UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY

Escuela de Ciencias Biológicas e Ingeniería

TÍTULO: Evolutionary dynamics of toxin-antitoxin system in *Ralstonia solanacearum*

Trabajo de integración curricular presentado como requisito para la obtención del título de Bióloga

Autor:

Maldonado Carrión Stephanie

Tutor:

Ph. D. José Antonio Castillo Morales

Urcuquí, marzo 2021

Urcuquí, 28 de mayo de 2021

SECRETARÍA GENERAL
(Vicerrectorado Académico/Cancillería)
ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA
CARRERA DE BIOLOGÍA
ACTA DE DEFENSA No. UITEY-BIO-2021-00005-AD

A los 28 días del mes de mayo de 2021, a las 09:30 horas, de manera virtual mediante videoconferencia, y ante el Tribunal Calificador, integrado por los docentes:

Presidente Tribunal de Defensa	Dr. ALVAREZ BOTAS, FRANCISCO JAVIER , Ph.D.
Miembro No Tutor	Dr. SANTIAGO VISPO, NELSON FRANCISCO , Ph.D.
Tutor	Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.

El(la) señor(ita) estudiante **MALDONADO CARRION, STEPHANIE** , con cédula de identidad No. **1719221440**, de la **ESCUELA DE CIENCIAS BIOLÓGICAS E INGENIERÍA**, de la Carrera de **BIOLOGÍA**, aprobada por el Consejo de Educación Superior (CES), mediante Resolución **RPC-SO-37-No.438-2014**, realiza a través de videoconferencia, la sustentación de su trabajo de titulación denominado: **EVOLUTIONARY DYNAMICS OF TOXIN ANTITOXIN SYSTEM IN RALSTONIA SOLANACEARUM**, previa a la obtención del título de **BIÓLOGO/A**.

El citado trabajo de titulación, fue debidamente aprobado por el(los) docente(s):

Tutor	Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.

Y recibió las observaciones de los otros miembros del Tribunal Calificador, las mismas que han sido incorporadas por el(la) estudiante.

Previamente cumplidos los requisitos legales y reglamentarios, el trabajo de titulación fue sustentado por el(la) estudiante y examinado por los miembros del Tribunal Calificador. Escuchada la sustentación del trabajo de titulación a través de videoconferencia, que integró la exposición de el(la) estudiante sobre el contenido de la misma y las preguntas formuladas por los miembros del Tribunal, se califica la sustentación del trabajo de titulación con las siguientes calificaciones:

Tipo	Docente	Calificación
Presidente Tribunal De Defensa	Dr. ALVAREZ BOTAS, FRANCISCO JAVIER , Ph.D.	10,0
Tutor	Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.	10,0
Miembro Tribunal De Defensa	Dr. SANTIAGO VISPO, NELSON FRANCISCO , Ph.D.	10,0

Lo que da un promedio de: **10 (Diez punto Cero)**, sobre 10 (diez), equivalente a: **APROBADO**

Para constancia de lo actuado, firman los miembros del Tribunal Calificador, el/la estudiante y el/la secretario ad-hoc.

Certifico que *en cumplimiento del Decreto Ejecutivo 1017 de 16 de marzo de 2020, la defensa de trabajo de titulación (o examen de grado modalidad teórico práctica) se realizó vía virtual, por lo que las firmas de los miembros del Tribunal de Defensa de Grado, constan en forma digital.*

MALDONADO CARRION, STEPHANIE
Estudiante

FRANCISCO JAVIER ALVAREZ BOTAS
Digitally signed by FRANCISCO JAVIER ALVAREZ BOTAS
Date: 2021.06.09 12:09:51 -05'00'

Dr. ALVAREZ BOTAS, FRANCISCO JAVIER , Ph.D.
Presidente Tribunal de Defensa

Dr. CASTILLO MORALES, JOSE ANTONIO , Ph.D.
Tutor

CERTIFICACIÓN ELECTRÓNICA
BANCO CENTRAL DEL ECUADOR
Firmado Digitalmente por: JOSE ANTONIO CASTILLO MORALES
Hora oficial Ecuador: 28/05/2021 15:49

NELSON FRANCISCO SANTIAGO VISPO

Firmado digitalmente por NELSON FRANCISCO SANTIAGO VISPO
Fecha: 2021.06.09 13:34:00 -05'00'

Dr. SANTIAGO VISPO, NELSON FRANCISCO , Ph.D.
Miembro No Tutor

KARLA ESTEFANIA ALARCON FELIX

Firmado digitalmente por KARLA ESTEFANIA ALARCON FELIX
Fecha: 2021.05.28 10:48:17 -05'00'

ALARCON FELIX, KARLA ESTEFANIA
Secretario Ad-hoc

AUTORÍA

Yo, **STEPHANIE MALDONADO CARRIÓN**, con cédula de identidad 1719221440, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así cómo, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autora (a) del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, marzo, 2021.

Stephanie Maldonado Carrión

CI: 1719221440

AUTORIZACIÓN DE PUBLICACIÓN

Yo, **STEPHANIE MALDONADO CARRIÓN,** con cédula de identidad 1719221440, cedo a la Universidad de Investigación de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior

Urcuquí, marzo, 2021.

Stephanie Maldonado Carrión

CI: 1719221440

Dedicatoria

A mis padres Miguel y Betty,

mis tíos Bemjamín y Nelly,

mi hermana Michelle,

y mi sobrino Nicolás;

pilares fundamentales en mi vida

y mi mayor motivación.

Stephanie Maldonado Carrión

RESUMEN

El complejo de especies *Ralstonia solanacearum* (RSSC) es un grupo de patógenos bacterianos de plantas transmitidos por el suelo que afectan los cultivos al causar marchitez bacteriana y eventualmente provocar la muerte de la planta, lo que resulta en pérdidas económicas significativas en todo el mundo. Debido a su vasta distribución geográfica, amplia gama de hospedadores y tremenda diversidad fenotípica y genética, ha sido un desafío encontrar formas adecuadas de proteger los cultivos contra esta bacteria. Los estudios se han centrado en las interacciones planta-bacteria mientras se buscan tratamientos viables. No obstante, los estudios sobre las interacciones bacterias-fagos podrían proponer soluciones novedosas contra estas bacterias persistentes. El sistema de defensa bacteriano llamado toxina-antitoxina (TA) participa en funciones bacterianas esenciales, como la protección contra el ataque de fagos, la resistencia al estrés mediante la formación de persistentes bacterianos y la formación de biopelículas. En este estudio, me enfoco en determinar el contenido genético y las fuerzas evolutivas que actúan sobre los genes TA. Se utilizaron BLASTn y BLASTp para determinar la existencia y diversidad de los sistemas TA en quince cepas del RSSC. Las tasas de ganancia, duplicación y pérdida de genes TA se calcularon utilizando COUNT. Además, las presiones selectivas positivas que actúan sobre los genes TA se analizaron utilizando BUSTED y MEME. Finalmente, los TA Pfam sujetos a transferencia horizontal de genes se identificaron utilizando NOTUNG para reconciliar árboles filogenéticos especie-gen. Los resultados sugieren que los sistemas de TA están ampliamente diversificados en todas las cepas de RSSC estudiadas. Además, se encontró selección positiva en un solo sitio de cinco genes TA que podría ser necesario para mantener la función del gen y causar fenotipos ventajosos para controlar las infecciones con fagos. Además, las fuerzas evolutivas que afectan a los sistemas de TA en RSSC son principalmente la duplicación de genes y la ganancia de genes. En consecuencia, se encontraron eventos HGT en todos los sistemas de TA analizados, lo que implica que los sistemas de TA son tanto ancestrales como de reciente obtención.

Palabras clave: RSSC, sistemas de toxina-antitoxina, selección positiva, tasas de ganancia-duplicación-pérdida de genes, transferencia horizontal de genes.

ABSTRACT

Ralstonia solanacearum species complex (RSSC) is a group of soil-borne bacterial plant pathogen that affects crops by causing bacterial wilt and eventually plant death, resulting in significant economic losses worldwide. Due to its vast geographical distribution, extensive host range, and tremendous phenotypic and genetic diversity, it has been challenging to find proper ways to protect crops against it. Studies have focused on plant-bacteria interactions while searching for viable treatments. Nonetheless, studies on bacteria-phage interactions could propose novel solutions against these persistent bacteria, because different phage therapy strategies can be applied. However, bacteria have numerous and diverse mechanisms to defend from phage damage. The bacterial defense system called toxin-antitoxin (TA) is involved in essential bacterial functions, such as protection from phage attack, stress resistance by forming bacterial persistors, and biofilm formation. In this study, I focus on determining the gene content and evolutionary forces acting on TA genes. BLASTn and BLASTp were used to determine TA systems' existence and diversity across fifteen strains from the RSSC. Gene gain, duplication, and loss rates of TA genes were calculated using COUNT. Moreover, positive selective pressures acting on TA genes were analyzed using BUSTED and MEME. Finally, TA Pfams subject to horizontal gene transfer were identified using NOTUNG to reconcile species-gene phylogenetic trees. Results suggest that TA systems are widely diversified across all studied RSSC strains. Besides, positive selection was found in a single site of five TA genes, which could be necessary to maintain gene function and cause advantageous phenotypes to control phage infections. Also, the evolutionary forces that affect TA systems in RSSC are mainly gene duplication and gene gain. Accordingly, in all the analyzed TA systems HGT events were found, implying that TA systems are both ancestral and recently obtained.

Keywords: RSSC, Toxin-antitoxin systems, positive selection, gene gain-duplication-loss rates, horizontal gene transfer

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

ABBREVIATIONS

RSSC	*Ralstonia solanacearum* Species Complex
TA	Toxin-Antitoxin
EPS	Extracellular polysaccharides
Pfam	Protein family
HGT	Horizontal gene transfer
CRISPR	Clustered regularly interspaced short palindromic repeats
R-M	Restriction-modification
NCBI	National Center for Biotechnology Information
BLAST	Basic Local Alignment Search Tool
DTL	Duplication-transfer-loss

CHAPTER 1. INTRODUCTION

1.1. *Ralstonia solanacearum* species complex

R. solanacearum species complex (RSSC) is a group of gram-negative bacteria found in soil. It is responsible for causing bacterial wilt disease in more than 200 plant species, including important crops from the Solanaceae family (Leonard, Hommais, Nasser, & Reverchon, 2017). Its vast geographical distribution, followed by an extensive host range, and tremendous phenotypic and genetic diversity have made it difficult to find proper ways to protect crops against it (Genin & Boucher, 2002). Moreover, according to a survey in 2012, RSSC is considered the second most crucial bacterial plant pathogen because it affects a wide range of crops, causing high economic impacts worldwide (Mansfield et al., 2012). Even if the actual economic impact is unknown, a study calculated that in potato crops, RSSC causes US$1 billion in losses each year globally (Mansfield et al., 2012). This could be significantly harmful to developing countries like Ecuador, whose economy relies mainly on agricultural practices (Pacheco, Ochoa-Moreno, Ordoñez, & Izquierdo-Montoya, 2018).

Based on an analysis of 140 strains sampled globally, RSSC has been classified into four phylotypes or genetic groups related to their geographic origin (Castillo & Greenberg, 2007; Peeters, Guidot, Vailleau, & Valls, 2013). Phylotype I includes strains from Asia, phylotype II, which could be further divided into IIA and IIB (Castillo & Greenberg, 2007), those from America, phylotype III strains from Africa, and phylotype IV, those from Indonesia, Japan, and Australia (Peeters et al., 2013). Additionally, RSSC can be taxonomically classified into three different species: *R. pseudosolanacearum* which includes phylotypes I and III, *R. solanacearum* which refers to phylotype II, and *R. syzygii* which refers to phylotype IV (Safni, Cleenwerck, De Vos, Fegan, Sly, & Kappler, 2014).

Bacteria of RSSC invade the plant's xylem vessels through root wounds, areas of secondary root emergence, and aerial transmission by insects (Genin & Boucher, 2002; Leonard et al., 2017). Then, they produce large quantities of extracellular polysaccharide (EPS), blocking water movement and resulting in wilting and plant death (Genin & Boucher, 2002; Leonard et al., 2017).

The ability to live in soil, plants, and insects, puts RSSC at an increased risk of phage attack (Castillo, Secaira-Morocho, Maldonado, & Sarmiento, 2020). Indeed, multiple phages, many of which are lytic, belonging to four viral families: Inoviridae, Podoviridae, Myoviridae, and Siphoviridae, infect this pathogen (Castillo et al., 2020).

1.2. Bacteria-phage interactions

Bacteriophages, also known as phages, are viruses that infect bacteria and use their resources to replicate (Monk, Rees, Barrow, Hagens, & Harper, 2010). Although bacteriophages were discovered in 1915, nowadays, 13 families have been described (Ackermann, 2003). Furthermore, phages can be found in a wide range of hosts (Monk et al., 2010); accordingly, they have been discovered in over 140 bacteria genera so far (Ackermann, 2003).

Bacteriophages and their hosts are a genetically large and diverse group of organisms (Weitz et al., 2013). Research suggests that phage-bacteria interactions play an essential role in microbial ecology and bacterial genomes' evolution (Chibani-Chennoufi, Bruttin, Dillmann, & Brüssow, 2004). Bacteria have evolved multiple tactics to prevent phage infection; nevertheless, phages have also evolved to counter-attack bacterial defense systems (Labrie, Samson, & Moineau, 2010). Next, I will describe bacterial defense systems and the mechanism phages use to avoid them.

First, bacteria can prevent phage adsorption in multiple ways. Such as by adapting the cell's surface to block phage receptors, producing an extracellular matrix to obstruct receptors, and producing competitive inhibitors (Dy, Richter, Salmond, & Fineran, 2014; Labrie et al., 2010). Still, phages have evolved to avoid some of these defensive mechanisms, like recognition and degradation of extracellular polymers or by altering their tail fibers to recognize altered receptors (Dy, Richter, et al., 2014; Labrie et al., 2010). Bacteria are also able to avoid phage DNA entry. For instance, bacteria have *Sie, Im, and sp systems*, which cause rapid inhibition of DNA injection into cells (Dy, Richter, et al., 2014; Labrie et al., 2010). Bacteria can also cut phage nucleic acids by using CRISPR-Cas and restriction-modification (R-M) systems (Dy, Richter, et al., 2014; Labrie et al., 2010). R-M systems are the bacterial innate immune system against phages (Dy, Richter, et al., 2014). R-M systems protect the cell by using a *methyltransferase* to add a methyl group to specific bacterial sequences that need to be conserved and a restriction endonuclease that cleaves nonmethylated foreign sequences (Dy, Richter, et al., 2014; Labrie et al., 2010). To avoid R-M systems, phages have evolved through point mutations to eliminate endonuclease recognition sites

in their genome (Labrie et al., 2010). On the other hand, CRISPR-Cas systems confer phage resistance by aiding in recognizing and degrading the phage genome, helping gain immunity (Dy, Richter, et al., 2014). Phages can avoid CRISPR-Cas immunity by evolving anti-CRISPR genes by point mutation (Dy, Richter, et al., 2014; Labrie et al., 2010).

Lastly, bacteria also use abortive infection (Abi) systems that provide resistance to phage attacks by leading to the infective cell's death (Labrie et al., 2010). Abi proteins are activated by phages and later inhibit essential cellular processes, resulting in cell death (Dy, Richter, et al., 2014). Toxin-antitoxin systems also result in altruistic cell death and will be discussed further below.

1.3. Toxin-antitoxin (TA) systems

Toxin-antitoxin systems can be found in the majority of bacterial genomes. TA systems consist of a toxin which is primarily a protein, and its related antitoxin, which could either be RNA or protein that counteracts the toxic activity (Guglielmini & Van Melderen, 2011; Song & Wood, 2020; Unterholzner, Poppenberger, & Rozhon, 2013). Toxins act by interfering with vital cellular processes like translation, replication, cytoskeleton formation, membrane, and cell wall biosynthesis, leading to cell dormancy and persistence (Guglielmini & Van Melderen, 2011; Song & Wood, 2020; Unterholzner et al., 2013). However, they do not typically result in cell death while maintaining toxin-antitoxin equilibrium (Dy, Richter, et al., 2014; Song & Wood, 2020; Unterholzner et al., 2013). Antitoxins neutralize toxins; however, they are highly unstable. For this reason, antitoxin activity could easily be affected by phage attacks, which would result in a toxin takeover in the cell, leading to cell death and consequently phage destruction (Dy, Richter, et al., 2014).

TA systems can be classified into five TA classes related to their genetic structure and antitoxin type (Unterholzner et al., 2013). **Type I TA systems** have unstable antisense sRNAs antitoxins that inhibits translation of the toxin mRNA (Unterholzner et al., 2013). In **Type II TA systems,** both toxin and antitoxin are small proteins (Unterholzner et al., 2013). When type II toxins are freed from antitoxin activity, bacterial cell growth is inhibited, aiding in cell persistence and survival (Guglielmini & Van Melderen, 2011). **Type III TA systems** have sRNA antitoxins that neutralize the toxin by binding directly to it; nevertheless, when toxins are freed, they lead to cell death (Dy, Richter, et al., 2014; Unterholzner et al., 2013). **Type V TA systems** have an antitoxin

that cleaves toxin mRNA, preventing toxin translation (Unterholzner et al., 2013). It neutralizes toxicity by interacting with toxin targets and catalyzing opposing reactions in the cell (Dy, Richter, et al., 2014). **Type V TA systems** have an antitoxin that cleaves toxin mRNA, preventing toxin translation (Unterholzner et al., 2013).

Ultimately, TA systems include several adaptive functions such as phage protection, stress resistance by forming bacterial persistors, and biofilm formation (Unterholzner et al., 2013; Van Melderen, 2010).

1.4. Positive selection in bacterial genomes

Natural selection is one of the most powerful forces driving the evolution of living organisms on Earth. Researchers have stated that natural selection, both positive and negative, has been universal in most genomes (Booker, Jackson, & Keightley, 2017). Moreover, a positive selection model is commonly used to explain genetic variations and diversity in genomes (Booker et al., 2017). While negative selection is considered a purifying process, in which the spread of disadvantageous alleles is prevented, positive selection encourages the spread of new and advantageous alleles in an organism (Zhang, 2008).

Positive selection in bacteria results in higher substitution rates, which aid in adaptive changes in function to adjust to their environment (Petersen, Bollback, Dimmic, Hubisz, & Nielsen, 2007). Studies have found that positive selection is involved in both ionizing-radiation (Sghaier, Ghedira, Benkahla, & Barkallah, 2008) and antibiotic resistance (Farhat et al., 2013); and in horizontal gene transfer (Petersen et al., 2007).

Furthermore, genome-wide studies have discovered that positive selection is strictly involved in host-pathogen dynamics, including defense system genes (Djordjevic, O'Sullivan, Walker, Conkling, & Klaenhammer, 1997; Petersen et al., 2007).

1.5. Horizontal gene transfer (HGT)

Horizontal gene transfer occurs when foreign DNA from a different organism is inserted into a microbial genome (Gyles & Boerlin, 2014a). Mobile genetic elements (MGE) are foreign DNA inserted into the host's genome, mainly by three mechanisms (Gyles & Boerlin, 2014b).

First, conjugation in which donor and recipient bacteria must be in physical contact and genetic material is transferred by a pilus (Soucy, Huang, & Gogarten, 2015). Next, transduction, in which DNA is inserted into the bacterial genome by bacteriophages, resulting in either the activation of lysogenic or lytic cycles (Gyles & Boerlin, 2014a). Finally, transformation, where foreign DNA passes through the cell wall and membrane and later enters the bacterial chromosome by homologous recombination (Gyles & Boerlin, 2014a; Soucy et al., 2015)

Studies have shown that multiple traits have been acquired through HGT in bacterial genomes, including antibiotic resistance, virulent factors, adherence to host cells, manipulation of host signal transduction, and toxin production (Gyles & Boerlin, 2014a; Ochman, Lawrence, & Grolsman, 2000).

1.6. Gene content (gene gain, loss, duplication)

The gene content in the genomes of microbes refers to the number of genes that conform the genome. The gene content of a genome is determined by evolutionary history, function selection, and genome size (Snel, Bork, & Huynen, 2002). By studying reconstructed ancestral genomes and present genomes, it is possible to infer how evolutionary processes like gene loss, duplication, and gene gain have shaped their gene content over time (Snel et al., 2002).

Gene duplication is essential for acquiring new genes and creating genetic novelty (Magadum, Banerjee, Murugan, Gangapur, & Ravikesavan, 2013). In bacteria, studies suggest that gene duplication resulted in the appearance of protein families during evolution (Serres, Kerr, McCormack, & Riley, 2009).

Gene loss occurs when coding sequences with no functional value are lost over time (D'Souza et al., 2014). In bacteria, gene loss may result in decreased energy expenditure on DNA, RNA, and proteins, which would result in higher fitness by increasing bacterial growth rates (Koskiniemi, Sun, Berg, & Andersson, 2012).

Gene gain can occur by either gene genesis (origin of a new gene) or HGT (Snel et al., 2002). The appearance of new genes can result in new bacterial functions that ensure adaptation to environmental changes (Li et al., 2017).

1.7. Pfam system

Pfam is an extensive online database that contains detailed information of more than 13 000 protein families (Punta et al., 2012). Pfam is used by experimental, computational, and evolutionary biologists while researching proteins, organizing protein sequences, and studying the origins and evolution of proteins (Finn et al., 2009).

All Pfam families on the database are identified by Markov model strategies, which can find occurrences of Pfams while searching against a protein sequence (Punta et al., 2012). Results show homologous proteins, which are evolutionary related and probably share structural and functional characteristics (Punta et al., 2012).

1.8. Problem statement

Different strains of RSSC are widely studied due to the economic footprint they leave behind each year on agricultural production (Mansfield et al., 2012). Factors like wide geographic distribution, extensive host range, and genomic diversity (Genin & Boucher, 2002) have made them complicated to find feasible control methods so far (Yuliar, Nion, & Toyota, 2015).

Most *in silico* studies seeking for effective control of bacterial wilt caused by RSSC focus on plant-pathogen interactions (Z. G. Li, He, Zhang, & Peng, 2012; Yao & Allen, 2007; Yuliar et al., 2015). Their results may be used to create effective antibacterial drugs by having a better understanding of RSSC pathogenesis. However, antibiotic resistance and the rise of consumers' preference for antibiotic-free produce may require the use of different control approaches like phage therapy. A better understanding of TA systems present in RSSC could help select the appropriate phages to avoid resistance.

My research will analyze RSSC's TA systems, a bacterial defense mechanism that is constantly evolving to overcome phage attacks. By analyzing the evolutionary dynamics of TA systems in RSSC, I will have a better understanding of how these genes have expanded across RSSC, the evolutionary forces responsible for this expansion, and their role in the bacteria's physiology. With these results, I could aid in proposing novel therapies to treat these persistent bacteria.

1.9. General and specific objectives

1.9.1. General objective

Describe the diversity and evolutionary dynamics of Toxin-antitoxin systems among fifteen representative strains of all the phylogenetic diversity of RSSC.

1.9.2. Specific objectives

- Report the presence of TA genes and their respective Pfams across the studied strains of RSSC using BLASTn and BLASTp
- Analyze positive selection pressure for TA genes in RSSC strains using BUSTED and MEME algorithms.
- Compare gene gain, loss, and duplication rates for TA systems in RSSC using COUNT software.
- Analyze HGT events of TA Pfams in RSSC using NOTUNG program.

CHAPTER 2. METHODS

2.1. Sequence data and determining the presence of TA defense system in *Ralstonia solanacearum* species complex (RSSC)

Considering that to date (June 2019) RSSC's phylotype III only had three available genomic sequences in NCBI's FTP server, only three genomic sequences from each remaining phylotype (I, IIA, IIB, and IV) were used to work with an equal number of sequences in further studies (Castillo et al., 2020). Ultimately, I selected the best 15 genomic sequences of RSSC (both megaplasmid and chromosome when available) available in NCBI's FTP server and downloaded them in FASTA format (see Table 1. Proteomes for the 15 RSSC mentioned above were also downloaded from UniProt (Bateman et al., 2021), in which protein names, loci, and Pfams were included for future analyses.

Table 1. Genomes of 15 RSSC retrieved from NCBIs FTP server

Phylotype	**Strain**	**NCBI Accession Number**
I	GMI1000	NC_003295/NC_003296
	FQY_4	NC_021745/NC_020799
	HA4-1	NZ_CP022481/NZ_CP022482
IIA	CFBP2957	NC_014307
	RS489	NZ_CP021766/NZ_CP021767
	CIP120	NZ_JXAY01000001
IIB	Po82	NC_017574/NC_017575
	UW551	NZ_AAKL01000577
	IBSBF1503	NZ_CP012943/NZ_CP012944
III	CMR15	NC_017559/NC_017589

	CFBP3059	NZ_JXBA01000001
	UW386	NZ_CP039339/NZ_CP039340
IV	T98	NZ_CP022759/NZ_CP022760
	PSI07	NC_014310/NC_014311
	KACC10722	NZ_CP014702/NZ_CP014703

Using TextPad (TextPad 8.2.0 (64-bit Edition) ("TextPad - the text editor for Windows," n.d.), genomes were searched for keywords such as: "toxin", "antitoxin", "antidote", and "addiction". For a TA pair to be annotated in our TA database, toxin and antitoxin genes had to be found consecutively in the genome (Gupta, Venkataraman, Vasudevan, & Gopinath Bankar, 2017). Besides, Pfams belonging to confirmed TA systems from our previous analysis were searched for along the proteomes of our RSSC, and, after confirming TA pairs by locating their loci in the genomes, they were added to our database.

Furthermore, TA systems (genes and Pfams) present in strain GMI1000 (reference genome) were identified (see Table 2 and Table 3) (Salanoubat et al., 2002). Next, NCBI's BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul, Gish, Miller, Myers, & Lipman, 1990; Camacho et al., 2009) was used to search for TA protein/gene homologs in all three species (*Ralstonia solanacearum*, *Ralstonia pseudosolanacearum*, *Ralstonia syzygii*) that make up the RSSC. Aligned search results with >95% identity, > 40-bit score, and <0.001 E-value (Pearson, 2013) belonging to the 14 remaining strains' genomes were downloaded in FASTA format and added to our database.

Table 2. TA genes identified in RSSC and their respective locus tag in the genome of model organism strain GMI1000

TA GENES	**Locus tag**	**References**
AbiEii toxin	RSp0266	(Salanoubat et al., 2002)
AbiEii antitoxin	RSp0267	(Salanoubat et al., 2002)
HipA toxin	RSc1446	(Salanoubat et al., 2002)
HipB antitoxin	RSc1447	(Salanoubat et al., 2002)
HicA toxin	RS08575	(Salanoubat et al., 2002)
HicB antitoxin	RS08570	(Salanoubat et al., 2002)
RatA toxin	RSc1425	(Salanoubat et al., 2002)
RatB antitoxin	RSc1426	(Salanoubat et al., 2002)
RelE/ParE toxin	RS02441	(Salanoubat et al., 2002)
ParD antitoxin	RS02442	(Salanoubat et al., 2002)

VapC toxin	RSc0871	(Salanoubat et al., 2002)
MazE antitoxin	RSc3056	(Salanoubat et al., 2002)
Phd/YefM antitoxin	RSc0872	(Salanoubat et al., 2002)

Table 3. TA Pfams identified in RSSC and their respective gene locus tag in the genome of the model organism strain GMI1000

Pfam	**Locus Tag**	**References**
PF01381	RSp0076	(Salanoubat et al., 2002)
PF01850	RSc0871	(Salanoubat et al., 2002)
PF02604	RSc0872	(Salanoubat et al., 2002)
PF03364	RSc1425	(Salanoubat et al., 2002)
PF03658	RS05266	(Salanoubat et al., 2002)
PF03693	RS02442	(Salanoubat et al., 2002)
PF04014	RSc3056	(Salanoubat et al., 2002)
PF05016	RSc0263	(Salanoubat et al., 2002)
PF05534	RSc1697	(Salanoubat et al., 2002)
PF05973	RSp0077	(Salanoubat et al., 2002)
PF06296	RSc0124	(Salanoubat et al., 2002)

PF08843	RSp0267	(Salanoubat et al., 2002)
PF11459	RSp0202	(Salanoubat et al., 2002)
PF13657	RSc1446	(Salanoubat et al., 2002)

Ultimately, two different databases were created for all subsequent studies. The first one consisted of gene datasets for each TA gene sequence in each RSSC strain genome. The other database consisted of a dataset of amino acid sequences for individual TA Pfams present in considered RSSC strains. A summary table for the presence or absence of TA is shown in **A**PPENDIX 1.

2.2. Selection pressure

Nucleotide and amino acid datasets were independently aligned using the MAFFT server (https://mafft.cbrc.jp/alignment/server/) (Katoh, Rozewicki, & Yamada, 2019), the 'Adjust direction according to the first sequence (accurate enough for most cases)' option was selected in case the downloaded nucleotide sequences were complementary strands. For this, I had to be sure that the first sequence was correct. In most cases, sequences belonging to strain GMI1000 were used. All other settings were left as default.

For the next part of the analysis, all stop codons from the gene datasets were removed manually using TextPad (TextPad 8.2.0 (64-bit Edition) ("TextPad - the text editor for Windows," n.d.). Datamonkey Adaptive Evolution Server (https://www.datamonkey.org/), which has multiple bioinformatic methods and tools, was used to detect selection. Specifically, BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification), which provides a gene-wide test for positive selection in at least one branch at the gene level (Murrell et al., 2015) and MEME (Mixed Effects Model of Evolution), which detects sites under positive selection utilizing a mixed-effects maximum likelihood approach (Murrell et al., 2012). Before the analysis, all phylogeny branches were selected in BUSTED, and results with a p-value <0.05 were considered significant. All TA gene alignments with more than two unique sequences were analyzed.

2.3. Horizontal gene transfer (HGT) events

Notung (Notung v2.9.1.2) software is used to reconcile a species tree with a gene tree utilizing the parsimony principle to infer gene duplication, transfer, and loss (DTL) events in phylogenic analysis. In order to work, Notung requires both trees to be rooted to reconcile them (Darby, Stolzer, Ropp, Barker, & Durand, 2017; Durand, Halldórsson, & Vernot, 2005; Lai, Stolzer, & Durand, 2017; Stolzer et al., 2012; Vernot, Stolzer, Goldman, & Durand, 2008). For executing the analysis in Notung, the species tree and gene tree must be uploaded. Then, I select the "Prefix of the gene label (i.e., SPECIESGENE)" option in the reconcile tab to obtain results.

The species tree for this analysis was provided by Dr. J.A. Castillo (Castillo et al., 2020), who used the Bayesian Evolutionary Analysis Sampling Tree (BEAST) v.1.10.4 (Suchard et al., 2018) coupled with BEAGLE library v2.1 for accelerated calculations (Suchard et al., 2018). Ultimately, an aligned amino acid sequence corresponding to the fifteen RSSC strains' core genomes was analyzed. The best model selection was estimated using SMS software (http://www.atgc-montpellier.fr/phyml/) (Guindon et al., n.d.). The following parameters were used in BEAUti: strict clock, constant growth for tree prior, (α) parameter of the gamma distribution, lognormal distribution with mu (μ) equal to 0.5, 25 million iterations, and 2000 as echo state. Tracer v1.7.1 (Rambaut, Drummond, Xie, Baele, & Suchard, 2018) was utilized to analyze the convergence of the Markov Chain Monte Carlo (MCMC) by evaluating the Effective Sample Size (ESS). Lastly, the maximum clade credibility tree was obtained using TreeAnnotator v1.10.4.

Next, aligned amino acid sequences from TA datasets were used for obtaining the gene trees. These followed the same processes as the species tree mentioned above, with the following parameters in BEAUti: JTT substitution model, Gamma+Invariant site heterogeneity model, four gamma categories, strict clock, coalescent-constant tree prior size, Lognormal (α) parameter with an initial mu(μ) equal to 0.5, 20 million iterations, and 1000 as echo state. Convergence was verified using Tracer v1.7.1 (Rambaut et al., 2018), with an ESS value over 200 for all parameters. TreeAnnotator v.1.10.4. generated the Maximum clade credibility tree as a TREE file. Gene trees were obtained for all Pfam alignments that contained more than five sequences (See Table 4).

Table 4. TA Pfams used for HGT events in NOTUNG, with related TA gene description and number of used sequences.

Pfam (PF)	**Related TA gene**	**Number of sequences**
01381	HigA Antitoxin	15
01850	VapC Toxin	11
02604	Phd_YefM Antitoxin	14
03364	RatA Toxin	15
03658	RatB Antitoxin	15
03693	ParD Antitoxin	8
04014	MazEF Antitoxin	10
04221	RelB antitoxin	4*
05015	HigB toxin	2*
05016	RelE/ParE Toxin	14
05973	RelE/ParE Toxin	14
06296	RelE/ParE Toxin	9
06414	Zeta Toxin	10

08843	AbiEii/AbiGii Toxin	13
13560	HipB Antitoxin	13
13657	HipA Toxin	15
15738	YafQ Toxin	5
05534	HicB Antitoxin	3*
07927	HicA Toxin	3*
08845	SymE Toxin	3*
11459	AbiEi Antitoxin	4*
11663	YhaV Toxin	3*
15937	PrlF Antitoxin	3*

* indicates Pfams with less than 5 sequences that were not analyzed.

2.4. Gene gain, loss and duplication rates

COUNT v10.04 (Csuos, 2010) was used to infer protein family rates of gain, loss, and duplication using a phylogenetic birth-and-death model included in the software. COUNT requires a presence/absence table of all TA proteins in .TXT format (see table 2.4) and a rooted species tree in .NWK format (mentioned above), both provided by Dr. J.A. Castillo (Castillo et al., 2020).

While performing rate optimization, the following parameters were selected: Gain-loss-duplication model type, Poisson family size distribution at root, lineage-specific variations: gain-loss/duplication-loss ratios same in all lineages, and four gamma categories for maximum gamma variation. Furthermore, convergence criteria consisted of a maximum number of optimization rounds of 100, with a 0.1 convergence threshold on the likelihood.

CHAPTER 3. RESULTS

3.1. Presence of toxin-antitoxin proteins in 15 RSSC strains

Toxin-antitoxin systems, which are encoded by contiguous genes, are commonly found across the majority of bacterial genomes (Goeders & Van Melderen, 2014). I was able to identify twenty-eight protein families (Pfams) and twelve TA gene pairs across the fifteen RSSC strains analyzed, using the methods mentioned above.. Appendix 1 shows the relation of all TA systems found in the 15 RSSC strains.

Overall, the highest presence of TA systems was identified in phylotype III with 42 TA systems in total. The two strains with the highest number of TA systems were CMR15 and UW386, both from phylotype III, with 15 TA systems each. Contrarily, the lowest presence of TA systems was found in phylotype IIA with 31 TA systems. Moreover, the strains with the lowest number of TA systems were HA4-1, CIP120, and T98, with 9 TA systems each.

Furthermore, the TA systems with the highest presence among the 15 strains of RSSC were: HipA toxin, RelE/ParE toxin, RatA toxin, and RatB antitoxin, which were found in all of them. On the contrary, the one with the lowest presence was Epsilon antitoxin which was only found in strain CMR15.

Lastly, as seen in Table 4, the Pfams with the highest presence among the 15 RSSC strains are PF01381, PF03364, PF03658, and PF13657, which were found in all of them. Alternatively, the Pfams with the slightest presence was PF05015 which was only found in 2 RSSC strains.

3.2. Selection pressure

BUSTED is a branch-site unrestricted statistical test for episodic diversification that identifies gene-wide (not site-specific) evidence of positive selection. It can be applied on either the entire phylogeny or a specific subset of branches, in which all branches are treated as foreground branches (Murrell et al., 2015). Unlike other tests where ω (nonsynonymous versus synonymous substitutions, dN/dS, ratio) is averaged over branches or codon sites (Kosakovsky Pond et al., 2011; Murrell et al., 2013; Muse & Gaut, 1994; Yang, Nielsen, Goldman, & Pedersen,

2000), BUSTED can more accurately detect localized selection events by allowing ω to vary from branch to branch (Murrell et al., 2015). This model indicates whether the positive selection is present in at least one foreground branch at a given moment in time, ignoring results in other background or foreground branches (Murrell et al., 2015).

MEME is a mixed-effects model of evolution which detects positive selection at individual sites for pervasive and episodic selection (Murrell et al., 2012; Spielman et al., 2019). MEME allows ω to vary from site to site and from branch to branch at a site, making it more powerful than existing methods (Murrell et al., 2012).

After applying BUSTED and MEME to the TA gene sequences, results with a *p*-value <0.05 were considered significant for selective pressure at either gene-wide level or individual sites and are shown in Table 5. It is necessary to highlight that two antitoxins PrlF and Epsilon, were not analyzed with either BUSTED or MEME because there were not enough unique sequences (more than two different sequences).

Consequently, according to BUSTED, there were no entire TA genes under positive selective pressure since none of them obtained a *p*-value under 0.05. Whereas, MEME results indicated individual sites with positive selection in AbiEi antitoxin at 1 site, HipA toxin at 1 site, RelE/ParE toxin at 1 site, ParD antitoxin at 1 site, and Zeta toxin at 1 site as well.

Table 5. Summary statistics from BUSTED and MEME showing evidence of positive selection in TA genes belonging to RSSC

GENE	**FUNCTION**	**BUSTED[1]**	**MEME[2]**
AbiEii	T	0.249	0
AbiEi	A	0.301	1
HipA	T	0.480	1
HipB	A	0.500	0
HicA	T	0.431	0

HicB	A	0.500	0
RatA	T	0.500	0
RatB	A	0.500	0
RelE/ParE	T	0.500	1
RelE/ParE	A	0.500	0
ParD	A	0.083	1
SymE	T	0.500	0
SymR	A	0.500	0
VapC	T	0.500	0
MazE	A	0.339	0
Phd/YefM	A	0.500	0
YafQ	T	0.500	0
RelB/DinJ	A	0.500	0
YhaV	T	0.250	0
PrlF	A	-	-
Zeta	T	0.500	1
Epsilon	A	-	-

[1]*p*-value provided by BUSTED. A *p*-value < 0.05 indicates evidence of positive selection of the gene
[2]Number of significant sites under positive selection provided by MEME, significant at a *p*-value < 0.05

3.3. Horizontal gene transfer

Vertical gene transfer is considered to be the dominant form of genetic transmission among organisms; however, studies have proven that horizontal gene transfer is responsible for radical adaptations in eukaryotes and prokaryotes (Zhaxybayeva & Doolittle, 2011). There is a better

comprehension of HGT events in prokaryotes, proving that such events lead to essential functions such as antibiotic resistance, virulence, and defense mechanisms (Andersson, 2009; Gyles & Boerlin, 2014a; Zhaxybayeva & Doolittle, 2011).

NOTUNG applies the duplication-transfer-loss-reconciliation model on species trees and gene trees to discover HGT events by inferring the most likely parsimonious events (Bansal, Wu, Alm, & Kellis, 2015; Chen, Durand, & Farach-Colton, 2000). Species trees denote the evolutionary history of organisms, and gene trees help understand the evolution of a specific gene family (Bansal et al., 2015).

For HGT events to be considered, donors and recipients of a transfer had to co-exist at a given time; therefore, Notung analyzes all temporarily feasible events. (Darby et al., 2017; Durand et al., 2005; Lai et al., 2017; Stolzer et al., 2012; Vernot et al., 2008). In Notung, green circles at nodes represent alternative optimal event histories. Also, yellow lines with a T represent an HGT event, while the yellow triangle implies direction. Red boxes and Ds represent duplication events. Finally, losses are shown in grey font.

Reconciled phylogenetic trees for each Pfam with evidence of HGT are shown in Figure 1. It is important to emphasize that green circles represent multiple solutions; only one of them is shown below. Furthermore, according to NOTUNG, there were no feasible solutions for PF01381 and PF13560.

Figure 1a shows that in PF01850 there were 6 possible HGT events: between the clade 'n11' (CFBP3059) and the clade 'n16' (HA4-1, FQY_4); between the clade 'n13' (HA4-1, FQY_4, CFBP3059, CMR15) and the clade 'n4' (UW551, RS489, Po82); between the strains UW551 and RS489; between the clade 'n18' (GMI1000) and the strain PSI07; between the clade 'n18' (PSI07, GMI1000) and the strain IBSBF1503; between the strains IBSBF1503 and UW386. In addition, there were 3 possible losses: strain UW386 from clade ‘n11’ (CFBP3059); strain IBSBF1503 from clade ‘n2’ (Po82); and clade ‘n16’ (HA4-1, FQY_4) from clade ‘n18’ (GMI1000).

Figure 1b indicates that in PF02604 there were 5 possible HGT events: between clade ‘n15’ (KACC10722, T98) and clade ‘n25’ (CFBP2957, CIP120); between clade ‘n15’ (CFBP2957, CIP120, KACC10722, T98) and strain FQY_4; between the strains FQY_4 and CFBP3059;

between clade ‘n9’ (CMR15, UW386) and clade ‘n25’ (IBSBF1503, Po82, RS489); between the strains GMI1000 and PSI07. In addition, there were 7 possible losses: strain RS489 from clade ‘n24’ (CFBP2957); clade ‘n19’ (IBSBF1503, Po82) from clade ‘n21’ (CIP120); strain PSI07 from clade ‘n13’ (T98); strain CFBP3059 from clade ‘n7’(UW386); strain CIP120 from clade ‘n21’ (IBSBF1503, Po82); strain CFBP2957 from clade ‘n24’ (RS489); and strain FQY_4 from clade ‘n2’ (HA4-1).

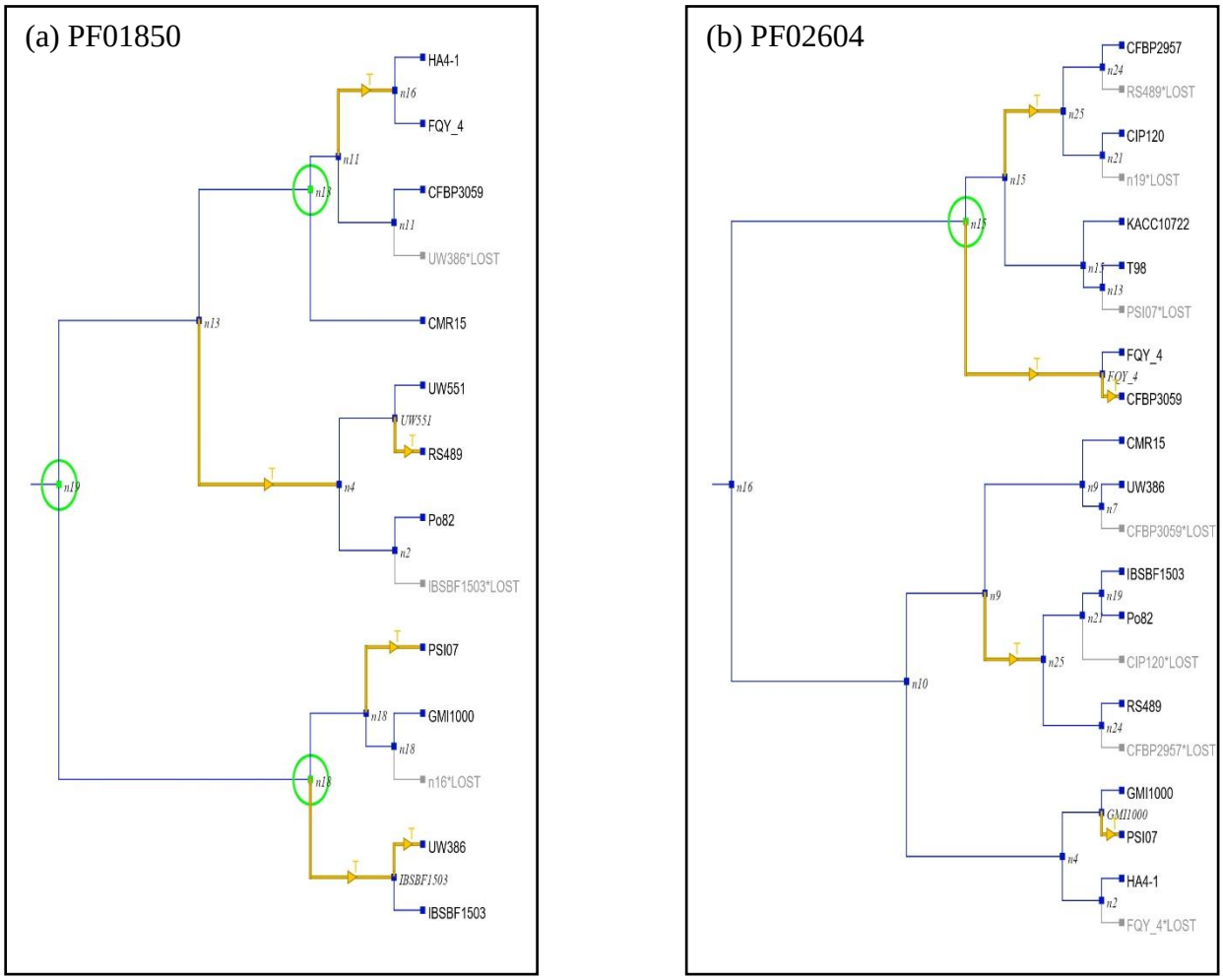

Figure 1. Reconciled phylogenetic trees between Pfam PF01850 and PF02604 and species trees, displaying HGT, duplication, and loss events across 15 RSSC strains by Notung. HGT events are colored yellow, duplication events are red, and losses are colored grey

Figure 2a shows, that PF03364 has 6 possible HGT events: between the strains CIP120 and CFBP2957; between the strains PSI07 and KACC10722; from clade ‘n2’ (PSI07, KACC10722, T98) to clade ‘n21’ (GMI1000, UW386, CFBP3059, FQY_4, HA4-1, CMR15); from clade ‘n19’

(GMI1000, UW386, CFBP3059, FQY_4) to strain HA4-1; between the strains UW386 and GMI1000; between the strains CFBP3059 and FQY_4. Also, there were 2 possible losses: strain CFBP2957 from clade ‘n7’ (RS489) and strain KACC10722 from clade ‘n4’ (GMI1000, UW386, CFBP3059, FQY_4, HA4-1, CMR15, PSI07, KACC10722, T98).

Figure 2b shows that in PF03658 there were 4 possible HGT events: between clade ‘n4’ (T98, PSI07, KACC10722) and clade ‘n26’ (CFBP3059, UW386, HA4-1, FQY_4, GMI1000, CMR15); between clade ‘n24’ (CFBP3059, UW386) and clade ‘n21’ (HA4-1, FQY_4, GMI1000); between the strains UW551 and RS489; and between clade ‘n9’ (CIP120, CFBP2957) and strain UW551. Furthermore, there were 2 possible losses: strain RS489 from clade 'n7' (CFBP2957) and strain UW551 from clade 'n14' (IBSBF1503, Po82).

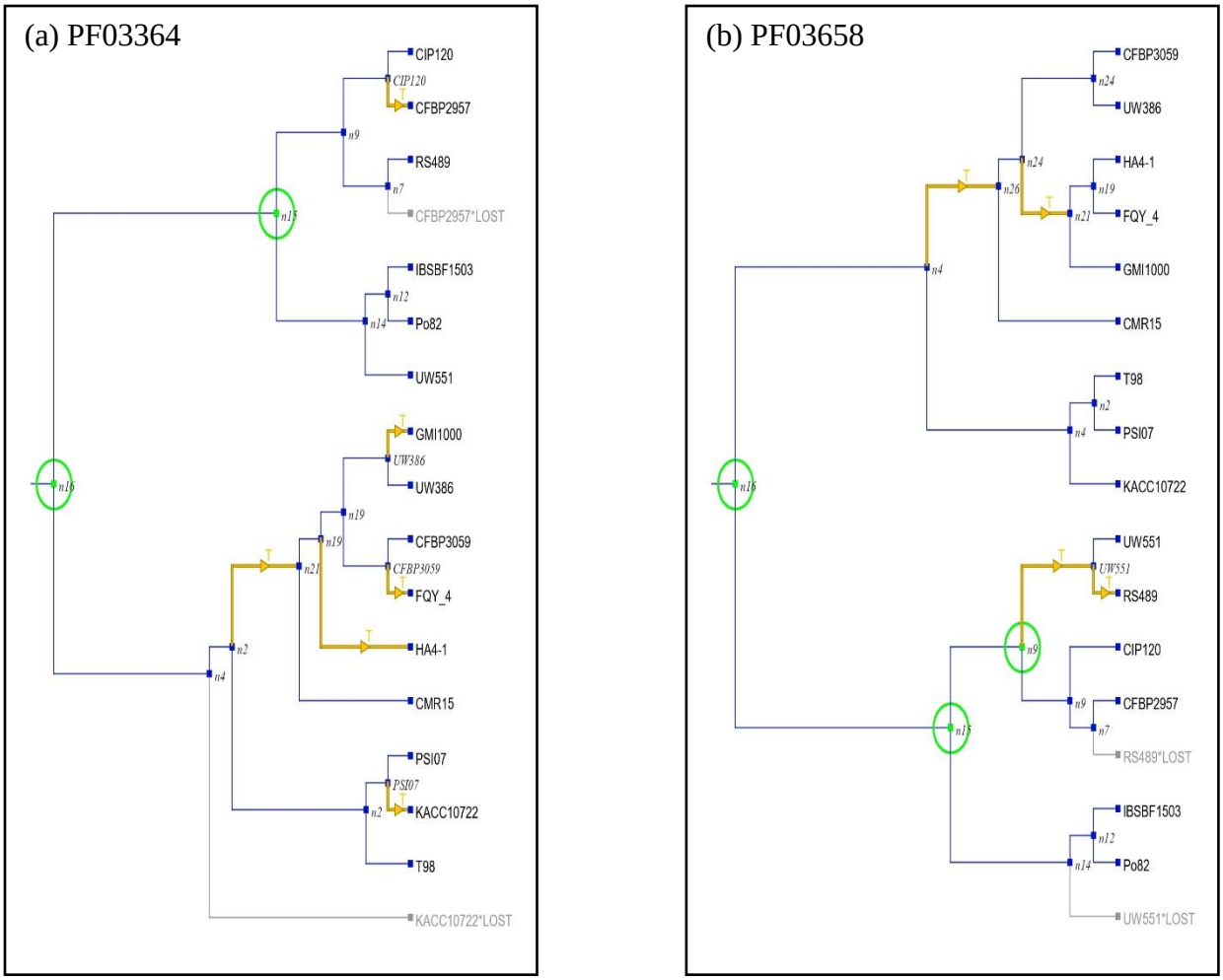

Figure 2. Reconciled phylogenetic trees between Pfam PF03364 and PF03658 and species trees, displaying HGT, duplication, and loss events across 15 RSSC strains by Notung. HGT events are colored yellow, duplication events are red, and losses are colored grey

Figure 3a shows that PF03693 had 3 possible HGT events: between the strains FQY_4 and KACC10722; between the strains GMI1000 and PSI07; and between clade 'n95' (CFBP3059, UW386) and clade 'n92' (Po82, IBSBF1503). There were no possible losses.

Figure 3b displays that in PF04014 there were 3 possible HGT events: between the strains UW551 and RS489; between the strains CMR15 and UW386; and between the strains GMI1000 and FQY_4. In addition, there were 3 possible losses: strain RS489 from clade ‘n6’ (UW551, IBSBF1503, Po82); strain UW386 from clade ‘n14’ (CFBP3059); and train FQY_4 from clade ‘n9’ (HA4-1).

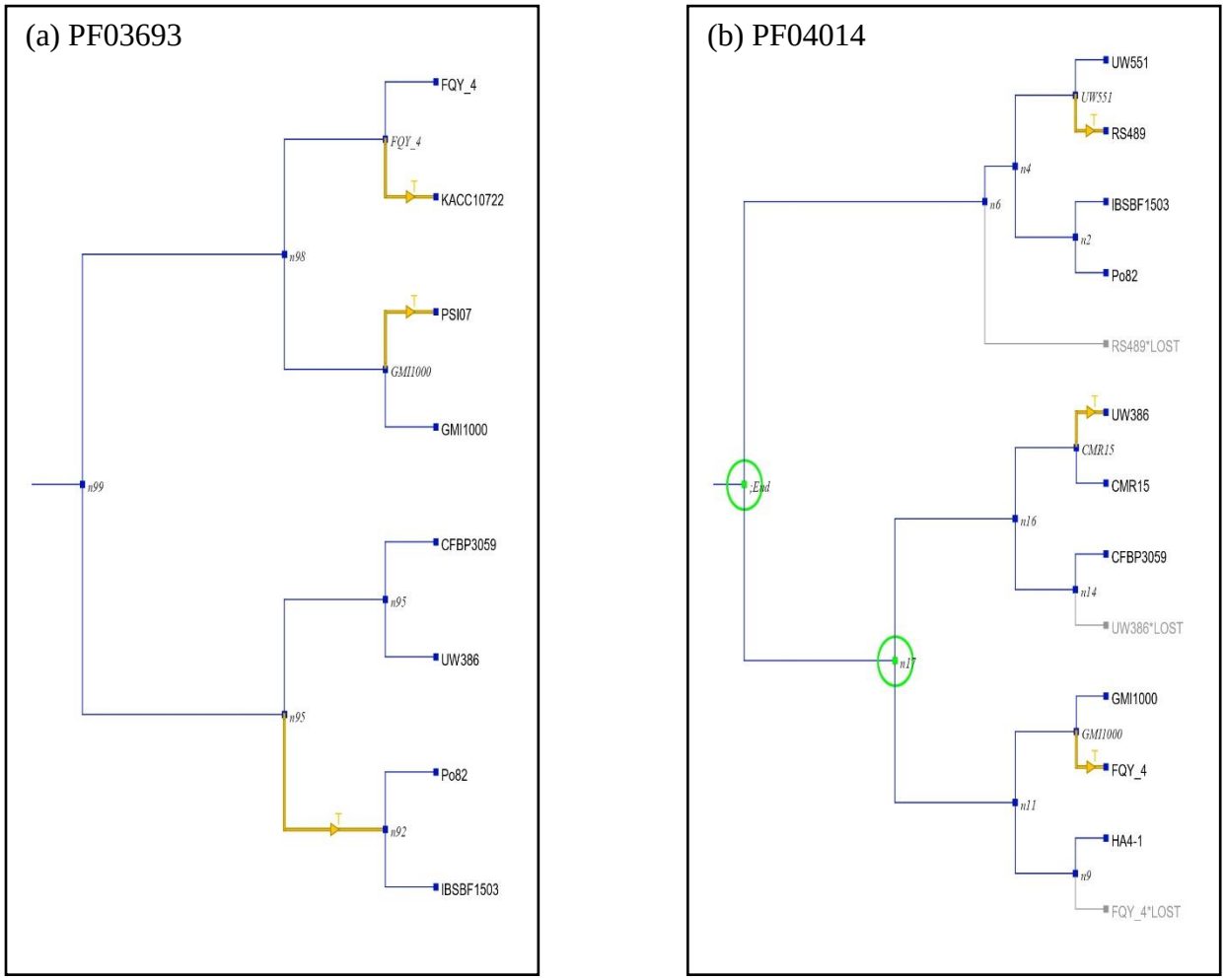

Figure 3. Reconciled phylogenetic trees between Pfam PF03693 and PF04014 and species trees, displaying HGT, duplication, and loss events across 15 RSSC strains by Notung. HGT events are colored yellow, duplication events are red, and losses are colored grey

Figure 4a shows that in PF05016 there were 3 possible HGT events: between clade ‘n8’ (CIP120, CFBP2957, CMR15, PSI07, T98, KACC10722, GMI1000, UW386, CFBP3059, FQY_4) and strain Po82; between the strains CMR15 and clade ‘n18’ (CIP120, CFBP2957); and between the strains GMI1000 and clade ‘n13’ (PSI07, T98, KACC10722). Also, there were 9 possible losses: strain Po82 from clade ‘n21’ (IBSBF1503); strain CFBP2957 from clade ‘n16’(RS489); strain CIP120 from clade ‘n18’ (RS489); clade ‘n13’ (PSI07, T98, KACC10722) from clade ‘n25’ (UW551, IBSBF1503, RS489); strain RS489 from clade ‘n16’ (CFBP2957); clade ‘n2’ (UW386, CFBP3059) from clade ‘n4’ (CIP120, CFBP2957, CMR15); strain FQY_4 from clade ‘n7’ (PSI07, T98, KACC10722, GMI1000); strain CMR15 from clade ‘n4’ (UW386, CFBP3059); and strain GMI1000 from clade ‘n7’ (FQY_4). Furthermore, there was 1 duplication event in node ‘n132’ giving rise to clade ‘n8’ (CIP120, CFBP2957, CMR15, PSI07, T98, KACC10722, GMI1000) and clade ‘n8’ (UW386, CFBP3059, FQY_4).

Figure 4b indicates that in PF05973 there were 4 possible HGT events: between clade ‘n13’ (PSI07, T98, KACC10722) and clade ‘n25’ (UW386, CFBP3059, CMR15, GMI1000, HA4-1, FQY_4, CIP120, RS489, IBSBF1503, UW551); between the strains HA4-1 and GMI1000; between the strains CIP120 and RS489; and between clade ‘n22’ (GMI1000, HA4-1, FQY_4) and clade ‘n8’ (CIP120, RS489, IBSBF1503, UW551). Also, there were 6 possible losses: strain Po82 from clade ‘n5’ (IBSBF1503); strain RS489 from clade ‘n2’ (UW551); strain GMI1000 from clade ‘n24’ (GMI1000, HA4-1, FQY_4, CIP120, RS489, IBSBF1503, UW551); strain IBSBF1503 from clade ‘n5’ (Po82); strain CIP120 from clade ‘n7’ (Po82); and clade ‘n2’ (UW551) from clade ‘n8’ (Po82).

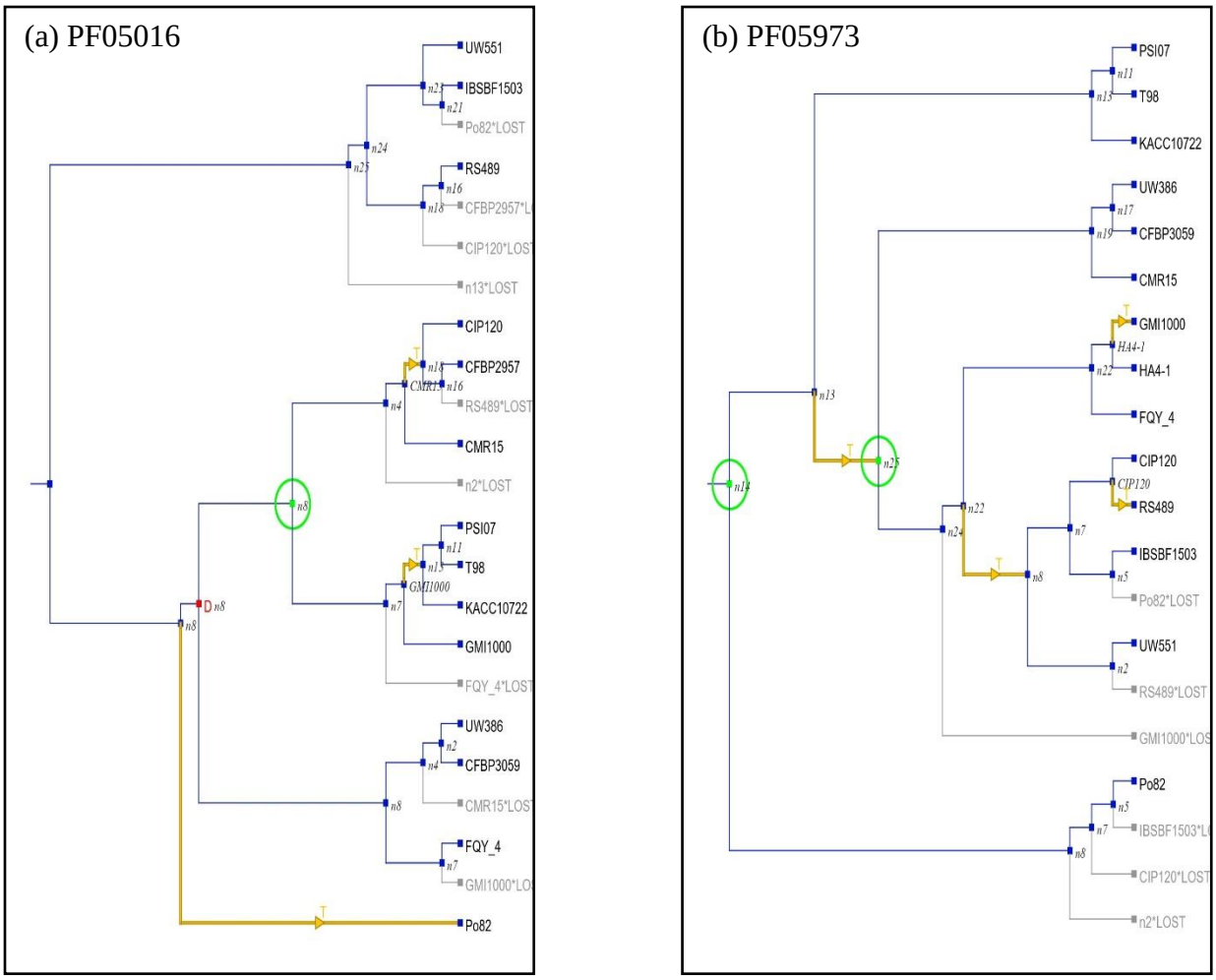

Figure 4. Reconciled phylogenetic trees between Pfam PF05016 and PF05973 and species trees, displaying HGT, duplication, and loss events across 15 RSSC strains by Notung. HGT events are colored yellow, duplication events are red, and losses are colored grey

Figure 5a shows that PF06296 had 3 possible HGT events: between clade 'n4' (GMI1000, FQY_4, HA4-1) and clade 'n15' (RS489, UW551, IBSBF1503, Po82, PSI07); between the strains UW551 and RS489; and between the strains GMI1000 and FQY_4. Furthermore, there were 2 possible losses: strain RS489 from clade 'n13' (RS489, UW551, IBSBF1503, Po82) and strain FQY_4 from clade 'n2' (HA4-1).

Figure 5b indicates that in PF06414 there were 4 possible HGT events: between the strains UW386 and HA4-1; between the strains UW386 and IBSBF1503; between the strains UW386 and PSI07; and between the strain HA4-1 and clade 'n12' (KACC10722, T98, RS489, UW551,

CIP120). Also, there were 2 possible losses: strain PSI07 from clade 'n9' (T98) and strain IBSBF1503 from clade 'n2' (CIP120).

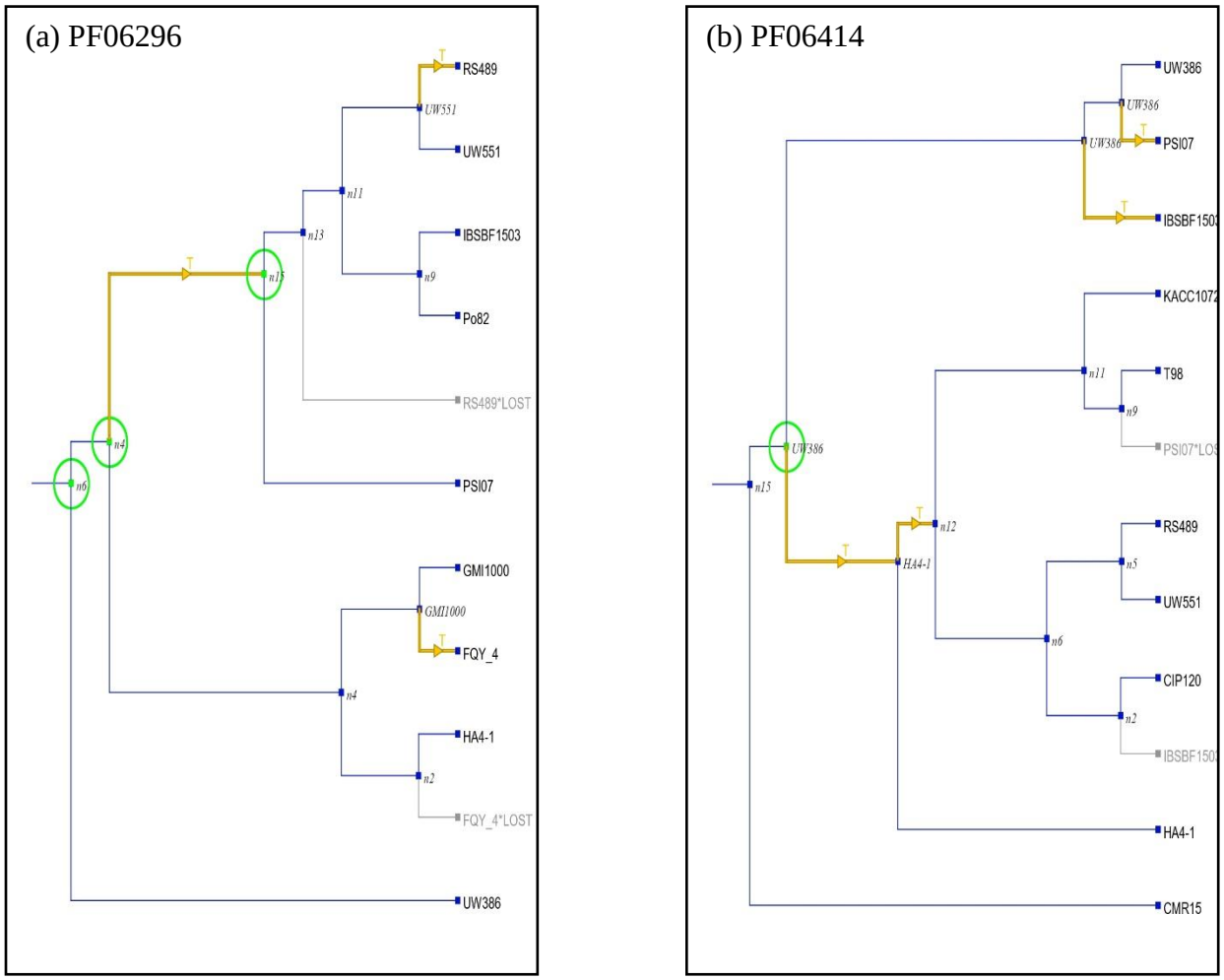

Figure 5. Reconciled phylogenetic trees between Pfam PF06296 and PF06414 and species trees, displaying HGT, duplication, and loss events across 15 RSSC strains by Notung. HGT events are colored yellow, duplication events are red, and losses are colored grey

Figure 6a shows that in PF08843 there were 4 possible HGT events: between the strain CMR15 and clade 'n7' (UW551, RS489); between the strain UW386 and clade 'n14' (FQY_4, HA4-1, GMI1000, IBSBF1503, Po82, CIP120, KACC10722, PSI07); between clade 'n10' IBSBF1503, Po82) and clade 'n22' (FQY_4, HA4-1, GMI1000); and between the strains UW386 and T98. Also, there were 2 possible losses: clade 'n7' (UW551, RS489) from clade 'n13' (FQY_4, HA4-1, GMI1000, IBSBF1503, Po82, CIP120) and strain T98 from clade 'n2' (PSI07).

Figure 6b indicates that in PF13657 there were 3 possible HGT events: between clade ‘n4’ (UW386, CFBP3059, GMI1000, HA4-1, CMR15) and clade ‘n27’ (IBSBF1503, UW551, RS489, CFBP2957, CIP120, PSI07, T98, KACC10722); between clade ‘n2’ (UW386, CFBP3059, GMI1000, HA4-1) and clade ‘n9’ (GMI1000, HA4-1); and between the strains FQY_4 and Po82. Furthermore, there were 4 possible losses: strain FQY_4 from clade ‘n7’ (HA4-1); strain Po82 from clade ‘n13’ (IBSBF1503); strain HA4-1 from clade ‘n7’ (FQY_4, Po82); and strain GMI1000 from clade ‘n9’ (FQY_4, Po82).

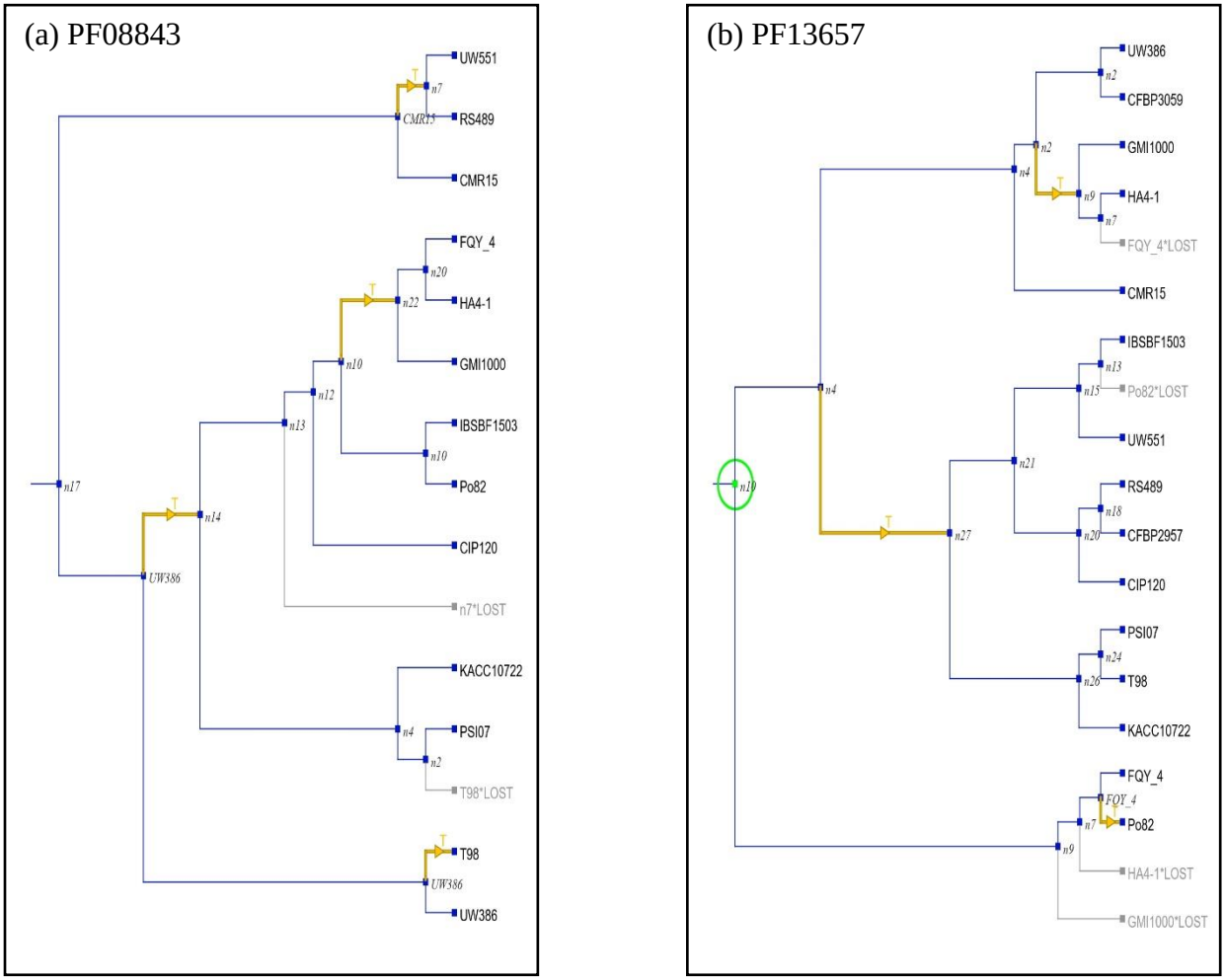

Figure 6. Reconciled phylogenetic trees between Pfam PF08843 and PF13657 and species trees, displaying HGT, duplication, and loss events across 15 RSSC strains by Notung. HGT events are colored yellow, duplication events are red, and losses are colored grey

Figure 7 shows that in PF15738 there was 1 possible HGT event between the strains RS489 and IBSBF1503; and there was 1 possible loss: strain IBSBF1503 from clade ‘n2’ (Po82).

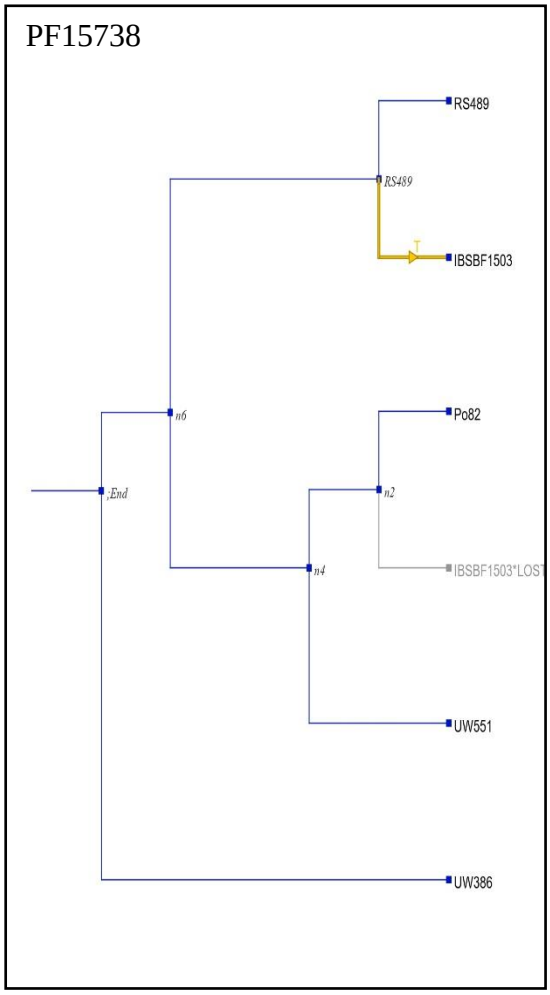

Figure 7. Reconciled phylogenetic trees between Pfam PF15738 and species trees, displaying HGT, duplication, and loss events across 15 RSSC strains by Notung. HGT events are colored yellow, duplication events are red, and losses are colored grey

3.4. Gene gain, loss, and duplication rates

COUNT v10.04 software package was used to perform an evolutionary study of TA gene content evolution using phylogenetic birth-and-death and maximum likelihood models (Csuos, 2010). Generally, gene duplication is the primary process of genome evolution; nonetheless, gene loss in which a gene is inactivated and gene gain where new genes are obtained through HGT is equally essential in genomic evolution history (Karev, Wolf, & Koonin, 2007). I used Pfam accessions for TA estimation of gene content in genomes of RSSC strains. Rates of Pfam duplication, gain, and loss was calculated for TA systems in the fifteen genomes of RSSC analyzed in this study (see Table 6).

As expected, duplication rates are the highest among the three analyzed processes (gain, loss, and duplication), with 3.3-fold higher vs. loss; and 1.7-fold higher vs. gain. Nevertheless, contrary to previous studies in which defense systems evolution is primarily driven by gene loss (Puigbò, Makarova, Kristensen, Wolf, & Koonin, 2017), our results indicate that gene gain is 1.9-fold higher than gene loss. As mentioned before in section 3.3, multiple HGT events responsible for gene gain can be observed among the 15 strains of RSSC.

Table 6. Evolutionary dynamic events (gene gain, loss, and duplication rates) from TA systems across studied RSSC

STRAIN	***LOSS RATE***	***DUPLICATION RATE***	***GAIN RATE***
PSI07	0.23	0.75	0.45
T98	0.15	0.51	0.31
KACC10722	0.13	0.43	0.26
GMI1000	0.11	0.37	0.22
HA4-1	0.16	0.52	0.31
FQY4	0.09	0.28	0.17
CMR15	0.07	0.24	0.15
CFBP3059	-	-	-
UW386	0.24	0.78	0.47
RS489	0.22	0.72	0.43
CIP120	0.13	0.42	0.25
CFBP2957	0.11	0.37	0.22
UW551	0.02	0.06	0.04
IBSBF1503	0.10	0.34	0.20
Po82	-	-	-
AVERAGE	0.135384615	0.445384615	0.267692308

CHAPTER 4. DISCUSSION

In this study, I detailed the presence of toxin-antitoxin systems across fifteen RSSC genomes belonging to the five phylotypes. Results indicate that TA systems can be found broadly in all of them. This high conservation of TA systems, especially type II, could indicate that they are part of the resistance against phage damage, and they would be transferred by HGT from genome to genome (Van Melderen, 2010). Moreover, the high conservation of TA systems could be related to the important biological roles related to them; such as stress-coping, guarding against DNA loss, protection against invading DNA, gene regulation, and growth control (Magnuson, 2007; Unterholzner et al., 2013; Van Melderen, 2010; Van Melderen & Saavedra De Bast, 2009).

Phylotype III had the highest presence of TA systems among the five phylotypes. The strains CMR15 and UW386 were the ones with the highest presence of TA systems among all strains. This could possibly be related to the fact that phylotype III has the highest levels of recombination of all strains (Castillo & Greenberg, 2007).

TA genes with the highest incidence among the fifteen RSSC strains were HipA toxin, RelE/ParE toxin, RatA toxin, and RatB antitoxin. HipA and RelE/ParE toxins belong to the type II toxin superfamilies (Unterholzner et al., 2013). Type II toxins are present in plasmids, making them prone to HGT, which could be why the high prevalence in all the studied strains (Leplae et al., 2011; Van Melderen, 2010). Alternatively, RatA toxin and RatB antitoxin belong to the type I toxin system, in which the antitoxin is a small RNA that interacts with the toxin's mRNA (Brantl & Jahn, 2015).

Epsilon antitoxin was only found in strain CMR15 which belongs to phylotype III; given that this phylotype has high levels of recombination (Castillo & Greenberg, 2007), which could imply the occurrence of horizontal gene transfer (Ely, 2020), I could assume that this gene was gained by an HGT event at the tip of the gene-species reconciled phylogenetic tree. However, this could not be proven due to not having enough sequences for our analysis.

I used multiple methods to analyze the TA systems' evolutionary dynamics in RSSC, focusing on selective pressure, gene gain, loss and duplication rates, and HGT events.

According to our selective pressure analysis, there were no wholly positively selected TA genes in the fifteen studied RSSC genomes. Nonetheless, AbiEi antitoxin, HipA toxin, RelE/ParE toxin, ParD antitoxin, and Zeta toxin were positively selected at a particular site in the gene sequence. AbiEii toxin and AbiEi antitoxin are considered a part of the Type IV toxin-antitoxin systems (Dy, Przybilski, Semeijn, Salmond, & Fineran, 2014). AbiEii toxins are part of the abortive infection (Abi) systems, which prevent viral replication by producing cell suicide (Dy, Przybilski, et al., 2014; Hampton et al., 2018). Furthermore, AbiEii antitoxins have a conserved N-terminal winged helix-turn-helix domain (WHTH), which has been proven to be essential for the transcriptional repression of the AbiE operon, resulting in toxin neutralization (Dy, Przybilski, et al., 2014; Hampton et al., 2018). I could hypothesize that the positively selected sites could be related to this specific WHTH domain. However, further research is needed to support this assertion.

HipA toxin is part of the Type II toxin-antitoxin system and is known for aiding in multidrug resistance by increasing persistence (Korch & Hill, 2006; Moyed & Bertrand, 1983; Schumacher et al., 2015) and biofilm formation (Zhao et al., 2013). Previous mutagenetic studies indicate that position 22 is necessary for higher toxicity and conferring persistence (Korch, Henderson, & Hill, 2003; Schumacher et al., 2015). I could assume this is the site that was found to be positively selected.

The RelE/ParE toxin family is part of the type II toxin-antitoxin system (Gotfredsen & Gerdes, 1998). RelE toxin is an mRNA endoribonuclease that promotes amino acid starvation, leading to inhibition of bacterial growth and resulting in persistence which increases multidrug resistance (Keren, Shah, Spoering, Kaldalu, & Lewis, 2004; Maisonneuve, Shakespeare, Jørgensen, & Gerdes, 2011; Tashiro et al., 2012). In addition, ParE toxin inhibits DNA gyrase, resulting in plasmid DNA partition, which affects cell growth. (Jiang, Pogliano, Helinski, & Konieczny, 2002; Johnson, Ström, & Helinski, 1996; R. C. Roberts & Helinski, 1992; Richard C. Roberts, Ström, & Helinski, 1994). Mutagenetic studies of the RelE toxin reveal several positions (52, 54, 61, 81, and 87) that are necessary to maintain high mRNA cleavage rates (Griffin, Davis, & Strobel, 2013; Neubauer et al., 2009; Pedersen, Christensen, & Gerdes, 2002). Possibly one of the mentioned sites is under positive selection; yet, further research is needed to confirm this assumption. The ParD antitoxin inhibits the anti-DNA gyrase activity of ParE by the post

segregational killing (PSK) of plasmid-free cells (Jiang et al., 2002; Johnson et al., 1996; R. C. Roberts & Helinski, 1992; Richard C. Roberts et al., 1994). Mutagenetic studies reveal that position 12 is necessary to assure proper DNA binding activity of ParD antitoxin for plasmid stabilization (R. C. Roberts, Spangler, & Helinski, 1993), which could be the site under positive selection.

Zeta toxin is also a part of the type II toxin-antitoxin system; it inhibits MurA, necessary for cell wall synthesis and induces programmed cell death (Mutschler, Gebhardt, Shoeman, & Meinhart, 2011). Mutagenetic studies revealed that multiple sites (46, 67, 158, and 171) are necessary to maintain toxic activity (Meinhart, Alonso, Sträter, & Saenger, 2003), one of these sites could be under positive selection according to our results.

The obtained average rates for gene gain, loss, and duplication for the studied TA systems present on the RSSC genomes indicate that gene duplication and gain are the primary evolutionary processes acting on these systems. Gene duplication is a major evolutionary force in prokaryotic genomes (Hooper & Berg, 2003). For instance, resulting paralogs are highly expressed and persistent; nevertheless, they evolve slowly and are not closely related to adaptive roles (Treangen & Rocha, 2011). In addition, gene gain by HGT is a predominant force in bacterial genome evolution (Marri, Hao, & Golding, 2006), where resulting xenologs persist in prokaryotic genomes because they confer higher adaptive roles, making them essential for microbial survival (Koonin, Makarova, & Aravind, 2001; Treangen & Rocha, 2011). Previous studies state gene loss as the primary evolutionary force in bacterial defense systems (Puigbò et al., 2017); hence, it increases fitness by deleting unessential genes (Serres et al., 2009; Koskiniemi et al., 2012). Nonetheless, defense systems evolution dynamics in RSSC are primarily driven by gene gain and duplication (Castillo et al., 2020) which is in accordance with our results overall.

HGT events confer innovative functions in bacterial genomes (Puigbò et al., 2017); hence, it is not surprising that HGT is the primary mechanism of gene gain in RSSC (Castillo et al., 2020; Coupat-Goutaland et al., 2011; Guidot, Coupat, Fall, Prior, & Bertolla, 2009). According to our HGT analysis results, in concordance with previous studies, I could assume that TA genes in RSSC have been highly influenced by at least one HGT between strains and phylotypes.

According to our HGT results, the majority of TA Pfams (PF01850, PF02604, PF03364, PF03658, PF03693, PF05016, PF05973, PF06296, PF06414, PF08843, PF13657) have undergone multiple HGT events between strains of RSSC along their evolutionary history. While Pfams PF04014 (MazEF antitoxin) and PF15738 (YafQ toxin) only have had HGT events at the tips of the reconciled gene-species phylogenetic trees (Figure 3b and Figure 7), demonstrating recent evolutionary events in their phylogeny. In concordance with our selective pressure results and gene gain rates, I could probably assume that HGT played an essential part in TA system evolution throughout their history; hence, some TA systems could be considered ancestral in RSSC, while others could be considered novel. Consequently, various TA genes that HGT events have driven at some point in history are also positively selected according to our results: Zeta toxin (PF06414), HipA toxin (PF13657), and RelE/ParE toxin (PF05016). All these genes are part of the toxin-antitoxin type II systems associated with mobile genetic elements (Fraikin, Goormaghtigh, & Van Melderen, 2020), known for promoting HGT events (Rodríguez-Beltrán et al., 2020; Van Melderen, 2010). The remaining type II toxin-antitoxin systems should be further analyzed to prove this hypothesis.

CHAPTER 5. CONCLUSIONS

TA systems are ubiquitously found in all phylotypes of RSSC. While the majority belong to the type II TA systems, type I and type IV TA systems can be found as well. Phylotype III has the highest amount of TA systems, which could be explained by its high recombination rates (Castillo & Greenberg, 2007) that may be related to a high incidence of HGT events (Ely, 2020).

AbiEi antitoxin, HipA toxin, RelE/ParE toxin, ParD antitoxin, and Zeta toxin genes exhibit positive selection at one site. Our literature review revealed possible sites in each gene that are directly involved in conferring T/A function; therefore, they would need to be positively selected.

Gene duplication rates were the highest evolutionary force in TA systems across the studied RSSC strains; followed by gene gain and gene loss. In bacterial evolutionary studies, gene duplication has high persistence; however, it is not closely involved in adaptive functions (Hooper & Berg, 2003; Treangen & Rocha, 2011). On the contrary, even if gene gain has a lower incidence, it is closely related to adaptive roles and is crucial for survival (Treangen & Rocha, 2011).

HGT events were found across all the analyzed Pfam reconciled phylogenetic trees of TA systems. The majority of Pfams had multiple HGT events in their evolutionary history, implying that specific TA genes are ancestral in RSSC. Nonetheless, Pfams PF04014 and PF15738 only had HGT events at the tips, suggesting they are novel TA genes in certain RSSC strains.

Results indicate that some TA genes that were positively selected at one site have also undergone HGT events. These genes belong to the type II toxin-antitoxin system, known for having mobile genetic elements (Fraikin et al., 2020).

CHAPTER 6. RECOMMENDATIONS AND FURTHER RESEARCH

Evolutionary analyses should be performed again once RSSC genomes are updated with an increased quality in online databases to assure more accurate results. Furthermore, our study helps understand the evolutionary dynamics of TA systems in RSSC. Given that TA system functions include programmed cell death and growth repression, it would be feasible to use this knowledge in designing antibacterial drugs (Unterholzner et al., 2013). In addition, phage-RSSC interactions should be studied to create better phage therapy strategies (Castillo et al., 2020) by avoiding host resistance (Koskella, Lin, Buckling, & Thompson, 2012).

LIST OF REFERENCES

Ackermann, H. W. (2003, May). Bacteriophage observations and evolution. *Research in Microbiology*, Vol. 154, pp. 245–251. https://doi.org/10.1016/S0923-2508(03)00067-6

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andersson, J. O. (2009). Horizontal gene transfer between microbial eukaryotes. *Methods in Molecular Biology (Clifton, N.J.)*, Vol. 532, pp. 473–487. https://doi.org/10.1007/978-1-60327-853-9_27

Bansal, M. S., Wu, Y. C., Alm, E. J., & Kellis, M. (2015). Improved gene tree error correction in the presence of horizontal gene transfer. *Bioinformatics*, *31*(8), 1211–1218. https://doi.org/10.1093/bioinformatics/btu806

Bateman, A., Martin, M. J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., … Zhang, J. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. https://doi.org/10.1093/nar/gkaa1100

Booker, T. R., Jackson, B. C., & Keightley, P. D. (2017, October 30). Detecting positive selection in the genome. *BMC Biology*, Vol. 15, p. 98. https://doi.org/10.1186/s12915-017-0434-y

Brantl, S., & Jahn, N. (2015). sRNAs in bacterial type I and type III toxin-antitoxin systems. *FEMS Microbiology Reviews*, *39*(3), 413–427. https://doi.org/10.1093/femsre/fuv003

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*. https://doi.org/10.1186/1471-2105-10-421

Castillo, J. A., & Greenberg, J. T. (2007). Evolutionary dynamics of Ralstonia solanacearum. *Applied and Environmental Microbiology*, *73*(4), 1225–1238. https://doi.org/10.1128/AEM.01253-06

Castillo, J. A., Secaira-Morocho, H., Maldonado, S., & Sarmiento, K. N. (2020). Diversity and

Evolutionary Dynamics of Antiphage Defense Systems in Ralstonia solanacearum Species Complex. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00961

Chen, K., Durand, D., & Farach-Colton, M. (2000). NOTUNG: A program for dating gene duplications and optimizing gene family trees. *Journal of Computational Biology*, *7*(3–4), 429–447. https://doi.org/10.1089/106652700750050871

Chibani-Chennoufi, S., Bruttin, A., Dillmann, M. L., & Brüssow, H. (2004, June 15). Phage-host interaction: An ecological perspective. *Journal of Bacteriology*, Vol. 186, pp. 3677–3686. https://doi.org/10.1128/JB.186.12.3677-3686.2004

Coupat-Goutaland, B., Bernillon, D., Guidot, A., Prior, P., Nesme, X., & Bertolla, F. (2011). Ralstonia solanacearum virulence increased following large interstrain gene transfers by natural transformation. *Molecular Plant-Microbe Interactions*, *24*(4), 497–505. https://doi.org/10.1094/MPMI-09-10-0197

Csuos, M. (2010). Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics*, *26*(15), 1910–1912. https://doi.org/10.1093/bioinformatics/btq315

D'Souza, G., Waschina, S., Pande, S., Bohl, K., Kaleta, C., & Kost, C. (2014). LESS IS MORE: SELECTIVE ADVANTAGES CAN EXPLAIN THE PREVALENT LOSS OF BIOSYNTHETIC GENES IN BACTERIA. *Evolution*, *68*(9), 2559–2570. https://doi.org/10.1111/evo.12468

Darby, C. A., Stolzer, M., Ropp, P. J., Barker, D., & Durand, D. (2017). Xenolog classification. *Bioinformatics*, *33*(5), 640–649. https://doi.org/10.7488/ds/1503

Djordjevic, G. M., O'Sullivan, D. J., Walker, S. A., Conkling, M. A., & Klaenhammer, T. R. (1997). A triggered-suicide system designed as a defense against bacteriophages. *Journal of Bacteriology*, *179*(21), 6741–6748. https://doi.org/10.1128/jb.179.21.6741-6748.1997

Durand, D., Halldórsson, B. V., & Vernot, B. (2005). A hybrid micro-macroevolutionary approach to gene tree reconstruction. *Lecture Notes in Bioinformatics (Subseries of Lecture Notes in Computer Science)*, *3500*, 250–264. https://doi.org/10.1007/11415770_19

Dy, R. L., Przybilski, R., Semeijn, K., Salmond, G. P. C., & Fineran, P. C. (2014). A widespread

bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic Acids Research*, *42*(7), 4590–4605. https://doi.org/10.1093/nar/gkt1419

Dy, R. L., Richter, C., Salmond, G. P. C., & Fineran, P. C. (2014). Remarkable mechanisms in microbes to resist phage infections. *Annual Review of Virology*, *1*(1), 307–331. https://doi.org/10.1146/annurev-virology-031413-085500

Ely, B. (2020). Recombination and gene loss occur simultaneously during bacterial horizontal gene transfer. *PLOS ONE*, *15*(1), e0227987. https://doi.org/10.1371/journal.pone.0227987

Farhat, M. R., Shapiro, B. J., Kieser, K. J., Sultana, R., Jacobson, K. R., Victor, T. C., … Murray, M. (2013). Genomic analysis identifies targets of convergent positive selection in drug-resistant Mycobacterium tuberculosis. *Nature Genetics*, *45*(10), 1183–1189. https://doi.org/10.1038/ng.2747

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., … Bateman, A. (2009). *The Pfam protein families database*. https://doi.org/10.1093/nar/gkp985

Fraikin, N., Goormaghtigh, F., & Van Melderen, L. (2020). Type II Toxin-Antitoxin Systems: Evolution and Revolutions. *Journal of Bacteriology*, *202*(7). https://doi.org/10.1128/jb.00763-19

Genin, S., & Boucher, C. (2002). *Ralstonia solanacearum* : secrets of a major pathogen unveiled by analysis of its genome. *Molecular Plant Pathology*, *3*(3), 111–118. https://doi.org/10.1046/j.1364-3703.2002.00102.x

Goeders, N., & Van Melderen, L. (2014). Toxin-Antitoxin Systems as Multilevel Interaction Systems. *Toxins*, *6*(1), 304–324. https://doi.org/10.3390/toxins6010304

Gotfredsen, M., & Gerdes, K. (1998). The Escherichia coli relBE genes belong to a new toxin-antitoxin gene family. *Molecular Microbiology*, *29*(4), 1065–1076. https://doi.org/10.1046/j.1365-2958.1998.00993.x

Griffin, M. A., Davis, J. H., & Strobel, S. A. (2013). Bacterial toxin RelE: A highly efficient ribonuclease with exquisite substrate specificity using atypical catalytic residues. *Biochemistry*, *52*(48), 8633–8642. https://doi.org/10.1021/bi401325c

Guglielmini, J., & Van Melderen, L. (2011). Bacterial toxin-antitoxin systems. *Mobile Genetic Elements*, *1*(4), 283–306. https://doi.org/10.4161/mge.18477

Guidot, A., Coupat, B., Fall, S., Prior, P., & Bertolla, F. (2009). Horizontal gene transfer between ralstonia solanacearum strains detected by comparative genomic hybridization on microarrays. *ISME Journal*, *3*(5), 549–562. https://doi.org/10.1038/ismej.2009.14

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., … Bioinformatique, L. (n.d.). *New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0*. Retrieved from http://www.lirmm.fr/~gascuel

Gupta, A., Venkataraman, B., Vasudevan, M., & Gopinath Bankar, K. (2017). Co-expression network analysis of toxin-antitoxin loci in Mycobacterium tuberculosis reveals key modulators of cellular stress. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-06003-7

Gyles, C., & Boerlin, P. (2014a, March 6). Horizontally Transferred Genetic Elements and Their Role in Pathogenesis of Bacterial Disease. *Veterinary Pathology*, Vol. 51, pp. 328–340. https://doi.org/10.1177/0300985813511131

Gyles, C., & Boerlin, P. (2014b, March 6). Horizontally Transferred Genetic Elements and Their Role in Pathogenesis of Bacterial Disease. *Veterinary Pathology*, Vol. 51, pp. 328–340. https://doi.org/10.1177/0300985813511131

Hampton, H. G., Jackson, S. A., Fagerlund, R. D., Vogel, A. I. M., Dy, R. L., Blower, T. R., & Fineran, P. C. (2018). AbiEi Binds Cooperatively to the Type IV abiE Toxin–Antitoxin Operator Via a Positively-Charged Surface and Causes DNA Bending and Negative Autoregulation. *Journal of Molecular Biology*, *430*(8), 1141–1156. https://doi.org/10.1016/j.jmb.2018.02.022

Hooper, S. D., & Berg, O. G. (2003). On the nature of gene innovation: Duplication patterns in microbial genomes. *Molecular Biology and Evolution*, *20*(6), 945–954. https://doi.org/10.1093/molbev/msg101

Jiang, Y., Pogliano, J., Helinski, D. R., & Konieczny, I. (2002). ParE toxin encoded by the

broad-host-range plasmid RK2 is an inhibitor of Escherichia coli gyrase. *Molecular Microbiology*, *44*(4), 971–979. https://doi.org/10.1046/j.1365-2958.2002.02921.x

Johnson, E. P., Ström, A. R., & Helinski, D. R. (1996). Plasmid RK2 toxin protein ParE: Purification and interaction with the ParD antitoxin protein. *Journal of Bacteriology*, *178*(5), 1420–1429. https://doi.org/10.1128/jb.178.5.1420-1429.1996

Karev, G. P., Wolf, Y. I., & Koonin, E. V. (2007). Birth and Death Models of Genome Evolution. In *Power Laws, Scale-Free Networks and Genome Biology* (pp. 65–85). https://doi.org/10.1007/0-387-33916-7_6

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160–1166. https://doi.org/10.1093/bib/bbx108

Keren, I., Shah, D., Spoering, A., Kaldalu, N., & Lewis, K. (2004). Specialized persister cells and the mechanism of multidrug tolerance in Escherichia coli. *Journal of Bacteriology*, *186*(24), 8172–8180. https://doi.org/10.1128/JB.186.24.8172-8180.2004

Koonin, E. V., Makarova, K. S., & Aravind, L. (2001). Horizontal gene transfer in prokaryotes: Quantification and classification. *Annual Review of Microbiology*, Vol. 55, pp. 709–742. https://doi.org/10.1146/annurev.micro.55.1.709

Korch, S. B., Henderson, T. A., & Hill, T. M. (2003). Characterization of the hipA7 allele of Escherichia coli and evidence that high persistence is governed by (p)ppGpp synthesis. *Molecular Microbiology*, *50*(4), 1199–1213. https://doi.org/10.1046/j.1365-2958.2003.03779.x

Korch, S. B., & Hill, T. M. (2006). Ectopic overexpression of wild-type and mutant hipA genes in Escherichia coli: Effects on macromolecular synthesis and persister formation. *Journal of Bacteriology*, *188*(11), 3826–3836. https://doi.org/10.1128/JB.01740-05

Kosakovsky Pond, S. L., Murrell, B., Fourment, M., Frost, S. D. W., Delport, W., & Scheffler, K. (2011). A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution*, *28*(11), 3033–3043. https://doi.org/10.1093/molbev/msr125

Koskella, B., Lin, D. M., Buckling, A., & Thompson, J. N. (2012). The costs of evolving resistance in heterogeneous parasite environments. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1735), 1896–1903. https://doi.org/10.1098/rspb.2011.2259

Koskiniemi, S., Sun, S., Berg, O. G., & Andersson, D. I. (2012). Selection-Driven Gene Loss in Bacteria. *PLoS Genetics*, *8*(6), e1002787. https://doi.org/10.1371/journal.pgen.1002787

Labrie, S. J., Samson, J. E., & Moineau, S. (2010, March 29). Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, Vol. 8, pp. 317–327. https://doi.org/10.1038/nrmicro2315

Lai, H., Stolzer, M., & Durand, D. (2017). Fast heuristics for resolving weakly supported branches using duplication, transfers, and losses. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10562 LNBI*, 298–320. https://doi.org/10.1007/978-3-319-67979-2_16

Leonard, S., Hommais, F., Nasser, W., & Reverchon, S. (2017). Plant-phytopathogen interactions: bacterial responses to environmental and plant stimuli. *Environmental Microbiology*, *19*(5), 1689–1716. https://doi.org/10.1111/1462-2920.13611

Leplae, R., Geeraerts, D., Hallez, R., Guglielmini, J., Drze, P., & Van Melderen, L. (2011). Diversity of bacterial type II toxin-antitoxin systems: A comprehensive search and functional analysis of novel families. *Nucleic Acids Research*, *39*(13), 5513–5525. https://doi.org/10.1093/nar/gkr131

Li, N., Wang, K., Williams, H. N., Sun, J., Ding, C., Leng, X., & Dong, K. (2017). Analysis of gene gain and loss in the evolution of predatory bacteria. *Gene*, *598*, 63–70. https://doi.org/10.1016/j.gene.2016.10.039

Li, Z. G., He, F., Zhang, Z., & Peng, Y. L. (2012). Prediction of protein-protein interactions between Ralstonia solanacearum and Arabidopsis thaliana. *Amino Acids*, *42*(6), 2363–2371. https://doi.org/10.1007/s00726-011-0978-z

Magadum, S., Banerjee, U., Murugan, P., Gangapur, D., & Ravikesavan, R. (2013). Gene duplication as a major force in evolution. *Journal of Genetics*, *92*(1), 155–161. https://doi.org/10.1007/s12041-013-0212-8

Magnuson, R. D. (2007, September 1). Hypothetical functions of toxin-antitoxin systems. *Journal of Bacteriology*, Vol. 189, pp. 6089–6092. https://doi.org/10.1128/JB.00958-07

Maisonneuve, E., Shakespeare, L. J., Jørgensen, M. G., & Gerdes, K. (2011, August 9). Bacterial persistence by RNA endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108, pp. 13206–13211. https://doi.org/10.1073/pnas.1100186108

MANSFIELD, J., GENIN, S., MAGORI, S., CITOVSKY, V., SRIARIYANUM, M., RONALD, P., … FOSTER, G. D. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*, *13*(6), 614–629. https://doi.org/10.1111/j.1364-3703.2012.00804.x

Marri, P. R., Hao, W., & Golding, G. B. (2006). Gene gain and gene loss in Streptococcus: Is it driven by habitat? *Molecular Biology and Evolution*, *23*(12), 2379–2391. https://doi.org/10.1093/molbev/msl115

Meinhart, A., Alonso, J. C., Sträter, N., & Saenger, W. (2003). Crystal structure of the plasmid maintenance system ε/ζ: Functional mechanism of toxin ζ and inactivation by ε2ζ2 complex formation. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(4), 1661–1666. https://doi.org/10.1073/pnas.0434325100

Monk, A. B., Rees, C. D., Barrow, P., Hagens, S., & Harper, D. R. (2010). Bacteriophage applications: where are we now? *Letters in Applied Microbiology*, *51*(4), 363–369. https://doi.org/10.1111/j.1472-765X.2010.02916.x

Moyed, H. S., & Bertrand, K. P. (1983). hipA, a newly recognized gene of Escherichia coli K-12 that affects frequency of persistence after inhibition of murein synthesis. *Journal of Bacteriology*, *155*(2), 768–775. https://doi.org/10.1128/jb.155.2.768-775.1983

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: A fast, unconstrained bayesian AppRoximation for inferring selection. *Molecular Biology and Evolution*, *30*(5), 1196–1205. https://doi.org/10.1093/molbev/mst030

Murrell, B., Weaver, S., Smith, M. D., Wertheim, J. O., Murrell, S., Aylward, A., …

Kosakovsky Pond, S. L. (2015). Gene-Wide Identification of Episodic Selection. *Molecular Biology and Evolution*, *32*(5), 1365–1371. https://doi.org/10.1093/molbev/msv035

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, *8*(7), 1002764. https://doi.org/10.1371/journal.pgen.1002764

Muse, S. V., & Gaut, B. S. (1994). A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution*, *11*(5), 715–724. https://doi.org/10.1093/oxfordjournals.molbev.a040152

Mutschler, H., Gebhardt, M., Shoeman, R. L., & Meinhart, A. (2011). A novel mechanism of programmed cell death in bacteria by toxin-antitoxin systems corrupts peptidoglycan synthesis. *PLoS Biology*, *9*(3). https://doi.org/10.1371/journal.pbio.1001033

Neubauer, C., Gao, Y. G., Andersen, K. R., Dunham, C. M., Kelley, A. C., Hentschel, J., … Brodersen, D. E. (2009). The Structural Basis for mRNA Recognition and Cleavage by the Ribosome-Dependent Endonuclease RelE. *Cell*, *139*(6), 1084–1095. https://doi.org/10.1016/j.cell.2009.11.015

Ochman, H., Lawrence, J. G., & Grolsman, E. A. (2000, May 18). Lateral gene transfer and the nature of bacterial innovation. *Nature*, Vol. 405, pp. 299–304. https://doi.org/10.1038/35012500

Pacheco, J., Ochoa-Moreno, W.-S., Ordoñez, J., & Izquierdo-Montoya, L. (2018). Agricultural Diversification and Economic Growth in Ecuador. *Sustainability*, *10*(7), 2257. https://doi.org/10.3390/su10072257

Pearson, W. R. (2013). An Introduction to Sequence Similarity ("Homology") Searching. *Current Protocols in Bioinformatics*, *42*(1), 3.1.1-3.1.8. https://doi.org/10.1002/0471250953.bi0301s42

Pedersen, K., Christensen, S. K., & Gerdes, K. (2002). Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. *Molecular Microbiology*, *45*(2), 501–510. https://doi.org/10.1046/j.1365-2958.2002.03027.x

Peeters, N., Guidot, A., Vailleau, F., & Valls, M. (2013). Ralstonia solanacearum, a widespread bacterial plant pathogen in the post-genomic era. *Molecular Plant Pathology*, *14*(7), 651–662. https://doi.org/10.1111/mpp.12038

Petersen, L., Bollback, J. P., Dimmic, M., Hubisz, M., & Nielsen, R. (2007). Genes under positive selection in Escherichia coli. *Genome Research*, *17*(9), 1336–1343. https://doi.org/10.1101/gr.6254707

Puigbò, P., Makarova, K. S., Kristensen, D. M., Wolf, Y. I., & Koonin, E. V. (2017). Reconstruction of the evolution of microbial defense systems. *BMC Evolutionary Biology*, *17*(1), 94. https://doi.org/10.1186/s12862-017-0942-y

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., … Finn, R. D. (2012). *The Pfam protein families database*. https://doi.org/10.1093/nar/gkr1065

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Roberts, R. C., & Helinski, D. R. (1992). Definition of a minimal plasmid stabilization system from the broad-host- range plasmid RK2. *Journal of Bacteriology*, *174*(24), 8119–8132. https://doi.org/10.1128/jb.174.24.8119-8132.1992

Roberts, R. C., Spangler, C., & Helinski, D. R. (1993). Characteristics and significance of DNA binding activity of plasmid stabilization protein ParD from the broad host-range plasmid RK2. *Journal of Biological Chemistry*, *268*(36), 27109–27117. https://doi.org/10.1016/s0021-9258(19)74224-0

Roberts, Richard C., Ström, A. R., & Helinski, D. R. (1994). The parDE operon of the broad-host-range plasmid RK2 specifies growth inhibition associated with plasmid loss. *Journal of Molecular Biology*, *237*(1), 35–51. https://doi.org/10.1006/jmbi.1994.1207

Rodríguez-Beltrán, J., Sørum, V., Toll-Riera, M., de la Vega, C., Peña-Miller, R., & Millán, Á. S. (2020). Genetic dominance governs the evolution and spread of mobile genetic elements in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(27), 15755–15762. https://doi.org/10.1073/pnas.2001240117

Safni, Irda; Cleenwerck, Ilse; De Vos, Paul; Fegan, Mark; Sly, Lindsay; Kappler, U. (2014). Polyphasic taxonomic revision of the Ralstonia solanacearum species complex: proposal to emend the descriptions of Ralstonia solanacearum and Ralstonia syzygii and reclassify current R. syzygii strains as Ralstonia syzygii subsp. syzygii subsp. nov., R. s. *Int. J. Syst. Evol. Micro*, (64), 3087–3103. https://doi.org/10.1099/ijs.0.066712-0.

Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Mangenot, S., Arlat, M., … Boucher, C. A. (2002). Genome sequence of the plant pathogen ralstonia solanacearum. *Nature*, *415*(6871), 497–502. https://doi.org/10.1038/415497a

Schumacher, M. A., Balani, P., Min, J., Chinnam, N. B., Hansen, S., Vulić, M., … Brennan, R. G. (2015). HipBA-promoter structures reveal the basis of heritable multidrug tolerance. *Nature*, *524*(7563), 59–66. https://doi.org/10.1038/nature14662

Serres, M. H., Kerr, A. R. W., McCormack, T. J., & Riley, M. (2009). Evolution by leaps: Gene duplication in bacteria. *Biology Direct*, *4*, 46. https://doi.org/10.1186/1745-6150-4-46

Sghaier, H., Ghedira, K., Benkahla, A., & Barkallah, I. (2008). Basal DNA repair machinery is subject to positive selection in ionizing-radiation-resistant bacteria. *BMC Genomics*, *9*(1), 297. https://doi.org/10.1186/1471-2164-9-297

Snel, B., Bork, P., & Huynen, M. A. (2002). Genomes in flux: The evolution of Archaeal and Proteobacterial gene content. *Genome Research*, *12*(1), 17–25. https://doi.org/10.1101/gr.176501

Song, S., & Wood, T. K. (2020, August 13). A Primary Physiological Role of Toxin/Antitoxin Systems Is Phage Inhibition. *Frontiers in Microbiology*, Vol. 11, p. 1895. https://doi.org/10.3389/fmicb.2020.01895

Soucy, S. M., Huang, J., & Gogarten, J. P. (2015, August 21). Horizontal gene transfer: Building the web of life. *Nature Reviews Genetics*, Vol. 16, pp. 472–482. https://doi.org/10.1038/nrg3962

Spielman, S. J., Weaver, S., Shank, S. D., Magalis, B. R., Li, M., & Kosakovsky Pond, S. L. (2019). Evolution of viral genomes: Interplay between selection, recombination, and other forces. In *Methods in Molecular Biology* (Vol. 1910, pp. 427–468).

https://doi.org/10.1007/978-1-4939-9074-0_14

Stolzer, M., Lai, H., Xu, M., Sathaye, D., Vernot, B., & Durand, D. (2012). Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. *Bioinformatics*, *28*(18). https://doi.org/10.1093/bioinformatics/bts386

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1). https://doi.org/10.1093/ve/vey016

Tashiro, Y., Kawata, K., Taniuchi, A., Kakinuma, K., May, T., & Okabe, S. (2012). RelE-mediated dormancy is enhanced at high cell density in escherichia coli. *Journal of Bacteriology*, *194*(5), 1169–1176. https://doi.org/10.1128/JB.06628-11

TextPad - the text editor for Windows. (n.d.). Retrieved March 5, 2021, from https://www.textpad.com/home

Treangen, T. J., & Rocha, E. P. C. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genetics*, *7*(1). https://doi.org/10.1371/journal.pgen.1001284

Unterholzner, S. J., Poppenberger, B., & Rozhon, W. (2013). Toxin–antitoxin systems. *Mobile Genetic Elements*, *3*(5), e26219. https://doi.org/10.4161/mge.26219

Van Melderen, L. (2010, December 1). Toxin-antitoxin systems: Why so many, what for? *Current Opinion in Microbiology*, Vol. 13, pp. 781–785. https://doi.org/10.1016/j.mib.2010.10.006

Van Melderen, L., & Saavedra De Bast, M. (2009). Bacterial Toxin–Antitoxin Systems: More Than Selfish Entities? *PLoS Genetics*, *5*(3), e1000437. https://doi.org/10.1371/journal.pgen.1000437

Vernot, B., Stolzer, M., Goldman, A., & Durand, D. (2008). Reconciliation with non-binary species trees. *Journal of Computational Biology*, *15*(8), 981–1006. https://doi.org/10.1089/cmb.2008.0092

Weitz, J. S., Poisot, T., Meyer, J. R., Flores, C. O., Valverde, S., Sullivan, M. B., & Hochberg,

M. E. (2013, February). Phage-bacteria infection networks. *Trends in Microbiology*, Vol. 21, pp. 82–91. https://doi.org/10.1016/j.tim.2012.11.003

Yang, Z., Nielsen, R., Goldman, N., & Pedersen, A. M. K. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, *155*(1), 431–449. https://doi.org/10.1093/genetics/155.1.431

Yao, J., & Allen, C. (2007). The plant pathogen Ralstonia solanacearum needs aerotaxis for normal biofilm formation and interactions with its tomato host. *Journal of Bacteriology*, *189*(17), 6415–6424. https://doi.org/10.1128/JB.00398-07

Yuliar, Nion, Y. A., & Toyota, K. (2015). Recent Trends in Control Methods for Bacterial Wilt Diseases Caused by <i>Ralstonia solanacearum</i> *Microbes and Environments*, *30*(1), 1–11. https://doi.org/10.1264/jsme2.ME14144

Zhang, J. (2008, October 21). Positive selection, not negative selection, in the pseudogenization of rcsA in Yersinia pestis. *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 105, pp. E69–E69. https://doi.org/10.1073/pnas.0806419105

Zhao, J., Wang, Q., Li, M., Heijstra, B. D., Wang, S., Liang, Q., & Qi, Q. (2013). Escherichia coli toxin gene hipA affects biofilm formation and DNA release. *Microbiology (United Kingdom)*, *159*(PART3), 633–640. https://doi.org/10.1099/mic.0.063784-0

Zhaxybayeva, O., & Doolittle, W. F. (2011, April 12). Lateral gene transfer. *Current Biology*, Vol. 21, pp. R242–R246. https://doi.org/10.1016/j.cub.2011.01.045

APPENDIX 1. Presence of TA genes across 15 RSSC strains

T/A	Phylotype I			Phylotype IIA			Phylotype IIB			Phylotype III			Phylotype IV		
	STRAINS			STRAINS			STRAINS			STRAINS			STRAINS		
	GMI1000	FQY_4	HA4-1	CFBP2957	RS489	CIP120	Po82	UW551	IBSBF1503	CMR15	CFBP3059	UW386	T98	PSI07	KACC10722
AbiEii/AbiGii Toxin	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1
AbiEi Antitoxin	1	1	0	1	0	0	1	0	0	0	0	0	1	1	1
Epsilon Antitoxin	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Zeta Toxin	0	0	1	0	1	1	0	1	1	1	0	1	1	1	1
HicA Toxin	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0
HicB Antitoxin	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0
HigB-like_toxin	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
HigA Antitoxin	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
HipA Toxin	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
HipB Antitoxin	1	1	0	1	1	0	0	0	0	1	0	0	1	1	1
MazEF Antitoxin	1	1	1	0	1	0	1	1	1	1	1	1	0	0	0
VapC Toxin	1	1	1	0	1	0	1	1	1	1	1	1	0	1	0
Phd/YefM Antitoxin	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
ParD Antitoxin	1	1	0	0	0	0	1	0	1	0	1	1	0	1	1
RelE/ParE Toxin	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PrlF Antitoxin	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
YhaV Toxin	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
RatA Toxin	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
RatB Antitoxin	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
RelB/DinJ Antitoxin	0	0	0	1	1	0	1	1	1	0	0	0	0	0	0
YafQ Toxin	0	0	0	0	1	0	1	1	1	0	0	1	0	0	0
SymE Toxin	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0
SymR Antitoxin	0	0	0	0	0	1	1	0	1	0	0	0	0	1	0
Total TA per strain	**13**	**11**	**9**	**10**	**12**	**9**	**14**	**10**	**14**	**15**	**12**	**15**	**9**	**13**	**10**
Total TA per phylotype	**33**			**31**			**38**			**42**			**32**		

1 indicates presence of TA gene and 0 indicates absence of TA gene

APPENDIX 2. Presence of TA Pfams across 15 RSSC strains

	STRAINS														
Pfam	**GMI1000**	**FQY4**	**HA4-1**	**CFBP2957**	**RS489**	**CIP120**	**Po82**	**UW551**	**IBSBF1503**	**CMR15**	**CFBP3059**	**UW386**	**T98**	**PSI07**	**KACC10722**
PF08843	1	2	1	0	1	1	2	1	2	1	0	1	2	1	1
PF09952	1	1	2	0	1	1	1	1	1	0	1	0	1	1	1
PF17194	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0
PF11459	1	0	1	0	0	0	0	0	0	0	0	1	1	0	0
PF13338	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0
PF01381	11	12	12	9	15	16	16	11	21	9	11	24	13	14	12
PF01850	2	1	2	0	3	0	2	1	3	2	3	4	0	2	0
PF02604	2	2	2	1	1	1	1	0	1	2	4	5	1	2	1
PF02794	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1
PF03364	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PF03658	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1
PF03693	1	1	0	0	0	0	1	0	1	0	1	2	0	1	1
PF04014	1	0	1	0	1	0	1	1	1	2	1	2	0	0	0
PF04221	0	0	0	0	1	0	1	1	1	0	0	0	0	0	0
PF05015	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
PF05016	2	2	0	1	3	2	2	2	3	3	4	5	2	3	1
PF05534	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0
PF05973	1	1	0	0	2	2	2	2	2	1	1	3	1	1	1
PF06296	1	2	1	0	0	0	1	1	1	0	0	1	0	1	0
PF06414	0	0	1	0	3	1	0	1	2	1	0	2	1	2	2
PF07804	4	3	3	2	3	3	4	3	4	3	3	3	4	4	4
PF07927	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
PF08845	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0
PF11663	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
PF13560	0	0	0	0	14	0	0	0	20	0	0	18	0	14	12
PF13657	3	3	3	2	3	3	4	3	4	3	3	3	4	4	4
PF15738	0	0	0	0	1	0	1	1	1	0	0	1	0	0	0
PF15937	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0