505/1518957-1520333(-)/Proteobacteria/Vibrio_cholerae/Tandem/TP
NZ_CP009264/1181847-1183220(-)/Proteobacteria/Vibrio_coralliilyticus/Tandem/TP
NZ_CP014035/1315090-1316475(+)/Proteobacteria/Vibrio_fluvialis/Tandem/TP
NC_016602/2046361-2047746(+)/Proteobacteria/Vibrio_furnissii/Tandem/TP
NZ_CP014043/2210139-2211515(-)/Proteobacteria/Vibrio_mimicus/Tandem/TP
NC_022528/1969866-1971242(+)/Proteobacteria/Vibrio_nigripulchritudo/Tandem/TP
NC_004603/1856266-1857642(+)/Proteobacteria/Vibrio_parahaemolyticus/Tandem/TP
NZ_CP009261/1621127-1622503(-)/Proteobacteria/Vibrio_vulnificus/Tandem/TP

Table 3.1: 80 ykkC riboswitch sequences predating subtyping (2014).

NC_006322/1406463-1406565/1
NC_015410/1521902-1522002/-1
NC_009253/2382760-2382861/-1
NC_005090/1067279-1067380/1
NC_008781/42765-42882/-1
NC_009328/583649-583751/1
NC_014964/1749307-1749406/-1
NC_018520/1355827-1355938/1
NC_005296/1561347-1561451/-1
NC_006582/4004578-4004674/-1

NC_004757/2634101-2634203/-1
NC_008344/2598146-2598249/-1
NC_015740/2010288-2010390/-1
NC_013421/2939936-2940037/-1
NC_009089/2762255-2762349/-1
NC_008379/312516-312624/1
NC_007348/18312-18427/-1
NC_009439/1531564-1531681/-1
NC_006177/3084909-3085016/-1
NC_009617/3993812-3993909/-1
NC_008786/4178508-4178625/1
NC_007348/1610941-1611056/-1
NC_009338/84804-84923/-1
NC_008344/2598462-2598562/1
NC_009254/127094-127215/-1
NC_009633/951602-951713/1
NC_007005/4719777-4719879/1
NC_007644/2145827-2145928/-1
NC_007298/86805-86928/1
NC_007958/2437685-2437791/1
NC_006510/632114-632216/1
NC_009706/2546053-2546156/-1
NC_007765/146041-146152/-1
NC_012988/4578710-4578819/-1
NC_002570/741875-741978/-1
NC_008343/1970241-1970354/-1
NC_008726/851507-851616/1
NC_009922/585550-585652/1
NC_006177/808865-808978/-1
NC_007492/1520287-1520387/-1
NC_010512/472628-472749/1
NC_008062/82819-82950/-1
NC_009720/1413177-1413287/-1
NC_009253/3007686-3007792/-1
NC_009922/1397256-1397351/1
NC_007907/4836749-4836847/-1
NC_009445/7058665-7058767/-1
NC_009484/2993982-2994092/-1
NC_007350/2423973-2424072/-1

NC_009719/2046573-2046691/1
NC_004757/2634421-2634524/1
NC_007912/1455387-1455513/1
NC_010676/1408092-1408226/-1
NC_000911/801280-801401/-1
NC_017178/512915-513021/1
NC_007925/5067660-5067771/-1
NC_003869/586451-586555/1
NC_007952/1848294-1848426/-1
NC_005070/1383854-1383952/1
NC_011898/2951117-2951220/1
NC_008062/83383-83507/1
NC_012586/393914-394020/-1
NC_005125/1020399-1020511/-1
NC_005966/1073618-1073719/1
NC_017922/69423-69558/1
NC_008789/2583259-2583363/1
NC_015740/2020690-2020787/-1
NC_011595/2400442-2400537/-1
NC_016023/926937-927033/-1
NC_004193/3401639-3401735/-1
NC_012658/2486145-2486240/-1
NC_017195/3790541-3790643/-1
NC_008435/1370978-1371085/1
NC_016023/927248-927350/1
NC_009253/2597929-2598025/-1
NC_007614/1087459-1087562/1
NC_006322/3988810-3988916/-1
NC_008346/2034312-2034410/-1
NC_008531/1882072-1882171/1
NC_007516/1021560-1021660/-1

Table 3.2: 167 ykkC riboswitch sequences gathered using RFAM2014 model.

NC_009925/2875834-2875970/+
NZ_LN606600/523538-523438/-
NC_014378/2441135-2441053/-
NC_009484/2994092-2993982/-
NZ_CM001194/1643333-1643436/+

NZ_CP010368/2461269-2461174/-
NC_022657/7558152-7558053/-
NC_011989/2057123-2057231/+
NC_011983/2633328-2633438/+
NC_008260/2160939-2161078/+
NC_017167/753291-753393/+
NC_009633/951602-951713/+
NZ_CP008953/6257347-6257211/-
NZ_CP014140/3299715-3299613/-
NC_021878/1377631-1377533/-
NZ_FO818640/3756887-3756739/-
NC_009937/4992010-4992119/+
NC_013856/442291-442169/-
NC_018665/391120-391208/+
NC_017188/3739365-3739264/-
NZ_CP012946/3565146-3565043/-
NC_020453/732166-732268/+
NZ_CM001230/86252-86152/-
NC_012491/931130-931233/+
NZ_CP009322/2730431-2730306/-
NZ_CM003769/314712-314581/-
NZ_CP011371/3600517-3600647/+
NC_003869/586451-586555/+
NZ_CP010796/2128797-2128894/+
NZ_CP011427/3988230-3988136/-
NC_010995/3193184-3193297/+
NC_019697/4814084-4813965/-
NC_019695/1896963-1897120/+
NZ_CP009687/387633-387730/+
NZ_AP014568/144816-144716/-
NZ_CP010869/2663550-2663444/-
NZ_CP012268/4227966-4227872/-
NC_007348/18427-18312/-
NC_019776/3816086-3815958/-
NC_019675/985788-985932/+
NC_014501/2260736-2260866/+
NC_019780/3663911-3663774/-
NC_007298/86805-86928/+
NC_018017/1998365-1998464/+

NC_018068/4466994-4466901/-
NC_013216/736608-736730/+
NZ_CM001904/54696-54602/-
NZ_CP012487/2331309-2331407/+
NC_015968/2137130-2137228/+
NC_020063/2830592-2830494/-
NC_014828/2596987-2597091/+
NZ_AP017305/4222049-4222159/+
NC_014394/823426-823537/+
NC_019703/2860848-2860718/-
NC_006510/632114-632216/+
NZ_CP014335/558593-558695/+
NC_005125/1020511-1020399/-
NC_019745/2159920-2159791/-
NC_019396/3193599-3193699/+
NC_008343/1970354-1970241/-
NZ_CP007142/3985045-3984937/-
NC_014654/2476545-2476452/-
NC_019978/2611487-2611399/-
NZ_CP011052/2939698-2939820/+
NC_008789/2583259-2583363/+
NZ_CP015243/3483938-3484072/+
NC_019779/2128785-2128676/-
NC_011899/2415882-2415768/-
NC_010337/3064147-3064036/-
NZ_CP013737/5183627-5183726/+
NZ_CM002917/3005378-3005490/+
NC_015717/2212961-2213068/+
NZ_CP009416/3414823-3414729/-
NZ_CCSE01000001/1388102-1388004/-
NC_016612/4376383-4376481/+
NZ_CP011602/2858385-2858287/-
NZ_CP014007/3594207-3594109/-
NC_014098/2032540-2032430/-
NZ_CP011927/3136385-3136500/+
NZ_CP015398/70497-70590/+
NZ_CP013990/2541756-2541658/-
NC_019683/524566-524686/+

NC_008531/1882072-1882171/+
NZ_CP011112/3470029-3469926/-
NZ_CP010820/3523057-3522963/-
NZ_FO538765/2745108-2745228/+
NC_023065/106903-106794/-
NC_017506/3311458-3311566/+
NC_015559/3650363-3650472/+
NZ_CP015318/5451337-5451452/+
NC_012808/4179541-4179432/-
NC_011666/2840436-2840544/+
NZ_CM001475/678109-678206/+
NZ_CP014360/3863290-3863190/-
NZ_CP012020/446351-446254/-
NC_014207/2461464-2461359/-
NZ_CP012697/4187453-4187347/-
NC_019738/55200-55067/-
NZ_CP011304/2706187-2706321/+
NC_007644/2145928-2145827/-
NC_009338/84923-84804/-
NC_008595/2148471-2148364/-
NC_010718/477956-478058/+
NZ_CP011451/2425204-2425098/-
NZ_CP012371/796231-796128/-
NZ_LN885086/1127535-1127638/+
NC_004193/3401735-3401639/-
NC_019693/57118-56984/-
NZ_CM001633/1425799-1425911/+
NZ_CP011058/5217937-5217805/-
NZ_CP013068/5226361-5226254/-
NC_013956/4526513-4526415/-
NZ_CP013103/1054758-1054877/+
NC_009719/2046573-2046691/+
NZ_CP009761/721437-721530/+
NC_004547/2166770-2166871/+
NZ_CP010978/4511963-4511868/-
NC_009089/578842-578948/+
NZ_CM002803/1472057-1471917/-
NZ_CP016543/1431514-1431615/+
NC_019689/1756657-1756527/-

NC_008781/42882-42765/-
NZ_CP014135/1720693-1720591/-
NC_016147/67473-67369/-
NZ_CP010557/3054141-3054239/+
NC_010168/527597-527716/+
NC_007765/146152-146041/-
NC_007493/243850-243747/-
NC_008435/1370978-1371085/+
NC_011898/2951117-2951220/+
NC_007912/1455387-1455513/+
NC_019673/1511044-1511174/+
NZ_CP011366/208563-208661/+
NC_013517/1681184-1681281/+
NC_013959/82172-82284/+
NC_012586/394021-393914/-
NC_013165/2079115-2079218/+
NZ_CP011280/475171-475082/-
NZ_CP011477/1710086-1710247/+
NC_019748/1114454-1114310/-
NZ_CP009623/2154778-2154881/+
NC_014217/87994-87894/-
NC_014762/132180-132088/-
NZ_AP014724/324430-324525/+
NC_006177/808978-808865/-
NC_019702/737568-737429/-
NZ_CP007542/3197895-3197776/-
NC_015172/773462-773569/+
NC_008346/2034410-2034312/-
NC_014220/621404-621502/+
NC_012997/3854160-3854312/+
NZ_CP008876/1063337-1063432/+
NC_020888/2537439-2537550/+
NZ_CP004388/788176-788069/-
NC_014152/827070-827214/+
NC_013921/613962-614062/+
NC_014410/2110964-2110830/-
NC_014145/3597854-3597744/-
NC_012691/3187332-3187227/-
NC_008786/4178508-4178625/+

NZ_CP016178/125843-125940/+
NZ_CP007161/3295913-3295815/-
NZ_CP012873/2202128-2202222/+
NC_005090/1067279-1067380/+
NC_009720/1413287-1413177/-
NZ_CP014544/3080917-3080785/-
NC_014219/400499-400400/-

Table 3.3: Cluster stability after 100 bootstrap replicates for PRPP cluster.

	RNAmountAlign	RNApdist	Muscle
cluster_walktrap	0.82	0.27	0.97
cluster_leading_eigen	0.66	0.3	1
cluster_edge_betweenness	0.53	0.41	1
cluster_fast_greedy	0.59	0.3	1

Table 3.4: 105 ppGpp riboswitch sequences taken from published literature used for graph clustering analysis.

NZ_AKVO01000033.1/47488-47583/+
NZ_KB894584.1/37082-37183/-
NZ_ASXP01000002.1/549561-549662/-
NZ_AKVN01000035.1/9089-9184/-
JGI20133J14441_1000263/8482-8600/+
NZ_CP008852.1/4786376-4786485/+
NZ_CYSP01000007.1/150083-150211/+
JGI12269J14319_10140190/486-607/-
PRSSGFe2_Sequence0000005665/4866-4975/-
JGI12269J14319_10083514/1159-1279/-
JGI12269J14319_10018912/1293-1411/-
NZ_AAWL01000017.1/34692-34810/-
3300000589_100257/1933-2046/-
NZ_AFGF01000074.1/24694-24806/-
NC_014098.1/1810613-1810721/-
NZ_AUPA01000230.1/10819-10908/+
BS_KBA_SWE02_21m_10017557/1907-2007/+

NC_017672.3/7477830-7477969/-
NC_015690.1/7301535-7301674/-
NZ_CP013653.1/2205541-2205696/+
2200592311/421-551/-
NZ_LMRV01000097.1/7383-7501/-
NZ_KE993473.1/205154-205259/-
BMHBC_168058/6-149/+
PRSSGFe2_Sequence0000000212/11450-11554/-
PRSSGFe2_Sequence0000002333/1598-1713/-
NZ_AFHW01000085.1/40220-40340/+
MA40A_F51MHD001CDJ3Q/1-109/+
NZ_AKVO01000077.1/190931-191027/+
NC_015757.1/623941-624039/+
NZ_ASXP01000002.1/145273-145391/-
NZ_AQYX01000129.1/10395-10511/-
NZ_ASLS01000001.1/32920-33026/+
NC_016884.1/2934649-2934747/-
NC_011830.1/2677591-2677690/+
NZ_AGJB01000011.1/26362-26462/-
NZ_AUBR01000010.1/69700-69820/-
NZ_JMLG01000011.1/23594-23720/+
NZ_AUMW01000006.1/157793-157919/-
NZ_BAMO01000005.1/136903-137047/-
NZ_LFVV01000018.1/194699-194810/+
NZ_JYNH01000051.1/25199-25309/-
NZ_JPST01000003.1/81680-81799/-
NZ_KV440964.1/3193866-3194044/-
NZ_JMGA01000032.1/40575-40703/+
NZ_LBMD01000016.1/37671-37850/+
NZ_LDZY01000005.1/14616-14727/-
NZ_LKET01000032.1/189560-189674/-
NZ_JYNH01000120.1/1791-1899/-
NC_014831.1/2119334-2119522/+
NZ_LFEL01000003.1/275926-276104/+
NZ_AXAR01000001.1/86593-86710/+
NZ_LDXJ01000064.1/112242-112378/+
NZ_LGRO01000001.1/724969-725079/+
NZ_JMGA01000012.1/25563-25672/+

NZ_AUCA01000015.1/56713-56832/+
NZ_JQKL01000037.1/23078-23300/+
NZ_AAWL01000017.1/30778-30902/-
MA55A_contig00035/27082-27180/+
NC_014377.1/866146-866258/+
NZ_LDZY01000015.1/6926-7035/-
NZ_JH976535.1/833534-833756/-
bisonPool14jan08_BXAW24855_x2/20-145/+
NZ_KE993508.1/538719-538840/-
NC_015519.1/423274-423384/-
SL_1KL_011_SED_10036648/1463-1589/+
NZ_AFGF01000085.1/82297-82412/-
NZ_AFGF01000065.1/16774-16889/+
NZ_JH806609.1/57239-57347/+
NZ_GL892032.1/3633718-3633836/+
MA55A_contig01460/1575-1701/+
NZ_AAWL01000031.1/24358-24474/+
NZ_AQXL01000093.1/12583-12702/-
SL_1KL_011_SED_10114630/706-828/+
NZ_AAWL01000001.1/34755-34867/-
ASPA01000009.1/30799-30927/-
SakLake120mDRAFT_c0060996/773-904/-
JGI20133J14441_1001635/2067-2179/+
NC_007907.1/1602834-1602948/+
PRSSGFe2_Sequence0000002463/704-817/+
PRSSGFe2_Sequence0000001307/5211-5320/+
JGI12269J14319_10001093/8643-8752/+
NZ_GL892032.1/148483-148604/+
NC_013216.1/608725-608855/+
SakLake120mDRAFT_c0156973/272-404/-
JGI12269J14319_10003118/4270-4381/+
ASA129_GJG7ZZE02HAPB0/204-336/+
NZ_HG764817.1/1540082-1540192/+
SakLake120mDRAFT_c0072242/291-422/-
SakLake120mDRAFT_c0187404/123-254/+
NZ_ACPC01000009.1/302082-302260/+
NC_015573.1/2893883-2894011/-
NC_018068.1/4466887-4466994/-

NZ_CM001441.1/5165054-5165165/-
NZ_AGAF01000167.1/18389-18498/-
NC_016584.1/3661414-3661525/-
NC_018017.1/1998365-1998479/+
NC_018515.1/4384712-4384822/-
NC_016584.1/5416978-5417087/-
NZ_AFCE01000155.1/42320-42467/+
NZ_AGAF01000281.1/21490-21601/-
NC_018515.1/2933436-2933564/-
NC_018068.1/3401209-3401321/-
NZ_CM001441.1/3442349-3442483/-
NZ_AFGF01000135.1/6970-7085/-

Table 3.5: 253 PRPP riboswitch sequences taken from published literature used for graph clustering analysis.

NZ_JRMY01000001.1/9230092408/+
NZ_CP009283.1/7081989-7082116/-
NZ_CCDG010000038.1/57043-57170/-
NZ_CP009284.1/6725061-6725188/-
NZ_JQCR01000002.1/22775292277654/+
NZ_CP009428.1/6600509-6600627/-
NZ_LCZJ02000037.1/383986-384107/-
NZ_CP009281.1/6557535-6557656/-
NZ_CP009279.1/6544295-6544416/-
NZ_LIPY01000121.1/66234-66360/-
NZ_CP009285.1/7902582-7902707/-
NZ_ASPX01000009.1/30328-30453/-
NZ_CP009241.1/7385984-7386110/-
NZ_ASPS01000012.1/6686866993/+
NZ_CP009282.1/7458095-7458220/-
NZ_JUEI01000059.1/86278752/+
NZ_LN831199.1/403289-403419/-
NZ_AGBD01000380.1/21942-22067/-
NZ_LIRB01000149.1/92099-92225/-
NZ_CP009287.1/6891887-6892013/-
SRS047113_C5407849/245367/+

SL_8KL_010_SED_10008549/6511-6772/-
NZ_AFUS01000004.1/626182626290/+
NZ_AFII01000001.1/320513-320621/-
SRS023958_Baylor_scaffold_41336/14-36/-
SRS022719_LANL_scaffold_67508/33983521/+
SRS013533_C5506489/391-514/-
SRS045197_C3607724/18902013/+
SRS064329_LANL_scaffold_42366/102463-102586/-
NZ_AENT01000025.1/32626-32741/-
SRS013506_C1116491/6158-6271/-
SRS019125_WUGC_scaffold_6919/11103-11231/-
SRS023468_LANL_scaffold_20263/1471-1575/-
NC_010718.1/477956478073/+
NZ_CYUJ01000003.1/377225-377339/-
NZ_AFII01000002.1/496807-496929/-
SRS044662_LANL_scaffold_30258/11409-11532/-
NZ_AXUQ01000034.1/18611967/+
NZ_AQWN01000002.1/218893219015/+
NZ_JONT01000010.1/29363050/+
NZ_JMLG01000005.1/140063-140187/-
NC_015573.1/1707449-1707578/-
NZ_JHVU01000026.1/105733-105844/-
NZ_AUMW01000001.1/53835-53961/-
SRS013705_Baylor_scaffold_11019/217-323/-
NZ_AJZT01000010.1/205231205337/+
SRS054687_C3397657/14-40/-
SRS077736_C1758051/161-267/-
SRS018791_WUGC_scaffold_20417/334-440/-
SRS014689_WUGC_scaffold_22209/85508656/+
NZ_AFIH01000001.1/2234718-2234824/-
2205844349/275389/+
NZ_AFZD01000020.1/111314-111435/-
SRS015209_WUGC_scaffold_31606/34173538/+
SRS011086_WUGC_scaffold_28361/529650/+
SRS024081_LANL_scaffold_45637/553-674/-
NZ_AGJQ01000004.1/11989-12136/-
SRS023604_Baylor_scaffold_10032/1508015193/+
SRS011111_Baylor_scaffold_11533/5865-5984/-

NZ_CVPE01000004.1/203425-203537/-
NZ_KQ955328.1/90089111/+
NZ_AEEH01000043.1/1475314870/+
NZ_AEAA01000041.1/60494-60600/-
NZ_CP007032.1/893257893412/+
NZ_LDZY01000008.1/3883938941/+
NZ_AGAF01000124.1/36165-36272/-
JGI12269J14319_10010382/11971302/+
NZ_JYNH01000001.1/151499-151646/-
NC_016584.1/713070713199/+
NC_018068.1/911937912042/+
NZ_CYUO01000004.1/320058320168/+
NZ_ABXA01000036.1/5314353249/+
NZ_HE978616.1/626865-626970/-
NZ_AEXN01000015.1/56855-56961/-
NZ_CWHU01000005.1/1113391-1113494/-
longitudinal_123_257/74797583/+
NZ_LN913016.1/1173836-1173939/-
NZ_HE578907.1/867528-867631/-
longitudinal_207_107/21432-21550/-
NZ_ACXU01000005.1/8591786035/+
longitudinal_447_312/4023840357/+
NC_018515.1/759601759708/+
NZ_HG764817.1/23004932300610/+
NZ_LN874948.1/553435-553552/-
NZ_GL638023.1/1281012942/+
NZ_AEFH01000078.1/1281012942/+
JGI996J12353_1000374/9167-9277/-
JGI994J12354_10002612/170280/+
2227080785/139243-139353/-
PCEOT_contig00605/139240/+
NZ_AGJB01000015.1/46455-46579/-
NC_019903.1/12511121251256/+
SL_8KL_010_SED_10022983/29413115/+
SL_8KL_010_SED_10042056/363536/+
NC_014220.1/621404621517/+
NZ_BBQT01000008.1/6405164164/+
NC_008346.1/2034297-2034410/-
PRSSGFe2_Sequence0000006776/1957-2069/-

JGI994J12354_10482035/239-370/-
NZ_BBCE01000001.1/165346-165459/-
PCEOT_contig01528/69017015/+
SL_8KL_010_SED_10031631/651816/+
NZ_CGIH01000047.1/3755737674/+
UnmappedStool_Broad_scaffold_685386/16791797/+
2016906512/342444/+
GBSCECS77_contig00594/181-283/-
NZ_ADDO01000063.1/106547-106653/-
NZ_JMLS01000019.1/105243-105368/-
NZ_BAMO01000056.1/16613-16751/-
2033701275/323-439/-
NZ_AUFT01000023.1/22352365/+
NZ_AYXG01000013.1/32243359/+
SRS050244_C4575790/22972425/+
NZ_AFCE01000103.1/17028-17223/-
NZ_KE387173.1/190191-190309/-
NZ_JYBN01000008.1/159868-159981/-
NZ_KE952809.1/3856538679/+
NZ_CP014140.1/3299599-3299715/-
NZ_BBWZ01000010.1/4657146686/+
NZ_ACJM01000008.1/145315145422/+
SL_5KL_010_SED_10000537/1025010421/+
NC_009253.1/2597914-2598025/-
NZ_CAOS01000014.1/34593576/+
NC_015589.1/12977761297890/+
NZ_AEVP01000023.1/3413934267/+
NC_013216.1/736608736745/+
SL_3KL_010_SED_10003418/343499/+
SL_9KL_010_SED_10020703/3029-3170/-
SL_8KL_010_SED_10028073/19662138/+
SL_1KL_011_SED_10003395/7915-8058/-
SL_5KL_010_SED_10012239/1667-1831/-
SL_9KL_010_SED_10000015/674830/+
SL_1KL_011_SED_10011804/11201287/+
SL_3KL_010_SED_10001626/11051287/+
SL_9KL_010_SED_10004010/114-971/-
SL_5KL_010_SED_10002709/57905955/+
SL_1KL_011_SED_10000130/37806-37971/-

NC_006177.1/3084894-3085016/-
NC_014152.1/827070827229/+
NZ_LGTE01000017.1/2005120211/+
SL_8KL_010_SED_10000172/29131-29310/-
SL_5KL_010_SED_10009162/2110-2274/-
SL_5KL_010_SED_10001041/1108711263/+
SL_5KL_010_SED_10016439/18582016/+
SL_8KL_010_SED_10000991/19853-20044/-
SL_8KL_010_SED_10000266/18840-19031/-
SL_3KL_010_SED_10066664/449-641/-
SL_8KL_010_SED_10017205/26652843/+
SL_9KL_010_SED_10016461/2789-2906/-
NC_009454.1/2654383-2654559/-
NC_014378.1/2441038-2441135/-
NZ_KB900621.1/6841468652/+
P_1C_Sed_2_UnCty_1006277/11801286/+
NC_017455.1/177388177556/+
SL_8KL_010_SED_10033978/14081619/+
SL_8KL_010_SED_10022742/15861809/+
NC_014654.1/2476437-2476545/-
NZ_CAUI01000019.1/8165481812/+
SL_8KL_010_SED_10001500/1690-1831/-
SL_6KL_011_BRINE_1003705/3873-4030/-
2211348873/170303/+
NC_019978.1/2611384-2611487/-
NZ_KI912094.1/668311-668510/-
NZ_CP009281.1/16801441680267/+
NZ_CP008852.1/1401716-1401847/-
NZ_JXAK01000005.1/49714-49835/-
NZ_JYCD01000011.1/21367-21505/-
NZ_KE695649.1/474674474812/+
NZ_JQMH01000018.1/472516472654/+
NZ_LGIQ01000017.1/277540-277653/-
NZ_KI629785.1/1135762-1135901/-
NZ_LJIZ01000005.1/7105571172/+
NZ_AXAR01000006.1/43212-43341/-
NZ_JMIR01000036.1/3030130425/+
NZ_CP013652.1/2949233-2949361/-
NZ_JNIL01000001.1/2559059-2559172/-

NZ_AQXL01000109.1/2306223212/+
NZ_LN876586.1/613393613504/+
NZ_JNFS01000002.1/1378113912/+
NZ_AFRV01000006.1/544244-544389/-
NZ_KB894300.1/59967-60101/-
IMNBL1_c0107919/383513/+
NZ_AFGF01000017.1/5780157943/+
NZ_JMIR01000033.1/13745-13873/-
NZ_AKIX01000068.1/2235-2373/-
NZ_LDZV01000001.1/4146688-4146825/-
NZ_LDCN01000005.1/196955197093/+
NC_012491.1/4667470-4667623/-
NZ_AKYF01000026.1/2094-2228/-
SRS023468_C1213219/21-92/-
NZ_KQ957097.1/4004740158/+
NZ_CYSP01000003.1/264381-264500/-
NZ_KE386494.1/684934-685058/-
NZ_CP009170.1/556732556852/+
NZ_CM001486.1/16122261612356/+
NZ_ACXY01000003.1/20537-20666/-
NC_014209.1/657666657796/+
NC_003869.1/586436586570/+
NZ_KK211288.1/138607-138718/-
NZ_JXLL01000037.1/19112043/+
NZ_CCEZ01000007.1/57653-57769/-
NC_009922.1/585550585667/+
NZ_CP008852.1/14637331463869/+
NZ_AKVN01000062.1/6687667026/+
NZ_AKVO01000066.1/104625-104775/-
NZ_AAWL01000019.1/2691827047/+
NZ_AZTB01000004.1/6981069942/+
NZ_KB900618.1/318000-318218/-
NC_009633.1/951602951728/+
NC_017992.1/2639227-2639374/-
NZ_CELZ01000042.1/161468-161584/-
NC_007644.1/2145812-2145943/-
NZ_BBKT01000052.1/24352567/+
MA55A_contig07155/649-798/-
NC_019970.1/2157170-2157317/-

MA55A_contig03190/1043-1190/-
NC_015555.1/748565748698/+
JGI20130J15133_1039652/670-814/-
NZ_JH932301.1/481348481454/+
NZ_AGZE01000033.1/6012260243/+
NZ_JHVL01000031.1/37996-38113/-
NZ_JYHU01000002.1/413238-413350/-
longitudinal_307_167/14791-14915/-
longitudinal_154_167/2727427382/+
NZ_BAEW01000001.1/4967349781/+
NZ_AENP01000008.1/79850-79960/-
longitudinal_153_212/6315-6422/-
SRS019016_C497669/3724-3836/-
NZ_FCEX01000012.1/669395-669507/-
NZ_ASXP01000001.1/474895475031/+
NZ_KI259821.1/168295-168400/-
NZ_CYSP01000010.1/6646766605/+
NZ_KB898186.1/247208247334/+
NZ_AUFO01000015.1/149890150016/+
NZ_AUCA01000022.1/3163431754/+
NZ_JH165061.1/456105456213/+
NZ_AGBB01000034.1/2868328806/+
NZ_AURB01000093.1/49895102/+
SakLake120mDRAFT_c0028370/14671621/+
SakLake120mDRAFT_c0619158/22-68/-
ASMM170b_GJFD58A01CETJT/212-367/-
SaLa_GLZXY3L01C2NYU/168322/+
SakLake120mDRAFT_c0097431/584-739/-
SakLake120mDRAFT_c0014727/333486/+
SakLake120mDRAFT_c0203832/254-408/-
SakLake120mDRAFT_c0621178/1661/+
SakLake120mDRAFT_c0618394/1659/+
ASOZ01000001.1/34354-34494/-
taComm3_FNHH33628_b1/14-21/-
3300000501_104317/381551/+
JGI12104J13512_1005492/38984060/+
NZ_HE978546.1/17823-17968/-
NZ_ARZA01000203.1/41471-41614/-

Table 3.6: 45 ADP-CDP riboswitch sequences taken from published literature used for graph clustering analysis.

NZ_DS981517.1/878682878787/+
NC_017297.1/3026211-3026316/-
NZ_KK211278.1/226948227069/+
NZ_AUCA01000061.1/10625-10754/-
NZ_BCRQ01000071.1/7817-7946/-
NC_015589.1/1715918-1716038/-
NZ_JAEK01000007.1/19909-20025/-
NZ_KB898625.1/86190-86308/-
NZ_CP014140.1/18227481822873/+
NZ_LAKJ01000048.1/33883506/+
NZ_JYBN01000090.1/4396544089/+
NZ_BAZO01000041.1/43074423/+
NZ_LQYJ01000027.1/7844-7963/-
NC_014829.1/1047293-1047407/-
NZ_LBMQ01000098.1/1081962/+
NZ_LN831786.1/1662188-1662305/-
NZ_JRJU01000047.1/12276-12393/-
NZ_AYYI01000075.1/4653-4771/-
NZ_AYOS02000059.1/2062-2180/-
NZ_LFVV01000035.1/26642-26762/-
NZ_JHVL01000017.1/3472134841/+
NZ_AJLR01000115.1/1814618263/+
NZ_LAGD01000033.1/31314-31434/-
NZ_BAXQ01000037.1/1561-1679/-
NZ_ATAE01000011.1/18321-18438/-
NZ_KB889752.1/841920842036/+
3300000053_316411/291-408/-
NZ_JAEL01000019.1/32033-32151/-
NZ_LFPH01000002.1/9370793827/+
NZ_JH976434.1/58610-58722/-
NZ_ALPT02000018.1/44860-44977/-
NZ_LAGJ01000036.1/3586-3706/-
NZ_KE952849.1/4241-4362/-
NZ_LAGF01000025.1/18629-18749/-
NZ_DF384213.1/3211486-3211606/-
NZ_JPVQ01000004.1/2947929598/+

NZ_LITJ01000001.1/34793-34912/-
NZ_AZRS01000124.1/8708-8827/-
NZ_LFQE01000003.1/371580-371700/-
NZ_LAGH01000196.1/11024-11142/-
NC_017299.1/3023256-3023376/-
NZ_KB946251.1/854896-855011/-
NZ_ASWH01000002.1/531422-531536/-
NZ_BCPU01000001.1/130376-130490/-
NZ_BBPK01000014.1/31518-31631/-