Simple sequence repeat analysis and qualitative pathway analysis of *Rathayibacter toxicus*

by

Kseniya Chumachenko

B.S., Kansas State University, 2017

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Plant Pathology
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2020

Approved by:

Major Professor
Dr. James P. Stack

Copyright

Abstract

Rathayibacter toxicus is a plant pathogenic bacterium vectored by *Anguina* species (nematodes) and hosted by grass species in several genera in the *Poaceae* family including, *Lolium*, *Lacnagrostis*, and *Polypogon*. *Rathayibacter toxicus* forms bacterial galls in developing grass ovules where it produces a dangerous neurotoxin, tunicamycin. Grazing animals that consume toxic galls develop convulsions and staggers, often resulting in death. *Rathayibacter toxicus* is indigenous to Australia, but concern for global spread and potential impact to livestock industries resulted in its designation as a U.S. Select Agent pathogen. Several genetic and genomic studies including MLST, ISSR, and WGS support genetic population groupings and identified a newly emerged population (RT-I) in South Australia. Unique attributes of RT-I preclude determination of its geographic or phylogenetic origin. Simple Sequence Repeats (SSRs) are short tandem nucleotide repeats that vary in size and number of repeats among and within genomes and have been useful in determining evolutionary heritage. Unlike other genome components, SSRs generally have faster mutation rates and have been used as the basis of phylogenetic and phylogeographic analyses in many organisms including bacteria and plants. SSRs in *R. toxicus* were identified and analyzed to help determine the origin and evolutionary trajectory of the bacterium. Whole genome sequences of 16 strains representing the five previously identified genetic populations of *R. toxicus* were used for SSR discovery and analysis. Ten SSRs were selected based on common parameters and two were PCR-verified on all available strains for confirmation of the sequence data. Individual SSRs greatly varied across genomes and the results of different combinations of concatenated SSRs generated different phylogenies. Some but not all SSR combinations corresponded to previous studies' results. SSR analyses were inconclusive and not predictive for phylogeny or geographic origin. To summarize, SSRs may lead to different conclusions depending on which SSRs are included in the analysis and how many SSRs are

analyzed. Caution should be applied when using SSR analysis for understanding evolutionary dynamics of *R. toxicus*. For diagnostic purposes, different SSR markers within *R. toxicus* had varying levels of specificity, with one SSR being able to differentiate to strain level, some SSRs to population level, and others only to species level, *R. toxicus* generally. The analyzed SSR markers were all unique to *R. toxicus* and created a spectrum of diagnostic specificity. *Rathayibacter toxicus* and its toxin has been a threat to Australian agriculture for decades and now threatens other countries because of the increase in globalized trade markets including hay exports from Australia. Countries including the U.S. may be at risk of *R. toxicus* introduction, which may include potential establishment, spread, and ultimate disease outbreaks. It is necessary to understand how the pathogen survives and spreads within and from the source country, how it is moved from source to destination country, and how environmental conditions at the destination may facilitate pathogen establishment. This qualitative pathway analysis was developed to study the possible avenues through which *R. toxicus* may elude containment and reach areas outside of outbreak centers in Australia. A scenario of *R. toxicus* spread was analyzed which included assuming a range of concentration of bacterial galls within a hay field harvested for hay bales for export out of the country. It is evident that risk exists for *R. toxicus* gall release at multiple points throughout its journey from a grass panicle in an Australian paddock, all the way through to the arrival on a packaged bale of hay reaching a feeding pasture in a U.S. livestock farm.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank many people for helping and supporting me throughout my time as a graduate research assistant. Dr. Stack has been a great advisor and I am extremely grateful for the opportunity to be a part of his Plant Biosecurity Lab. I learned very much about biosecurity, plant pathology, research, science and much more from him, and it has been an invaluable experience. The Plant Biosecurity Lab has had wonderful lab members throughout the years. Dr. Jarred Yasuhara-Bell and Javier Kiyuna have been great coworkers who are supportive and knowledgeable in many ways, and all other lab members have been great as well.

Also, big thank you to my committee members for their time, guidance, thoughtful questions and input: Dr. Sanzhen Liu and Dr. Thomas Platt. Dr. Stack's NPDN team, including Lindsey Ashmore and Chandler Day have also been helpful. I am also appreciative of the KSU diagnostic lab in letting me train in diagnostic skills, particular big thanks to Christian Webb and Judy O'Mara.The department of plant pathology at KSU has had extremely helpful and friendly professors and staff members. Several other professors helped me with understanding various scientific methods, aspects of my research, and administrative things, including but not limited to, Tim Todd, Dr. Cook, Dr. Kennelly, and Dr. Mehra.

Finally, I give a big thank you to my parents and brother for their endless support and encouragement, I could not have had success without their support throughout my entire life and education experience.

Chapter 1: Literature Review

Overview

Rathayibacter toxicus is a plant-pathogenic bacterium with a complex life cycle and the ability to cause death for animals that consume its galls. The *R. toxicus* host range includes many grass species in several genera, it is vectored by several species in the genus *Anguina*, and currently exists mostly in Australia, with occasional reports in South Africa. Understanding the origin, evolution, and dissemination pathway of *R. toxicus* is important for managing its movement and spread, and critical to developing effective prevention strategies.

Rathayibacter toxicus background

Rathayibacter toxicus was previously known as *Clavibacter toxicus, Corynebacterium sp*., and *Corynebacterium rathayi* (Davis et al. 1984). The earliest classification of coryneform phytopathogenic bacteria was the genus *Corynebacterium* (Dowson 1942). Species of *Corynebacterium* containing 2,4-diaminobutyric acid (DAB), a peptidoglycan cell wall group B component were separated into the genus *Clavibacter* (Davis et al. 1984, Carlson et al. 1982). Species within the genus *Clavibacter* containing the L-isomer of DAB were then separated into the genus *Rathayibacter* (Evtushenko and Takeuchi 1994; Rainey 1994; Goodfellow et al. 2012).

The genus *Rathayibacter* comprises nine gram-positive bacterial species; the distinctness of these species has been described (Zgurskaya et al. 1993). Species of *Rathayibacter* differ significantly among each other based on physiological phenotypes, cell-wall composition, multilocus enzyme profiles, and pigmentation. These species include *R. agropyri, R. caricis, R. festucae, R. iranicus, R. oskolensis, R. rathayi, R. tanaceti, R. toxicus,* and *R. tritici* (Zgurskaya et al. 1993; Murray 1986; Dorofeeva et al. 2002; Riley and Ophel 1992; Sasaki et al. 1998; Vasilenko et al. 2016; Dorofeeva et al. 2018; Schroeder et al. 2018). *Rathayibacter* species are vectored by

various *Anguina* species (Bertozzi and McKay 1995; McKay et al. 1993; Riley 1992; Riley and McKay 1991; Riley and McKay 1990), with the most common one associated with *R. toxicus* being *A. funesta*. *Rathayibacter toxicus* was distinguished, classified and named by Sasaki and colleagues in 1998. Its various genetic components, specific aspects of life cycle including toxicity and agricultural impacts, and several other features described later in this review significantly separate *R. toxicus* from the rest of the genus.

The exact historical and evolutionary origins of *Rathayibacter toxicus* are unknown. One of *R. toxicus*' major hosts, annual ryegrass (*Lolium rigidum*), was deliberately introduced as a pasture plant into Australia in 1880 (Kloot 1983). It is possible the bacterium existed in the environment prior to introduction of the host, adapted to the introduced grass species, and evolved over time. A more recent introduction of the bacterium into Australia is possible as well. *Rathayibacter toxicus* and livestock toxicities have also been reported in South Africa (Schneider 1981, Grewar et al. 2009), but not as widespread or frequent as in Australia. Several studies have analyzed genetic and genomic variation of the pathogen, revealing several distinct genetic populations, but have not determined the geographic or phylogenetic origin of the bacterium (Arif et al. 2016, Yasuhara-Bell et al. 2020). This information and an additional genetic analysis based on microsatellites are described in this study as a potential approach to investigate the geographic and phylogenetic origins of *R. toxicus*.

Life Cycle

Rathayibacter toxicus is associated with nematode vectors, specifically in the *Anguina* genus; the specific nematode species determines the plant host that *R. toxicus* infects. This association makes for a complex life cycle that involves multiple species of grass hosts, several species of the nematode vector, and different genetic populations of the pathogen with conducive

environmental factors. There are several grass species that the bacterium can infect and result in toxic bacterial galls. These include annual ryegrass, *Lolium rigidum*; annual beard grass, *Polypogon monspeliensis*; bent grass or blown grass, *Agrostis acenacea* (McKay and Ophel 1993a,b). Several species from the *Anguina* genus can vector this bacterium, including *Anguina agrostis* and *Anguina funesta* (Riley 1992, 1991) and several other less commonly known species; there are 14 *Anguina* species known to exist across Australia (Atlas of Living Australia 2020).

Rathayibacter toxicus and its nematode vectors overwinter in the soil (Price et al. 1979). As *Anguina* nematodes migrate through the soil toward their plant host, they may acquire the bacteria and vector them into flowering structures of grasses, where they colonize the ovules of the plant The nematodes and *R. toxicus* may compete at the infection site and transform potential seed into either a nematode gall or a bacterial gall. The bacterium produces toxins within bacterial galls during seed set at end of a season, and these are a type of corynetoxin now referred to as tunicamycins (Price et al. 1979; Stynes and Bird 1982; Government of Australia, 2018). The tunicamycins that *R. toxicus* produces are responsible for the poisoning of animals that graze on the diseased grasses. Infected grasses remain toxic after senescence and harvested as hay if grown in a hay production field (Price et al. 1979; Stynes and Bird 1982; Government of Australia 2018). Of all the *Rathayibacter* species, *R. toxicus* is the only one to be shown to produce toxins within the plant (Cunningham and Hartley 1959; Galey et al. 1997; Galloway 1961; Haag 1945; Jensen 1961; Kurochkina and Chizhov 1980; Shaw and Muth 1949). However, recent evidence indicates that the genes for the tunicamycin production are housed in the genome of at least one other species (Sechler et al. 2017). There are also studies that suggest the involvement of a bacteriophage in toxin production (Bird et al. 1980; Bird 1983, Ophel et al. 1993), although at least one study refutes the involvement of a bacterial virus (Kowalski et al. 2007).

Genetic and genomic analysis

Rathayibacter toxicus genetics and genomic studies have shown various unique features in this pathogen. The whole genome is approximately 2.3 megabases in length and is comprised of 61.5% GC content, while other *Rathayibacter* species have larger genome sizes and higher GC contents (Davis et al. 2018; Arif et al. 2016; Tancos et al. 2020, NCBI database). *Rathayibacter toxicus* has a CRISPR system, specifically the type I-E CRISPR-Cas system (Davis et al. 2018, Yasuhara-Bell et al. 2020). *Rathayibacter toxicus* has a tunicamycin gene cluster similar to the tunicamycin biosynthetic cluster from *Streptomyces chartreusis* (Sechler et al. 2017). The tunicamycin gene cluster in *R. toxicus* has 14 genes in two transcriptional units, all the elements necessary for tunicamycin biosynthesis (Sechler et al. 2017). In the tunicamycin gene cluster region of the genome, the GC content is significantly lower than the GC content in the rest of the genome (Sechler et al. 2017; Yasuhara-Bell et al. 2020), suggesting horizontal transfer events may have occurred in its evolutionary history resulting in the acquisition of the tunicamycin cluster (Sechler et al. 2017; Yasuhara-Bell et al. 2020).

Several genetic analyses were done on *R. toxicus* to determine variation and population structure. Strains included were isolated from various host species across several regions of Australia in different years and decades. These studies aimed to identify genomic and genetic differences across geographical, biological, and environmental factors. Genetic structure of bacterial populations can result from factors such as gene flow, host range and variation, and geographic spacing and isolation of the pathogen, its hosts, or its vectors (Fraser et al. 2009, Guerrero-Ferreira and Nishiguchi 2011).

Several studies identified distinct genetic populations of *R. toxicus* with varying geographical and temporal distributions. Multiple *R. toxicus* sub-specific groups were identified

through amplified fragment-length polymorphism (AFLP) analysis and pulse-field gel electrophoresis (PFGE) (Agarkova et al. 2006). They differentiated Western Australian isolates, named Group A, from South Australian isolates, named Group B and C. Group A isolates were later re-named by Arif et al. as RT-III, Group B as RT-II, and Group C comprised an outlier, FH100 (Agarkova et al. 2006; Johnston et al. 1996; Riley 1987; Riley et al. 1988, Arif et al. 2016) that was ultimately placed in the RT-IV population (Yasuhara-Bell and Stack 2019). The Arif et al. population genetic study of *R. toxicus* using multi-locus sequence typing (MLST) and inter-simple sequence repeat (ISSR) analysis confirmed the different populations and identified a newly emergent and genetically distinct population from South Australia, designated RT-I (Arif et al. 2016). The same group also described two additional genetic populations, RT-IV and RT-V, which came from New South Wales, Australia and South Australia respectively. Populations RT-I, RT-II, RT-III, RT-IV, and RT-V are backed by Multi Locus Sequence Typing (MLST), Inter Simple Sequence Repeats (ISSRs), and whole genome sequence analyses (Yasuhara-Bell et al. 2020, Yasuhara-Bell and Stack 2019; Arif et al. 2016). Population RT-1 was reported as a newly emerged population in South Australia of uncertain origin (Arif et al. 2016).

These studies used *R. toxicus* isolates from a period of forty years from four distinct geographic locations in Australia but were not able to determine the evolutionary origin of each *R. toxicus* population. Comparative genomic analyses determined that the genomes of *R. toxicus* populations contained genes unique to each population as well as genes shared among some but not all populations (Yasuhara-Bell et al. 2020). Fewer unique genes were identified between RT-II and RT-III, suggesting that these two populations were more closely related (Yasuhara-Bell et al. 2020). Twelve of the genes identified in RT-I and RT-IV are identical between those two populations and not present in RT-V, which may indicate that those two populations are closely

related to each other. The horizontal gene transfer analysis on population RT-V identified a thiopeptide-type bacteriocin biosynthesis protein, which may have come from a Streptomyces sp. (Yasuhara-Bell et al. 2020).

Another analysis often used by scientists to study genetic structure, population dynamics, and evolution of organisms is Simple Sequence Repeats (SSRs); short tandem nucleotide repeats that vary in size and number of repeats among and within genomes. They mutate at rates between 10^{-3} and 10^{-6} per cell generation, which is approximately 10 orders of magnitude greater than point mutations and mutations of the total genome (Gemayel et al., 2012). Repeat polymorphisms, or the differences in number of repeats of the same SSR motif among multiple genomes of organisms in question, usually result from addition or deletion of an entire repeat unit or multiple ones. This variation is usually a result of polymerase strand-slippage in DNA replication or by recombination errors. It is hypothesized that the longer and more "perfect" an SSR is, with perfect referring to the lack of single nucleotide polymorphisms (SNPs) within an SSR, the higher the mutation frequency, while shorter repeats with lower perfection may have a lower mutation frequency (Viera et al. 2016).

The SSRs are found in prokaryotic and eukaryotic organisms and are widely distributed through genomes and can occur in coding and non-coding nuclear and organellar DNA (Perez-Jimenez et al., 2013; Phumichai et al., 2015). They have been found in transcripts and regulatory regions of the genome, and there is evidence that microsatellites play a role in transcription and translation, organization of chromatin, genome size, and the cell cycle. (Nevo, 2001; Li et al., 2004; Gao et al., 2013).

SSRs can be very useful in studies of population structure, genetic mapping, and evolutionary processes (Viera et al. 2016). It is important to analyze SSRs in addition to the

existing genetic analyses because the faster mutation rate may aid in phylogenetic and phylogeographic determinations that the other methods could not fully answer. In addition to population genetics and phylogenetic analysis of *R. toxicus*, diagnostics are an important potential output of various genetic markers. SSRs may by useful for diagnostic purposes if the specific SSRs are unique to the level of classification in question, as far as species and perhaps sub-species levels. The accuracy of molecular-based diagnostics can be highly dependent on the level of understanding of genetic variation of any given pathogen species.

Agricultural Impacts

Outbreaks of toxicity and animal poisoning caused by *Rathayibacter toxicus* have been prevalent throughout fifty years in Australia, over several regions within the country, and are commonly known as "annual ryegrass toxicity" or "Flood plain staggers" outbreaks (McKay et al. 1993). Bacterial galls causing these outbreaks were described as early as 1956 (Fisher 1977) and was first attributed to *Corynebacterium* species (Fisher 1977). McIntosh et al. 1967 first concluded that there was a toxin involved with these galls. By 1967, there were large numbers of dying sheep and since they are very important to Australians and their economy, funding was allocated to begin researching this pathogen (Fisher 1977).

Throughout the last few decades, much research has been done to understand the impact of *R. toxicus* on farm animals. The corynetoxins, or specifically, tunicamycins, that *R. toxicus* produces can be lethal if a large enough number of bacterial galls/cells is consumed by an animal. For example, an adult sheep must consume around 20,000 galls for lethality. (Jago and Culvenor 1987; McKay and Ophel 1993). The tunicamycins of *R. toxicus* that cause toxicity can work cumulatively. When consumed, the toxin can accumulate in the animal up to a period of nine weeks and if the animal consumes enough toxic galls it can cause disease and death (Jago and Culvenor

1987; McKay and Ophel 1993; Grewar et al. 2009). Signs of *R. toxicus* toxicity may appear as soon as four days or as late as several weeks after animals are introduced to toxic feed through pasture grazing, hay, or grain. If the animal recovers, it takes up to six months to clear all effects of the toxin. If clinical signs appear sooner than four days after stock are introduced to a paddock, stock have ingested toxin before being moved to the new paddock (Government of Australia, 2018).

Multiple animal species can be affected if they consume toxic bacterial galls produced by *R. toxicus*. Sheep, goats, horses, cattle, and other livestock have experienced toxicity, often lethal, from consuming the galls (Government of Australia, 2018; Edgar et al. 1982; Finnie 2006; Eckardt 1983; Sechler et al. 2017; Jago et al. 1983). The list of vulnerable animal species is not exclusive to farm mammals and may include unverified examples of wildlife (Murray et al. 2015). Poisoned animals exhibit several different symptoms, mostly of a neurological nature, including convulsions, paralysis, excessive salivation, muscle and head tremors, abortions, and death (Bourke et al. 1992; Finnie 2006). The neurological symptoms are due to the toxins inhibiting N-linked glycosylation, which leads to oxygen deprivation and tissue damage; hepatic damage is another effect (Jago et al. 1983; Finnie and Jago, 1985; Bourke et al. 1992; Finnie 2006).

Rathayibacter toxicus poisoning does not have a treatment or cure, but some affected animals may recover after a long period of time of feeding on toxin-free feed and with abundant hydration. When affected animals are found, they should be immediately transferred to a 'safe' paddock with good water, safe feed and shade. A safe paddock with safe feed is one that has been tested negative for *R. toxicus* during the same season. Animals with the most severe symptoms should be protected from the extreme weather and predation. Animals that are unable to rise within 12 hours are usually humanely euthanized (Government of Australia, 2018).

The economic impact of this toxicity has been immense. Since detailed recordkeeping of the animal deaths began in the late 1960s, more than 500,000 total animals have died due to *R. toxicus* poisoning in Western Australia alone (Page 2004). *Rathayibacter toxicus* impact on plant agriculture economics is significant as well. It has caused disease in over 10 million hectares of grassland with losses that amount to an estimated 37 million USD for the plant loss only. (Carslake 2006; Kessell 2010).

Biosecurity implications

Due to the widespread prevalence of *R. toxicus* in Australia and the high frequency of toxicity outbreaks, the country has had phytosanitary inspection requirements for export of hay to other countries for several years (Government of Australia 2005). There are several major biosecurity concerns for *R. toxicus.* One is that exporting toxic hay and plant materials may result in livestock health issues, and another is that such export may lead to establishment of *R. toxicus* in a new location. It is essential to quickly and accurately detect such potentially dangerous pathogens before they are exported and get introduced into new locations (Ouyang et al. 2013). However, the possibility of false negatives in any pathogen detection test always exists and may lead to inaccurate conclusions made about the presence/absence of *R. toxicus* in any sampled material; often times the importing country does not retest following a negative test for certain pathogens.

Rathayibacter toxicus was designated a Plant Pathogen Select Agent by the U.S. Department of Agriculture (USDA) and Animal and Plant Health Inspection Service (APHIS) in 2008 (Murray et al. 2015). The Federal Select Agent program is managed by the Centers for Disease Control and Prevention (CDC) and USDA's APHIS, with the latter maintaining and designating plant and animal pathogens. The purpose of the U.S. Select agent program is to oversee

work with biological select agents and toxins that have the potential to threaten public, animal, or plant health to animal or plant products (selectagents.gov 2017). The Select Agent program sets standards and rules for how to handle select agent material, including conducting experimental research and maintaining inventories and collections of Select Agent strains (selectagents.gov 2017). At Kansas State University, Select Agent work is required to be conducted at the Biosecurity Research Institute, a Biosafety Level-3 laboratory facility. Conducting research at a BSL-3 laboratory coupled with the Select Agent guidelines specific to different types of pathogens further reduces the risk of spread and release of *R. toxicus* outside of the laboratory environment.

BSL-3 laboratories house research with infectious agents or toxins, including ones that that may be transmitted through the air. Experiments with live agents are performed primarily in biosafety cabinets that use high efficiency particulate air (HEPA)-filtered, carefully controlled air flow or sealed enclosures to prevent infection, as well as strict guidelines for movement of samples, secondary containers, detailed inventory procedures, and adherence to strict biosafety procedures specific for each experiment and pathogen. The laboratory parameters and conditions which are designed to keep any pathogen contained are extensive. Several of the factors of laboratory design include unidirectional air flow using room pressure gradients of negative pressure, exhaust air being HEPA -filtered and proper procedures for disposal of biomedical waste (Richmond, 2000).

Rathayibacter toxicus is one of only seven plant pathogens on the U.S. Select agent list, where there are a total of 60 animal, human pathogens and toxins. *Rathayibacter toxicus* is a U.S. Select Agent because of its ability to cause disease in plants and toxicities in animals which consume the affected plants (Riley and Ophel 1992). To date, there are no official reports of *R. toxicus* in the U.S. It has only been confirmed in the southern hemisphere, mostly in Australia, with some reports in South Africa (Schneider 1981). However, in the current globalized market

and movement of people, imports, and exports, it may be difficult to detect and stop the spread of pathogens and complete containment can never be fully guaranteed.

The United States has a significant livestock and hay industry. An introduction or establishment of *Rathayibacter toxicus* could be detrimental to the health and well-being of forage grasses and animals, as well as the livelihoods of farmers who depend on them. The cattle industry has an annual value of 67 billion USD, the hog industry at 22 billion USD, and sheep and goat industry while much smaller, is in the tens of millions USD (USDA ERS 2019). In terms of numbers of some of the most susceptible animals, there are approximately 95 million cattle, 6 million sheep, 9 million horses, all of which are important to farmers and their livelihoods (Murray et al. 2015). There is also a significant hay industry in the U.S. which may be susceptible to establishment of *R. toxicus* in the event of an introduction. The U.S. produced 131.5 million tons of hay from 53.8 million acres harvested valued at 16.2 billion dollars USD in 2017 (NASS 2017). The presence of grass hosts would increase the likelihood of establishment of *R. toxicus* following an introduction. Even without host grass species, *R. toxicus* could be introduced into any country along with host seeds present in the hay material or in seeds as a commodity. This could result in the accidental introduction of the host, pathogen, and vector simultaneously into the environment. Australia and the USA share many climate characteristics across both countries, including similar weather patterns (Weather History and Climate Data, Worldclimate.com). Many *Anguina* species of nematodes already exist in the US, which adds the other required component of the *Rathayibacter toxicus* life cycle besides the host, pathogen, and environment. The National Plant Disease Recovery System (NPDRS) developed a recovery plan for this pathogen (Murray et al. 2015), but as with any pathogen threat, it is best to seek ways of prevention through gaining understanding of the pathogen's movement and ability to establish in any location.

Pathogen Pathway Risk Analysis

People have transported and traded plants and animals for millennia, and currently there is a large global market of agriculture goods and trade occurs between all countries (Gordh and McKirdy 2014). This boosts economic growth and provides countries with supplies and sustenance they need (Koo and Kennedy 2005; Piggott 2013). The Industrial Revolution led to a period of increased international trade across most continents and facilitated by the construction of canals, highways, railways, and steamships (Findlay and O'Rourke 2007). Along with this comes an increase in risk of introducing exotic invasive species throughout the trade routes and destinations of goods (McCullough et al. 2006; Brasier 2008). Exotic or alien species comprise plants, animals, and microbial pathogens introduced in a region through human activities and are the focal point of regulatory activities addressing plant biosecurity (Gordh and McKirdy 2014). Invasive pests include weeds, plant pathogens and animals that act in a negative way to consume or impact agricultural commodities and plant resources (Gordh and McKirdy 2014).

Historically, introduction of plant pathogens by natural means into new locations has been uncommon, and human activities are most often the main cause of such introductions (NRS 2002). Pests and pathogens often travel as contaminants of plants and plant products that are traded between countries and various biogeographical zones (Brasier 2008; NRC 2002). There are enormous amounts of plants and plant products being traded across the world and pathogens are often present within such commodities which makes and raises a risk factor of introducing unwanted invasive pests (Brasier 2008). Trade facilitates spread of alien species directly through unintentional introduction as contaminants of cargo or stowaways within different modes of transport (Hulme 2009).

Entities such as the International Plant Protection Convention (IPPC), formed in 1951, exist to create International Standards for Phytosanitary Measures (ISPMs) in order to minimize plant pest risks associated with trade. The World Trade Organization is also involved in the creation of an Sanitary-Phytosanitary (SPS) Agreement, which is in coordination with IPPC for applying sanitary and phytosanitary measures by governments based on international standards or risk assessments (WTO 1995). There are certain risk analyses that can be made for assessing likelihoods of exotic pests and pathogens spread to and establishing in new locations. ISPMs contain guidelines for pest risk analysis in order to protect plants and plant products from pests and pathogens that can be moved in the course of trade (Hulme 2011). Main concerns of the SPS agreement included setting measures for risk assessments that include three steps: identifying pests of concern, evaluating their likelihood of entry, establishment, spread, potential consequences, and re-evaluating likelihoods and consequences against potential measures. Risk analysis for plant pests include identification of pathways that carry potential pest hazards and then calculating the probability of entry, establishment, and spread. The entry probability describes the likelihood that a pathogen can arrive in a new area in a viable state on an identified pathway, and establishment determines the conduciveness of the new location on the survival and propagation of the pathogen (2.2.1.1 ISPM 11, FAO 2004a,b).

Entry probability factors to consider include association of pathogen with the pathway and life cycle stages/survival, environmental conditions in country of origin, prevalence of pathogen through phytosanitary measures and cultivation within origin, and quantity of the associated trade. Assessing probability of entry can include developing scenarios with key steps in export and import of the commodity. Establishment probability factors include availability of hosts, vectors, conducive environment/habitat, climate conditions, abiotic factors such as soil

conditions and topography, presence of management practices, and more. Assessing establishment can be done using various methods including the use of Geographic Information Systems (GIS) data of environmental factors and host/vector presence, which can result in developing hypotheses on what specific areas may be most suitable for the pathogen (FAO 2004a,b; Sansford et al. 2009; FAO 2010).

Researching pathogen risk analysis can include qualitative or quantitative methods or a mix of both. Qualitative methods are answerable to data and fundamental rules of probability and formal logic, and quantitative methods rely on subjective judgment to formulate models and estimate parameters. Different combinations of qualitative and quantitative methods are used for pathogen risk analysis, which includes reasoning based on biological systems, trade pathways and impacts, and quantitative methods that can also be expressed qualitatively such as point-scoring systems, logical rules, and Monte Carlo simulation. The World trade organization SPS agreement does not require a specific method to be used and recommends that an analysis should be enough for its purpose and not any more complex than is necessary. Qualitative analysis is the most common type of assessment for routine decision making in most authorities and governments involved with this treaty and phytosanitary establishments. To summarize, there are many properties of pathogen risk analysis that are most desirable, including pathway assessments, volume of trade, defining terms related to likelihoods and consequences precisely, including assessments of economic, social and environmental impacts, highlighting uncertainties and gaps in information, and more (Gordh and McKirdy 2014, WTO 1995, Sansford 2002, OIE 2010, MacLeod et al. 2005; FAO 2004a, Biosecurity New Zealand 2006, USDA 2000). Qualitative pathway analysis and the aforementioned factors associated with the analysis are some of the most relevant to analyzing *Rathayibacter toxicus* risks.

References (Chapter 1)

Agarkova, I. V., Vidaver, A. K., Postnikova, E. N., Riley, I. T., & Schaad, N. W. (2006). Genetic characterization and diversity of *Rathayibacter toxicus*. *Phytopathology*, *96*(11), 1270-1277.

Arif, M., Busot, G. Y., Mann, R., Rodoni, B., Liu, S., & Stack, J. P. (2016). Emergence of a new population of *Rathayibacter toxicus*: an ecologically complex, geographically isolated bacterium. *PLoS One*, *11*(5), e0156182.

Atlas of Living Australia (2020). Online: https://www.csiro.au/en/Research/Collections/ALA

Biosecurity New Zealand (2006) risk analysis procedures version 1. Ministry of Agriculture and Forestry, Wellington

Bird, F. A., & Riddle, L. D. (1984). Effect of attachment of *Corynebacterium rathayi* on movement of *Anguina agrostis* larvae. *International journal for parasitology*, *14*(5), 503-511.

Bird, A. F., Stynes, B. A., & Thomson, W. W. (1980). A comparison of nematode and bacteria-colonized galls induced by *Anguina agrostis* in *Lolium rigidum*. *Phytopathology*, *70*(11), 04-11.

Bourke, C. A., Carrigan, M. J., & Love, S. C. J. (1992). Flood plain staggers, a tunicaminyluracil toxicosis of cattle in northern New South Wales. *Australian veterinary journal*, *69*(9), 228-229.

Brasier, C. M. (2008). The biosecurity threat to the UK and global environment from international trade in plants. *Plant Pathology*, *57*(5), 792-808.

Carlson, R. R., & Vidaver, A. K. (1982). Taxonomy of *Corynebacterium* plant pathogens, including a new pathogen of wheat, based on polyacrylamide gel electrophoresis of cellular proteins. *International Journal of Systematic and Evolutionary Microbiology*, *32*(3), 315-326.

Carslake, T. (2006). Livestock: Ryegrass toxicity. *Farming Ahead* 178:62-63.

Cunningham, I. J., & Hartley, W. J. (1959). Ryegrass staggers. *New Zealand veterinary journal*, *7*(1), 1-7.

Davis, M. J., Gillaspie Jr, A. G., Vidaver, A. K., & Harris, R. W. (1984). *Clavibacter*: a new genus containing some phytopathogenic coryneform bacteria, including *Clavibacter xyli* subsp. xyli sp. nov., subsp. nov. and *Clavibacter xyli* subsp. *cynodontis* subsp. nov., pathogens that cause ratoon stunting disease of sugarcane and bermudagrass stunting disease. *International Journal of Systematic and Evolutionary Microbiology*, *34*(2), 107-117.

Davis, E. W., Tabima, J. F., Weisberg, A. J., Lopes, L. D., Wiseman, M. S., Wiseman, M. S., Pupko, T., Belcher, M. S., Sechler, A. J., Tancos, M. A., Schroeder, B. K., Murray, T. D., Luster, D. G., Schneider, W. L., Rogers, E. E., Andreote, F. D., Grünwald, N. J., Putnam, M. L., Chang, J. H. (2018). Evolution of the U.S. Biological Select Agent *Rathayibacter toxicus*. *MBio*, 9(4).

Dorofeeva, L. V., Evtushenko, L. I., Krausova, V. I., Karpov, A. V., Subbotin, S. A., & Tiedje, J. M. (2002). *Rathayibacter caricis* sp. nov. and *Rathayibacter festucae* sp. nov., isolated from the phyllosphere of *Carex* sp. and the leaf gall induced by the nematode *Anguina graminis* on *Festuca rubra* L., respectively. *International journal of systematic and evolutionary microbiology*, *52*(6), 1917-1923.

Dorofeeva, L.V.; Starodumova, I.P.; Krauzova, V.I.; Prisyazhnaya, N.V.; Vinokurova, N.G.; Lysanskaya, V.Y.; Tarlachkov, S.V.; Evtushenko, L.I. (2018). *Rathayibacter oskolensis* sp. nov., a novel actinobacterium from Androsace koso-poljanskii Ovcz.(Primulaceae) endemic to the Central Russian Upland. *International journal of systematic and evolutionary microbiology*, *68*(5), 1442-1447.

Dowson WJ. (1942). On the generic name of the gram-positive bacterial plant pathogens. Trans Br Mycol Soc. 1942;25:311–314. doi: 10.1016/S0007-1536(42)80020-0.

Eckardt, K. (1983). Tunicamycins, streptovirudins, and corynetoxins, a special subclass of nucleoside antibiotics. *Journal of natural products*, *46*(4), 544-550.

Edgar, J.A.; Frahn, J.L.; Cockrum, P.A.; Anderton, N.; Jago, M.V.; Culvenor, C.C.J.; Jones, A.J.; Murray, K.; Shaw, K.J. (1982). Corynetoxins, causative agents of annual ryegrass toxicity; their identification as tunicamycin group antibiotics. *Journal of the Chemical Society, Chemical Communications*, (4), 222-224.

FAO (2004a). International standards for phytosanitary measures publication no. 21: pest risk analysis for regulated non-quarantine pests, including analysis of environmental risks and living modified organisms. Secretariat of the International Plant Protection Convention, Food and Agriculture Organisation of the United Nations, Rome.

FAO (2004b) International standards for phytosanitary measures publication no 21: pest risk analysis for regulated non-quarantine pests. Secretariat of the International Plant Protection Concention, Food and Agriculture Organisation of the United Nations, Rome

FAO (2010) International standards for phytosanitary measures publication no. 5. Glossary of phytosanitary terms. Secretariat of the International Plant Protection Convention, Food and Agriculture Organisation of the united Nations, Rome.

Findlay, R., & O'Rourke, K. H. (2007). Power and Plenty: Trade. *War, and the World Economy in the Second Millennium, Princeton.*

Finnie, J. W. (2006). Review of corynetoxins poisoning of livestock, a neurological disorder produced by a nematode-bacterium complex. *Australian Veterinary Journal*, *84*(8), 271-277.

Finnie, J. W., & Jago, M. V. (1985). Experimental production of annual ryegrass toxicity with tunicamycin. *Australian veterinary journal*, *62*(7), 248-249.

Fisher, J. M. (1977). Annual ryegrass toxicity. *Biennial Report of the Waite Agricultural Institute, 1976-77, Pages 7-11. University of Adelaide, Glen Osmond, South Australia.*

Fraser, C., Alm, E. J., Polz, M. F., Spratt, B. G., & Hanage, W. P. (2009). The bacterial species challenge: making sense of genetic and ecological diversity. *Science*, *323*(5915), 741-746.

Galey, F. D., Blanchard, P. C., Adaska, J., Puschner, B., Holstege, D., McDougald, N., Garthwaite, I., Miles, C., and Whitehead, G. C. (1997). Staggers in pastured beef cattle in the central valley of California. Proceedings 40th American Association Veterinary Laboratory Diagnosticians. Louisville, KY.

Galloway, J. H. (1961). Grass seed nematode poisoning in livestock. *Journal of the American Veterinary Medical Association*, *139*(11), 1212-1214.

Gao, C., Ren, X., Mason, A. S., Li, J., Wang, W., Xiao, M., & Fu, D. (2013). Revisiting an important component of plant genomes: microsatellites. *Functional Plant Biology*, *40*(7), 645-661.

Gemayel, R., Cho, J., Boeynaems, S., & Verstrepen, K. J. (2012). Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes*, 3(3), 461-480.

Gordh and S. McKirdy (eds.) (2014). The Handbook of Plant Biosecurity, DOI 10.1007/978-94-007-7365-3_1, (C) Springer Science+Business Media Dordrecht

Government of Australia (2005) Standard for Minimising the Risk of Corynetoxin Contamination of Hay and Straw for export. Australian Quarantine and Inspection Service. Department of Agriculture and Water Resources, Australian Government. http://www.agriculture.gov.au/biosecurity

Government of Western Australia, (2018). Hay Production in the Wheatbelt, Wheatbelt Development Commission. Online: https://www.wheatbelt.wa.gov.au/news/hay-production-wheatbelt/

Government of Western Australia (2018). Annual Ryegrass Toxicity in Livestock. Online: https://www.agric.wa.gov.au/livestock-biosecurity/annual-ryegrass-toxicity-livestock

Grewar, J. D., Allen, J. G., & Guthrie, A. J. (2009). Annual ryegrass toxicity in thoroughbred horses in ceres in the Western Cape Province, South Africa. *Journal of the South African Veterinary Association*, *80*(4), 220-223.

Guerrero-Ferreira, R. C., & Nishiguchi, M. K. (2011). Bacterial biodiversity in natural environments. The Importance of Biological Interactions in the Study of Biodiversity, 1.https://bie.ala.org.au/species/urn:lsid:biodiversity.org.au:afd.taxon:d002fea4-b6bb-442d-bf3c-870ae2c0fc19

Haag, J. R. (1945). Toxicity of nematode infested Chewings fescue seed. Science (Washington), 102(2651), 406-407.

K-State Biological Safety 2020. Online: https://www.k-state.edu/safety/lab/labsafety/topics/biologicalsafety/selectAgents.html

Hulme, P.E. (2011). Biosecurity: the changing face of invasion biology. In: DM Richardson (ed) Fifty years of invasion ecology: the legacy of Charles Elton, 1st edn. Blackwell, Oxford

IPPC (1997) International plant protection convention (New revised text). FAO, Rome

Jago, M. V., Payne, A. L., Peterson, J. E., & Bagust, T. J. (1983). Inhibition of glycosylation by corynetoxin, the causative agent of annual ryegrass toxicity: a comparison with tunicamycin. Chemico-biological interactions, 45(2), 223-234.

Jago, M.V., Culvenor, C.C.J. (1987). Tunicamycin and corynetoxin poisoning in sheep. Aust. Vet. J. 64:232-35.

Jensen, H. J. (1961). Nematodes affecting Oregon Agriculture. Oregon State Univ. Agric. Exp. Stn. Bull, 579.

Johnston, M. S., Sutherland, S. S., Constantine, C. C., & Hampsons, D. J. (1996). Genetic analysis of *Clavibacter toxicus*, the agent of annual ryegrass toxicity. Epidemiology & Infection, 117(2), 393-400.

Kessell, D. (2010). Annual ryegrass toxicity-Current situation. W.A. Department of Agriculture and Food, ed. Western Australian Agriculture Authority, South Perth, Western Australia.

Kloot, P. M. (1983). The genus *Lolium* in Australia. Australian Journal of Botany, 31(4), 421-435.

Koo, W., & Kennedy, P. L. (2005). International trade and agriculture. *Management of Environmental Quality: An International Journal*.

Kowalski, M. C., Cahill, D., Doran, T. J., & Colegate, S. M. (2007). Development and application of polymerase chain reaction-based assays for *Rathayibacter toxicus* and a

bacteriophage associated with annual ryegrass (*Lolium rigidum*) toxicity. Australian Journal of Experimental Agriculture, 47(2), 177-183.

Kurochkina, K. G., & Chizhov, V. N. (1980). On the toxicity of *Anguina agrostis* galls to animals. *IX Konferentsiya Ukrainskogo Parazitologicheskogo Obshchestva. Tezisy dokladov. Chast'2*.

Li, Y. C., Korol, A. B., Fahima, T., & Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. Molecular biology and evolution, 21(6), 991-1007.

MacLeod, A., Baker, R. H. A., & Hoddinott, R. (2005, October). The analysis of detections in consignments to identify and target pests' entry pathways. In *Proceedings of the British crop protection council international congress, crop science and technology, Glasgow* (pp. 1013-1018).

McCullough, D. G., Work, T. T., Cavey, J. F., Liebhold, A. M., & Marshall, D. (2006). Interceptions of nonindigenous plant pests at US ports of entry and border crossings over a 17-year period. *Biological invasions*, *8*(4), 611.

McIntosh, G. H., & Thomas, M. R. (1967). Toxicity of parasitised Wimmera ryegrass, *Lolium rigidum*, for sheep and cattle. Australian Veterinary Journal, 43, 349-353.

McKay, A. C., & Ophel, K. M. (1993). Toxigenic *Clavibacter/Anguina* associations infecting grass seedheads. Annual Review of Phytopathology, 31(1), 151-167.

Murray, T. D. (1986). Isolation of *Corynebacterium agropyri* from 30-to 40-year-old herbarium specimens of *Agropyron* species. Plant disease, 70(5), 378-380.

Murray, T.D., Agarkova, I., Alderman, S., Allen, J., Bulluck, R., Chitamber, J., Divan, C., Riley, I., Schroeder, B., Sechler, A., and Subbotin, S. (2015). Recovery plan for *Rathayibacter* poisoning caused by *Rathayibacter toxicus* (syn. *Clavibacter toxicus*). National Plant Disease Recovery System, a cooperative project of The American Phytopathological Society and The United States Department of Agriculture

National Agricultural Statistics Service (NASS), (2017). Online: https://www.nass.usda.gov/Charts_and_Maps/Crops_County/index.php#oh

Nevo, E. (2001). Evolution of genome–phenome diversity under environmental stress. *Proceedings of the National Academy of Sciences*, *98*(11), 6233-6240.

OIE (2010) Handbook on import risk analysis for animals and animal products, vol 1. Introduction and qualitative risk analysis, 2nd edn. Office International des Epizooties, Paris

Ouyang, P., Arif, M., Fletcher, J., Melcher, U., & Corona, F. M. O. (2013). Enhanced reliability and accuracy for field deployable bioforensic detection and discrimination of *Xylella*

fastidiosa subsp. pauca, causal agent of citrus variegated chlorosis using Razor Ex technology and TaqMan quantitative PCR. *PloS One*, *8*(11), e81647.

Page, S. (2004). Annual Ryegrass Toxicity - Review of Current Management Options and Research Needs.

Phumichai, C., Phumichai, T., & Wongkaew, A. (2015). Novel chloroplast microsatellite (cpSSR) markers for genetic diversity assessment of cultivated and wild Hevea rubber. *Plant molecular biology reporter*, *33*(5), 1486-1498.

Piggott, N. 2013. Agricultural Economic and Policy Outlook. Presentation to: 8th Annual Ag Development Forum. Governor Holshouer Building, Raleigh.

Price, N. P., & Tsvetanova, B. (2007). Biosynthesis of the tunicamycins: a review.

Price, P. C., Fisher, J. M., & Kerr, A. (1979). Annual ryegrass toxicity: parasitism of *Lolium rigidum* by a seedgall forming nematode (*Anguina* sp.). *Annals of Applied Biology*, *91*(3), 359-369.

Rainey, F., Weiss, N., Prauser, H., & Stackebrandt, E. (1994). Further evidence for the phylogenetic coherence of actinomycetes with group B-peptidoglycan and evidence for the phylogenetic intermixing of the genera *Microbacterium* and *Aureobacterium* as determined by 16S rDNA analysis. *FEMS microbiology letters*, *118*(1-2), 135-139.

Riley, I. T. (1987). Serological relationships between strains of coryneform bacteria responsible for annual ryegrass toxicity and other plant-pathogenic corynebacteria. *International Journal of Systematic and Evolutionary Microbiology*, *37*(2), 153-159.

Riley, I. T., & Ophel, K. M. (1992). *Clavibacter toxicus* sp. nov., the bacterium responsible for annual ryegrass toxicity in Australia. *International Journal of Systematic and Evolutionary Microbiology*, *42*(1), 64-68.

Riley, I. T., Reardon, T. B., & Mckay, A. C. (1988). Genetic analysis of plant pathogenic bacteria in the genus *Clavibacter* using allozyme electrophoresis. *Microbiology*, *134*(11), 3025-3030.

Sansford, C. (2002, March). Quantitative versus qualitative: Pest risk analysis in the UK and Europe including the European and Mediterranean Plant Protection (EPPO) system. In *NAPPO International Symposium on Pest Risk Analysis Puerto Vallarta, Mexico*.

Sansford, C. E., Inman, A. J., Baker, R., Brasier, C., Frankel, S., De Gruyter, J., ... & Steeghs, M. (2009). Report on the risk of entry, establishment, spread and socio-economic loss and environmental impact and the appropriate level of management for Phytophthora ramorum for the EU. *Deliverable report*, *28*.

Sasaki, J., Chijimatsu, M., & Suzuki, K. I. (1998). Taxonomic significance of 2, 4-diaminobutyric acid isomers in the cell wall peptidoglycan of actinomycetes and reclassification of *Clavibacter toxicus* as *Rathayibacter toxicus* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*, *48*(2), 403-410.

Schneider, D. J. (1981). First report of annual ryegrass toxicity in the Republic of South Africa. *Onderstepoort Journal of Veterinary Research.* 48:251-255

Schroeder, B. K., Schneider, W. L., Luster, D. G., Sechler, A., & Murray, T. D. (2018). *Rathayibacter agropyri* (non O'Gara 1916) comb. nov., nom. rev., isolated from western wheatgrass (*Pascopyrum smithii*). *International journal of systematic and evolutionary microbiology*, *68*(5), 1519-1525.

Sechler, A. J., Tancos, M. A., Schneider, D. J., King, J. G., Fennessey, C. M., Schroeder, B. K., Murray, T.D., Luster, D. G., Schneider, W. L., Rogers, E. E. (2017). Whole genome sequence of two *Rathayibacter toxicus* strains reveals a tunicamycin biosynthetic cluster similar to *Streptomyces chartreusis*. *PloS one*, *12*(8), e0183005.

Shaw, J. N., & Muth, O. H. (1949). Some types of forage poisoning in Oregon cattle and sheep. *Journal of the American Veterinary Medical Association*, *114*, 315-317.

Stynes, B. A., & Bird, A. F. (1982). Development of galls induced in *Lolium rigidum* by *Anguina agrostis*. *Phytopathology*, *72*(3), 336-346.

Tancos, M. A., Sechler, A. J., & Davis, E. W. (2019). The Identification and Conservation of Tunicaminyluracil-Related Biosynthetic Gene Clusters in Several *Rathayibacter* Species Collected From Australia, Africa, Eurasia, and North America. *Frontiers in microbiology*, *10*.

USDA ERS (2019). Animal products. Online: https://www.ers.usda.gov/topics/animal-products/

USDA (2000). Guidelines for pathway-initiated pest risk asessments, version 5.02 Animal and Plant Health Inspection Service, US Department of Agriculture, Riverdale.

Vasilenko, O. V., Starodumova, I. P., Tarlachkov, S. V., Dorofeeva, L. V., Avtukh, A. N., & Evtushenko, L. I. (2016). Draft genome sequence of "*Rathayibacter tanaceti*" strain VKM Ac-2596 isolated from *Tanacetum vulgare* infested by a foliar nematode. *Genome announcements*, *4*(3).

Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. D. F. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetics and molecular biology*, *39*(3), 312-328.

WTO (1995) SPS Agreement. World Trade organization agreement on the application of sanitary and phytosanitary measures. World Trade Organization, Geneva, Switzerland.

Yasuhara-Bell, J., & Stack, J. P. (2019). Panel of three loop-mediated isothermal amplification assays differentiates *Rathayibacter toxicus* populations RT-I, RT-II, RT-III, RT-IV and RT-V. *Journal of Plant Pathology*, *101*(3), 707-717.

Yasuhara-Bell, J., Arif, M., Busot, G. Y., Mann, R., Rodoni, B., & Stack, J. P. (2020). Comparative Genomic Analysis Confirms Five Genetic Populations of the Select Agent, *Rathayibacter toxicus*. *Microorganisms*, *8*(3), 366.

Zgurskaya, H. I., Evtushenko, L. I., Akimov, V. N., & Kalakoutskii, L. V. (1993). *Rathayibacter* gen. nov., including the species *Rathayibacter rathayi* comb. nov., *Rathayibacter tritici* comb. nov., *Rathayibacter iranicus* comb. nov., and six strains from annual grasses. *International Journal of Systematic and Evolutionary Microbiology*, *43*(1), 143-149.

Chapter 2: Simple Sequence Repeat analysis of *Rathayibacter toxicus*

Abstract

Rathayibacter toxicus is a plant pathogenic bacterium vectored by *Anguina* species (nematodes) and hosted by grass species in several genera in the *Poaceae* family including, *Lolium*, *Lacnagrostis*, and *Polypogon*. *Rathayibacter toxicus* forms bacterial galls in developing grass ovules where it produces a dangerous neurotoxin, tunicamycin. Grazing animals that consume toxic galls develop convulsions and staggers, often resulting in death. *Rathayibacter toxicus* is indigenous to Australia, but concern for global spread and potential impact to livestock industries resulted in its designation as a U.S. Select Agent pathogen. Several genetic and genomic studies including MLST, ISSR, and WGS support genetic population groupings and identified a newly emerged population (RT-I) in South Australia. Unique attributes of RT-I preclude determination of its geographic or phylogenetic origin. Simple Sequence Repeats (SSRs) are short tandem nucleotide repeats that vary in size and number of repeats among and within genomes and have been useful in determining evolutionary heritage. Unlike other genome components, SSRs generally have faster mutation rates and have been used as the basis of phylogenetic and phylogeographic analyses in many organisms including bacteria and plants. SSRs in *R. toxicus* were identified and analyzed to help determine the origin and evolutionary trajectory of the bacterium. Whole genome sequences of 16 strains representing the five previously identified genetic populations of *R. toxicus* were used for SSR discovery and analysis. Ten SSRs were selected based on common parameters and two were PCR-verified on all available strains for confirmation of the sequence data. Individual SSRs greatly varied across genomes and the results of different combinations of concatenated SSRs generated different phylogenies. Some but not all SSR combinations corresponded to previous studies' results. SSR analyses were inconclusive and

not predictive for phylogeny or geographic origin. To summarize, SSRs may lead to different conclusions depending on which SSRs are included in the analysis and how many SSRs are analyzed. Caution should be applied when using SSR analysis for understanding evolutionary dynamics of *R. toxicus*. For diagnostic purposes, different SSR markers within *R. toxicus* had varying levels of specificity, with one SSR being able to differentiate to strain level, some SSRs to population level, and others only to species level, *R. toxicus* generally. The analyzed SSR markers were all unique to *R. toxicus* and created a spectrum of diagnostic specificity.

Introduction

Simple sequence repeats (SSRs) are short tandem nucleotide repeats that vary in numbers of repeats and size of repeat motifs and are found in genomes of all known organism types. The three main classifications of SSRs are microsatellites, minisatellites, and macrosatellites, and they differ in nucleotide motif repeat number range. Micro, mini, and macro satellites are SSRs that have repeat motifs of 1-10, 10-100, and over 100 nucleotides, respectively (Gemayel et al. 2012, Oliveira et al. 2006). When referring to any group of same type or different combinations of satellite types, the umbrella term SSRs can be used.

The SSRs are found in prokaryotic and eukaryotic organisms and are widely distributed through genomes and can occur in coding and non-coding nuclear and organellar DNA (Perez-Jimenez et al., 2013; Phumichai et al., 2015). They have been found in transcripts and regulatory regions of the genome, and there is evidence that some SSRs play a role in transcription and translation, organization of chromatin, genome size, and the cell cycle. (Nevo, 2001; Li et al., 2004; Gao et al., 2013). The great variation and polymorphism in DNA satellites has been the basis for studies on genetic structure, population genetics, evolutionary genomics, DNA fingerprinting, lineage analysis, and gene mapping (Fondon et al. 2008).

SSRs have been long known for their ability to vary greatly without causing alterations in the flanking regions around them (Richards and Sutherland 1994; Kovtun and McMurray 2008). Repeat polymorphisms, or the differences in number of repeats of the same SSR motif among multiple genomes of organisms in question, usually result from addition or deletion of an entire repeat unit or multiple ones. This variation is usually a result of polymerase strand-slippage in DNA replication or by recombination errors. Strand slippage replication is a DNA replication error in which the template and strands are mismatched, and then the template strand can loop

out, causing contractions and expansions, which results in repeats increasing or decreasing in number (Vieira et al. 2016). SSRs also can be considered “perfect” or “imperfect”. Perfect SSRs are ones where repeat motifs are identical throughout each consecutive repeat. They can also be imperfect if a nucleotide not belonging to the motif occurs between repeats, if a motif sequence is interrupted by an insertion of one or a few base pairs or replaces any of its nucleotides, or if deletion of nucleotides (but not an entire motif) occurs (Oliveira et al. 2006, Mason 2015). The imperfections of insertion, substitution, or deletion of a nucleotide is also generally referred to as single nucleotide polymorphisms (SNPs). SSRs with core repeat motifs 3-5 nucleotides long are most commonly preferred in research in forensics and parentage analysis, but many studies have also used SSRs with much longer motifs (Vieira et al. 2016, Oliveira et al. 2006).

The SSRs were long considered to be non-functional portions of DNA (genetic “junk”) with no effect on phenotype, but there have been many studies with evidence of the contrary (Kovtun and McMurray 2008; Hamada et al. 1984; Trifonov 1989; Gerber et al. 1994; Kashi et al. 1997; King 1997; Comings 1998, King 1999; Fondon 2004; Kashi 2006). The SSRs are found in prokaryotes and eukaryotes and are widely distributed throughout the genome including coding and non-coding nuclear and organellar DNA, and especially in the euchromatin of eukaryotes. There are many repeat-expansion diseases that have repetitive sequences in areas where slippage mutations have dramatic consequences (Perez-Jimenez et al. 2013; Phumichai et a 2015).

An important characteristic for SSRs is the quick mutation rate that ranges between 10^{-3} and 10^{-6} be as high as 10^{-2} per cell division and the variation in repeats across time and generations is reversible (Fondon et al. 2008; Gemayel et al., 2012). One of the key characteristics that is the focus of using SSRs for population genetic studies such as on *Rathayibacter toxicus* is their

faster than normal mutation rate. *Rathayibacter toxicus* has a relatively slow generation time compared to other bacterial species taking 7-14 days for colony development on artificial media. Due to its complex life cycle linked to seed set in its grass host, there is usually only one disease cycle per year. SSR analysis may be an ideal tool because the faster rate of mutation in satellite DNA may be more detectable across a smaller number of generations of *R. toxicus*. It is possible that analyses of SSRs in addition to the existing genetic analyses may aid in phylogenetic and phylogeographic determinations that the other methods could not fully resolve.

In addition to population genetic and phylogenetic applications of genetic analysis of *R. toxicus*, diagnostics are an important potential output of various genetic markers. SSRs may by useful for diagnostic purposes if the specific SSRs are unique to the level of classification in question, as far as species and perhaps to sub-species levels. The accuracy of molecular-based diagnostics can be highly dependent on the level of understanding of genetic variation of any given pathogen.

Materials and Methods

Ethics Statement

As all handling of *R. toxicus*-infected plants and live cultures of *R. toxicus*, and other *Rathayibacter* species, was conducted in Australia, specific permissions from U.S. government agencies or regulatory bodies were not required for the collection or processing of the plant materials or cultures used in this study. Prior to importation, the U.S. Select Agent Program was notified of the importation and analysis of extracted DNA of *R. toxicus* and other *Rathayibacter* species. Endangered or protected species were not collected or used in this study. No samples were collected from endangered or protected field sites.

Genomes used

Whole genome sequences (WGS) of 16 strains of *Rathayibacter toxicus* (Table 2.1) were analyzed using genomic analysis programs, primarily Geneious version 9.8.1. (Biomatters, New Zealand; http://www.geneious.com). WGS for strains 70137 and WAC3373 were obtained from the National Center for Biotechnology Information (NCBI) GenBank nucleotide database accession numbers ASM146585v1 and ASM87567v1(Arif et al. 2016). WGS for strains SA03-14, SA03-19, SA08-07, SA08-08, SA08-09, SA19-02, SA1906, SA19-07 were obtained by mapping Illumina MiSeq sequence to a reference genome (Illumina Inc., San Diego CA, USA; Arif et al. 2016, Yasuhara-Bell et al. 2020). The PacBio sequencing data was de novo assembled with HGAP using default parameters (500bp min. subread length; 6 kb min seed read length) and polished with Quiver. Illumina MiSeq Data was assembled by mapping to complete PacBio genomes using Bowtie in Geneious version 7.1.9 and de novo using the Geneious assembler (Arif et al. 2016, Yasuhara-Bell et al. 2020).

Table 2.1: List of Whole Genome Sequenced strains of *Rathayibacter toxicus* available

#	Name	Location collected	Plant Host	Year
1	SA 03-04	Corny Point, Southern Australia	*Lolium rigidum*	2014
2	SA 03-14	Corny Point, Southern Australia	*Lolium rigidum*	2014
3	SA 03-19	Corny Point, Southern Australia	*Lolium rigidum*	2014
4	SA 08-08	Corny Point, Southern Australia	*Lolium rigidum*	2014
5	SA 08-09	Lake Sunday, Southern Australia	*Lolium rigidum*	2014
6	SA 19-02	Yorktown, Southern Australia	*Lolium rigidum*	2013
7	SA 19-06	Yorktown, Southern Australia	*Lolium rigidum*	2013
8	SA 19-07	Yorktown, Southern Australia	*Lolium rigidum*	2013
9	WAC7056	Murray Bridge, Southern Australia	*Lolium rigidum*	1983
10	SA19-14	Yorktown, Southern Australia	*Lolium rigidum*	2013
11	CS28	Western Australia	*Lolium rigidum*	1978
12	WA40-23	Western Australia	*Lolium rigidum*	2015
13	WAC3373	Gnowangerup, Western Australia	*Phalaris paradoxa*	1978
14	CS36	New South Wales	*Lachnagrostis filiformis*	1990
15	CS39	New South Wales	*Polypogon monspeliensis*	1990
16	70137	Western Australia	*Dactylis glomerata*	1975

Table 2.1 notes: year means which year the R. toxicus sample was collected.

A total of 41 *R. toxicus* strains were used (including strains in list of 16) for testing two Simple Sequence Repeats (RTSatA and RTSatB) that have the largest motif size in nucleotides, 32 and 9 respectively, and the highest numbers and variability of repetitiveness (Table 2.2). These were chosen as the top two also because their expected PCR band sizes could be most easily distinguishable, even if different *R. toxicus* strains vary by only one repeat. Strains that were not WGS were used as genomic DNA extracts. These *R. toxicus* strains came from various regions of Australia, were isolated from multiple types of grass species, and collected at different times and years (Table 2.2).

Table 2.2: List of 41 *Rathayibacter toxicus* strains used for PCR testing two simple sequence repeats, RTSatA and RTSatB

#	Strain Name	Location collected	Plant Host species	Year
1	SA 03-02	Corny Point, Southern Australia	*Lolium rigidum* (annual ryegrass)	2014
2	SA 03-08	Corny Point, Southern Australia	*Lolium rigidum*	2014
3	SA 03-14	Corny Point, Southern Australia	*Lolium rigidum*	2014
4	SA 03-15	Corny Point, Southern Australia	*Lolium rigidum*	2014
5	SA 03-16	Corny Point, Southern Australia	*Lolium rigidum*	2014
6	SA 03-17	Corny Point, Southern Australia	*Lolium rigidum*	2014
7	SA 03-18	Corny Point, Southern Australia	*Lolium rigidum*	2014
8	SA 03-19	Corny Point, Southern Australia	*Lolium rigidum*	2014
9	SA 03-20	Corny Point, Southern Australia	*Lolium rigidum*	2014
10	SA 03-24	Corny Point, Southern Australia	*Lolium rigidum*	2014
11	SA 03-25	Corny Point, Southern Australia	*Lolium rigidum*	2014
12	SA 03-26	Corny Point, Southern Australia	*Lolium rigidum*	2014
13	SA 03-27	Corny Point, Southern Australia	*Lolium rigidum*	2014
14	SA 03-28	Corny Point, Southern Australia	*Lolium rigidum*	2014
15	SA 08-03	Lake Sunday, Southern Australia	*Lolium rigidum*	2014
16	SA 08-08	Lake Sunday, Southern Australia	*Lolium rigidum*	2014
17	SA 08-09	Lake Sunday, Southern Australia	*Lolium rigidum*	2014
18	SA 08-11	Lake Sunday, Southern Australia	*Lolium rigidum*	2014
19	SA 08-13	Lake Sunday, Southern Australia	*Lolium rigidum*	2014
20	SA 08-16	Lake Sunday, Southern Australia	*Lolium rigidum*	2014
21	SA 19-02	Yorketown, Southern Australia	*Lolium rigidum*	2013
22	SA 19-04	Yorketown, Southern Australia	*Lolium rigidum*	2013
23	SA 19-06	Yorketown, Southern Australia	*Lolium rigidum*	2013
24	SA 19-07	Yorketown, Southern Australia	*Lolium rigidum*	2013
25	SA 19-09	Yorketown, Southern Australia	*Lolium rigidum*	2013
26	SA 19-11	Yorketown, Southern Australia	*Lolium rigidum*	2013
27	SA 19-12	Yorketown, Southern Australia	*Lolium rigidum*	2013

28	SA 19-13	Yorketown, Southern Australia	*Lolium rigidum*	2013
29	FH141 (CS2)	Southern Australia	*Lolium rigidum*	1983
30	FH83 (CRS2)	Southern Australia	*Lolium rigidum*	1975
31	FH85 (CRS3)	Southern Australia	*Lolium rigidum*	1975
32	SA 19-03	Yorketown, Southern Australia	*Lolium rigidum*	2013
33	SA 19-10	Yorketown, Southern Australia	*Lolium rigidum*	2013
34	SA 19-14	Yorketown, Southern Australia	*Lolium rigidum*	2013
35	WAC3368 (SAC3368)	Southern Australia	*Lolium rigidum*	1981
36	WAC3387 (SAC3387)	Southern Australia	*Lolium rigidum*	1981
37	FH138 (CS28)	Western Australia	*Lolium rigidum*	1978
38	FH81 (CRK73)	Western Australia	*Lolium rigidum*	1973
39	FH87 (CRW1)	Western Australia	*Lolium rigidum*	1974
40	CS36	New South Wales	*Lachnarostis filiformis* (bentgrass)	1990
41	CS39	New South Wales (south of)	*Polypogon monspeliensis* (beardgrass)	1990
42	WAC9601 (CS21) [*R. tritici*]	South Perth, Western Australia	*Triticum sp./aestivum* (Wheat)	
43	ICMP 2579 [*R. rathayi*]	Iran	*Dactylis glomerata* (Oat)	1994
44	WAC3369 [*R. rathayi*]	Western Australia	*Lolium rigidum*	
45	ESC-10 [*P. syringae*]			
46	ddH20 (control)			

Table 2.2 Notes: year means year the sample was collected. Rows 42-45 represent controls.

Simple Sequence Repeats (SSR)

Classification of SSRs

Table 2.3: General Simple Sequence Repeat Categorization

SSR Categorization	
Macrosatellites	>100bp motif
Minisatellites	>10-100bp motif
Microsatelites	<9bp motif

Discovery of SSRs within *Rathaybacter toxicus* genomes

As a starting point, 16 whole genome sequenced strains of *R. toxicus* (listed in Table 1.1) were analyzed using the Geneious bioinformatics platform, version 9.8.1 (Biomatters, New Zealand; http://www.geneious.com), to mine for and discover potential SSR markers. The Phobos tandem repeat finder tool was particularly helpful with discovering genomic regions where motif sizes repeat (Mayer 2010). This process included first searching for tandem repeats of any size, and then focusing specifically on different types of motifs, and these parameters are further explained in Table 2.4. SSRs include microsatellites, minisatellites, and microsatellites, which are defined in Table 2.3. The nucleotide sequence of each SSR repeat is referred to as the "motif" that repeats consecutively. The SSR motifs varied in size and the number of repeats varied as well.

SSRs motifs were generally classified in this project consistent with the SSR literature (reviewed in Oliveira et al., 2006):

i) Perfect

   a. composed entirely of repeats of a motif which may vary in number of nucleotides, but not contain any SNPs (insertions, substitutions, deletions) anywhere amid the motif repeats.

ii) Imperfect

a. a nucleotide not belonging to the motif occurs between repeats or replaces any nucleotides within a repeat unit – a single nucleotide polymorphism, aka SNP, within the satellite

b. Interrupted: a sequence of a few nucleotides is inserted into the motif

c. Composite: formed by multiple, adjacent, repetitive motifs (SSRs?).

SSR determinations are based on a summary of various experiments throughout literature. The bigger microsatellite motifs tend to have more SNP polymorphisms. The bigger numbers of repeats help with being certain that they are actual SSRs. Microsatellites (SSRs with motif size 10 and smaller) are most often used by researchers for studying variation and evolution.

The in-depth parameters used for selecting SSRs for this analysis are based on the most common and frequently used parameters through literature. Microsatellites (SSRs with motif size 10 and smaller) are most often used by researchers for studying variation and evolution, and minisatellites are less commonly found but have also been used in analyses. The bigger microsatellite motifs tend to have more SNP polymorphisms. The bigger numbers of repeats help with being certain that they are actual SSRs (Gemayel et al. 2012, Oliveira et al. 2006, Viera et al. 2016).

Table 2.4: SSR search parameters

2 bp
• Must repeat at least 5 times, 100% perfection. • Could include 90%-99% perfection if a 2bp motif repeated many times (no results) • Rare for this to be used throughout literature, unless it had much more than 5 repeats. • Hard to be certain whether a 2-nucleotide motif repeated less than 5 times is a true SSR.

3 bp

- Must repeat at least 5 times, 100% perfection
- Could include 90%-99% perfection if a 3bp motif repeated many times (no results)

4bp

- Must repeat at least 5 times, 100% perfection
- Could include 90%-99% perfection if a 4bp motif repeated many times (no results)

5bp

- 1 - Must repeat at least 5 times, 90-100% perfection
- 2 – Repeat as little as 3 times but at 100% perfection

6bp

- 1 – Must repeat at least 5 times at 90-100% perfection, or
- 2 – Repeat as little as 3 times but at 100% perfection

7bp

- 1 - Must repeat at least 5 times, 90-100% perfection
- 2 – Repeat as little as 3 times but at 100% perfection

8bp

- Must repeat at least 5 times, 90-100% perfection
- 2 – Repeat as little as 3 times but at 100% perfection

9bp

- 1 - Must repeat at least 5 times, 90-100% perfection
- 2 – Repeat as little as 3 times but at 100% perfection

32bp

- 1 - Must repeat at least 5 times, 90-100% perfection
- 2 - repeat as little as 3 times but at 100% perfection

Polymerase Chain Reaction (PCR) Verification of *Rathayibacter toxicus* SSRs

Primer Design

In order to sequence verify, experiments were designed to use PCR for band presence and size and used to compare SSR bands across strains. PCR is commonly used to detect the presence of a specific nucleotide/base pair sequence within a genome.

The bioinformatics program Geneious, version 9.8.1 (Biomatters, New Zealand; http://www.geneious.com), was utilized to identify the genomic sequence of the flanking regions surrounding each satellite in order to determine ideal forward and reverse primer sequences.

Most primer pairs for each SSR were designed manually using a set of ideal conditions; for some, the Geneious Primer design tool was used.

There are several top considerations that were used when manually designing primers for RTSatA-J. Primers are always specified 5' to 3', left to right and it is important to verify that they are designed and ordered in the correct orientation. For most applications, PCR primers should be between 18 to 25 nucleotides in length and should have a GC content between 40 and 60%, with the 3′ of a primer ending in C or G to promote binding. The last 5 bases at the 3′ end of the primer, should contain at least 2 G or C bases (GC clamp). G-C base pairs have a stronger bond than A-T base pairs (3 hydrogen bonds versus 2). Nucleotide runs of 4 consecutive nucleotides or more and dinucleotide repeats (for example, ACCCC or ATATATAT) were avoided to minimize primer mispriming. Primer sequences were designed and checked (idtdna.com primer check software) to preclude hairpins and primer dimers that may cause primer self-amplification. If primer dimers or hairpins were indicated through this software, the primers were then redesigned from the beginning, by choosing a different genomic sequence in a different flanking area around the SSR. Primer sequences for each SSR region are listed and described in Table 2.6. Using the optimized primers, the SSR regions were PCR-amplified using genomic DNA of all strains.

Determination of annealing temperature for each primer set

To determine the optimum annealing temperature, each SSR primer set was tested using the gradient option in a thermocycler. A general protocol was used which begins with a denaturation step at 95º for two minutes and continues to 34-36 cycles of denaturation at 95º for 15 seconds. After this, the gradient test occurs and the thermocycler is set at a “gradient” range of temperatures, generally from 52-63 instead of a specific temperature, for 15 seconds. Then

there is elongation at 72º for 1 minute, and final elongation for another minute at the same temperature. One DNA strain was used to test each SSR, to ensure that variation between the different temperatures in the gradient was due to temperature and not strain or DNA quality differences.

To visualize amplicons from the PCR gradient test, the PCR products were electrophoresed on 2% agarose gels for 2 hours at 90 volts. The gels were stained with Ethidium bromide and viewed under UV light. The clearest and brightest band in each SSR gradient test was identified and the specific temperature it corresponded to was chosen as the annealing temperature for that SSR.

PCR Optimization

PCR conditions were optimized and then used to test for satellites RTSatA-RTSatJ presence in all *R. toxicus* genomes. The optimized PCR protocol had the following conditions: initial denaturation at 95º for two minutes, 34-36 cycles of denaturation at 95º for 15 seconds, annealing temperature of either 53 or 61, depending on the SSR, for 15 seconds, elongation at 72º for 1 minute, and a final elongation for 1 minute at the same temperature. PCR products were stored at -4 degrees C, and/or used immediately for gel electrophoresis. The products were visualized on 2% agarose gels stained with ethidium bromide which were ran for 2 hours at 90 volts. The band size and SSR length were estimated based on a gel ladder and several samples were sequenced for further verification.

For the SSRs, RTSatA and RTSatB, 41 *R. toxicus* strains were PCR verified using *R. toxicus* genomic DNA (Table 2.2). These were the top two also because their expected PCR band sizes could be most easily distinguishable, even if different *R. toxicus* strains vary by only one repeat. Satellite presence was determined based on estimated band size after subtracting primer

sequences and sequence extraneous to the SSR repeat (Figure 2.1). The approximated bands were compared to nucleotide count of the SSRs from the available whole genome sequences and used as references for other bands on the gel that were from strains without WGS data.

Phylogenetic analysis of *R. toxicus* SSRs

Geneious, version 9.8.1 (Biomatters, New Zealand; http://www.geneious.com), was used to generate phylogenetic trees based on concatenated SSR sequence data. Each SSR sequence (RTSatA-J, Table 2.5) was extracted from the genome sequences of the 16 strains (Table 2.1) and concatenated. This concatenated SSR sequence of each strain was used to create a phylogenetic tree. The phylogenetic common parameters included using the Jukes-Cantor with neighbor-joining method. Phylogenetic analyses often calculate the number of differences between sequences, but this does not fully reflect the evolutionary history of the sequence, because not all the mutation events are recorded in the current sequences; the Jukes-Cantor algorithm takes that into account.

R Analysis of *R. toxicus* SSRs: Poppr package

The R package Poppr, created for population genetics and genomics by members of the Grunwald group at Oregon State University was utilized for the genetic analysis of *R. toxicus* using SSR repeat variation (Kamvar et al. 2014). The SSR data across 16 strains (Table 2.6) was used for this analysis. Several analyses were done from the Poppr package, including a genotype accumulation curve, missing data chart, ploidy, multilocus genotype table, genotypic richness, and Bruvo's distance phylogenetic trees. A genotype accumulation curve is a tool to make sure there are enough loci analyzed, and important to make sure that the observed diversity in the analysis will not appreciably increase if an additional marker is added (Arnaud-Hanod et al., 2007). The genotype accumulation curve in *poppr* is constructed by randomly sampling *x* loci

and counting the number of observed multi-locus genotypes (MLGs), or SSR markers. Missing data plot shows missing data per population, so any *R. toxicus* strain that does not have one of the 10 SSRs. The Discriminant analysis of principal components (DAPC) is useful to analyze difference between groups or populations more than just between all the individuals, as variance is partitioned into between-group and within-group components (Jombart et al., 2010). The poppr package Bruvo analysis in this study uses UPGMA method as a base but is specifically designed for microsatellite genotype distances, which considers mutation processes, repeat numbers and length, and permits comparison of individuals with different ploidy levels (Bruvo et al. 2004).

Results

Discovery of SSRs within *Rathaybacter toxicus* genomes

Multiple microsatellites and one minisatellite were identified using the Phobos Tandem Repeat Finder extension tool in Geneious (Mayer 2010; Biomatters, New Zealand http://www.geneious.com). Macrosatellites (motif greater than 100 nucleotides) were not found.

SSRs with varied numbers of repeats and single nucleotide polymorphisms (SNPs) were discovered in the *R. toxicus* genomes. In total, 10 SSRs were chosen based on commonly used parameters including, size of motifs, numbers of repeats, and motif sequence; the satellites were named RTSatA-RTSatJ (Table 2.5).

Table 2.5: Top Selected Simple Sequence Repeats of *Rathayibacter toxicus*

#	Name of discovered SSR	SSR Motif Sequence
1	RTSatA	AAATCCCTGATTAGGTTGTGGTTGGGGACTGT
2	RTSatB copy 1	AAAGACCTC
	RTSatB copy 2	AAAGACCTC
3	RTSatC	ACTGAGCT
4	RTSatD	AGCCGG
5	RTSatE	AACCAGGC
6	RTSatF	AGCAGCG
7	RTSatG	AATACAG
8	RTSatH	AGCCGC
9	RTSatI	AGGCGC
10	RTSatJ	ATCCG

Table 2.5 Notes: "Motif" represents the "repeat unit" which can be adjacently repeated varying numbers of times. The sequence here is only of the repeat unit, not the entire SSR. RTSatB has two copies which means there are two SSRs which have the identical nucleotides in their repeat unit, but the two copies occur in different regions of the genome that are very far from one another, approximately 609k nucleotides apart.

Of all the mined SSR regions, there was only one SSR larger than 10bp in motif length: a unique 32bp motif SSR (minisatellite), which repeated many times (range: 13 to 29) in the 16 genomes, and varied in the number of repeats among the genomes (Table 2.6). Each individual SSR had a unique motif sequence and one SSR (RTSatB, motif: AAAGACCTC) was detected at two locations within the same genome; i.e., there were two copies within the same genome (Table 2.6). Each SSR motif was NCBI blasted to see if it occurred in other organisms' genomes, and the only motif that was unique to *R. toxicus* was the 32-nucleotide RTSatA motif. When NCBI Blast searching the entire SSR sequences, some were also found in other organisms, but some were not perfect matches, particularly if the SSR in *R. toxicus* was imperfect and contained SNPs.

Table 2.6: Number of Repeats of each SSR motif in whole genome sequenced *Rathaybacter toxicus* strains

RTSat -	A	B (1)	B (2)	C	D	D	F	G	H	I	J
SA 03-04	17	7	7	10	7	3	3	n/a	6	3	6
SA 03-14	17	7	7	10	7	3	3	n/a	6	3	6
SA 03-19	17	7	7	10	7	3	3	n/a	6	3	6
SA 08-08	17	7	7	10	7	3	3	n/a	6	3	6
SA 08-09	17	7	7	10	7	3	3	n/a	6	3	6
SA 19-02	17	7	7	10	7	3	3	n/a	6	3	6
SA 19-06	17	7	7	10	7	3	3	n/a	6	3	6
SA 19-07	17	7	7	10	7	3	3	n/a	6	3	6
WAC7056	16	8	9	4	11	3	3	3	n/a	3	n/a
SA19-14	16	8	9	3	11	3	3	3	n/a	3	n/a
CS28	29	10	7	4	10	3	3	3	6	3	6
WA40-23	24	9	7	4	10	3	3	3	6	3	6
WAC3373	26	9	7	4	10	3	3	3	6	3	6
CS36	13	6	11	7	9	3	n/a	3	10	3	6
CS39	18	6	9	3	9	3	n/a	n/a	n/a	3	6
70137	29	10	7	4	10	3	3	3	6	3	6

Table 2.6 Notes: table depicts the 10 Simple Sequence Repeats (SSRs) across 16 whole genome sequenced strains. The numbers within the table represent the number of repeats of each motif unit of the designated SSR. The names of the SSRs are "RTSatA" to "RTSatJ". RTSat B has two copies in genome.

Polymerase Chain Reaction (PCR) Verification of *Rathayibacter toxicus* SSRs

Based on the top considerations for designing primers manually and through use of Geneious 9.8.1 tools, the primer sequences listed in Table 2.7 were determined for each SSR region. Nine out of ten SSR annealing temperatures were set at 61 degrees. For SSR RTSatG 53-56 was the ideal annealing temperature range, with 53 as the optimal one. The primer pairs were NCBI blasted to check for any perfect matches among other organisms, and there were not primer pair sets which contained perfect matches of both forward and reverse region in any organism.

Table 2.7: Primer Sequences for SSRs of *Rathayibacter toxicus*

SSR ID	Forward Primer sequence	Reverse Primer sequence
RTSatA	5’-GCTATCGGATGCCCCCCCTAAAACG-3	5’-ACAGTGGATTGTGTTGTGGTTGG-3’
RTSatB	5’-CATTGGGTTTGAGTGTTCCGTC-3’	5’-CACTACGAACGTTTTGCTGCTC-3’
RTSatC	5’-GAAGACGCTTGCCTTTGCTC-3’	5’-AAAAAGCACGGCAACATCCC-3’
RTSatD	5’-CTGAGAGGTCGCGGATTCTC-3’	5’-CACCATCGTGCAGGTTCTCC-3’
RTSatE	5’-TCGCGATTACAGTCACGAGC-3’	5’-GCCGTCGAGAATTCTGGTGT-3’
RTSatF	5’-CATGATTCGTGCTGTTCGCC-3’	5’-TTTTCTAGGGCAGGTCAGGC-3’
RTSatG	5’-GCTGACTTACTACGCCGAGG-3’	5’-AGACCTCTTCTCCTTTGTGTTGA-3’
RTSatH	5’-CTGCTCCACGAACGGCTC-3’	5’-GCAAAGGAAATGGCGGTCTC-3’
RTSatI	5’-GGTATGGGTGTTCCGCAGAG-3’	5’-GTGCCGTTGATCGAGGTGAT-3’
RTSatJ	5’-CGAGTGCTCCGTGATTCCTT-3’	5’-GGTCACTGAACAGTCTCAGCA-3’

PCR Verification

The results of the RTSatA and RTSatB PCR verification are in Table 2.8 and include repeat numbers of each SSR based on the extrapolated band size showcasing the full SSR nucleotide count, divided by the number of nucleotides in one repeat motif. This was shown for both RTSatA and RTSatB of 41 *R. toxicus* strains in Table 2.8.

Table 2.8: RTSatA and RTSatB in 41 *Rathayibacter strains*, PCR results

		RTSatA			**RTSatB Copy 1**			**RTSatB Copy 2**		
#	Strain Name	PCR target band	Motif size (bp)	Repeats	PCR target band	Motif size (bp)	Repeats	PCR target band	Motif Size (bp)	Repeats
1	SA 03-02	610	32	19	67	9	7	67	9	7
2	SA 03-08	610	32	19	67	9	7	67	9	7
3	SA 03-14	610	32	19	67	9	7	67	9	7
4	SA 03-15	610	32	19	67	9	7	67	9	7
5	SA 03-16	610	32	19	67	9	7	67	9	7
6	SA 03-17	610	32	19	67	9	7	67	9	7
7	SA 03-18	610	32	19	67	9	7	67	9	7

8	SA 03-19	578	32	18	67	9	7	67	9	7
9	SA 03-20	578	32	18	67	9	7	67	9	7
10	SA 03-24	610	32	19	67	9	7	67	9	7
11	SA 03-25	610	32	19	67	9	7	67	9	7
12	SA 03-26	610	32	19	67	9	7	67	9	7
13	SA 03-27	610	32	19	67	9	7	67	9	7
14	SA 03-28	578	32	18	67	9	7	67	9	7
15	SA 08-03	674	32	21	67	9	7	67	9	7
16	SA 08-08	642	32	20	67	9	7	67	9	7
17	SA 08-09	642	32	20	67	9	7	67	9	7
18	SA 08-11	610	32	19	67	9	7	67	9	7
19	SA 08-13	610	32	19	67	9	7	67	9	7
20	SA 08-16	610	32	19	67	9	7	67	9	7
21	SA 19-02	610	32	19	67	9	7	67	9	7
22	SA 19-04	642	32	20	67	9	7	67	9	7
23	SA 19-06	642	32	20	67	9	7	67	9	7
24	SA 19-07	642	32	20	67	9	7	67	9	7
25	SA 19-09	610	32	19	67	9	7	67	9	7
26	SA 19-11	642	32	20	67	9	7	67	9	7
27	SA 19-12	610	32	19	67	9	7	67	9	7
28	SA 19-13	610	32	19	67	9	7	67	9	7
29	FH141 (CS2)	578	32	17	85	9	9	103	9	11
30	FH83 (CRS2)	482	32	15	85	9	9	94	9	10
31	FH85 (CRS3)	482	32	15	85	9	9	94	9	10
32	SA 19-03	514	32	16	85	9	9	94	9	10
33	SA 19-10	514	32	16	85	9	9	94	9	10
34	SA 19-14	514	32	16	85	9	9	94	9	10
35	WAC3368 (SAC3368)	514	32	16	85	9	9	94	9	10
36	WAC3387 (SAC3387)	552	32	17	94	9	10	94	9	10
37	FH138 (CS28)	930	32	29	67	9	7	103	9	11

38	FH81 (CRK73)	930	32	29	67	9	7	103	9	11
39	FH87 (CRW1)	930	32	29	67	9	7	103	9	11
40	CS36	450	32	14	58	9	6	112	9	12
41	CS39	710	32	22	58	9	6	94	9	10
42	WAC9601 (CS21) [R. tritici]	0	0	0	0	0	0	0	0	0
43	ICMP 2579 [R. rathayi]	0	0	0	0	0	0	0	0	0
44	WAC3369 [R. rathayi]	0	0	0	0	0	0	0	0	0
45	ESC-10 [P. syringae]	0	0	0	0	0	0	0	0	0
46	ddH20 (control)	0	0	0	0	0	0	0	0	0

Table 2.8 Notes: The gray area highlights the number of repeats, which is the number of times that the SSR motif unit is repeated, and is a measure of genetic variation (or lack thereof) across strains."PCR Target band" means the extrapolated target band based on approximated gel band size according to the ladder, and expected outcome based on referencing several available whole genome sequences (analyzed through Geneious) which were also PCR tested.

Figure 2.1: PCR testing of RTSatB and RTSatA using Strains from Table 2.2.

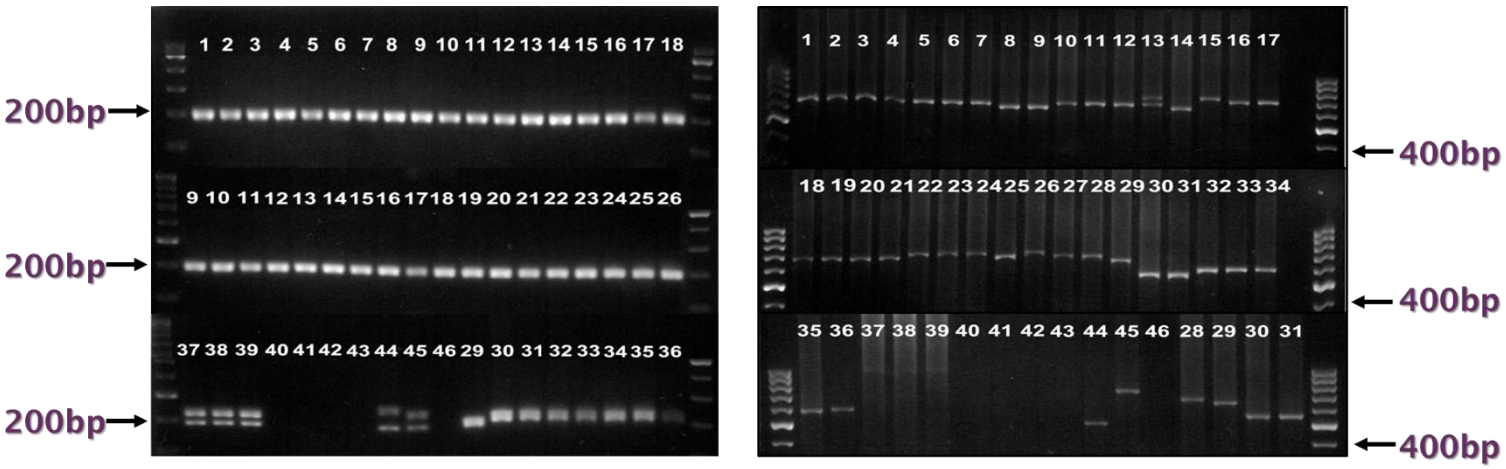


Figure 2.1 Notes: These are the PCR bands on gel electrophoresis from which data was extrapolated into Table 2.10. The image on the left is of RTSatB and on the right is RTSatA.

When looking at SSR groupings within populations derived from MLST, ISSR, and WGS analyses, the results in Table 2.9 are shown in order to see SSR repeat patterns or lack thereof between populations.

Table 2.9: How many times Motif of RTSatA-J are repeated within strains of RT-I-RT-V populations

SSR ID	Motif Repeat #				
	RT-I	RT-II	RT-III	RT-IV	RT-V
RTSatA	18-21	15-18	29	14	22
RTSatB (two copies)	7	9-10	7	6	6
	7	10-11	11	12	10
RTSatC	10	3-4	4	7	3
RTSatD	7	11	10	8	8
RTSatE	3	3	3	3	3
RTSatF	3	3	3	-	-
RTSatG	-	3	3	3	-
RTSatH	6	-	6	10	-
RTSatI	3	3	3	3	3
RTSatJ	6	6	6	6	6

Table 2.9 Notes: RTSatA and RTSatB also takes PCR results of 41 strains from Table 2 into consideration, and RTSatC-J are based on the 16 whole genome sequenced strains.

Phylogenetic analysis of *R. toxicus* SSRs

Multiple phylogenetic analyses were completed based on the satellite data for the 16 *Rathayibacter toxicus* strains (Table 2.6), using different combinations of SSRs commonly used in the literature.

The first phylogenetic analysis included all SSRs (Table 2.5), concatenated and aligned across all 16 strains (Figure 2.2, Table 2.1). This phylogenetic tree groupings supported the previously established populations established by MLST, ISSR, and comparative genomic analyses (Arif et al. 2016, Yasuhara-Bell and Stack 2019, Yasuhara-Bell et al. 2020).

The second analysis (Figure 2.3) included all SSRs (Table 2.5) except RTSatA, the minisatellite. Using only microsatellites is a common approach in population genetic studies using SSRs. Significant differences were evident between the analyses (Figure 2.3, Figure 2.2) with and without the minisatellite. Specifically, the relatedness between RT-IV and RT-V and among RT-IV and RT-V with other strains was very different as well as one RT-III strain split out from the rest of the RT-III strains. These groupings in this phylogenetic tree were largely the same as the populations based on other methods with those few significant relatedness differences.

The third phylogenetic analysis (Figure 2.4) was based on only SSRs that occurred in each of the 16 strains (Table 2.6), which excluded RTSatF, RTSatG, RTSatH, and RTSatJ. This phylogenetic tree contains groupings and distances similar to those of the previously established, with some variation. It also suggests a high degree of relatedness between RT-II + RTIII and between RT-IV + RT-V.

The final phylogenetic analysis (Figure 2.5) was the most stringent and exclusive, with the most commonly used SSR parameters throughout literature. It was based on using only microsatellites and only those SSRs that occur in each of the 16 strains, which excludes RTSatA (minisatellite), RTSatF, RTSatG, RTSatH, and RTSatJ. Groupings and distances changed the most with this analysis; RT-V was much closer to RT-I, RT-III split into two groups, RT-II strains were more closely related to two of the RT-III strains, and it suggests that RT-V may be the progenitor population.

Figure 2.2: Phylogenetic Tree of All SSRs across 16 strains of whole genome sequenced *Rathayibacter toxicus*

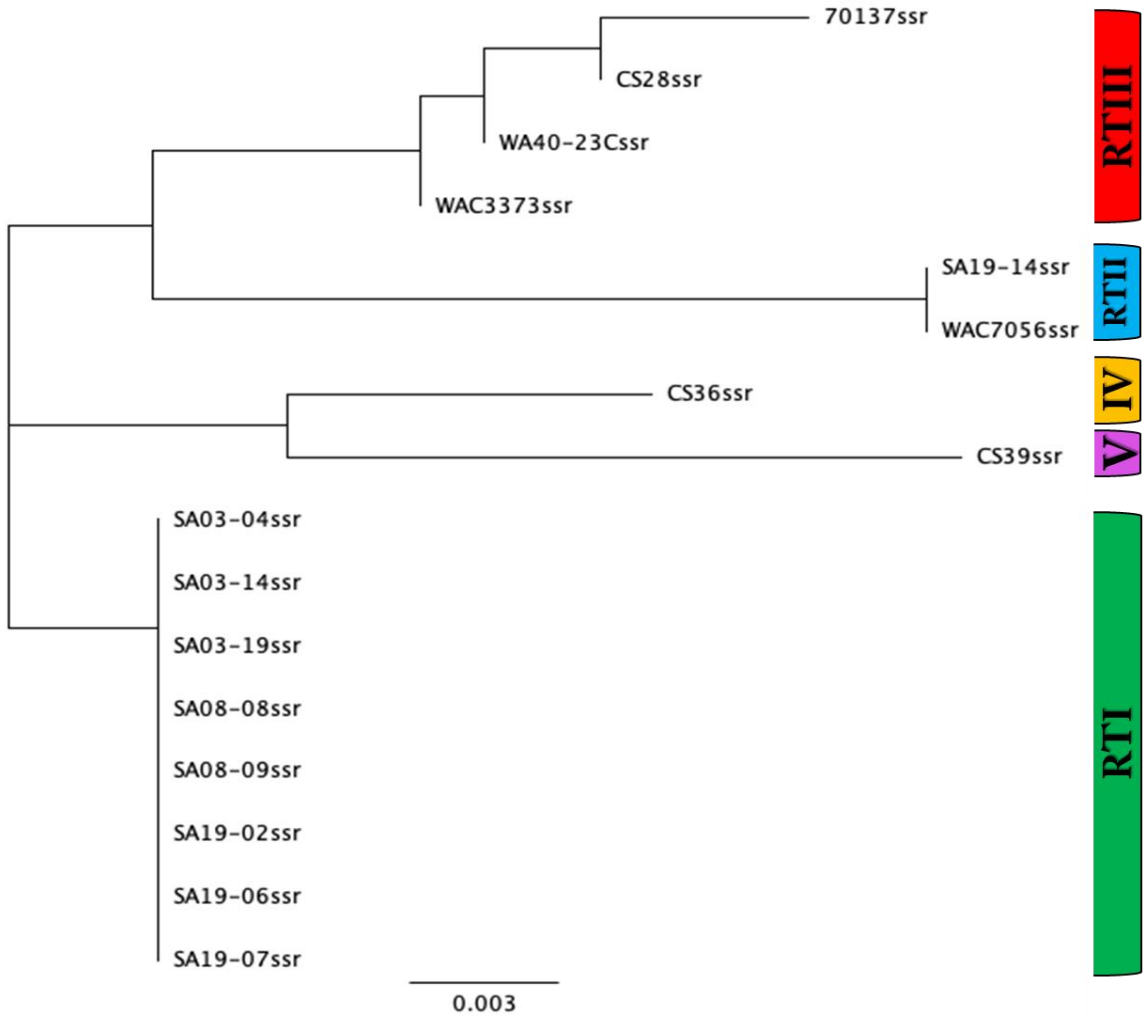


Figure 2.3: Concatenated all SSRs except the minisatellite, of the whole genome *Rathayibacter toxicus* strains

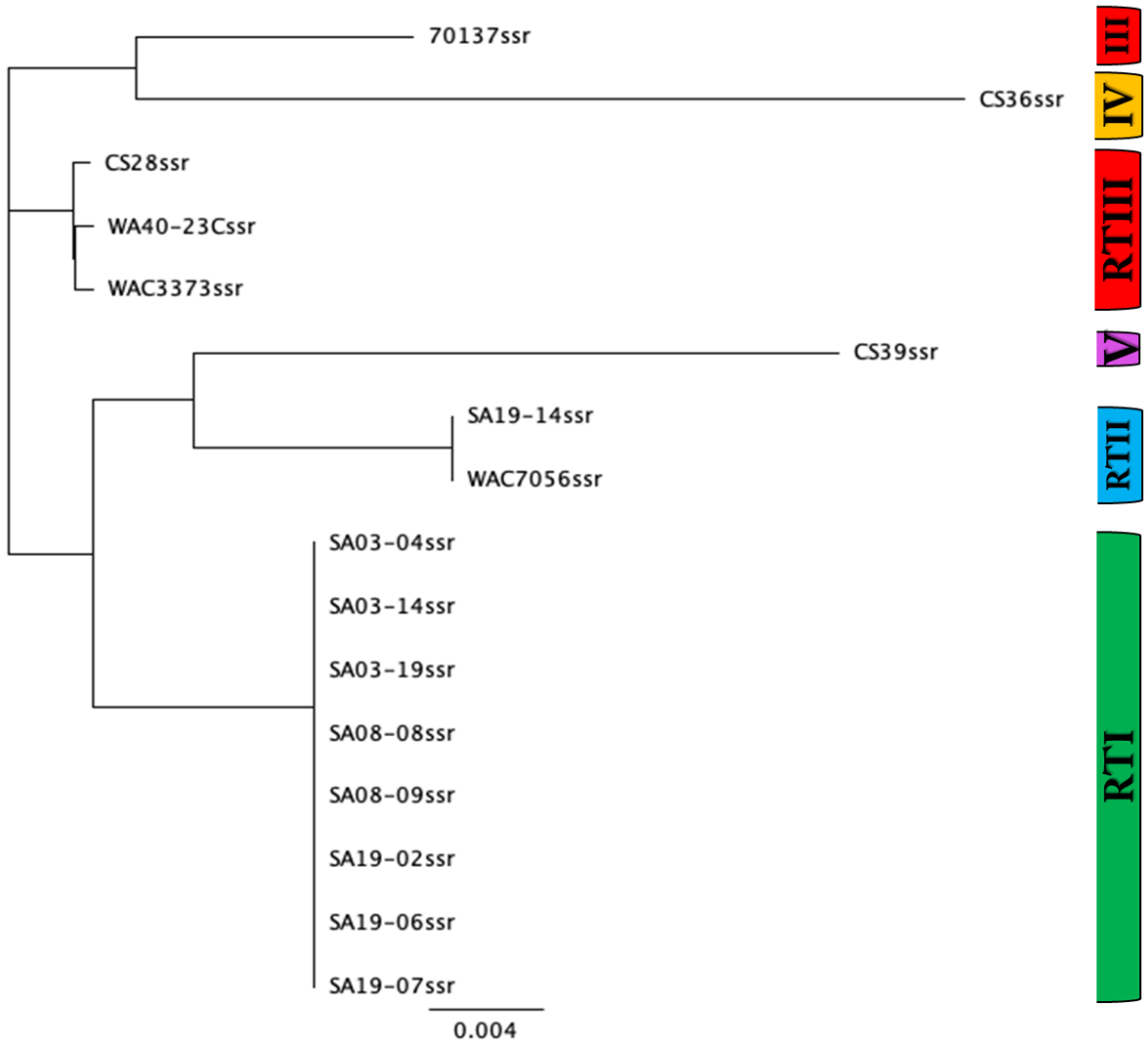
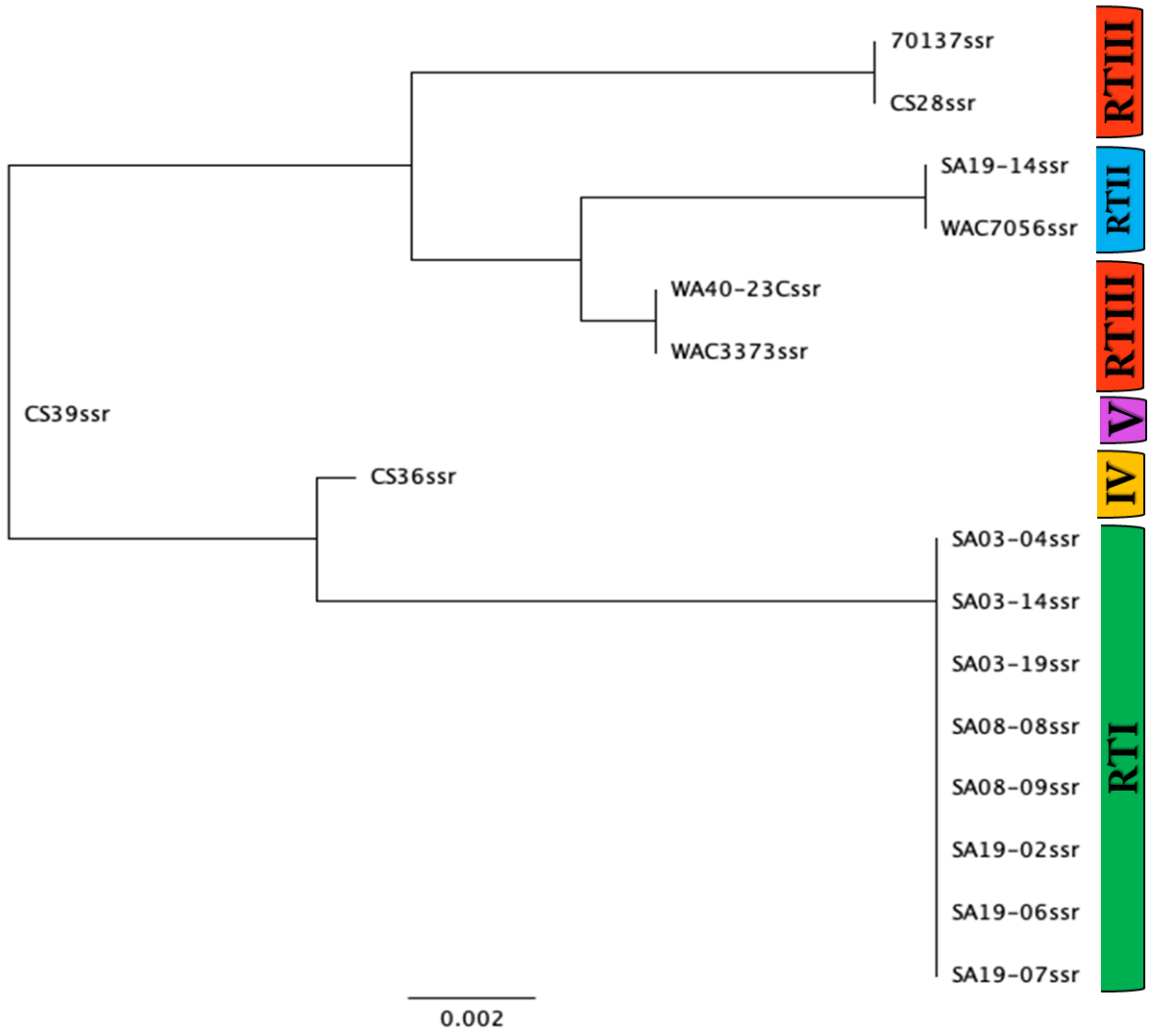


Figure 2.4: Concatenated only SSRs which occur in all of the available 16 whole genome sequence *Rathayibacter toxicus* strains

Figure 2.5: Concatenated all microsatellites, excluding minisatellite, that occur in each available *Rathayibacter toxicus* whole genome strain



R Analysis: Poppr package for population genetics

A genotype accumulation curve (Figure 2.6) is for assessing how much power there needs to be to discriminate between unique individuals given a sample of 10 loci (for the 10 SSRs). This analysis is particularly important for clonal organisms to confirm that a plateau has been reached in the number of loci necessary to discriminate individuals. For the five populations, or loci, of *R. toxicus*, there were enough markers to do the subsequent analysis, as the threshold was met at eight multilocus genotypes (MLG), and the analysis has ten SSRs.

Figure 2.6: Genotype accumulation curve.

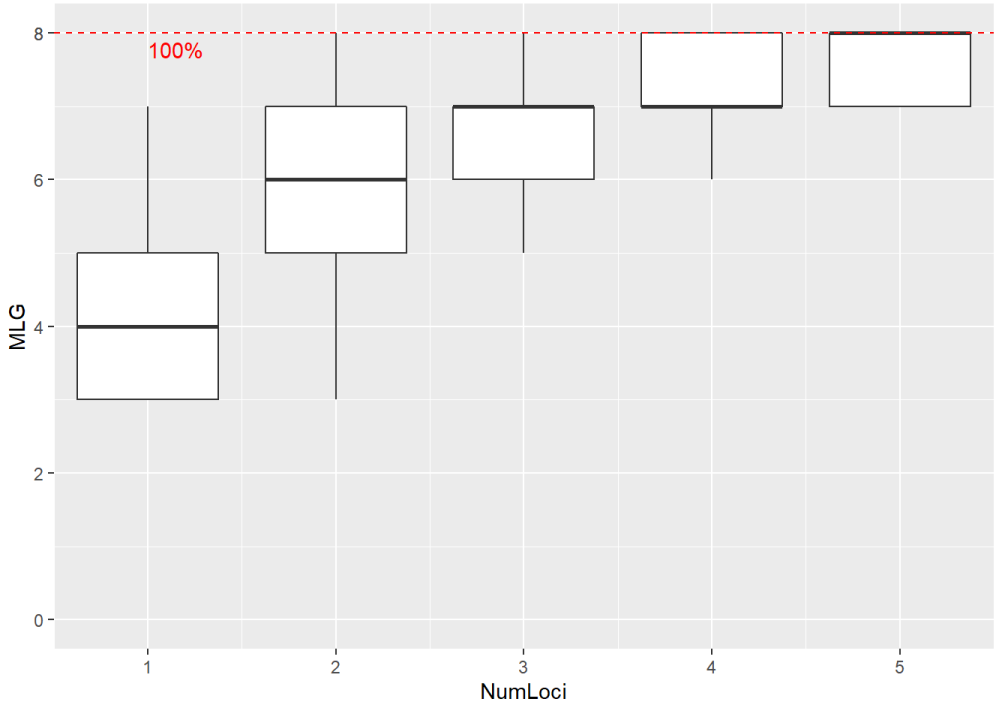
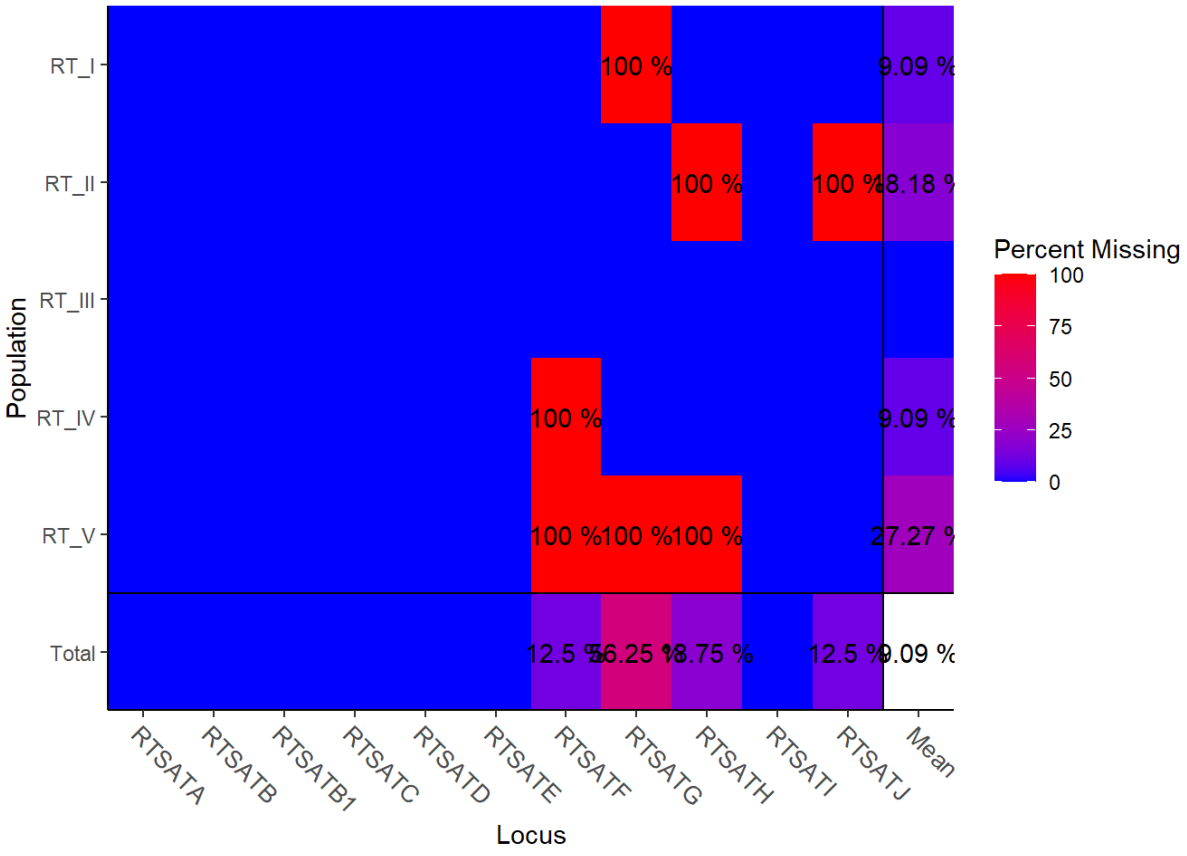


Figure 2.6 notes: NumLoci is referring to the number of loci or the populations in my SSR data set, which was five RT-I through RT-V populations and the MLG is referring to the multilocus genotypes, or the number of markers needed, which was the SSRs.

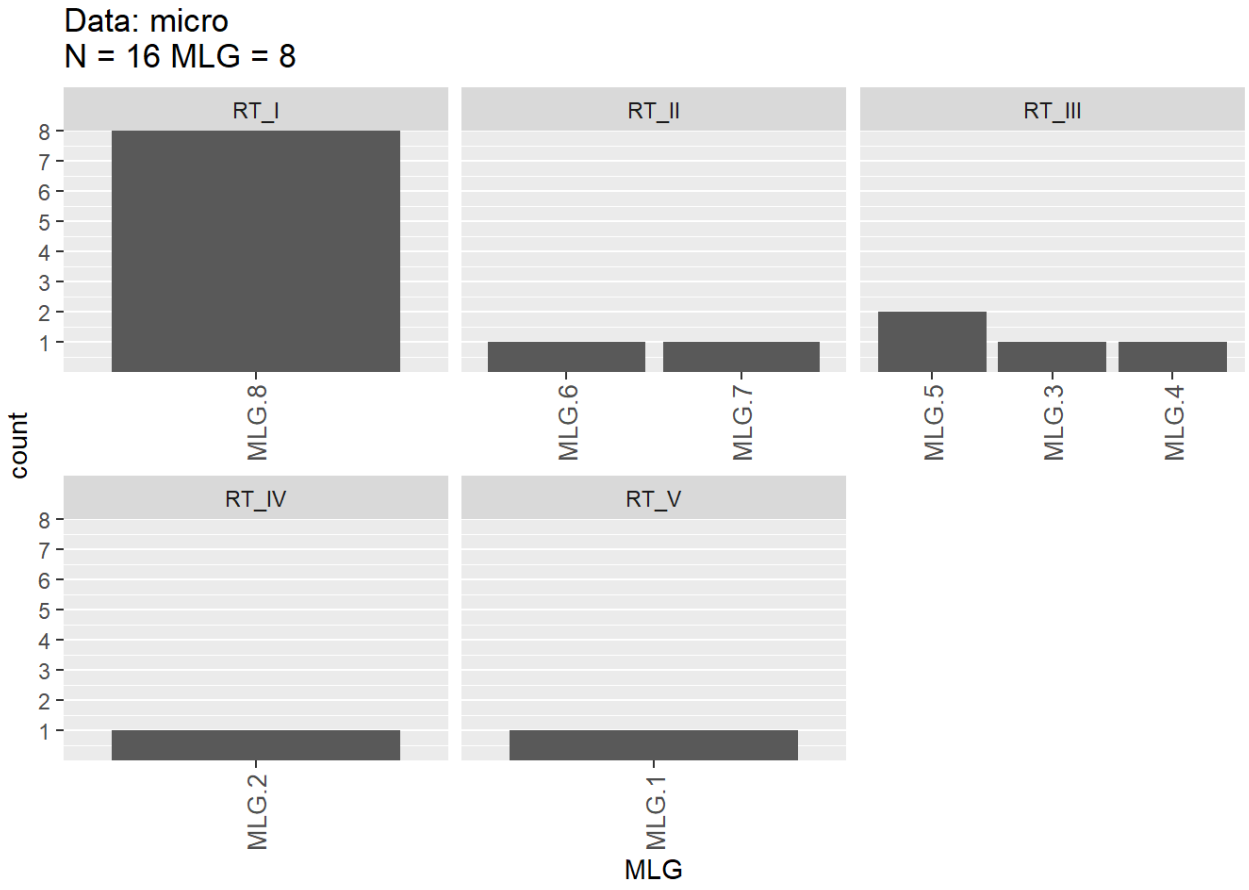
The missing data plot in the microsatellite analysis is intended for after loading data in order to look for missing data, rare alleles and overall quality of the data. Since the data consisted of polymorphic SSR markers, any instances where an SSR was not found in a strain, the unobserved allele would be designated as “0”. There is an overall average of 9% missing data, which was due to some of the SSRs not being found in some of the *R. toxicus* strains.

Figure 2.7: Missing data plot



A multilocus genotype analysis was performed to identify genotypic diversity (Figure 2.8). The number of individuals observed (N) and the number of multilocus genotypes (MLG) observed varied across populations. RT-III had the highest genotypic diversity and RT-I has the largest allelic occurrence. RT-I has the largest number of strains but only one multilocus genotype which is likely due to its clonal nature as it is a recently emerged population. This graph may help give insight into whether enough loci were sampled as well.

Figure 2.8: Multilocus genotype and genotypic diversity varied across the five *R. toxicus* populations



A discriminant analysis of principal components (DAPC) was conducted (Figure 2.9) to infer the number of clusters of genetically related individuals. It's a multivariate statistical approach where variance is partitioned into between-group and within-group components in order to maximize discrimination between groups. The DA stands for discriminant analysis. According to this analysis, RT-IV is the least related to the rest of the populations. It also suggests that RT-V may be the progenitor of the populations, or last common ancestor of the other four.

Figure 2.9: DAPC analysis of *R. toxicus* populations

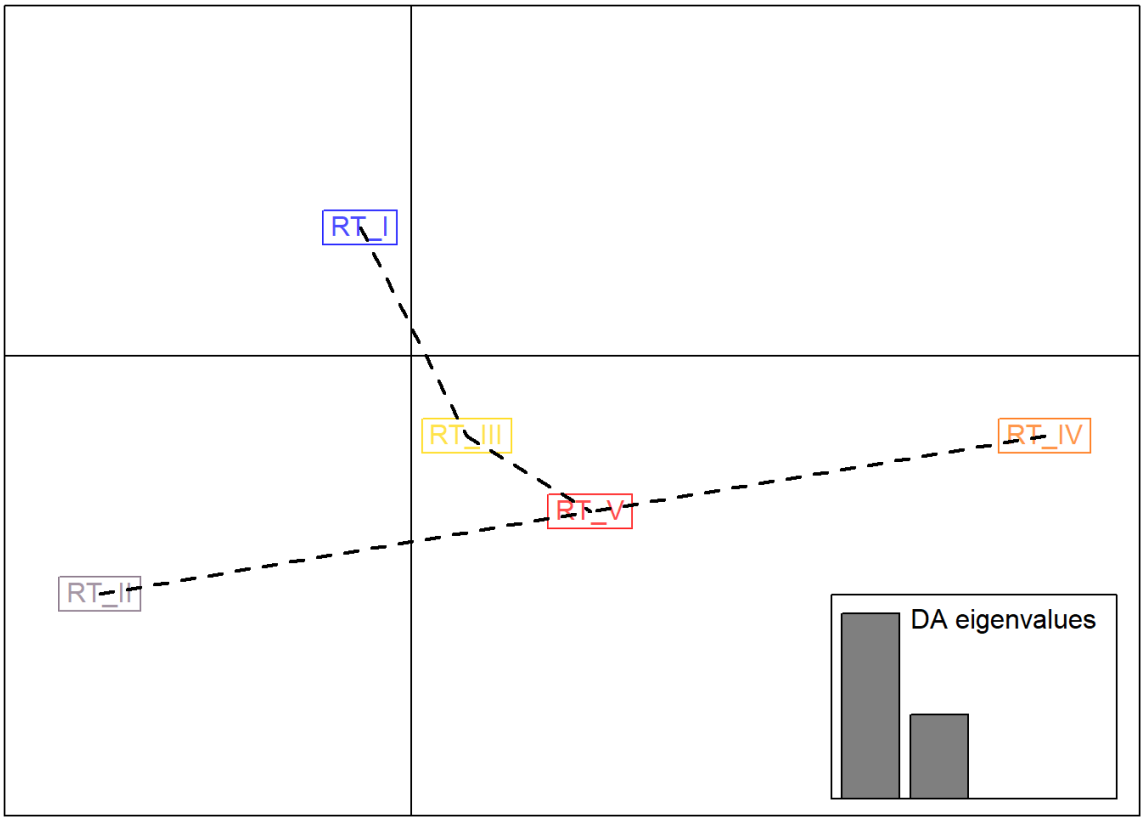
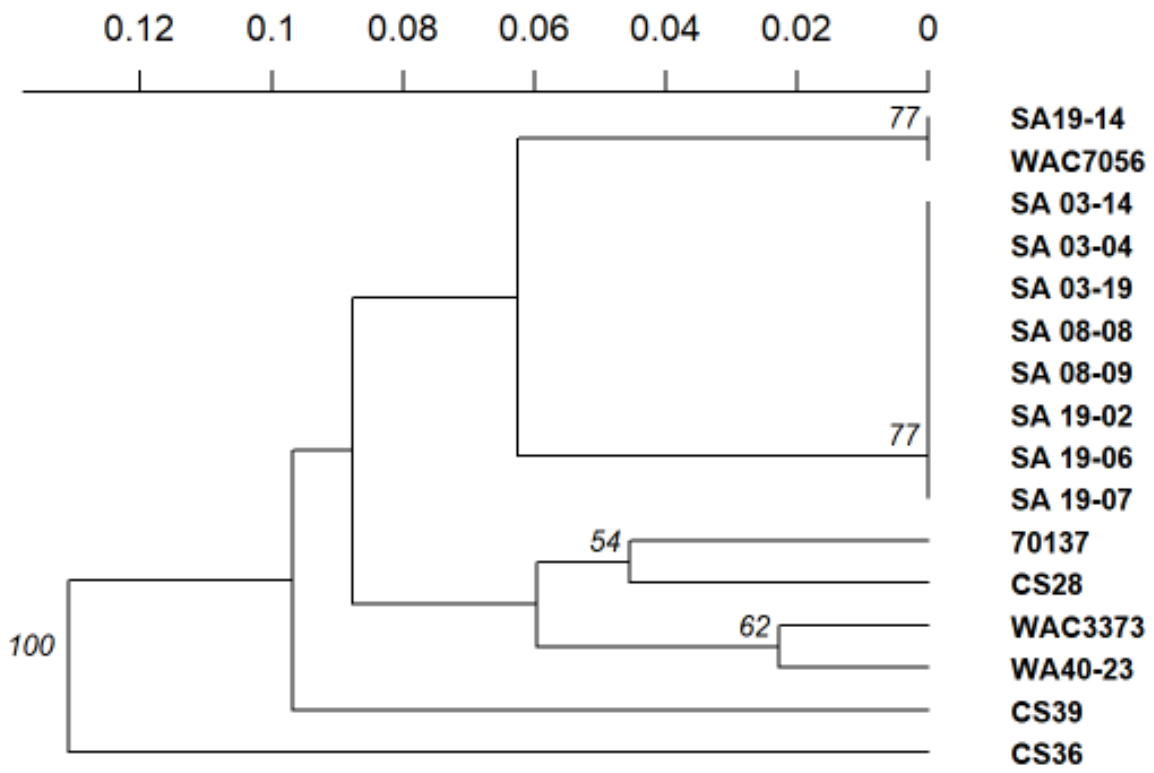


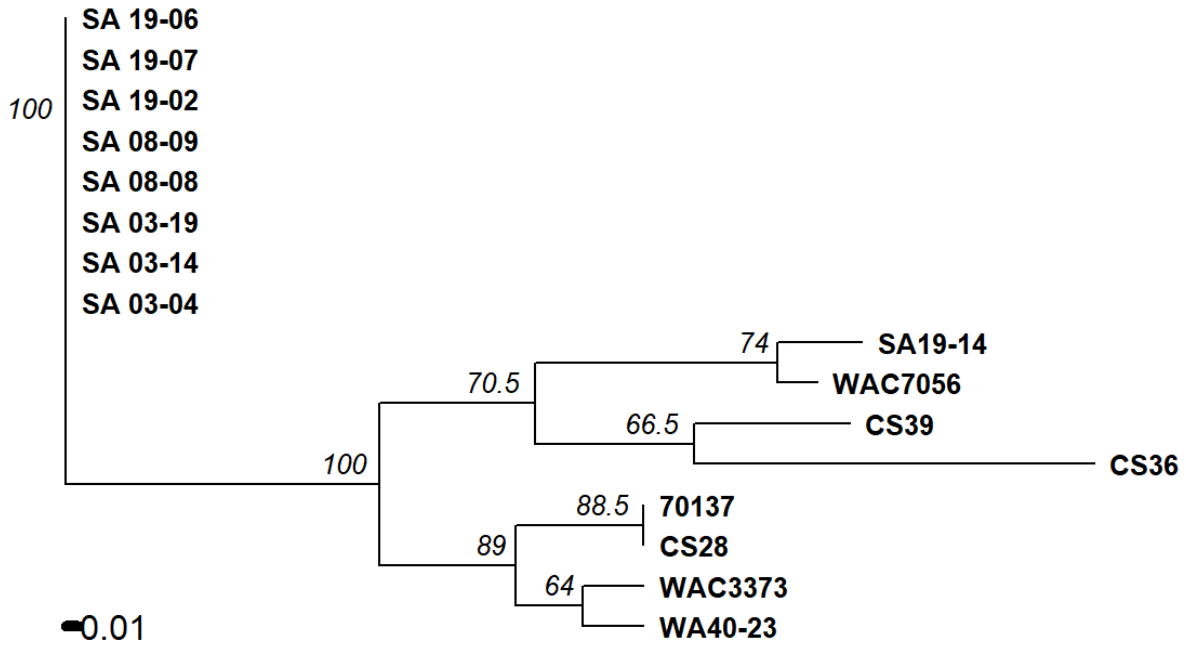
Figure 2.9 notes: RT-II is hidden behind the DA eigenvalues box, and RT-III is the population in yellow.

Figure 2.10: Bruvo's distance tree using SSRs across *R. toxicus* strains



Overall, the strain groupings that resulted (Figures 2.10-2.12) from the Bruvo's distance analysis in the R package were similar to the groupings that resulted from the MLST, WGS, and ISSR analyses as well as some of the other SSR analyses (Figure 2.2) that used all SSRs. The data input into Poppr was also analyzing all SSRs. However, the distances of the Bruvo trees from the Poppr package analyzing repeat numbers are not similar to those of the phylogenetic trees (Figures 2.2-2.5) derived from Geneious 9.8.1 analysis of sequence data. For example, in Poppr RT-I (the group of eight SA-) is more closely related to RT-II (SA-19 and WAC7056) while RT-I was more closely related to RT-IV (CS36) and RT-V (CS39) the sequence-based tree (Figure 2.2).

Figure 2.11: Neighbor joining phylogenetic tree using K means hierarchical method



The Poppr package contained an algorithm for phylogenetic analysis using the neighbor-joining algorithm with K means clustering. k-means clustering is a method that aims to partition genotype and samples into k clusters in which each observation belongs to the cluster with the nearest mean. This phylogenetic analysis resulted in population groupings that correspond to the Geneious Jukes Cantor neighbor joining method (Figure 2.2) and previous analyses, with differences in relative distances between populations such as relatedness of RT-II strains (SA19-14 and WAC7056) being closer to RT-IV (CS36) and RT-V (CS39) in this analysis rather than to RT-III in Geneious.

SSRs within annotated Genomes

To identify potential patterns in genome location among SSRs across *R. toxicus* strains, each SSR sequence was aligned in Geneious to the annotated genomes of the corresponding *Rathayibacter toxicus* strains in which it was found. Annotations were used to identify whether there were common genomic regions where the SSRs were located, whether they were in coding regions or non-coding regions, and if coding, the location in that sequence. The annotated genome data was obtained from Yasuhara-Bell et al. 2020 and included RAST annotations as well as the website mygene.info. The annotated genomes of strains representing genetic populations, RT-I-RT-V, were used to analyze each SSR, RTSatA-RTSatJ (Table 2.10).

Overall, there were no overarching patterns of genomic locations associating with SSRs. Some SSRs were located in non-coding regions, while others were in various coding regions, including within DNA and RNA and proteins sequences. This further shows how SSRs may occur in various genomic regions and may influence genomic processes or may be non-functional. Additionally, the results in Table 2.10 show conflicting annotations when two annotation software types are used for checking the same SSR in the same genome strain. For example, when looking at CS-39 MyGene.info versus RAST, the annotation of more than half of the SSRs does not match between the two, indicating possible issues with one or both of the annotation tools used.

Table 2.10: SSRs within Annotated *Rathayibacter toxicus* Genomes

	RTSatA	RTSatB	RTSatC	RTSatD	RTSatE	RTSatF	RTSatG	RTSatH	RTSatI	RTSatJ
RT-IV	**CS36** (RAST)	no annotation	Putative toxin component near putative ESAT-related proteins, repetitive / Repetitive hypothetical protein near ESAT cluster, SA0282 homolog. Region is 1452 bp long	Putative toxin component near putative ESAT-related proteins, repetitive / Repetitive hypothetical protein near ESAT cluster, SA0282 homolog. Region is 1398 bp long.	only partly in putative membrane protein. Region is 468 bp long.	in putative phosphatase. Region is 2334 bp long.	possible pre-pilin peptidase, region is 579 bp long.	this is not in CS36	no annotation	hypothetical protein, region is 1509 bp long	membrane associated zinc metalloprotease. Region is 1359 bp long	only partly in FIG007959 peptidase M16 family. Region is 1407 bp long
RT-V	**CS39** (RAST)	hypothetical protein, 1230 bp long, 1354212-1355441 (-)	Quinone oxidoreductase (EC 1.6.5.5), 978 bp long, 1572078-1573055 (+)	putative ABC transporter integral membrane protein, 1425 bp long, 2206250-2207674 (-)	no annotation	no annotation	putative integral membrane protein, 1242 bp long, 1865166-1866407 (-)	this is not in CS39	this is not in CS39	this is not in CS39	1-acyl-sn-glycerol-3-phosphate acyltranferase, 756 bp long, 810277-811032 (-)	cell division protein FtsK, 2841 bp long, 1549614-1552454 (-)

	CS39 (MyGene.Info)	no annotation	DNA/RNA nonspecific endonuclease, 1446 bp long, 1580647-1582092	DNA/RNA nonspecific endonuclease, 1482 bp long, 2218315-2219796	barely gets into propionyl-CoA carboylase beta chain, 1605 bp long, 550187-551791 (-)	hypothetical protein, 2334 bp long, 69496-71829 (-)	no annotation	this is not in CS39	this is not in CS39	this is not in CS39	1-acyl-sn-glycerol-3-phosphate acyltranferase, 756 bp long, 810277-811032(-)	S-DNA-T family DNA segregation ATPase FtsK/SpolllIE, 2841 bp long, 1549614-1552454 (-)
RT-I	**SA03-04** (RAST)	no annotation	Putative toxin component near ESAT related proteins, repetitive/repetitive hypothetical protein near ESAT cluster, SA0282 homolog, 1407 bp long, 1617957-1619363 (-)	Putative toxin component near ESAT related proteins, repetitive/repetitive hypothetical protein near ESAT cluster, SA0282 homolog, 1416 bp long, 2227720-2229135 (-)	partly gets into putative membrane protein, 2445 bp long, 552836-555256 (+)	putative phosphatase, 2322 bp long, 65600-67921 (-)	no annotation	no annotation	this is not in SA03-04	no annotation	membrane associated zinc metalloprotease, 1359 bp long, 825004-826362 (+)	barely in FIG007959: peptidase, M16 family, 1407 bp long, 1597610-1599016 (-)
	SA03-04 (My.Gene.Info)	no annotation	DNA/RNA non specific endonuclease, 1455 bp long, 1617957-1619411 (-)	DNA/RNA non specific endonuclease, 1464 bp long, 2227720-2229183 (-)	partly in lysylphosphatidylglycerol synthenase-like protein (DUF2156 family), 2538 bp long, 552743-555280 (+)	Hypothetical protein, 2322 bp long, 65600-67921 (-)	no annotation	no annotation	this is not in SA03-04	Cell wall associated NlpC family hydrolase, 1443 bp long, 235364-236806 (+)	membrane associated protease RseP (regulator of RpoE activity), 1440 bp long, 824923-826362 (+)	probable phosphoglycerate mutase, 603 bp long, 1596936-1597538 (-)

RT-III	**WA40-23** (RAST)	no annotation	Putative toxin component near ESAT related proteins, repetitive/repetitive hypothetical protein near ESAT cluster, SA0282 homolog, 1434 bp long, 11604979-1606412 (-)	Putative toxin component near putative ESAT-related proteins, repetitive / Repetitive hypothetical protein near ESAT cluster, SA0282 homolog, 2211410-2212249, 840 bp long	partly in putative membrane protein, 2445 bp long, 570824-573268	putative phosphatase, 2385 bp long, 69491-71875	no annotation	no annotation	no annotation	putative secreted protein, 1476 bp long, 257947-259422 (+)	membrane associated zinc metalloprotease, 1359 bp long, 834064-835422 (+)	barely in FIG007959: peptidase, M16 family, 1407 bp long, 1584743-1586149 (-)
	WA40-23 (MyGene.Info)	no annotation	DNA/RNA non specific endonuclease, 1482 bp long, 1604979-1606460 (-)	DNA/RNA non-specific endonuclease, 840 bp long, 2211410-2212249 (-)	partly in lysylphosphatidylglycerol synthenase-like protein (DUF2156 family), 2538 bp long, 570731-573268 (+)	hypothetical protein, 2340 bp long, 69491-71830 (-)	no annotation	no annotation	no annotation	cell wall associated NlpC family hydrolase, 1476 bp long, 257947-259422 (+)	membrane associated protease RseP (regulator of RpoE activity), 1359 bp long, 834064-835422 (+)	barely in predicted Zn-dependent peptidase, 1419 bp long, 1584743-1586161 (-)
RT-II	**WAC7056** (MyGene.Info)	no annotation	DNA/RNA non specific endonuclease, 1473 bp long, 1594068-1595540 (-)	no annotation	partly in lysylphosphatidylglycerol synthenase-like protein (DUF2156 family), 2538 bp long,	hypothetical protein, 2346 bp long, 69489-	no annotation	no annotation	no annotation	not in WAC7056	membrane-associated protease RseP (regulator of RpoE activity) 1359 bp	this is not in WAC7056

	559394-561931 (+)	71834 (-)	long, 827085-828443 (+)

Table 2.10 Notes: when the table says, "no annotation", that means that SSR exists in that genome but there was no annotation in the annotated genome data. When the table says, "not in 'name of strain'" or "this is not in 'name of strain", this means the SSR was inexistent in that specific strain in the first place. This part of the data is also described in Table 2.6 and indicated by "N/A"

Discussion and Conclusions

SSR analyses have often been used to identify population structure and phylogenetic relationships in various organisms. Their relatively fast mutation rate relative to the background genome may better highlight genetic differences in organisms that have a relatively slow generation time. *Rathayibacter toxicus* is a relatively slow regenerating bacterial pathogen in which we have samples from 1975 to 2014, and a generation is considered to be one entire season due to the complexity of its lifecycle. SSR analysis of *R. toxicu*s showed variation of individual and groups of SSRs across strains, populations that mostly corresponded with previous analyses, but did not yield consistent distance relationships between populations.

Single SSR analyses resulted in unique groupings; most SSRs yielded different phylogenetic population groupings. Individual SSRs varied across strains and phylogenetic trees showed conflicting results depending upon which combination of satellites were used in the analysis, therefore, caution should be exercised when interpreting SSR variance for evolutionary dynamics studies. It matters which SSRs are analyzed. For analyses with only the microsatellites without the minisatellite, the phylogenetic tree was slightly different, just as each of the combination are slightly different from each other, either in groupings or distances. Complementing SSR analysis with other population genetic analyses or using multiple satellites within a genome may better identify and provide better insight into phylogeographic relationships.

It is important to note that the SSR parameters used in this study were the same as those that are widely used and commonly published for population genetics and evolutionary dynamics studies. Studies vary in the number and combination of parameters used; e.g., some studies use a few parameters while others use most or all. The decisions sometimes seem arbitrary (Vieira et

al. 2016, Oliviera et al. 2006). For example, many studies focus solely on microsatellites which disqualify minisatellites and macrosatellites. The question becomes what is the most appropriate measure for satellite variation? How should we choose which SSRs to use and which to ignore? The relevance of this question becomes clear when comparing the differences observed among the phylogenetic trees generated based on different *R. toxicus* SSRs combinations. The four phylogenetic trees (figures 2.4-2.7) differ slightly but significantly among one another.

When all SSRs, including the minisatellite (32 nucleotide motif), Figure 2.2, were analyzed across all 16 *Rathayibacter toxicus* genomes, the phylogenetic tree groupings supported the previously identified populations established by MLST, ISSR, and comparative whole genome analyses (Arif et al. 2016, Yasuhara-Bell and Stack 2019, Yasuhara-Bell et al. 2020). Using only microsatellites in phylogenetic analysis is one of the most common practices reported in the literature.

When all of the concatenated SSRs were analyzed excluding the minisatellite, Figure 2.3, significant differences were observed compared to the analysis that included the minisatellite; specifically, the relatedness between RT-IV and RT-V and among RT-IV and RT-V with other strains was very different as well as one RT-III strain split out from the rest of the RT-III strains. These groupings in this phylogenetic tree were largely the same as the populations based on other methods with those few significant relatedness differences.

Another very common approach to satellite analysis is including only the SSRs that occur in every strain analyzed, shown in Figure 2.4. That approach excludes several of the SSRs used in this study. The resulting phylogenetic tree contains groupings and distances similar to those of the previously established, with some variation. It also suggests a high degree of relatedness between RT-II + RTIII and between RT-IV + RT-V.

The most exclusive SSR parameter set, which is also commonly used by researchers, concatenates only microsatellites that occur in every strain, shown in Figure 2.5. Groupings and distances changed with this analysis; RT-V was much closer to RT-I, RT-III split into two groups, RT-II strains were more closely related to two of the RT-III strains, and it suggests that RT-V may be the progenitor population.

A byproduct of this study was the analysis of the tools themselves; how do we determine whether the methods used and the parameters selected result in an accurate representation of reality? The fact that different parameter combinations yielded different results with respect to relatedness shows that using multiple methods may help in confirmation of any one method or in pointing out inconsistencies.

Another effect of checking which SSRs occur in which potential coding regions using annotated genomes from two different annotation sources is seeing how one cannot rely on one method alone for full certainty. There are differences among the tools used and this is another example. There was an instance where the same region in the same strain and genome was annotated slightly differently, and when a specific genomic location was searched for according to location of a specific SSR, the annotation was not the same. An example of this occurring is when looking in Table 2.10 at SSR – RTSatB copy 1, in strain CS39, using RAST versus MyGene.Info results, the annotation does not match. RAST says it is within Quinone oxidoreductase, and Gene Info says it's within a DNA/RNA nonspecific endonuclease. Therefore, checking with more resources may be necessary to determine which one is accurate, if either.

Rathayibacter toxicus also has the potential to be used for diagnostic purposes. Each of the individual SSRs' forward and reverse primer regions (Table 2.7) were BLAST searched

through the NCBI database, along with their in order to check for exact matches in other genomes. There were no results which had the same forward and reverse sequences. Forty-one *R. toxicus* genomes were tested for the presence of SSRs with the highest variability and largest motif sizes (RTSatA and RTSatB), with controls including a different species of Rathayibacter and another bacterium (*Pseudomonas syringae*); the controls remained negative which further indicates the specificity of these SSRs to *R. toxicus*. SSRs RTSatA-J show a spectrum of further specificity within *R. toxicus* from species to strain level. One SSR (RTSatA) was able to differentiate to strain level in RT-I, while other SSRs (RTSatB, RTSatC, RTSatD, RTSatF, RTSatG, RTSatJ) were only able to distinguish between the five previously established *R. toxicus* populations. Some SSRs (RTSatE and RTSatI) were unable to distinguish among populations but were unique to the *Rathayibacter toxicus* species. Overall, SSRs were suggestive but not conclusive for determining the geographic or phylogenetic origin of *R. toxicus*. This may also mean that one cannot rely solely on SSRs as a population genetics tool without other methods to verify the results.

References (Chapter 2)

Arif, M., Busot, G. Y., Mann, R., Rodoni, B., Liu, S., & Stack, J. P. (2016). Emergence of a new population of *Rathayibacter toxicus*: an ecologically complex, geographically isolated bacterium. *PLoS One*, *11*(5), e0156182.

Arnaud-Hanod, S., Duarte, C. M., Alberto, F., and Serrão, E. A. (2007). Standardizing methods to address clonality in population studies. *Molecular ecology*, *16*(24), 5115-5139.

Bruvo, R., Michiels, N. K., D'SOUZA, T. G., & Schulenburg, H. (2004). A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. *Molecular ecology*, *13*(7), 2101-2106.

Comings, D. E. (1998). Polygenic inheritance and micro/minisatellites. *Molecular psychiatry*, *3*(1).

Fondon III, J. W., Hammock, E. A., Hannan, A. J., & King, D. G. (2008). Simple sequence repeats: genetic modulators of brain function and behavior. *Trends in neurosciences*, *31*(7), 328-334.

Fondon, J. W., & Garner, H. R. (2004). Molecular origins of rapid and continuous morphological evolution. *Proceedings of the National Academy of Sciences*, 101(52), 18058-18063.

Gao, C., Ren, X., Mason, A. S., Li, J., Wang, W., Xiao, M., & Fu, D. (2013). Revisiting an important component of plant genomes: microsatellites. *Functional Plant Biology*, *40*(7), 645-661.

Gemayel, R., Cho, J., Boeynaems, S., & Verstrepen, K. J. (2012). Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes*, 3(3), 461-480.

Gerber, H. P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S., & Schaffner, W. (1994). Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science*, 263(5148), 808-811.

Hamada, H. Seidman, M., Howard, B. H., & Gorman, C. M. (1984). Enhanced gene expression by the poly (dT-dG). poly (dC-dA) sequence. *Molecular and cellular biology*, 4(12), 2622-2630.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, *11*(1), 94.

Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Frontiers in genetics*, *6*, 208.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

Kashi, Y., King, D., & Soller, M. (1997). Simple sequence repeats as a source of quantitative genetic variation. *Trends in genetics*, 13(2), 74-78.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Mentjies P., Drummond A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647-1649.

King, C. M., Bristow-Craig, H. E., Gillespie, E. S., & Barnett, Y. A. (1997). In vivo antioxidant status, DNA damage, mutation and DNA repair capacity in cultured lymphocytes from healthy 75-to 80-year-old humans. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 377(1), 137-147.

King, T. L., Lubinski, B. A., & Spidle, A. P. (2001). Microsatellite DNA variation in Atlantic sturgeon (*Acipenser oxyrinchus oxyrinchus*) and cross-species amplification in the *Acipenseridae*. *Conservation Genetics*, 2(2), 103-119.

Kovtun, I. V., & McMurray, C. T. (2008). Features of trinucleotide repeat instability in vivo. *Cell research*, 18(1), 198-213.

Li Y-C, Korol AB, Fahima T and Nevo E (2004) Microsatellites within genes: structure, function, and evolution. *Molecular biology and evolution*, *21*(6), 991-1007.

Mason AS (2015) SSR Genotyping. *Batley Journal ed) Plant Genotyping*, p 77-89.

Mayer, Christoph (2006-2010). Phobos 3.3.11.
Online: <http://www.rub.de/ecoevo/cm/cm_phobos.htm>.

Nevo E (2001) Evolution of genome-phenome diversity under environmental stress. *Proceedings of the National Academy of Sciences*, *98*(11), 6233-6240.

Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R and Vieira MLC (2006). Origin, evolution and genome distribution of microsatellites. *Genetic Molecular Biology*, 29:294-307.

Pérez-Jiménez, M., Besnard, G., Dorado, G., & Hernandez, P. (2013). Varietal tracing of virgin olive oils based on plastid DNA variation profiling. *PLoS One*, 8(8), e70507.

Phumichai C, Phumichai T and Wongkaew A (2015) Novel chloroplast microsatellite (cpSSR) markers for genetic diversity assessment of cultivated and wild Hevea rubber. *Plant Molecular Biology Reporter*, *33*(5), 1486-1498.

Richards, R. I., & Sutherland, G. R. (1994). Simple repeat DNA is not replicated simply. *Nature genetics*, 6(2), 114-116.

Trifonov, E. N. (1989). The multiple codes of nucleotide sequences. *Bulletin of mathematical biology*, 51(4), 417-432.

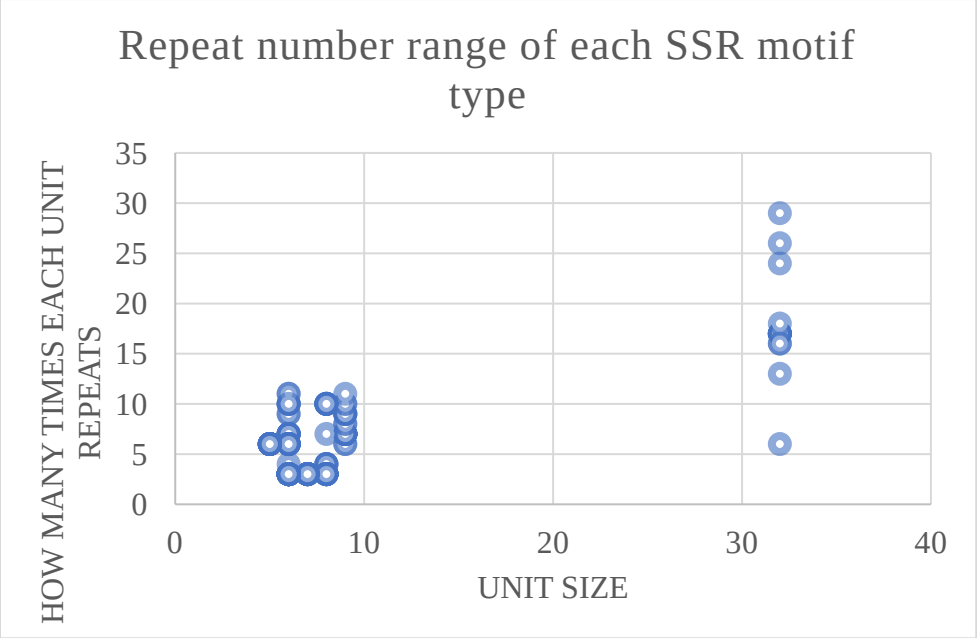
Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. D. F. (2016). Microsatellite markers: what they mean and why they are so useful. *Genetics and molecular biology*, *39*(3), 312-328.

Yasuhara-Bell, J., & Stack, J. P. (2019). Panel of three loop-mediated isothermal amplification assays differentiates *Rathayibacter toxicus* populations RT-I, RT-II, RT-III, RT-IV and RT-V. *Journal of Plant Pathology*, *101*(3), 707-717.

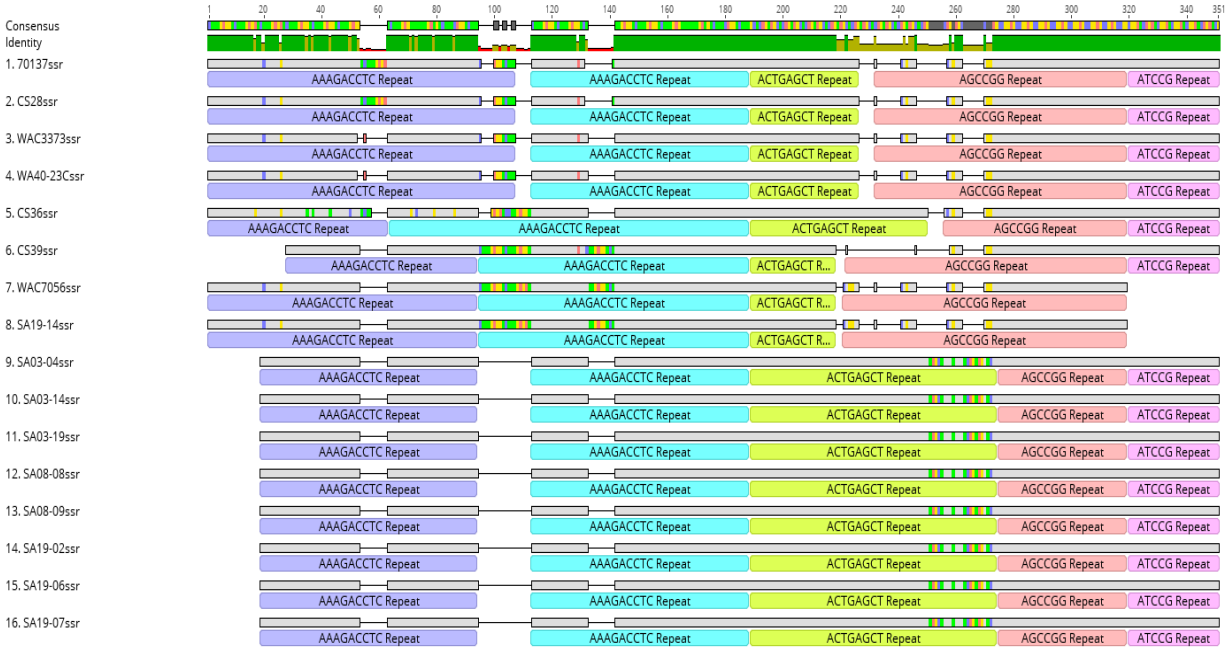
Yasuhara-Bell, J., Arif, M., Busot, G. Y., Mann, R., Rodoni, B., & Stack, J. P. (2020). Comparative Genomic Analysis Confirms Five Genetic Populations of the Select Agent, *Rathayibacter toxicus*. *Microorganisms*, *8*(3), 366.

Appendices (Chapter 2)

Appendix 1.1 Graph of range of SSR motif sizes and repeat numbers



Appendix 1.2 Visualization of an example of the concatenation process in Geneious: concatenation of 5 SSRs



Appendix 1.2 notes: This figure shows Purple is RTSatB copy 1, blue is RTSatB copy 2, green is RTSatC, pink is RTSatD, purple is RTSatE. These were first concatenated and aligned, for each of the 16 strains (named on the left side under "consensus identity").

Chapter 3: *Rathayibacter toxicus* Pathogen Pathway Risk Analysis

Abstract

Rathayibacter toxicus is a plant pathogenic bacterium that has caused the death of hundreds of thousands of animals in Australia since the 1950s. The bacterium involves a complex life cycle where a nematode vector carries the bacterium into grass ovules where it establishes galls within the seed structure and produces a toxin that is lethal to mammals. *Rathayibacter toxicus* and its toxin has been a threat to Australian agriculture for decades and now threatens other countries because of the increase in globalized trade markets including hay exports from Australia. Countries including the U.S. may be at risk of *R. toxicus* introduction, which may include potential establishment, spread, and ultimate disease outbreaks. It is necessary to understand how the pathogen survives and spreads within and from the source country, how it is moved from source to destination country, and how environmental conditions at the destination may facilitate pathogen establishment. This qualitative pathway analysis was developed to study the possible avenues through which *R. toxicus* may elude containment and reach areas outside of outbreak centers in Australia. A scenario of *R. toxicus* spread was analyzed which included assuming a range of concentration of bacterial galls within a hay field harvested for hay bales for export out of the country. It is evident that risk exists for *R. toxicus* gall release at multiple points throughout its journey from a grass panicle in an Australian paddock, all the way through to the arrival on a packaged bale of hay reaching a feeding pasture in a U.S. livestock farm.

Introduction

Rathayibacter toxicus is a complex plant pathogenic bacterium that has the potential to sicken and kill animals that it comes into contact, if enough of its toxic galls are consumed. Infected grass can remain toxic even when it has senesced and dried off and repurposed as hay (Price et al. 1979; Stynes and Bird 1982; Government of Australia 2018). Outbreaks of toxicity and animal poisoning caused by *R. toxicus* have been prevalent throughout fifty years in Australia, over several regions within the country, and are commonly known as "annual ryegrass toxicity" or "Flood plain staggers" outbreaks (McKay et al. 1993). *Rathayibacter toxicus* is one of only seven plant pathogens on the U.S. Select agent list because of its ability to cause disease in plants and toxicities in animals which consume the affected plants (Riley and Ophel 1992). So far, *R. toxicus* has been widespread only in Australia and detected in South Africa (Schneider 1981, Grewar et al. 2009). *Rathayibacter toxicus* detections have not occurred in the U.S. to date. However, through globalization and movement of people, imports, and exports, it may be difficult to detect and stop the spread of pathogens associated with these, and complete containment can never be guaranteed with full certainty.

Certain rules and regulations exist that aim to prevent the introduction and spread of pathogens at the global and international level. The International Plant Protection Convention (IPPC) and the Sanitary and Phytosanitary agreement (SPS) are two such organizations that strive to help plant protection throughout international trade markets (FAO 1996; Devorshak 2012). These entities and their regulations have existed for several years, but there are still issues with being able to fully implement the guidelines on a practical scale (Brasier 2008). One example is in the issue of sampling for detection of unwanted pests and pathogens; in the U.S. less than 2% of incoming containers are inspected at ports of entry (NRC 2002) and inspections

are often based on visual detection of disease symptoms caused by known and listed organisms, leaving much room for human error and unknown pathogens or asymptomatic diseases to pass through (Brasier 2008; USDA APHIS PPQ, 2012a). The sampling issue is also relevant specifically to *R. toxicus*. According to the guidelines of the Australian Quarantine and Inspection Service, there is a set standard for minimizing the risk of corynetoxin contamination of hay and straw for export, which similarly only requires a relatively small percentage of product to be tested. For example, it suggests that only 15% of all bales in a paddock should be sampled if there is a total of more than 12 bales, or one of a group of 12 or less bales should be sampled. And if the testing is negative, all bales in their groups meet the standard without needing further testing (Australian quarantine and Inspection Service 2019). There are additional guidelines as well, but they are not extremely thorough and some of the rules are rather arbitrary.

There are many phytosanitary measures that should be applied to the international movement of agricultural commodities, because imported agricultural products can hold non-indigenous pests that could threaten domestic agricultural industries (Reed, 2001). Efforts taken must be based on international standards and scientific principles (Griffin, 2012; Reed, 2001). There is a general standard that is the Pest Risk Analysis (PRA) and SPS-IPPC framework is part of the process (PRA, Griffin 2012). PRA analyzes information to determine the amount of risk that a pathogen may represent and how phytosanitary measures can be used to combat it (IPPC, 2004; IPPC, 2007). PRA is mainly a scientific decision-support tool used to justify phytosanitary measures, but it can aid in determining the potential of new pathogen risks of entrance and establishment in a new place for conducting surveillance programs (Devorshak, 2012). A PRA can be applied to consider pathways or means by which a pest can gain entry and spread from one location to another (Devorshak, 2012). There are various types of pathways such as through

natural spread, purposeful movement, and through countless commodities, with the latter being a major driver (Devorshak, 2012).

A pathway pest risk analysis considers important steps through transmission points that must occur for a pathogen to leave its origin, survive or escape during travel, become established, and spread in a new location. Analysis of the situation prior exporting a commodity (with pest or pathogen included) may include looking at pest prevalence and disease outbreaks at the place of origin, infestation at the field level, and amount of commodity for export. Some of the many scenarios that can be considered during such a pathway analysis (such as looking at the travel pathway of the infested commodity between source and sink) is commodity loss and spillage during shipment and transit, presence and amount of hosts, environmental conditions for pest overwintering survival, establishment, spread, and outbreak. A plant disease outbreak refers to a level of disease sufficient to cause economic loss or an epidemic greater than what would normally be expected in a particular geographical area or season (Adapted from McMichael et al., 2003). Pathways that could lead to pest arrival, establishment and spread are analyzed by pathway pest risk analyses (Devorshak, 2012). In some instances, only the likelihood of entry, establishment, or spread is analyzed. In general, conservative regulatory approaches are normal because studies often assume establishment of pathogen is possible just with its entry because of the complexity associated with establishment (Rafoss, 2003; Devorshak, 2012).

Rathayibacter toxicus is a pathogen that can have dire consequences, such as animal lethality and loss of forage crops, if there are ways in which it is released and spread past its origin. A pathway analysis can look at such possibilities. Several factors about *R. toxicus* can be taken into consideration, such as host range, vector range, as well as survivability in certain weather and climate environments, and more. *Rathayibacter toxicus* has many host possibilities

within the Poaceae family of plants. These are grasses that include *Lolium rigidum, Agrostis avanacea, Polypogon monospeliensis, Ehrharta longiflora, Austrodanthonia caespitosa, Avena sativa, Avena caespitosa, Danthonia caespitosa, Lolium multiflorum, Lolium perenne, Lolium persicum, Lolium strictum, Lolium temulentum, Phalaris species*, *Vulpia myuros,* and more (Bertozzi and Davies, 2009; Bertozzi and McKay, 1995; Chatel et al., 1979; Edgar et al., 1994; McKay et al., 1993; Riley, 1992a ; Riley, 1995; Riley, 1996; Riley et al., 2001; Riley and Barbetti, 2008). The known nematode vectors of *R. toxicus* are within the *Anguina* genus. These most commonly include *Aguina funesta, Anguina tritici, Anguina australis,* and *Anguina paludicola* (Riley 1995; Riley and McKay 1990, Riley and McKay 1991, Riley 1992a; Riley et al. 2001; Bertozzi and Davies 2009, Riley and Barbetti, 2008). *Rathayibacter toxicus* survivability is strong, as bacterial galls can survive for many years in a dry state. They can immediately be activated through rehydration in rainy conditions.

The climate of the Western Australia region is known to be mostly hot and dry. This is a conducive environment for *R. toxicus* senesced galls to survive and overwinter in. Arid climate is the most common condition in Australia. Many *Rathayibacter* species diseases happen in arid areas and there are biological characteristics of it that aids in their survival as well, such as production of extracellular polysaccharides (Murray et al. 2017). In the Western Australian arid environment, *Rathayibacter* genus can overwinter multiple decades. This also happens for nematode species. The nematodes drop to the soil surface near as plants senesce and survive anhydrobiotically until the following season (Murray et al. 2017). The freshly hatched juveniles cannot survive desiccation but mature physiologically to become the survival state as the host senesces and desiccates (Riley and Barbetti 2008). To summarize, the seed gall nematode vectors of *Rathayibacter* spp. (including *R. toxicus*) feed and reproduce within the developing ovaries of

plant seeds and overwinter in seed galls as anhydrobiotic juveniles or adults, and can survive within the seed gall for many years in this dormant stage. For example, *Anguina tritici* did not lose its ability to invade wheat seedlings after 32 years of dormancy (Limber 1973).

Characterization of the origin and destination can be done and they can be compared for similarities in conditions which would be favorable to establishment of a pathogen. Both Australia and U.S.A. have similar geographical and climate factors within the countries. U.S. land size is slightly larger than Australia, at almost 10 million square kilometers, while Australia is almost 8million square km. Both countries have large amounts of agricultural land, arable land, cropland, as well as pastures and forests. Concerning climate, there are parallel regions in U.S. and Australia. Both have temperate, arid, semiarid, and tropical regions. The proportion of these may vary, but both countries are large and diverse with many contrasting environments and varying land use.

Analyzing pathogen pathways and likelihoods of entry and establishment in new locations may take into account a combination of as many factors as possible. These factors include the basic information about the pathogen, such as the hosts, vectors, life cycle, survivability. Additionally, looking at the conditions of the source and destination locations adds more information to an analysis. The travel between source and destination can be described as well. This can include studying which points during a commodity's transfer is pathogen testing done, if any, the possible limitations in terms of sampling and error rates, and which entities in which countries are responsible for managing this process and approving the movement of materials. Overall, there are many avenues and factors to research and analyze. Assumptions can be made about how much invasive pathogen is overlooked and surpasses testing given certain known information; weaknesses in potentially important questions can be highlighted.

Materials and Methods

Rathayibacter toxicus pathway analysis chapter consisted of computer-based analyses made using Microsoft Office and Geographic Information System (GIS) software, ArcGIS Pro, and Quantum GIS 3.14 (QGIS). Microsoft Excel was used for data and calculations on various portions of the pathogen pathway scenario. Microsoft Office PowerPoint was used for the development of the pathway timeline flowcharts. This included creating all flowcharts of points of focus in the pathway of *R. toxicus* leaving the origin and establishing in its destination.

The climate maps of Australia and U.S were made with the help of Dr. Avocat of the Kansas State University Geography and Geospatial Sciences department. The Koppen-Geiger climate classification system files available in Google Earth format (kmz, kml), were converted to raster (geotiff) and vector format (shapefile). The shapefile geometry and topology was repaired using repair geometry tools in QGIS. in order to enable calculations and spatial analyses. Some calculations being impossible in a geographical (spherical) coordinate system relying on angular units of measures, the files were affected projected coordinate systems based on linear units of measures (NWS Lambert – EPSG 3308 for Australia, and Albers Equal area conic – EPGS 102003 for the contiguous United States). The shapefiles for the U.S. and Australian state outlines were provided by the U.S. Census Bureau, Natural Earth Data, and U.S. Geological survey.

An Excel database with the GPS coordinates of *R. toxicus* samples collected by various scientists was compiled and used to display the location of the sites along with climate zones, to visually assess a functional/spatial association between the presence of *R. toxicus* and the climate characteristics. Samples with missing GPS coordinates were applied with approximate coordinates of the closest town and specific area in the database. Latitude and longitude in form

of X, Y coordinates were retrieved for a total of 81 sites. The sites were initially displayed in WGS 84, which is the global geographical coordinate system used by the Global Positioning System technology, and then projected in NWS Lambert. A simple diamond shaped symbol was used for the Australian *R. toxicus* sites as well as for the top ports of entry of hay in the U.S. The U.S. climate map was established after creating an Excel database containing the approximate X,Y coordinates of the ten top ports of entry from WorldCity website and U.S. Census Bureau. These Australian and U.S. climate maps with *R. toxicus* sites and hay entry ports were created using Quantum GIS 3.14.

The host and vector presence information for each of 50 U.S. states was extrapolated from the websites of CABI Invasive species compendium, US Department of Agriculture (USDA) Fire Effects Information System, and USDA Natural Resource Conservation Service Plants Database, input into Microsoft Excel and Access tables compatible with applying into GIS programs. The data depicted a presence/absence as well as proportion data for each state: the rows depicted each state and the columns expressed the total of hosts and vectors respectively out of 15 and 4. This table was imported into ArcGIS Pro and QGIS. These data being qualitative (presence/absence) or standardized, a choropleth map with a single-color scheme was used to depict the proportion of hosts, respecting the low to high, light to dark rule of the international cartographic convections. The vectors were symbolized using different types of stripes, single stripes for one vector, cross stripe pattern for two vectors, which was the maximum number of vectors present in any one state.

A GIS analysis of the Western US (Washington, Oregon, and California) was performed in ArcGIS Pro using the United States Land Cover Database (NLCD) from the U.S. Geological Survey (Dewitz 2019, USGS 2016). This database contains 20 different land cover types in a

raster file with a spatial resolution of 30 meters. Three land cover types were extracted and clipped to the extent of the study area: grassland/herbaceous, pasture/hay, and cultivated crops. The number of each of these categories' pixels were calculated using the zonal statistics tools, then converted to square kilometers. Major highways were also added to this map. Much of these GIS analyses were done with the guidance of Dr Avocat, KSU Geography and Geospatial sciences.

Data/Results

Data and results are organized into sections of (1) origin/source, (2) transport/dissemination, and (3) destination/introduction. #1 breaks down into harvest and packing points, #2 breaks down into shipping and entry, #3 breaks down into escape and establishment. The Harvest, Pack, Ship, Entry, Escape, and Establish are the pathway factors known as "risk attrition" components, which were determined by USDA Plant Protection and Quarantine (APHIS) scientists, R. Sequeira and R. Griffin. This model is shown by Figure 3.4, Main Pathway Components of *Rathayibacter toxicus* establishment Scenario. "Scenario: Western Australia → U.S."

List of Assumptions for Maximum Likelihood *R. toxicus* escape scenario
1. *R. toxicus* travels with vector and host
2. *R. toxicus* will go undetected at some point throughout the process
3. Surveillance, diagnostic and certification testing have potential of error
a. human error
b. sampling error
c. equipment error
4. Known concentration of *R. toxicus* galls in outbreaks = 10k galls/kg of hay
5. 1 gall = 1 seed (1 seed is replaced by a gall when establishment occurs)
6. Source of *R. toxicus*: Western Australia (biggest hay producing state of Australia)
a. A total of approximately 55k value of hay is imported directly from Australia to U.S. and that is smaller than the possible hay yield from an average WA farm of around 4k hectares. This scenario assumes that one specific farm had an outbreak which was undetected and that the entirety of the Australian hay in U.S. is from this one location.

1A: Origin/Source - Harvest

Rathayibacter toxicus exists in Australia and may theoretically enter the United States; looking at a pathway scenario from a paddock in Western Australia to a pasture in U.S. using

known average information on location of *R. toxicus* and maximum likelihood factors of the process is important for the pathway analysis.

For this pathway analysis, a theoretical hay farm 1k hectares in size is the source of hay and *R. toxicus* outbreak, based on Australian grain crop farms being 1082 hectares on average (GrainCentral.com 2017). *Rathayibacter toxicus* has historically occurred in Western Australia (WA), so the type of environmental conditions there are conducive to maintaining *R. toxicus*. The weather in this location is mostly hot and dry, and this is important because it shows us a type of climate that has been conducive to the survival and spread of *R. toxicus* for many decades.

Western Australia was also chosen as an area of focus because it is the predominant Australian state for hay production and exports the largest percentage of Australian hay out of the country, at 48% of the total (Government of Western Australia 2017). In 2013, Australia exported approximately 300 million dollars' worth of hay, which equals to approximately 144 million dollars from Western Australia (Australian Fodder Industry Association). The WA scenario 1k hectare hay field can produce approximately 91k 50kg hay bales, based on one hectare yielding approximately 5 tons of hay, or 91 50kg hay bales (Government of Western Australia 2017). A theoretical localized outbreak can be assumed to make up an entire shipment of hay to the U.S., because it is equal to less than what an average size Australian hay farm is capable of producing, and this will be covered more in section 2A.

There are approximately 217,000 annual ryegrass (*Lolium rigidum*) seeds/pound, or 477,974 seeds per kilogram. Approximately 49% of the weight of *Lolium rigidum* hay is seed; therefore, the number of seeds per kilogram of hay is 236,194. Based on this, any number of *R. toxicus* galls will be divided by this to find the percentage seed infection rate in hay.

One of the major assumptions of *R. toxicus* is that one gall replaces one seed, therefore there is a 1:1 ratio of gall to seed, so when calculating % of seed infection of *R. toxicus* in a kilogram of hay, one can assume the rate is galls divided by total number of seeds in a kilogram. The seed infection rate of the known outbreak number of 10,361 galls per kilogram of hay is 4.39%, as highlighted in Table 3.1.

Table 3.1. Basic Information based on Outbreak *R. toxicus* concentration and Host (*Lolium rigidum*) information – basis for future calculations

Calculations and conversions

Outbreak # of galls/kg hay:	**10,361**
Kg of *Lolium rigidum* seed/ha	2,241
Lolium rigidum Seeds/kg	477,974
Lolium rigidum seeds/ha	1,071,138,767
Kg hay/ha in Australia	4535
50 Kg Bales hay/ha in Australia	91
# of seeds/kg hay	236,194
# of seeds/ 50 Kg Bale	11,710,352
# *R. toxicus* galls/ kg hay	10,361
Outbreak % seed infection rate:	**4.39%**
Outbreak % of hay	**2.11%**

Table 3.1 notes: ha=hectare, kg=kilograms. Sources of R. toxicus, seed, and hay information: Hart et al. 2011; USDA-SCS 1973; Allen 2012.

Table 3.2 shows the scenario calculations in the case that testing for *R. toxicus* in Australia failed to detect at various infection rates, from 1-20%. Failures to detect may be due to any sampling issues. It is important to note that the higher the infection rate, the less likely the possibility of not detecting the *R. toxicus* galls due to sampling or visual appearance, therefore much higher than 20% infection rate is less likely to evade visual detection or testing errors.

Table 3.2. Quantity of *Rathayibacter toxicus* galls in Australian Hay assuming various seed infection rates

%Seed Infection Rate	*R. toxicus* galls per 50kg bale	galls if entire 1k ha paddock is infected
0.001%	117	10,621,290
0.01%	1,171	106,212,896
0.1%	11,710	1,062,128,965
1%	117,104	10,621,289,648
4.39%	513,692	46,591,906,310
10%	1,171,035	106,212,896,476

Table 3.2 notes: known recorded outbreak level highlighted, the right column is a scenario assumption based on infection rate and average farm size information, not real-life record.

1B: Origin/Source - packing/testing

Initial *R. toxicus* detection process includes visual detection through field assessment. This entails scouting for symptomatic grazers and looking for galls and bacterial slime. Both of these have potential issues such as the fact that *R. toxicus* works cumulatively when consumed, and there is a certain threshold that must be reached for an animal to start exhibiting symptoms, and around 20,000 galls must be consumed for lethality (citation). Smaller infections might not be symptomatically visible. Scouting a field for visual signs of bacteria is difficult if the density of *R. toxicus* is low and if the sampling design does not take a scout near an outbreak localized to one small specific area.

If there is a positive field assessment, several different types of initial tests can be done. These include a mature ryegrass test, pre-flowering test, hay test, and toxin test. All four of these are ELISA based tests (Australian Quarantine and Inspection Service 2014, 2019). Such tests can also have technological, human, and sampling errors. If there is a negative field assessment, there are still official pest-free certifications that must occur before any export of hay can occur. There are four points at which this can occur. They include official testing of bales in the paddock,

bales in storage, bales on the processing line, and processed bales prepared for export. One of these options must be taken before the export goes through (Australian Quarantine Inspection Service 2014, 2019). Table 3.3 summarizes the sampling process of each of these four.

Table 3.3 Four pre-export certification methods for *Rathayibacter toxicus* contamination

Test Site	quantity for testing	# bales sampled	sample size	Action	
				negative	positive
paddock	<12 bales	1	?	no further testing needed	100% bales retested
	>12 bales	15%	?	no further testing needed	100% bales retested
storage	1-5 bales	1	?	no further testing needed	100% bales retested
	6-30 bales	1 from every 3rd; no less than 5 total	?	no further testing needed	100% bales retested
	>31 bales	1 from every 5th; no less than 10	?	no further testing needed	100% bales retested
processing line		Sample from container	450g	no further testing needed	100% bales retested
export	40 bale groups	1 10g sample from each 40	400g	no further testing needed	100% bales retested

Table 3 Notes: this information is from the Australian Quarantine and Inspection Service "Standard for Minimizing the Risk of Corynetosin Contamination of Hay and Straw for Export".

There are several possible error types during *R. toxicus* detection. These errors can occur during surveillance in the field, including with the design and execution of the surveillance, which includes human error. The testing technology also has potential for errors. All the in-field and testing sites utilize ELISA based assays, which have error rates for any organisms. One of the biggest issues with testing is the sampling error. For example, when 10g of hay is sampled from a 50kg hay bale being prepared for export, that is equal to 0.02% of the bale, and if a pathogen is randomly distributed, there may be low probability of taking a sample that contains *R. toxicus* galls.

2A: Transit/Dissemination - Shipping

Australia has a total of 13 major hay markets, the top ones include Japan, Korea, Taiwan, China, Saudi Arabia, United Arab Emirates, and Indonesia. U.S. is also an Australian hay importer, but Australia is not one of the top hay suppliers for the U.S. (Government of West Australia 2020, WorldCity, U.S. Census Bureau).

U.S. receives $55,074 worth of Australian hay per year (Table 4). In addition to the Australian hay, the U.S. imports approximately 90 million dollars' worth of hay from all other countries from June 2019-June2020. (WorldCity, U.S. Census bureau). The complexity of world trade is immense. It is important to consider the "indirect" entrance of *R. toxicus* by way of another country that received Australian hay directly and then re-exporting it to the U.S. China receives hay from Australia and exports hay to U.S, which is a potential indirect avenue of importing Australian hay (Table 3.4).

Table 3.4 Top Sources of U.S. Hay Imports (From June 2019-June 2020)

Rank	Country	YTD	Australian hay receiver?
1	Canada	$42.96 M	
2	Mexico	$8.95 M	
3	China	$157,457	yes
4	France	$83,170	
5	Australia	$55,074	
6	Netherlands	$46,990	
7	Kenya	$38,220	
8	Ecuador	$28,303	
9	Chile	$14,248	
10	Peru	$11,653	

Table 4 Notes: source: WorldCity Inc USTradeNumbers website, which uses data from U.S. Census bureau.

There are several specific ports of entry through which hay enters the U.S. Often, the country of origin location relative to these ports determines through which port the commodity enters. Based on these locations, the one with closest proximity to Australia would be the Calexico/Mexicali Border Crossing in California (see GIS map in section 3B).

Table 3.5 Top Gateways (ports of entry) for U.S. Hay Imports (June 2020)

Rank	Port	YTD
1	Buffalo Peace Bridge, NY	$8.85 M
2	Portal-North Portal Border Crossing, N.D.	$6.32 M
3	Calexico/Mexicali (East) Border Crossing, Calif.	$5.41 M
4	International Falls, Minn.	$5.1 M
5	Detroit Ambassador Bridge, Mich.	$4.72 M
6	Port Huron Blue Water Bridge, Mich.	$4.69 M
7	Derby Line, Vermont, border crossing	$3.46 M
8	Pembina Border Crossing, N.D.	$2.46 M
9	Rouses Point / Lacolle Border Crossing, N.Y.	$2.31 M
10	Highgate Springs Alburg, Vt.	$1.85 M

Table 5 Notes: Calexico is highlighted because it is in closest proximity to Australia, and no other west coast ports are in the top 10 for hay.

The total value of Australian hay imported into U.S. annually is approximately $55K USD (Table 3.5). At $175 per ton, the U.S. imports approximately 6,294 110-pound bales (~50kg bales) from Australia annually (Department of Primary Industries and Regional Development's Agriculture and Food Division of the Government of Western Australia).

The number of total bales of hay that can be made from the calculation of 91 bale yield/ha, WA 1k ha paddock scenario of 91k bales, is relevant for the assumption that the entire U.S. hay import from Australia (which is equal to 6,294 bales) can theoretically be from one location during an outbreak.

2B: Transit/Dissemination – Entry

Varying numbers of *R. toxicus* galls may enter the U.S. as a function of the level of *R. toxicus* infection in the hay (Table 6). The limit of detection (L.O.D.) is one gall per kilogram based on the *R. toxicus* ELISA test, suggesting that as many as 314,714 total galls could go undetected if the total import is infected at this L.O.D. rate. As many as approximately 3 billion galls could enter if the outbreak concentration (10,361 galls per kg) goes undetected in an entire import.

` According to the L.O.D., one gall per kilogram of hay equals to an infection rate of 0.00043%. At this rate, it would be likely that *R. toxicus* would go undetected. Calculating undetected galls from all Australian bales is based on the extrapolated number of how many hay bales enter the U.S. based on the value of the hay imports from Australia into U.S, which was established as 6,294 bales. This number of bales multiplied by total number of seeds per 50-kilogram bales, multiplied by the limit of detection rate, one can determine the theoretical number of *R. toxicus* galls entering undetected. Therefore, at a seed infection rate of 0.00043%, and theoretical uniform distribution, the total number of undetected galls from Australian hay imports may be 50 galls per bale or a total of **314,709 galls** per year. At the outbreak seed infection rate of 4.39% (part 1A), the total theoretical number of galls entering U.S. from a shipment directly from an outbreak paddock in Australia is approximately three billion.

Table 3.6: Scenarios of *R. toxicus* introduction quantities into the U.S.

% Seed Infection Rate	**Galls (50kg hay bale)**	**Total Galls entering U.S.** (from WA outbreak location scenario)	**Estimated Entrance likelihood** (evasion of detection)
0.00043%	50	314,709	L.O.D. level - highest
0.001%	117	737,070	High
0.01%	1,171	7,370,697	
0.1%	11,710	73,706,966	
1%	117,104	737,069,656	
2%	234,207	1,474,139,313	
4.39%	518,601	3,264,165,621	Low (outbreak)
10%	1,171,035	7,370,696,564	
50%	5,855,176	36,853,482,819	Extremely low
100%	11,710,352	73,706,965,639	Lowest

Table 3.6 Notes: Highlighted in yellow is the Limit of Detection (L.O.D.) number of galls entering the U.S. Highlighted in orange is the number of galls entering the U.S. if the outbreak concentration (4.39% of seed in hay) is undetected. This table also includes other theoretical amounts and conclusions on the likelihood of detection.

There is a spectrum of possible scenarios with varying *R. toxicus* gall numbers as a function of different seed infection rates (Table 3.6). The number of galls entering starts with the L.O.D. level of **314,709** galls and increases to the outbreak level of three billion galls; it can also be higher or lower. An interpretation of likelihood of entrance is useful given that an extremely high concentration of galls would be more detectable at any of the detection points summarized in section 1A.

3A: Destination/Introduction - Escape

There are many miles of major roads across Australia and U.S. through which shipments of hay may travel and accidentally disseminate material from the bales. These include the national highway system, interstate highways, railways, as well as rural and urban roadways. The U.S. has almost 16 times more miles of national highway system than Australia, at 160,955 miles

in U.S., and 47,432 miles of interstate highways. There are also many unpaved roadways (more than 1.4 million miles) throughout countless towns and cities in U.S., especially in the rural areas. When hay travels from a port of entry to a destination, this journey will likely comprise of a route that includes major and rural roads. Transport of bales across these roads often leads to disseminating hay along the roadways into conducive environments for the seeds, *R. toxicus* galls and nematode galls in the hay to enter and establish.

Hypothetical scenario-based numbers of *R. toxicus* galls escaping in the U.S. during transit were calculated (Table 3.7). The unofficial grain industry standard estimates that approximately 1% of plant material in shipments is lost during transit. The higher percentages in Table 3.7 are provided as reference information for more unusual cases, such as rougher terrain through which a truck travels with more bumps and wind dispersal, or events such as accidents. From a single truck, estimates indicate that between 15-154,108 galls (L.O.D. to outbreak concentration) may be lost. Assuming a worst-case scenario, a total of **3,147** to **32,332,684** galls may be lost during transit from all Australian hay imported into the U.S., representing a range of L.O.D. to outbreak levels of *R. toxicus*. This range can also be lowered to 1gall to 32M galls if considering that only one bale may be infected at LOD level. These estimates are based on LOD and Outbreak scenarios and the reality can be closer to the LOD levels because of the confidence in testing capabilities, but with realization that they may fail at one or several points.

Table 3.7: *Rathayibacter toxicus* and associated seeds/hay material potentially lost in transit

% lost along route	Seeds lost per bale (50kg)	Seeds lost per truck (30 bales)	L.O.D. gall loss (per bale)	L.O.D. gall loss (per truck)	L.O.D. gall loss (entire import)	Outbreak galls loss (per bale)	Outbreak gall loss (per truck)	Outbreak gall loss (entire import)
1%	117,104	3,513,106	1	15	3,147	5,181	154,108	32,332,684
2%	234,207	7,026,211	2	30	6,294	10,361	308,215	64,665,368
3%	351,311	10,539,317	3	45	9,441	15,542	462,323	96,998,053
4%	468,414	14,052,423	4	60	12,588	20,722	616,431	129,330,737
5%	585,518	17,565,529	5	75	15,735	25,903	770,539	161,663,421
10%	1,171,035	35,131,057	10	150	31,471	51,805	1,541,077	323,326,842

Table 3.7 Notes: the orange highlighted cells indicate the estimated number of galls lost per bale, per truck, and from all Australian hay imports in the U.S. for the hypothetical Outbreak rate undetected concentration scenario. The yellow highlighted cells indicate the estimated number of galls lost per bale, per truck, and out of all Australian hay imports in U.S. according to the most stringent L.O.D. concentration.

3B: Destination/Introduction - Establishment

Whether a pathogen can become established in a new location is a crucial point of analysis, and boils down to analyzing conduciveness of the destination's environment, as well as supporting factors which may increase chances of establishments, such as occurrence of hosts and vectors. For *R. toxicus*, host and vector occurrence in destination is a "supporting" factor because *R. toxicus* already travels with its host and vector, so the life cycle requirements are complete if there is a conducive environment.

Climate mapping and comparing Australian conditions with U.S. conditions is done in order to see where *R. toxicus* occurred historically in Australia and matching the climate of those locations with climate types in the U.S. (Figures 3.2 and 3.3). The Australian map (Figure 3.2) also indicates locations in which *R. toxicus* has been recorded. The U.S. map (Figure 3.3) includes the top 10 hay ports of entry. The climate zones are described in Figure 4 of the

appendix. Visual comparisons were made (Figure 3.2 and Figure 3.3), as well as calculations of how much of the U.S. land area matches the *R. toxicus* detection zones within Australia.

Figure 3.1: GIS Map: Climate types and *R. toxicus* detection locations across Australia

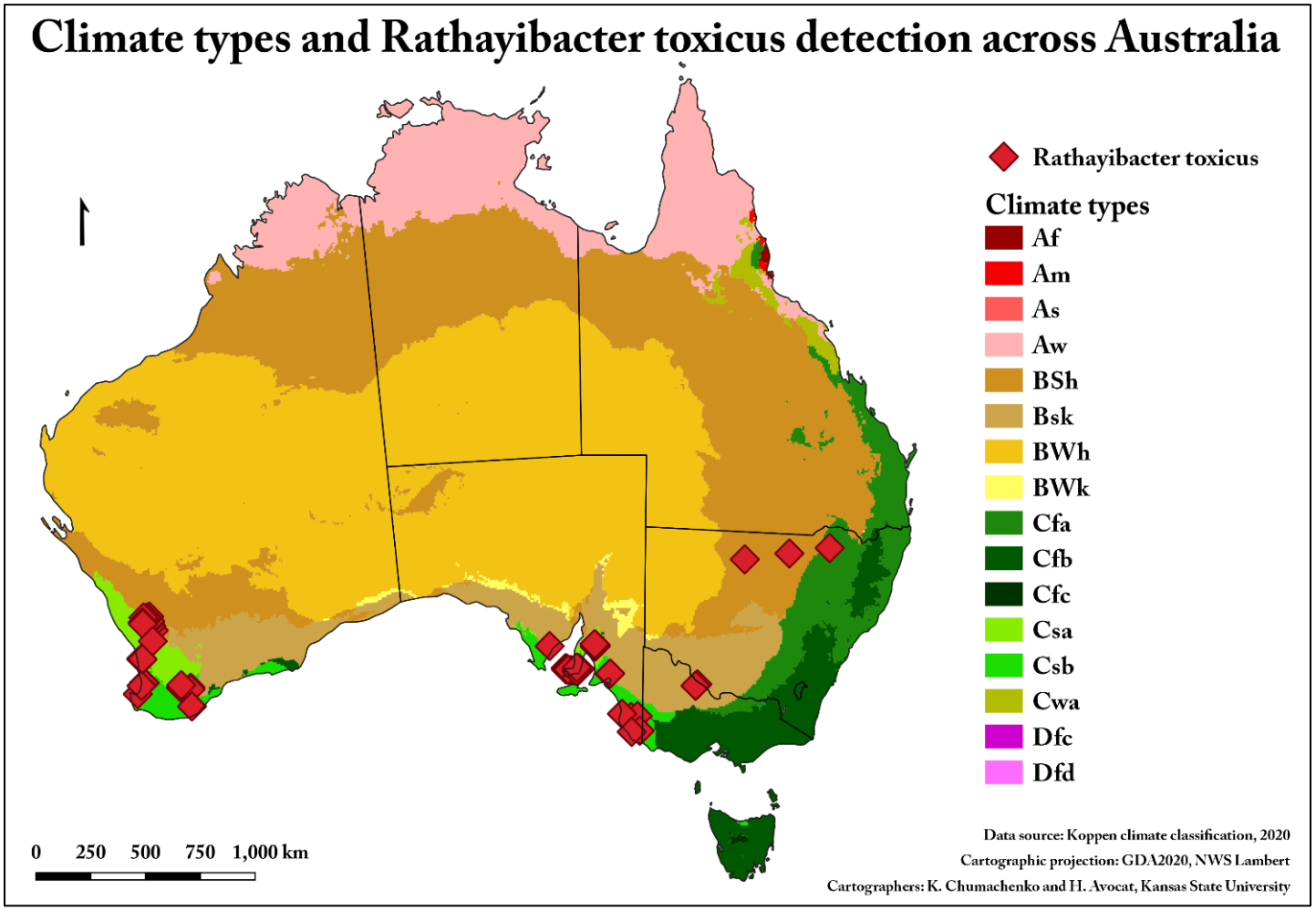
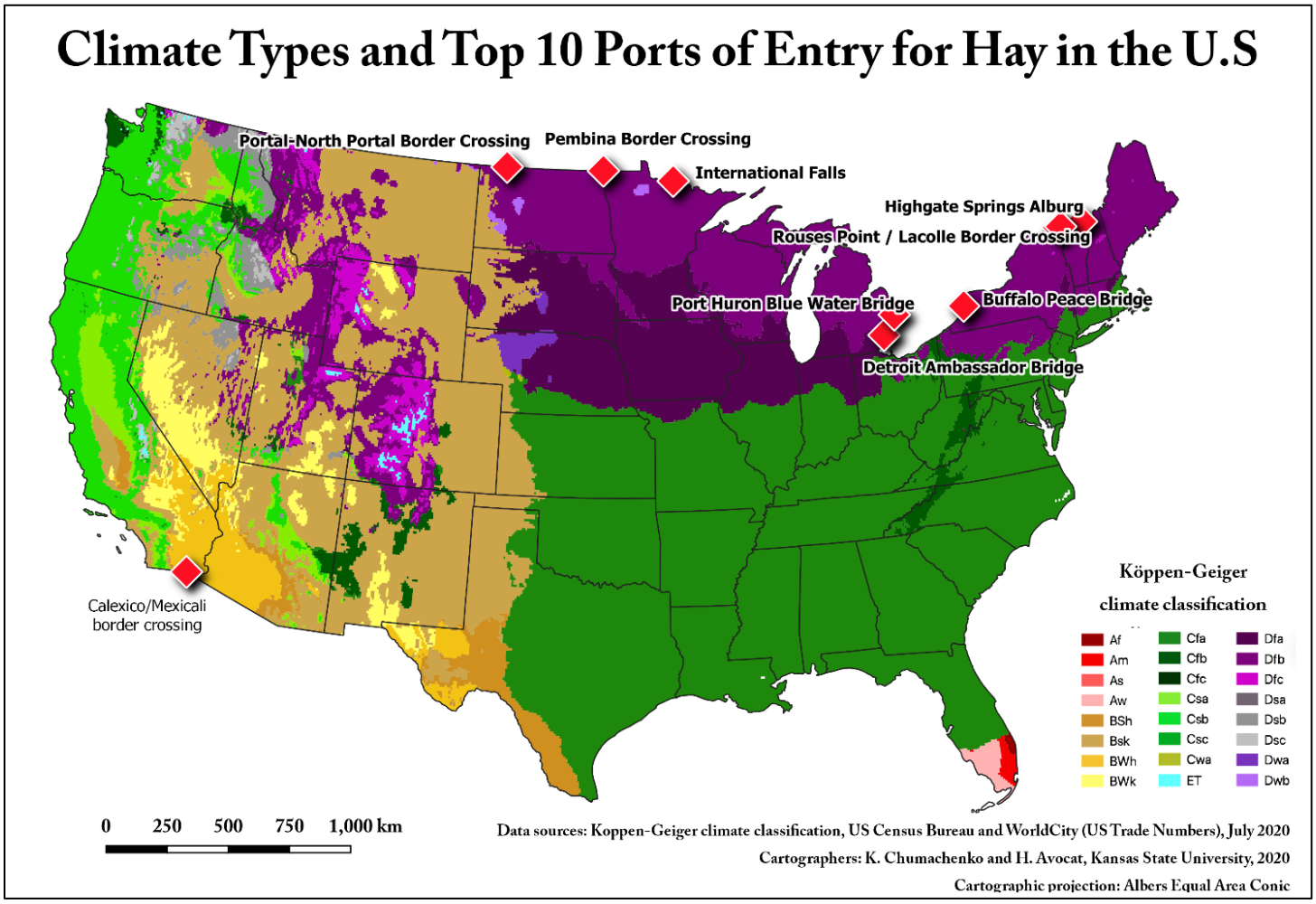


Figure 3.2: GIS Map: Climate Types and Top 10 Ports of Entry for Hay in the U.S.



Figures 3.1 and 3.2 show the climate types across the U.S. and similarities with Australia. There are widespread arid and temperate regions in U.S. similar to the areas in Australia where *R. toxicus* is prevalent. The Köppen-Geiger system uses colors and shades to classify the world into five climate zones based on criteria like temperature, which allows for different vegetation growth. These five main climate groups are Tropical, Dry, Temperate, Continental, and Polar. Dry zone is defined by little precipitation, with different types being dependent on the percentage of the precipitation occurring in the spring and summer, and average annual temperatures. Temperate zone has the coldest month averaging between zero and 18 degrees Celsius and second coldest month above 10 degrees Celsius; and differentiates into which season is dry versus wet or same across all seasons.

Using the Australia map (Figure 3.2), the *R. toxicus* geographic coordinates were within climate zones B and C, or Dry and Temperate respectively. *Rathayibacter toxicus* is prevalent specifically in the “hot desert” (BWh) “cold desert” (BWk), “humid subtropical temperate oceanic” (Cfa), “hot summer Mediterranean”(Csa), and “warm summer Mediterranean”(Csb), regions. With QGIS 3.14 geoprocessing tools, a calculation was performed that showed the percentage of the U.S. in the 5 *R. toxicus* climate zones (BWa, BWk, CFa, CSa, CSb) being 46% of the total area of the contiguous states.

It was assumed that *R. toxicus* occurs in any temperate and arid zones due to its large range of temperature survivability. It was also assumed that some tropical and continental zones may be conducive if the temperatures remain above 10 degree Celsius throughout year because they both have periods of humidity and precipitation, which helps activate *R. toxicus* galls. Table 8 summarizes the findings of Figure 3.1 and 3.2. The U.S. has similar conducive climates and environments as Australia, therefore there are widespread opportunities for *R. toxicus* to survive

and establish (Indexmundi Factbook, CIA World Factbook, Koppen-Geiger Climate Classification).

Table 3.8: Summary of Climate parallels between Australia and U.S.

Climate types	U.S.A.		Australia		*R. toxicus* occurance?
Temparate	✓	Mostly temperate	✓	South and east	✓
Semiarid	✓	Great plains west of Missippi River	✓	Mostly semiarid	✓
Arid	✓	Great Basin, southwest	✓	Mostly arid	✓
Tropical	✓	Hawaii and Florida	✓	North	

There are 15 reported hosts of *R. toxicus*, shown by Table 3.9. This may not be an exclusive list, because there may be other known *Poaceae* species of grasses that may host *R. toxicus*. (CABI Invasive Species Compendium, USDA Natural Resource Conservation Office – Plant Database, USDA - Fire Effects Information System, Bertozzi and Davies, 2009; Bertozzi and McKay, 1995; Chatel et al., 1979; Edgar et al., 1994; McKay et al., 1993; Riley, 1992a ; Riley, 1995; Riley, 1996; Riley et al., 2001; Riley and Barbetti, 2008)

There are at least 4 reported vectors of *R. toxicus*, all in the *Anguina* genus (Table 9, CABI Invasive Species Compendium, Riley 1995; Riley 1992a; Riley et al. 2001; Bertozzi and Davies 2009). This may not be an exclusive list, because there are likely other *Anguina* vectors.

Table 3.9: Geographic distribution of hosts and vectors of *Rathayibacter toxicus*

#	Host: Plant Species	Australia	U.S.	# of states	U.S. status
1	*Lolium rigidum*	✔	✔	7/50	invasive
2	*Agrostis avanacea (or Lachnagrostis filiformis)*	✔	✔	5/50	invasive
3	*Polypogon monospelensis*	✔	✔	41/50	invasive
4	*Ehrharta longiflora*	✔	✔	1/50	invasive
5	*Austrodanthonia caespitosa*	✔	Not detected		
6	*Avena sativa*	✔	✔	50/50	invasive
7	*Danthonia caespitosa*	✔	✔	49/50	native
8	*Lolium multiflorum*	✔	✔	50/50	invasive
9	*Lolium perenne*	✔	✔	50/50	invasive
10	*Lolium persicum*	✔	✔	6/50	invasive
11	*Lolium strictum*	✔	✔	7/50	invasive
12	*Lolium temulentum*	✔	✔	43/50	invasive
13	*Phalaris spp.*	✔	✔	50/50	native+invasive
14	*Vulpia myuros*	✔	✔	43/50	invasive
15	*Datylis glomerata*	✔	✔	50/50	invasive
#	**Vector: Nematode Species**	**Australia**	**U.S.**	**# of states**	**U.S. Status**
1	*Anguina funesta*	✔	✔	1/50	invasive
2	*Anguina tritici*	✔	✔	8/50	invasive
3	*Anguina australis*	✔	Not detected		
4	*Anguina paludicola*	✔	Not detected		

The U.S. contains fourteen out of fifteen known *R. toxicus* host species, and two out of four nematode vector species (Table 3.9). A GIS map of *R. toxicus* vectors and hosts occurrence across U.S. states illustrates areas of potential vulnerability for establishment of *R. toxicus*. Many of the hosts occur in every state across the U.S., while some are in most or several states. California has the most *R. toxicus* host species out of all the other states. Oregon is the only state that has two *R. toxicus* nematode species, the maximum number for any one U.S. state. Having the GIS visualization of vector and host data highlights areas with higher risk for *R. toxicus*

establishment; the more components of a pathogen system, such as pathogen+host+vector+conducive environment, the more likely the establishment of said pathogen.

Figure 3.3: Geographic distribution of *Rathayibacter toxicus* host and vector species in the U.S. by state

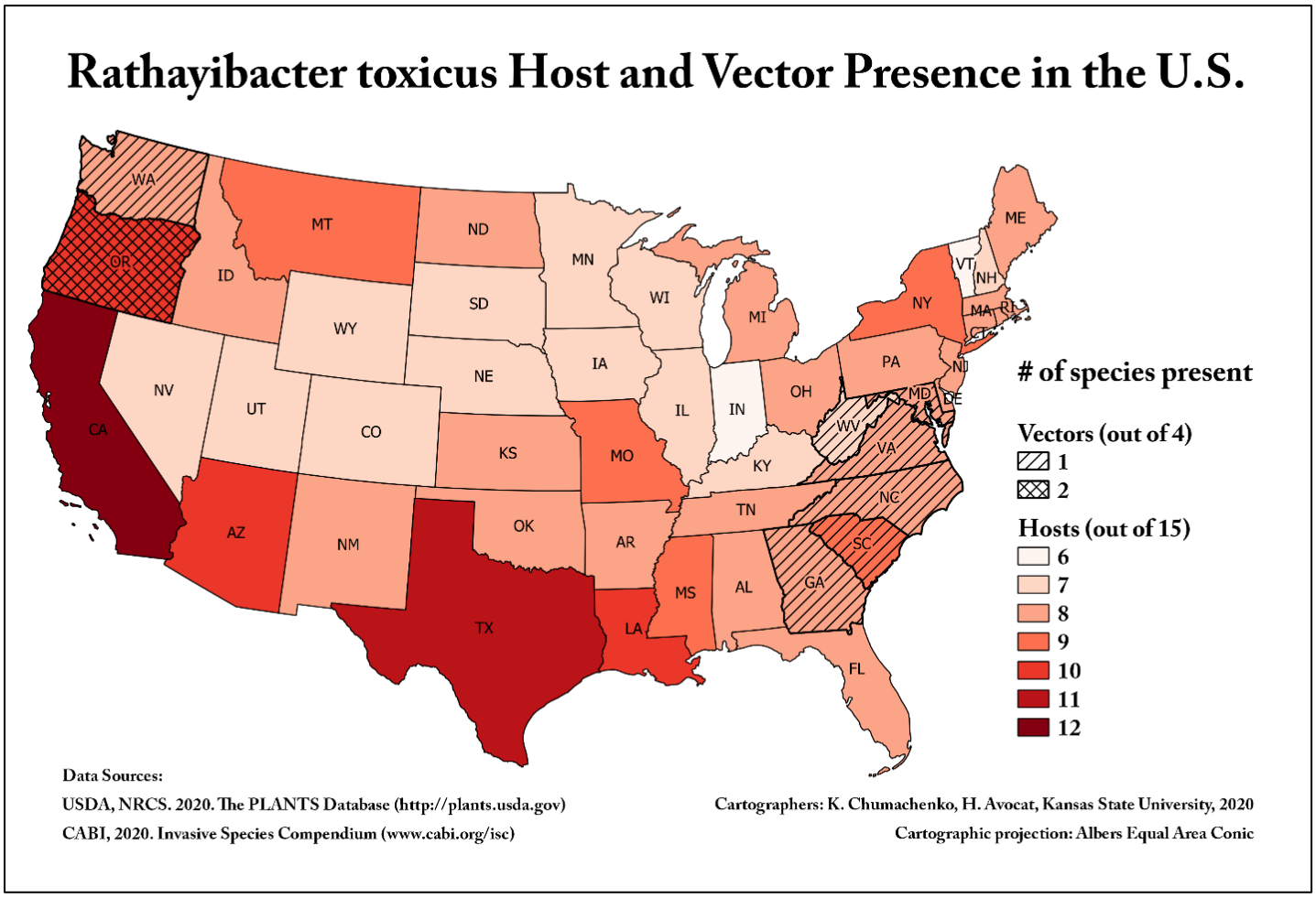


Figure 3.3 Notes: Geographic visualization of Rathayibacter toxicus hosts and vectors (Table 3.9). Comprehensive list of hosts and vectors in each state are provided in Appendices 3.5 and 3.6.

The geographic characteristics that allow for growth of grass and vegetation can be another important factor. Similarities between the U.S. and Australia in land usage and geographic characteristics indicate that both countries have agricultural, arable, and pastureland, with potentially conducive soil environments for *R. toxicus* given the correct climate. *Rathayibacter toxicus* may potentially survive and become established with its host and vectors in several environments (Table 3.10), where conducive weather and climates occur (Source: Indexmundi Factbook, CIA World Factbook).

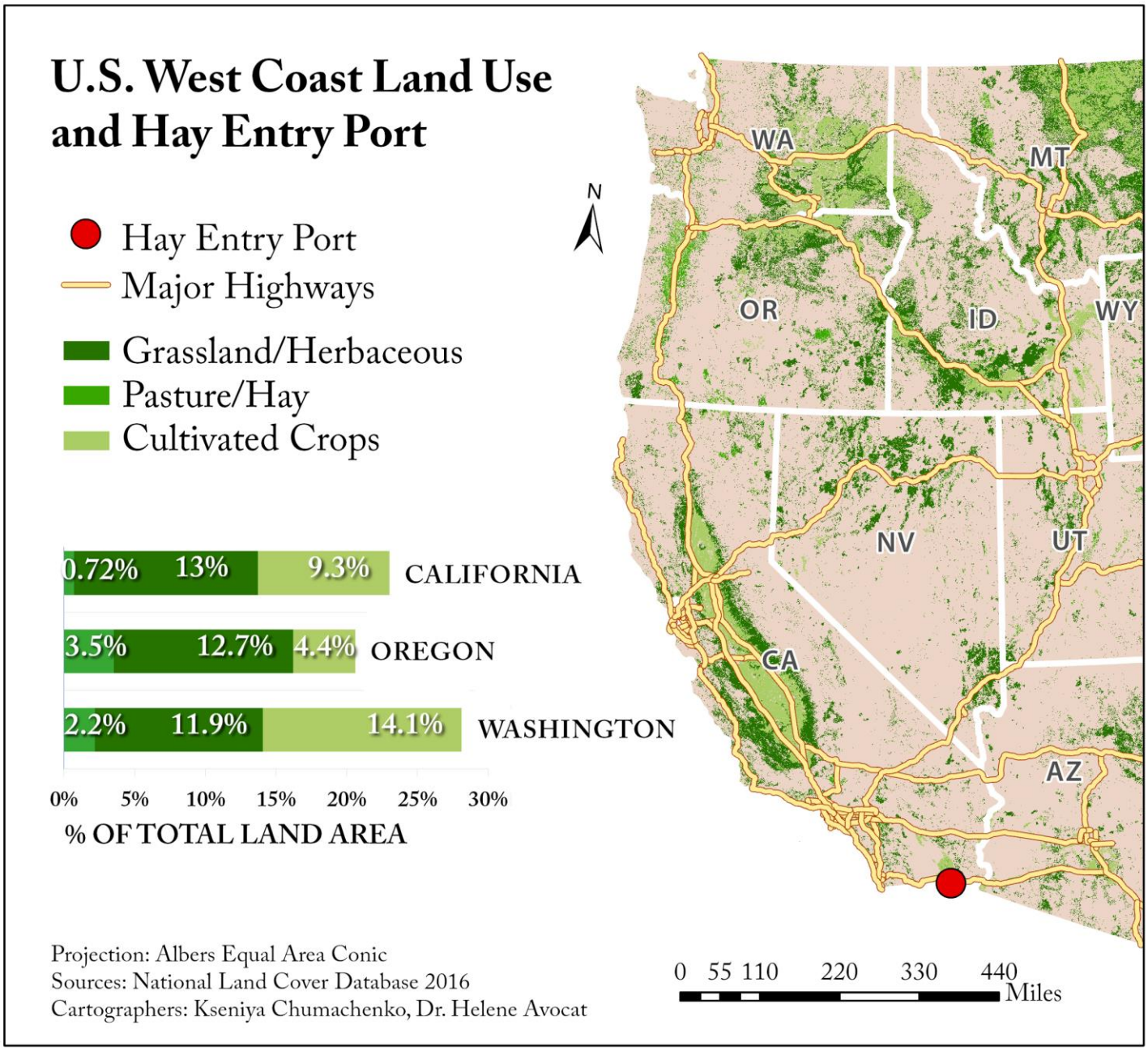
Table 3.10: Land Use Summary of Australia and U.S.

	U.S.	Australia
Total area (sq km)	9,833,517	7,741,220
Land (sq km)	9,147,593	7,682,300
Water (sq km)	685,924	58,920
Coastline (km)	19,924	25,760
Agricultural land	44.50%	52.90%
Arable land	16.80%	11.60%
Crops	0.30%	0.09%
Permanent pasture	27.40%	88.40%
Forest	33.30%	16.20%

Because of the likely Australian hay entry port being at the Calexico border crossing due to proximity, conducive climates in the Western U.S. near that port, the highest number of *R. toxicus* host species being present in California, and the highest vector species number in Oregon, a land use analysis was done after narrowing down to this West Coast area. The land area in closest proximity to the hay port would contain possible immediate routes of hay bales being driven to nearby farms or states. The transit dissemination pathway during hay transport would begin at the port of entry and continue until a destination location is reached. A GIS analysis of the Western U.S. was performed using United States National Land Cover Database

(Dewitz 2019, U.S. Geological Survey 2016) where three land cover types were contained in the U.S. map. These included the land type assumed to be most conducive for grass environments, “grassland/herbaceous”, “pasture/hay”, and “cultivated crops”, due to the fact that these areas already support grass or plant growth. Approximately 21-28% of California, Oregon, and Washington states comprise these three land cover types, and around 15% is hay or grassland. Additionally, major highways were included as a layer in the GIS map to illustrate transport routes and the proximity of different land cover to major roads. During this process, an area very close to the hay port of entry, the Imperial Valley, was found to be one of the largest hay producing areas of California, where 46% of the 116k ha of hay grown there are *Poaceae* or known *R. toxicus* host species (Riddel 2013).

Figure 3.4: Land Cover near likely port of entry for Australian hay import



Presence of farm animals can be another factor supporting establishment. All farm mammals were considered to be susceptible to *R. toxicus* poisoning. There are many millions of cattle, pigs, horses, sheep, goats, and other animals in the livestock industry in the U.S. Hay is usually received for the purpose of feeding, therefore the destination farm in the U.S. would be assumed to have animals which would feed on the hay. To optimize the chances of *R. toxicus* establishment, the placement of hay within a farm is taken into consideration. A farm with pasture grazing and hay influx is chosen as a realistic scenario of maximum likelihood establishment, as it would put the hay bales in contact with the environment – soil and other grasses growing in the area. Mammals are also assumed to be able to spread *R. toxicus* throughout a field in feces with the bacterium remaining viable.

Figure 3.5: Outline of main pathway components of *Rathayibacter toxicus* establishment

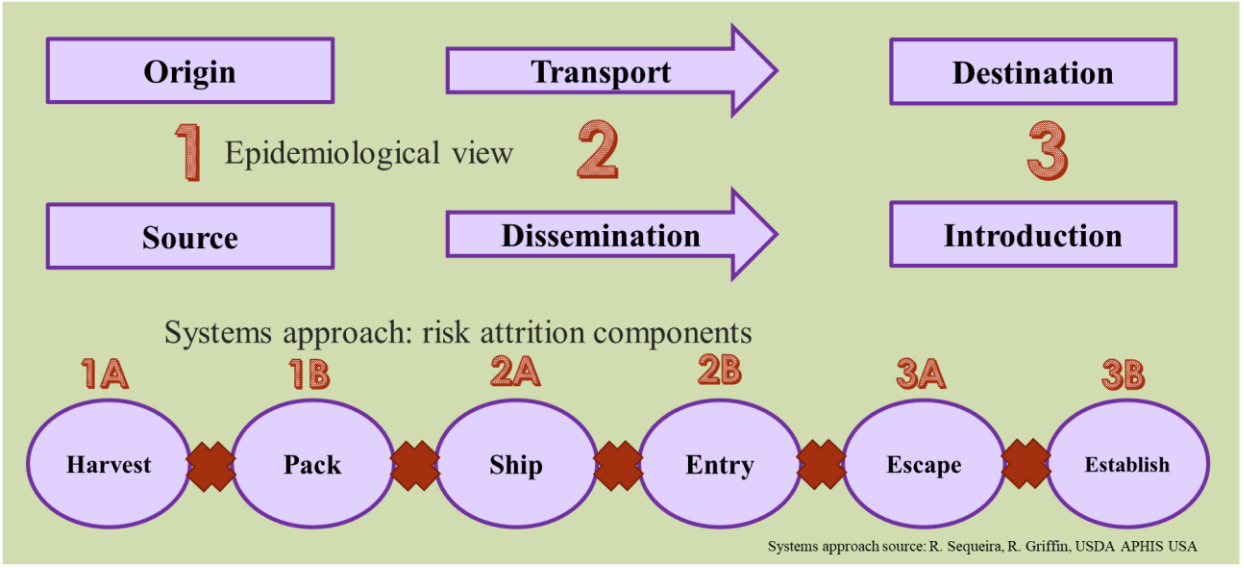


Figure 3.6: Pathway sub-components of focus within Origin, Transport, and Destination

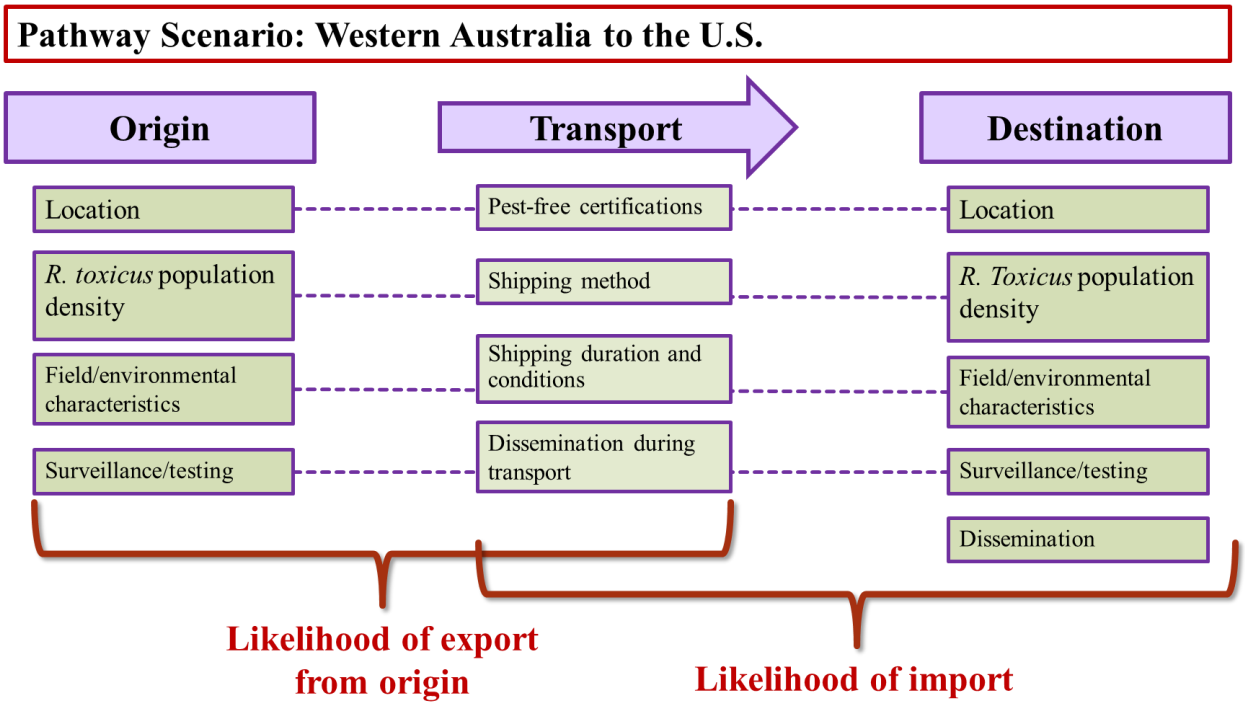
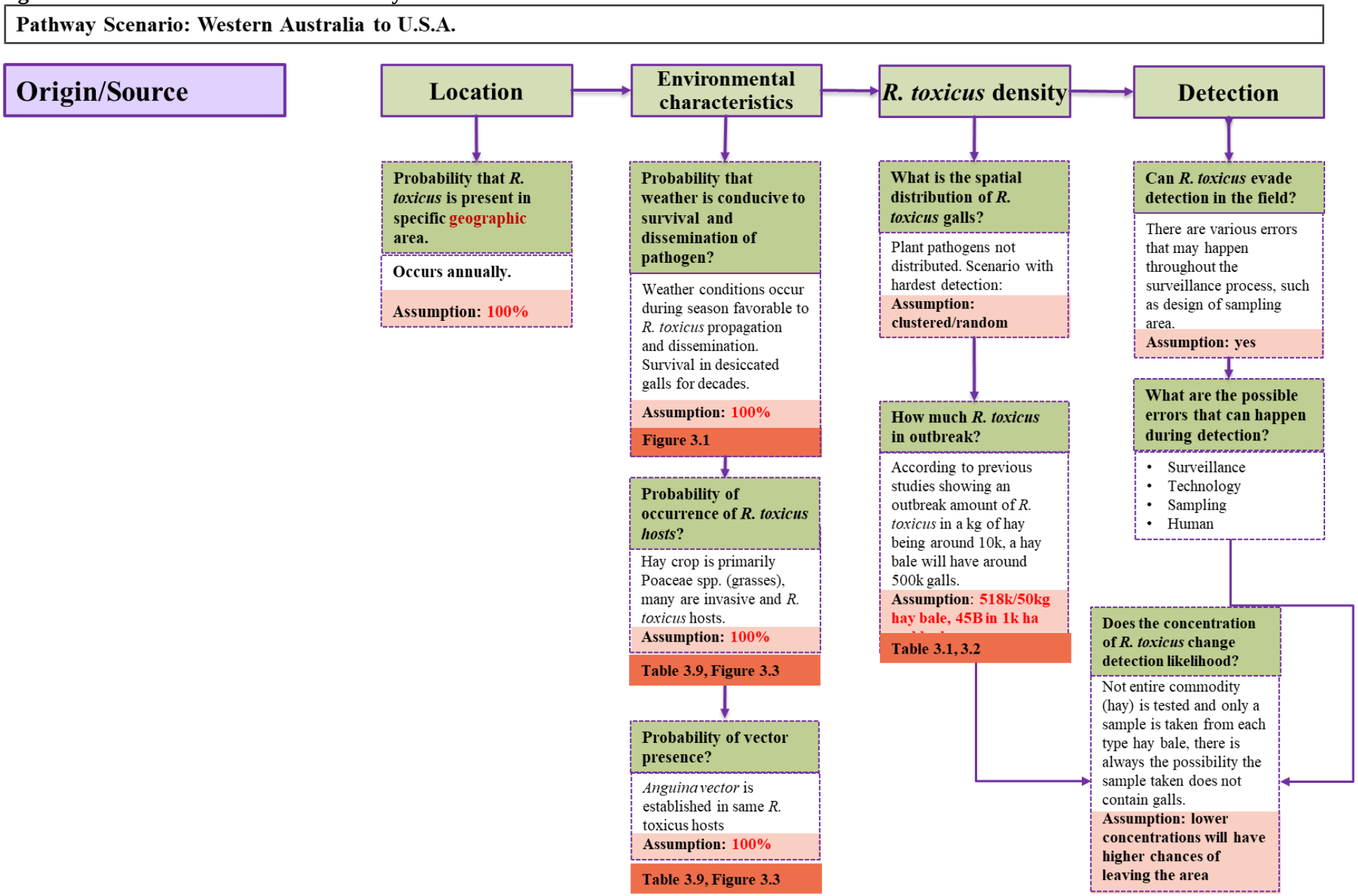
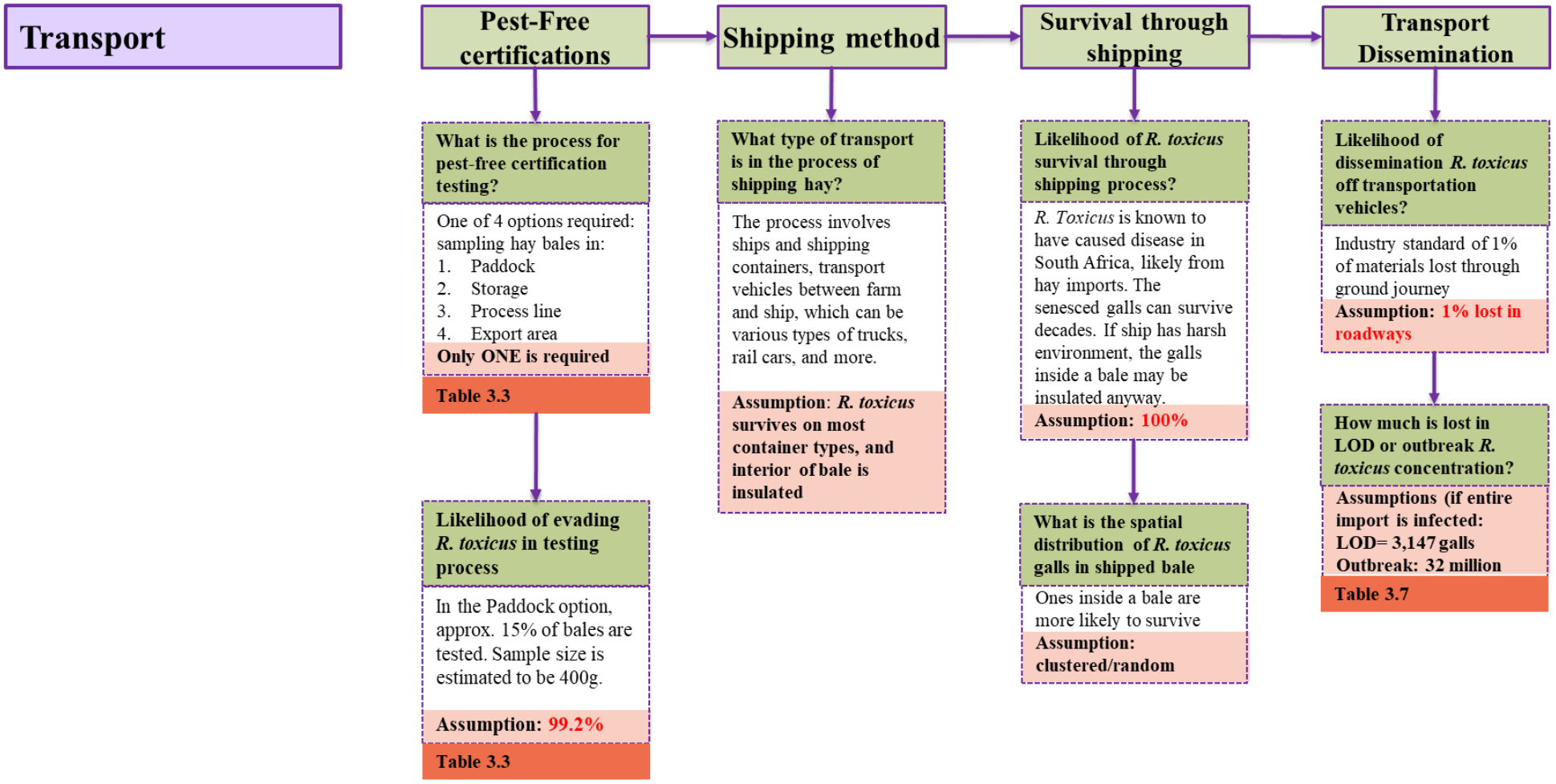


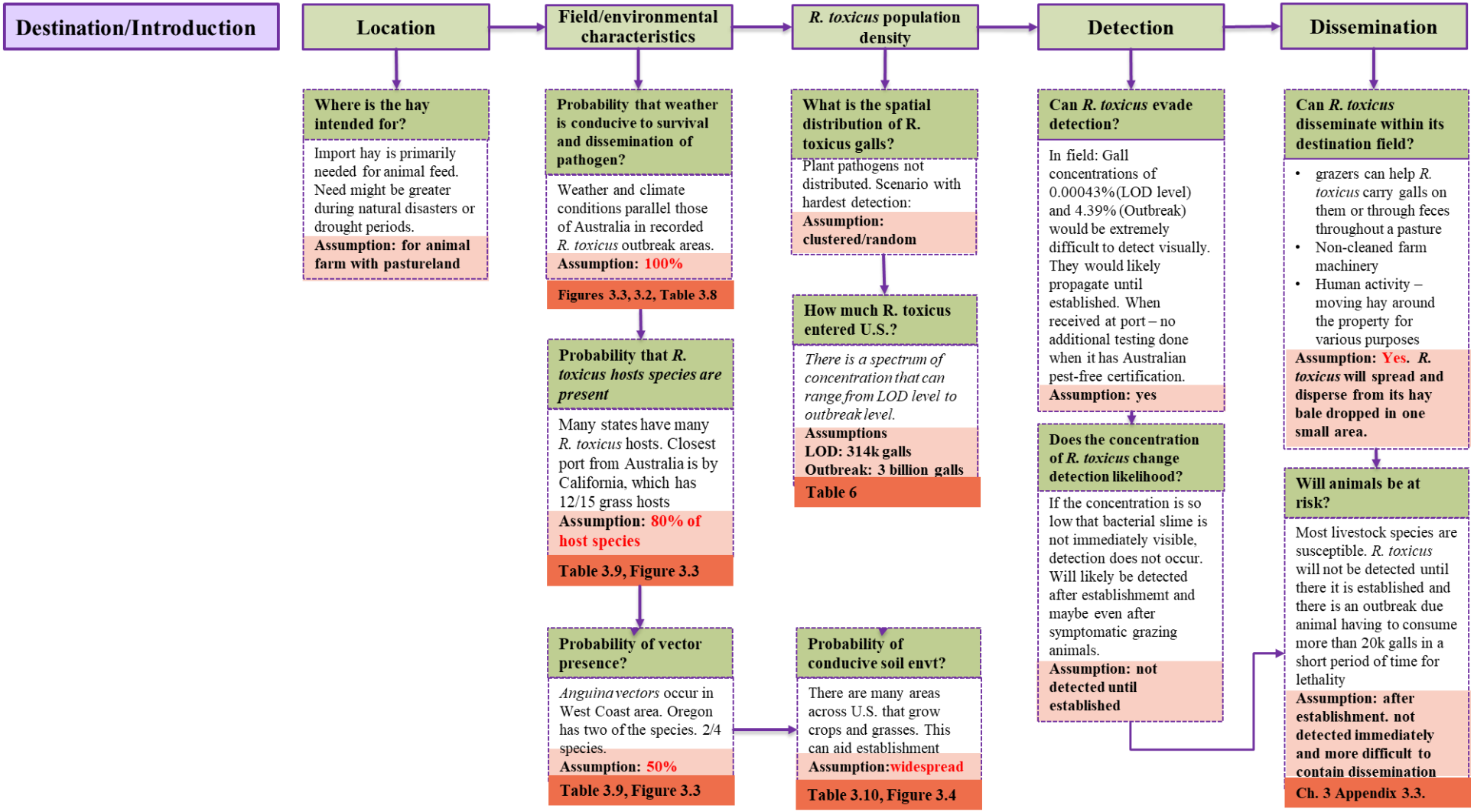
Figure 3.7: Focused Flowchart of Pathway Scenario



Pathway Scenario: Western Australia to U.S.A.



Pathway Scenario: Western Australia to U.S.A.



Discussion

Pathogen source/origin

Pathogen status within its geographic origin analyzed in a hypothetical scenario where the maximum amount of galls (based on previous outbreaks) surpassed detection and was transported in commodities intended for export. An average Australian farm at around 1 thousand hectares (Australian Bureau of Statistics) was assumed in the scenario. Western Australia was chosen as the region of origin of our potential escaped *R. toxicus* galls because of all Australian regions, it produces the highest percentage of Australian hay. This hypothetical farm primarily grew hay for export, and any hay field may have had a mixed variety of grasses. Grass fields are never uniformly one species. Therefore, even if a field is growing oats or Lucerne (alfalfa) there may be multiple other grass species mixed.

Lolium species and many others in the *Poaceae* family are known to be highly invasive and therefore easily established in new places. The hosts of *R. toxicus* have existed in Australia for a long time. Therefore, in terms of Australia, invasiveness of grasses is paddock and farm specific, and not a question of whether those grasses can establish in Australia generally, because they already occur in the country. Figure 3.3 and Table 3.9 show the occurrence of known *R. toxicus* hosts in Australia and US, and its status as native or invasive in the U.S. A hay field growing oats for export, which is very common in Australia, may have one or more *R. toxicus* hosts, and oats itself is within the *Poaceae* family and a potential host of the bacterium as well. Even a farm with other crops, not hay related, often has multiple invasive grass species present. These can also harbor *R. toxicus*, but this analysis focused on hay fields, for maximum *R. toxicus* escape scenario.

The climate of the Western Australia region is known to be mostly hot and dry (Figure 3.1, Table 3.8). This is a conducive environment for senesced galls of *R. toxicus* to survive and overwinter. An arid climate is the most common condition in Australia. Many *Rathayibacter* species and plant diseases occur in arid areas where their biological characteristics aid in their survival including, the production of extracellular polysaccharides (Murray et al. 2017). In the Western Australian arid environment, *Rathayibacter* and the nematode vector species may overwinter for decades. The nematodes drop to the soil surface as plants senesce and survive anhydrobiotically until the following season (Murray et al. 2017). The freshly hatched juveniles cannot survive desiccation but mature physiologically to become the survival state as the host senesces and desiccates (Riley and Barbetti 2008). To summarize, the seed gall nematode vectors of *Rathayibacter* spp. (including *R. toxicus*) feed and reproduce within the developing ovaries of plant seeds and overwinter in seed galls as anhydrobiotic juveniles or adults, and can survive within the seed gall for many years in this dormant stage. For example, *Anguina tritici* did not lose its ability to invade wheat seedlings after 32 years of dormancy (Limber 1973).

There are other questions that are important to analyzing the pathogen risk. These include questions of the density of *R. toxicus* within the field, such as number of galls, and dispersal. One issue that may arise is if there is a hay paddock with extremely non-uniform spatial distribution of *R. toxicus*. Combine this issue with a sampling strategy that does not take into account pathogen distribution, and detection may be missed by that sampling strategy. One example would be having clustered sampling coupled with a field where *R. toxicus* is concentrated in one corner of 20% size of the whole paddock. In a cluster sample, a population (the paddock) is divided into subgroups, and those subgroups are used as the sampling units, and combine that with a requirement of testing 15% of hay. You could have a scenario where a cluster or two are

chosen from the other side of the *R. toxicus* occurrence of a farm, equal to 15% of the field, and *R. toxicus* goes undetected. The same scenario may happen after hay bales are processed and prepared for movement out of a paddock. They may be randomly clustered, and the clean ones are randomly chosen as the testing sample, and then the whole shipment is deemed safe for export.

Transport/Dissemination

Testing and detection is the crucial step which may prevent spread of a pathogen or have a false negative which results in failed containment. Major concerns for testing include capability, robustness, and infrastructure. Australia has a strong biosecurity program across the country with the Australian Plant Protection and Quarantine setting standards for plant pests and pathogens. Other important factors include limitations in sampling and testing, as well as the significant issue of the detection threshold. Knowing the minimum number of *R. toxicus* galls needed for detection within a certain weight of grass material is important. There is an ELISA based test for *Rathayibacter toxicus* with limit of detection at 4×10^4 cells, or 1 gall per kg (Luster et al. 2020). However, all diagnostic tests have false negative and false positive error rates. Additionally, a diagnostic test uses a small sample and if the sample taken from a hay bale did not have bacterial galls directly in it or bacteria released from galls, that may complicate detection as well, so the design of the sampling strategy is crucial. Sampling may be an issue throughout the evaluation and testing process, even within the diagnostic test itself. Therefore, if a diagnostic test does not detect based on one or several samples from hay bales coming from a specific field in Australia, *R. toxicus* could still be there and leave the farm with the commodity.

When hay is harvested and baled, bacterial and nematode galls as well as grass seeds cannot be separated from the bales resulting in the possibility that they are disseminated with the

bales. This is a crucial component of the pathway: export of hay bales from Australia have the ability to not only carry *R. toxicus* galls, but also the seeds of the host plants and the vectors necessary for *R. toxicus* survival and establishment. This issue only increases the chances of *R. toxicus* survival throughout its transport and establishment in a new location such as the United States. The disease triangle could be potentially completed with every hay shipment, the only limiting factor being a conducive environment.

Transporting goods, especially plant materials which may or may not be covered on a transport vehicle presents the possibility of losing portions of that material during transport. Based on industry standard estimates that 1% of grain/plant material is lost throughout its transport, *R. toxicus*, its associated plant host, and vector nematodes may be disseminated in the U.S. during transport from port of entry to a destination pasture. The calculations in this study indicate that a truck of 30 50kg bales lose 1% of its material equating to many thousands of seeds and associated *R. toxicus* galls (Table 3.6). Potentially from a single transport vehicle, up to 1.5 million seeds and 150k *R. toxicus* galls could be disseminated at a 4.39% outbreak seed infection rate. A 1% transit loss from all the 6,294 (50kg) bales imported from Australia, potentially 3,147 (L.O.D.) to 32 million (4.39% outbreak level) galls could be disseminated along the roads. The roads that a truck travels through are risk areas, and the risk increases if they are rural farm roads with conducive soil environment close enough to where the seeds and galls may drop. If nematodes are disseminated simultaneously, then that provides all the factors needed for *R. toxicus* establishment (pathogen, host, vector, conducive environment), just from a 1% loss of hay from a transport truck.

The potential for this scenario happening multiple times is another important aspect. There are approximately 20,000 hay producers in Australia, and 110,000 truckloads of hay are

transported each year. The Australian hay industry is constantly expanding, and exports have doubled in just the last five years, 40% of these exports are from Western Australia, the location of the scenario of *R. toxicus* escape. That equates to 44,000 truckloads of hay. When taking into consideration the transport of *R. toxicus* infested hay through the U.S. after importation, it is reasonable to estimate a similar 1% loss of material from transport vehicles.

Trade is complex. Between source and destination many things impact a given commodity. Imports are often re-traded to other countries depending on market demands. This is not always tracked, and items can be repackaged obscuring the country of origin. This issue is sometimes addressed but not always solved, such as the U.S. trying proposed measures for meat labeling with Canada and Mexico; the latter two countries protested for financial reasons, so the issue was ultimately not addressed. These issues arise throughout the global trade system. Therefore, when we consider where Australia exports hay, we must also consider possible indirect pathways of that hay arriving in the U.S. One of Australia's largest hay export markets is China (Table 3.3), and China is #2 on the list of top sources of hay imports into the U.S. (Table 3.4), bringing in around 160k worth of hay in 2019 (WorldCity.com 2020, U.S. Census Bureau). There are other possible indirect trade routes for hay being re-traded.

Pathogen destination/introduction

An important part of pathway analysis are the many parallels for origin and destination in terms of the ability of a pathogen to survive. When analyzed, the likelihood of establishment of the potentially dangerous U.S. Select Agent, *Rathayibacter toxicus* in the U.S., it became evident that there are several pathways for this possibility. Despite the complexity of the *R. toxicus* life cycle, several ways by which an *R. toxicus* establishment in the U.S. could result were identified.

The U.S. and Australia have many regions with similar climate conditions. The most likely port of entry into the U.S. for Australian hay is at the Calexico/Mexicali border, which has the same arid type of climate as Australia. More climate similarities were identified throughout California and the Western U.S. that also have recorded presence of many *R. toxicus* host species and some nematode vectors which would facilitate establishment. Additionally, land cover of land types already containing grass and crop growth was quantified to determine what percentage of the land of the nearest states to the hay entry port may help facilitate *R. toxicus* and associated host establishment even more. An average of 21-28% of the West Coast states contained such land, and areas near the port of entry was discovered as a top producer of hay in California. This contributes to the understanding of risks associated with dissemination during transport of hay and the potential for establishment of the pathogen.

Even in the absence of vector and host populations at the destination location, the likelihood of establishment exists since *R. toxicus* galls are transported with their host and vector. Establishment of invasive grass species is a common problem in many countries including the U.S. Almost all *R. toxicus* host grasses entered the U.S. and became established and invasive in many states. These invasive species within *Poaceae*, colonized grasslands and farmland pastures. In particular, the establishment of invasive ryegrasses in new areas is strong (Fensham et al. 2013). Certain factors facilitate establishment, such as symbiosis with certain endophytes(Catford et al. 2009), presence of trees nearby and the associated litter fall as nutrient enrichment for the soil (Fensham et al. 2013), animal grazing (Neto etal 2014), and rainfalls after dry conditions (Fensham et al. 2013; Bock et al. 2007; Tozer et al. 2008), all of which increase establishment of invasive grasses. Invasive grasses also often displace native species (Fairfax and Fensham 2000; Jackson 2005). Invasions of exotic grasses into new areas occurs globally

and constantly, including in the U.S. To summarize, there are many ways in which the plant hosts that are co-transported with *R. toxicus* may also become established and thus help *R. toxicus* establish along with it.

References (Chapter 3)

Allen, J. (2012). Annual ryegrass toxicity–an animal disease caused by toxins produced by a bacterial plant pathogen. Microbiology Australia, 33(1), 18-21.

Australian Fodder Industry Association, (2020). Australian Export Hay Levy. Online: https://www.afia.org.au/index.php/afia-sponsors/305-australian-export-hay-levy#:~:text=Australia's%20fodder%20industry%20has%20a,a%20significant%20potential%20for%20growth.

Australian quarantine and Inspection Service. (2014). https://www.agriculture.gov.au/sites/default/files/style%20library/images/daff/__data/assets/pdffile/0016/114037/corynetoxin-hay-straw.pdf

Australian quarantine and Inspection Service. (2019). https://www.agriculture.gov.au/sites/default/files/style%20library/images/daff/__data/assets/pdffile/0016/114037/corynetoxin-hay-straw.pdf

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific data*, *5*, 180214.

Bertozzi, T., & Davies, K. A. (2009). *Anguina paludicola* sp. n.(*Tylenchida*: *Anguinidae*): the nematode associated with *Rathayibacter toxicus* infection in *Polypogon monspeliensis* and *Lachnagrostis filiformis* in Australia. *Zootaxa*, *2060*(1), 33-46.

Bertozzi, T., & McKay, A. C. (1995). Incidence on *Polypogon monspeliensis* of *Clavibacter toxicus* and *Anguina* sp., the organisms associated with flood plain staggers' in South Australia. *Australian journal of experimental agriculture*, *35*(5), 567-569.

Bock, C. E., Bock, J. H., Kennedy, L., & Jones, Z. F. (2007). Spread of non-native grasses into grazed versus ungrazed desert grasslands. *Journal of Arid Environments*, *71*(2), 229-235.

Brasier, C. M. (2008). The biosecurity threat to the UK and global environment from international trade in plants. *Plant Pathology*, *57*(5), 792-808.

CABI, (2020). Invasive Species Compendium. Wallingford, UK: CAB International. Online: www.cabi.org/isc.IndexMundi:

Compare United States and Australia. (2020). Online: https://www.indexmundi.com/factbook/compare/united-states.australia

Carey, Jennifer H. 1995. *Lolium perenne* subsp. multiflorum. In: Fire Effects Information System, [Online]. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (Producer). Online: https://www.fs.fed.us/database/feis/plants/graminoid/lolperm/all.htmlFire Effects Information System, U.S. Department of Agriculture, 2020. Online: https://www.feis-crs.org/feis/

Catford, J. A., Jansson, R., & Nilsson, C. (2009). Reducing redundancy in invasion ecology by integrating hypotheses into a single theoretical framework. *Diversity and distributions*, *15*(1), 22-40.

Catford, J. A., Vesk, P. A., White, M. D., & Wintle, B. A. (2011). Hotspots of plant invasion predicted by propagule pressure and ecosystem characteristics. *Diversity and Distributions*, *17*(6), 1099-1110.

Chatel, D. L., Wise, J. L., & Marfleet, A. G. (1979). Ryegrass toxicity organism found on other grasses. *Journal of Agriculture Western Australia*, *20*(3).

Data.gov website (2016). Online: https://catalog.data.gov/dataset/tiger-line-shapefile-2016-nation-u-s-primary-roads-national-shapefile

Devorshak, C. (2012). Uncertainty in Pest Risk Analysis. In: Plant Pest Risk Analysis: Concepts and Applications. Ed.: Devorshak, C. CAB International, Oxford, UK

Dewitz, J., (2019). National Land Cover Database (NLCD) 2016 Products: U.S. Geological Survey data release, Online: https://doi.org/10.5066/P96HHBIE.

Edgar, J. A., Cockrum, P. A., Stewart, P. L., Anderton, N. A., & Payne, A. L. (1994). Identification of Corynetoxins as the cause of poisoning associated with Annual Beardgrass (*Polypogon monspeliensis* (L) Desf.) and Blown Grass (*Agrostis avenacea C. Gemelin*).

Elmer, W. (2001). Seeds as vehicles for pathogen importation. *Biological Invasions 3*: 263–271.

Entegra Signature Structures (2020). Online: https://www.entegra.com.au/australias-hay-exports-where-does-the-hay-go-to/ (the fact that Western Australia exports 48% of Australia's hay exports)

Fairfax, R. J., & Fensham, R. J. (2000). The effect of exotic pasture development on floristic diversity in central Queensland, Australia. *Biological Conservation*, *94*(1), 11-21.

Fensham, R. J., Donald, S., & Dwyer, J. M. (2013). Propagule pressure, not fire or cattle grazing, promotes invasion of buffel grass *Cenchrus ciliaris*. *Journal of Applied Ecology*, *50*(1), 138-146.

Filipjev, (1936). *Anguina agrostis* and its ability to invade wheat seedlings after thirty-two years of dormancy. *Helminthol Soc Wash Proc*.

Findlay, R., & O'Rourke, K. H. (2007). Power and Plenty: Trade. *War, and the World Economy in the Second Millennium, Princeton*.

Gordh and S. McKirdy (eds.) (2014). The Handbook of Plant Biosecurity, DOI 10.1007/978-94-007-7365-3_1, (C) Springer Science+Business Media Dordrecht

Government of Australia (2005) Standard for Minimising the Risk of Corynetoxin Contamination of Hay and Straw for export. Australian Quarantine and Inspection Service. Department of Agriculture and Water Resources, Australian Government. Online: http://www.agriculture.gov.au/biosecurity

Government of Western Australia (2017) Department of Primary Industries and Regional Development: Agriculture and Food. Online: https://agric.wa.gov.au/n/2181

Government of Western Australia (2017). Oats: Hay Quality for Export and Domestic Use. Agriculture and Food. Online: https://www.agric.wa.gov.au/hay-production/oats-hay-quality-export-and-domestic-markets?page=0%2C3

Government of Western Australia (2018). Hay Production in the Wheatmbelt, Wheatbelt Development Commission. Online: https://www.wheatbelt.wa.gov.au/news/hay-production-wheatbelt/

GrainCentral.com (2017). Fewer, larger farms as consolidation continues in Australian grain industry. https://www.graincentral.com/cropping/fewer-larger-farms-as-consolidation-continues-in-australian-grain-industry/

Grewar, J. D., Allen, J. G., & Guthrie, A. J. (2009). Annual ryegrass toxicity in thoroughbred horses in ceres in the Western Cape Province, South Africa. *Journal of the South African Veterinary Association*, *80*(4), 220-223.

Griffin, R. (2012). 16 Uncertainty in Pest Risk Analysis. *Plant Pest Risk Analysis, 209.*

Griffin, R. (2012). Uncertainty in Pest Risk Analysis. In: Plant Pest Risk Analysis: Concepts and Applications. Ed.: Devorshak, C. CAB International, Oxford, UK

Griffin, R. (2014). Center for Plant Health Science and Technology, Plant Protection and Quarantine, APHIS, USDA, Raleigh, NC 27606, USA

Hannaway, D., Fransen, S., Cropper, J., Teel, M., Chaney, M., Griggs, T., Halse, R., Hart, J., Cheeke, P., Hansen, D. and Klinger, R., (1999). Annual ryegrass. *Oregon State University, USA Consultado el*, *11*(05), p.2005.

Hart, J. M., Mellbye, M. E., Young, W. C., & Silberstein, T. (2003). Nutrient management for annual ryegrass grown for seed: western Oregon.

Homer, C., Dewitz, J., Jin, S., & Xian, G. Z. (2019). Completion of the 2016 National Land Cover Database, Revealing Patterns of Conterminous US Land Cover Change from 2001 to 2016. *AGUFM*, *2019*, B24A-03.

Hulme, P. E. (2014). An introduction to plant biosecurity: past, present and future. In *The handbook of plant biosecurity* (pp. 1-25). Springer, Dordrecht.

International Plant Protection Convention. (2004). ISPM No. 11: Pest risk analysis for quarantine pests including analysis of environmental risks and living modified organisms. International standards for phytosanitary measures. Online. Available at http://www.fao.org/docrep/008/y5874e/y5874e00.HTM.

International Plant Protection Convention. (2007). International Standards For Phytosanitary Measures ISPM No. 2 Framework For Pest Risk Analysis. Secretariat of the International Plant Protection Convention (IPPC), Food and Agriculture Organization of the United Nations, Rome.

Jackson, J. (2005). Is there a relationship between herbaceous species richness and buffel grass (*Cenchrus ciliaris*)?. *Austral Ecology*, *30*(5), 505-517.

Koo, W., & Kennedy, P. L. (2005). International trade and agriculture. *Management of Environmental Quality: An International Journal*.

Limber, D. P. (1973). Notes on the longevity of *Anguina tritici* (Steinbuch, 1799)

Luster, D. G., McMahon, M. B., Carter, M. L., Sechler, A. J., Rogers, E. E., Schroeder, B. K., & Murray, T. D. (2020). Immunoreagents for development of a diagnostic assay specific for *Rathayibacter toxicus*. *Food and Agricultural Immunology*, *31*(1), 231-242.

McCullough, D. G., Work, T. T., Cavey, J. F., Liebhold, A. M., & Marshall, D. (2006). Interceptions of nonindigenous plant pests at US ports of entry and border crossings over a 17-year period. *Biological invasions*, *8*(4), 611.

McKay AC, Ophel KM, Reardon TB, Gooden JM (1993b) Livestock deaths associated with *Clavibacter toxicus/Auguina* sp. infection in seedheads of *Agrostis auenacea* and *Polypogon monspeliensis*. *Plant Disease 77:* 635–641

McMichael, A. J., Campbell-Lendrum, D. H., Corvalán, C. F., Ebi, K. L., Githeko, A., Scheraga, J. D., & Woodward, A. (2003). *Climate change and human health: risks and responses*. World Health Organization.

Murray, T. D., Schroeder, B. K., Schneider, W. L., Luster, D. G., Sechler, A., Rogers, E. E., & Subbotin, S. A. (2017). *Rathayibacter toxicus*, other *Rathayibacter* species inducing bacterial head blight of grasses, and the potential for livestock poisonings. *Phytopathology*, *107*(7), 804-815.

National Research Council, (1994). Science and Judgment in Risk Assessment. National Academy Press. Washington, D.C.

National Research Council (2009). Science and Decisions: Advancing Risk Assessment. The National Academies Press. Washington, D.C.

National Research Council (2002). Predicting invasions of nonindigenous plants and plant pests. National Academy Press. Washington, D.C.

Natural Earth Data (2020). Online: https://www.naturalearthdata.com/downloads/50m-cultural-vectors/50m-admin-1-states-provinces/

Neto, A.B., Savian, J.V., Schons, R.M.T., Bonnet, O.J.F., Do Canto, M.W., de Moraes, A., Lemaire, G. and de Faccio Carvalho, P.C. (2014). Italian ryegrass establishment by self-seeding in integrated crop-livestock systems: effects of grazing management and crop rotation strategies. *European Journal of Agronomy*, *53*, 67-73.

Piggott, N. (2013). Agricultural Economic and Policy Outlook. Presentation to: 8th Annual Ag Development Forum. Governor Holshouer Building, Raleigh.

Price, P. C., Fisher, J. M., & Kerr, A. (1979). Annual ryegrass toxicity: parasitism of *Lolium rigidum* by a seedgall forming nematode (*Anguina* sp.). *Annals of Applied Biology*, *91*(3), 359-369.

Rafoss, T. (2003). Spatial stochastic simulation offers potential as a quantitative method for pest risk analysis. *Risk Analysis: An International Journal*, *23*(4), 651-661.

Reed, M. R. (2001). *International trade in agricultural products/por Michael R. Reed* (No. 338.140973 R4.).

Riddel, B. (2013). Alternative Hay Crops for the Australian Hay Industry. Online: https://farmtable.com.au/build/alternative-hay-crops-australian-hay-industry/

Riley IT (1992) *Anguina tritici* is a potential vector of *Clavibacter toxicus*. *Australasian Plant Pathology 21*: 147–148.

Riley IT, Ophel KM (1992). *Clavibacter toxicus* sp. nov., the bacterium responsible for annual ryegrass toxicity in Australia. *International Journal of Systematic Bacteriology 42:* 64–68.

Riley, I. T., & Barbetti, M. J. (2008). Australian anguinids: their agricultural impact and control. *Australasian Plant Pathology*, *37*(3), 289-297.

Riley, I. (1992). Paddock sampling for management of annual ryegrass toxicity. *Journal of the Department of Agriculture, Western Australia, Series 4*, *33*(2), 51-56.

Riley, I. T. (1995). *Vulpia myuros* and the annual ryegrass toxicity organisms, *Anguina funesta* and *Clavibacter toxicus*. *Fundamental and Applied Nematology*, *18*(6), 595-598.

Riley, I. T. (1996). *Dilophospora alopecuri* on *Lolium rigidum* and *Holcus lanatus* in south-eastern Australia. *Australasian Plant Pathology*, *25*(4), 255-259.

Riley, I. T. (1992). *Anguina tritici* is a potential vector of *Clavibacter toxicus*. *Australasian Plant Pathology*, *21*(4), 147-149.

Riley, I. T., Schmitz, A., & de Silva, P. (2001a). *Anguina australis*, a vector for *Rathayibacter toxicus* in *Ehrharta longiflora*. *Australasian Plant Pathology*, *30*(2), 171-175.

Riley, I. T., & McKay, A. C. (1991). Invasion of some grasses by *Anguina funesta* (Nematoda: *Anguinidae*) juveniles. *Nematologica*, *37*(1-4), 447-454.

Riley, I. T., & McKay, A. C. (1990). Specificity of the adhesion of some plant pathogenic micro-organisms to the cuticle of nematodes in the genus *Anguina* (Nematoda: *Anguinidae*). *Nematologica*, *36*(1-4), 90-103.

Rubel, F., Brugger, K., Haslinger, K., & Auer, I. (2017). The climate of the European Alps: Shift of very high resolution Köppen-Geiger climate zones 1800–2100. *Meteorologische Zeitschrift*, *26*(2), 115-125.

Schneider, D. J. (1981). First report of annual ryegrass toxicity in the Republic of South Africa. *Onderstepoort Journal of Veterinary Research, 48*:251-255

Sequeira, R. (2014). Center for Plant Health Science and Technology, Plant Protection and Quarantine, APHIS, USDA, Raleigh, NC 27606, USA

Stynes, B. A., and Bird, A. F. (1982). Development of galls induced in *Lolium rigidum* by *Anguina agrostis*. Phytopathology 72, 336–346.

Subottin, S. A., & Riley, I. T. (2012). Stem and gall forming nematodes. Pp.: 521-577. *Practical Plant Nematology. Biblioteca Básica de Agricultura (BBA). Montecillo, Texcoco, Estado de México. México. 883p*.

The World Factbook 2020. Washington, DC: Central Intelligence Agency, 2020.

Tozer, K. N., Chapman, D. F., Quigley, P. E., Dowling, P. M., Cousens, R. D., & Kearney, G. A. (2008). Effect of grazing strategy, ryegrass overdrilling and herbicide application on vulpia content, tiller density and seed production in perennial pastures. *Australian Journal of Experimental Agriculture*, *48*(5), 632-640.

U.S. Census Bureau, (2020). Cartographic Boundary Files. Online: https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html

U.S. Geological Survey (2016). National Land Cover Database (NLCD) 2016 Land Cover. Science Product. Online: https://www.sciencebase.gov/catalog/item/5dadee41e4b09fd3b0c9d8ed

USDA APHIS (2008). Quantitative pathway analysis: likelihood of karnal bunt, (*Tilletia indica* M.), introduction as a result of the importation of United States wheat for Grain into the European Union and Desert durum wheat for grain into Italy. *United States Department of Agriculture-Animal and Plant Health Inspection Service.*

USDA APHIS PPQ (United States Department of Agriculture Animal and Plant Health Inspection Service Plant Protection and Quarantine). (2012a). Seeds Not for Planting. Online: http://www.aphis.usda.gov/import_export/plants/manuals/ports/downloads/seeds_not_for_planting.pdf

USDA APHIS PPQ (United States Department of Agriculture Animal and Plant Health Inspection Service Plant Protection and Quarantine). (2012b). Treatment Manual. Online: http://www.aphis.usda.gov/import_export/plants/manuals/ports/downloads/treatment.pdf

USDA, NRCS. 2020. The PLANTS Database (http://plants.usda.gov, 2 October 2020). National Plant Data Team, Greensboro, NC 27401-4901 USA.

WorldCity Trade Numbers. U.S. Census Bureau, 2020. Online: https://www.ustradenumbers.com/import/hay-including-alfalfa-hay-etc/

Appendices (Chapter 3)

Appendix 3.1: Summary of Australia and U.S.'s Koppen Geiger Climate classifications (includes zones common to Australia and U.S.) and zones of *R. toxicus* occurrence

Zone	Description	Specific Zone Classification	Zone Classification description	AU	U.S.	*R. toxicus* found
A	**tropical** or equatorial	**Af**	Tropical rainforest	✓	✓	
		Am	Tropical monsoon	✓	✓	
		As	Tropical Savanna, dry winter	✓	✓	
		Aw	Tropical Savanna, dry summer	✓	✓	
B	**arid or dry**	**BSh**	Hot desert	✓	✓	✓
		BSk	Cold desert	✓	✓	✓
		BWh	Hot semi-arid	✓	✓	
		BWk	Cold semi-arid	✓	✓	
C	**warm/mild temperate**	**Cfa**	Humid subtropical	✓	✓	✓
		Cfb	temperate oceanic	✓	✓	
		Cfc	subpolar oceanic	✓	✓	
		Csa	hot-summer mediterranean	✓	✓	✓
		Csb	warm-summer mediterranean	✓	✓	✓
		Csc	cold-summer meditrranean		✓	
		Cwa	monsoon-influenced subtropical	✓	✓	
D	**continental**	**Dfa**	hot-summer humid continental		✓	
		Dfb	warm-summer humid continental		✓	
		Dfc	subtropic climate	✓	✓	
		Dfd	extremely cold subarctic	✓		
		Dsa	mediterranean-influenced hot-summer humid continental		✓	
		Dsb	mediterranean-influenced warm-summer humid continental		✓	
		Dsc	mediterranean-influenced subarctic		✓	
		Dwa	monsoon-influenced hot-summer humid continental		✓	
		Dwb	monsoon-influenced warm-summer humid continental		✓	

Appendix 3.2: Roadways – many miles of possible routes for shipments of hay

Type	USA (mi)	Australia (mi)
National Highway system	160,955	10,500
Interstate Highways	47,432	8,469
Railways (freight)	138,000	21,014
Total navigatable roadways	4,090,000	567,311
Paved	2,658,500	219,550
Unpaved	1,431,500	347,762

Appendix 3.3: U.S. Animal species susceptible to *Rathayibacter toxicus* poisoning

Animal	*How many in the U.S.?*
Cattle	93,480,957
Horses	9,670,498
Sheep	5,146,648
Pigs	64,988,711
Goats	2,697,105
Buffalo	1,066,997
Other(mammals)	

Appendix 3.4: Detailed Data of which *R. toxicus* hosts are in each state for Figure 3.3 GIS map

State	*Lolium_rigidum*	*Agrostis_avanacea_Lachnagrostis_filiformis*	*Datylus_glomerata*	*Polypogon_monospelensis*	*Ehrharta_longiflora*	*Austrodanthonia_caespitosa*	*Avena_sativa*	*Danthonia_caespitosa*	*Lolium_multiflorum*	*Lolium_perenne*	*Lolium_persicum*	*Lolium_strictum*	*Lolium_temulentum*	*Phalaris_spp*	*Vulpia_myuros*
Alabama	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Alaska	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Arizona	1	0	1	1	0	0	1	0	1	1	0	1	1	1	1
Arkansas	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
California	1	1	1	1	1	0	1	0	1	1	0	1	1	1	1
Colorado	0	0	1	1	0	0	1	0	1	1	1	0	0	1	0
Connecticut	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Delaware	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Florida	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Georgia	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Hawaii	1	1	1	1	0	0	1	0	1	1	0	0	1	1	1
Idaho	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Illinois	0	0	1	0	0	0	1	0	1	1	0	0	1	1	1
Indiana	0	0	1	0	0	0	1	0	1	1	0	0	0	1	1
Iowa	0	0	1	0	0	0	1	0	1	1	0	0	1	1	1
Kansas	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Kentucky	0	0	1	0	0	0	1	0	1	1	0	0	1	1	1
Louisiana	1	0	1	1	0	0	1	0	1	1	0	1	1	1	1
Maine	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Maryland	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Massachussetts	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Michigan	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Minnesota	0	0	1	1	0	0	1	0	1	1	0	0	1	1	0
Mississippi	1	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Missouri	1	0	1	0	0	0	1	0	1	1	1	0	1	1	1
Montana	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1
Nebraska	0	0	1	1	0	0	1	0	1	1	0	0	1	1	0
Nevada	0	0	1	1	0	0	1	0	1	1	0	0	0	1	1

New Hampshire	0	0	1	1	0	0	1	0	1	1	0	0	0	1	1
New Jersey	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
New Mexico	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
New York	0	0	1	1	0	0	1	0	1	1	1	0	1	1	1
North Carolina	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
North Dakota	0	0	1	1	0	0	1	0	1	1	1	0	1	1	0
Ohio	0	1	1	0	0	0	1	0	1	1	0	0	1	1	1
Oklahoma	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Oregon	1	0	1	1	0	0	1	0	1	1	0	1	1	1	1
Pennsylvania	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Rhode Island	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
South Carolina	0	1	1	1	0	0	1	0	1	1	0	0	1	1	1
South Dakota	0	0	1	1	0	0	1	0	1	1	0	0	1	1	0
Tennessee	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Texas	1	1	1	1	0	0	1	0	1	1	0	1	1	1	1
Utah	0	0	1	1	0	0	1	0	1	1	0	0	0	1	1
Vermont	0	0	1	0	0	0	1	0	1	1	0	0	1	1	0
Virginia	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
Washington	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1
West Virginia	0	0	1	0	0	0	1	0	1	1	0	0	1	1	1
Wisconsin	0	0	1	1	0	0	1	0	1	1	0	0	0	1	1
Wyoming	0	0	1	1	0	0	1	0	1	1	0	0	1	1	0

Note: Zero means no detection of the host in the state, 1 means host species have been detected. No data were found to quantify presence in each state. These data were input into QGIS 3.14 for subsequent analysis.

Appendix 3.5. Vectors in U.S. states data used in Figure 3.3 GIS map.

	Anguina funesta	*Anguina tritici*	*Anguina australis*	*Anguina paludicola*
Alabama	0	0	0	0
Alaska	0	0	0	0
Arizona	0	0	0	0
Arkansas	0	0	0	0
California	0	0	0	0
Colorado	0	0	0	0
Connecticut	0	0	0	0
Delaware	0	0	0	0
Florida	0	0	0	0
Georgia	0	1	0	0
Hawaii	0	0	0	0
Idaho	0	0	0	0
Illinois	0	0	0	0
Indiana	0	0	0	0
Iowa	0	0	0	0
Kansas	0	0	0	0
Kentucky	0	0	0	0
Louisiana	0	0	0	0
Maine	0	0	0	0
Maryland	0	1	0	0
Massachussetts	0	0	0	0
Michigan	0	0	0	0
Minnesota	0	0	0	0
Mississippi	0	0	0	0
Missouri	0	0	0	0
Montana	0	0	0	0
Nebraska	0	0	0	0
Nevada	0	0	0	0
New Hampshire	0	0	0	0
New Jersey	0	0	0	0
New Mexico	0	0	0	0
New York	0	0	0	0

North Carolina	0	1	0	0
North Dakota	0	0	0	0
Ohio	0	0	0	0
Oklahoma	0	0	0	0
Oregon	1	1	0	0
Pennsylvania	0	0	0	0
Rhode Island	0	0	0	0
South Carolina	0	1	0	0
South Dakota	0	0	0	0
Tennessee	0	0	0	0
Texas	0	0	0	0
Utah	0	0	0	0
Vermont	0	0	0	0
Virginia	0	1	0	0
Washington	0	1	0	0
West Virginia	0	1	0	0
Wisconsin	0	0	0	0
Wyoming	0	0	0	0

Note: Zero means no detection of the host in the state, 1 means host species have been detected. No data were found to quantify presence in each state. These data were input into QGIS 3.14 for subsequent analyses.