University of Texas Rio Grande Valley
ScholarWorks @ UTRGV

Theses and Dissertations - UTB/UTPA

5-2014

Rates of evolution and point mutations of bacterial plant pathogens compared to bacterial vertebrate pathogens

Alejandro Cantu
University of Texas-Pan American

Follow this and additional works at: https://scholarworks.utrgv.edu/leg_etd

Part of the Biology Commons, and the Microbiology Commons

Recommended Citation

Cantu, Alejandro, "Rates of evolution and point mutations of bacterial plant pathogens compared to bacterial vertebrate pathogens" (2014). *Theses and Dissertations - UTB/UTPA*. 899.
https://scholarworks.utrgv.edu/leg_etd/899

This Thesis is brought to you for free and open access by ScholarWorks @ UTRGV. It has been accepted for inclusion in Theses and Dissertations - UTB/UTPA by an authorized administrator of ScholarWorks @ UTRGV. For more information, please contact justin.white@utrgv.edu, william.flores01@utrgv.edu.

RATES OF EVOLUTION AND POINT MUTATIONS OF BACTERIAL PLANT PATHOGENS COMPARED TO BACTERIAL VERTEBRATE PATHOGENS

A Thesis

by

ALEJANDRO CANTU

Submitted to the Graduate School of the
University of Texas-Pan American
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2014

Major Subject: Biology

RATES OF EVOLUTION AND POINT MUTATIONS OF BACTERIAL PLANT PATHOGENS COMPARED TO BACTERIAL VERTEBRATE PATHOGENS

A Thesis
by
ALEJANDRO CANTU

COMMITTEE MEMBERS

Dr. Erin L. Schuenzel
Chair of Committee

Dr. Robert K. Dearth
Committee Member

Dr. Matthew D. Terry
Committee Member

May 2014

Copyright 2014 Alejandro Cantu

All Rights Reserved

ABSTRACT

Cantu, Alejandro, Rates of Evolution and Point Mutations of Bacterial Plant Pathogens Compared to Bacterial Vertebrate Pathogens. Master of Science (MS), May, 2014, 67 pp., 14 tables, 22 illustrations, 39 references, 5 titles.

The relationship between a pathogen and its host is a constant interaction; as the host evolves defense mechanisms to protect itself from the pathogen, the pathogen will evolve in order to evade them. During this evolutionary process, the genomes of bacterial pathogens can change due to different selection pressures exerted by the immune system of their respective hosts. This study aimed to determine how differences in selection pressures affect the evolution of plant and vertebrate pathogens. The focus of this study was to analyze population level changes in conserved genes, as well as species level changes in the Type III Secretion System (T3SS) and in conserved genes of various pathogens. Results suggest that, at the population level, a difference in the life history of plant and vertebrate pathogens exists. At the species level, various genes are seemingly undergoing differential and/or positive selection regardless of the host.

DEDICATION

The completion of my graduate studies would not have been possible without the support and encouragement of my family and friends. My mother, Aida Cantu and my father, Alejandro Cantu, continually inspired and motivated me to accomplish this degree. Thank you for your always believing in me.

ACKNOWLEDGMENTS

I am very grateful to Dr. Erin L. Schuenzel, chair of my thesis committee, for all her mentoring and advice. From research design, data analysis, to manuscript editing, she encouraged me to complete this process through her patience and guidance. I am also grateful to Dr. Robert K. Dearth and Dr. Matthew D. Terry for all their advice and guidance throughout the course of my presentations.

TABLE OF CONTENTS

Page

ABSTRACT………………………………………………………………………………. iii

DEDICATION…………………………………………………………………………….. iv

ACKNOWLEDGEMENTS……………………………………………………………….. v

TABLE OF CONTENTS…………………………………………………………………. vi

LIST OF TABLES………………………………………………………………………… viii

LIST OF FIGRUES……………………………………………………………………….. x

CHAPTER I. INTRODUCTION…………………………..…………………………..…. 1

Innate and Adaptive Immune Systems………………………………………………. 3

Plant Immune System…………………………………………………………….... 4

Vertebrate Immune System……………………………………………………….. 5

CHAPTER II. ANALYSIS OF THE POINT MUTATION RATE IN CONSERVED GENES WITHIN BACTERIAL PATHOGEN POPULATIONS………………………….. 8

Materials and Methods..……………………………………………………………... 9

Results………………………………………………………………………………. 10

Discussion…………………………………………………………………………. 11

CHAPTER III. ADAPTIVE EVOLUTION OF THE TYPE III SECRETION SYSTEM (T3SS) IN PLANT AND VERTEBRATE PATHOGENS……………………………….... 13

Materials and Methods..……………………………………………………... 15

Results………………………………………………………………………….. 16

Discussion……………………………………………………………………… 17

CHAPTER IV. ANALYSIS OF CONSERVED GENES IN PLANT AND VERTEBRATE PATHOGENS TO DETERMINE DIFFERENCES IN EVOLUTIONARY RATES..……... 19

Materials and Methods..……………………………………………………... 20

Results………………………………………………………………………….. 22

Discussion……………………………………………………………………… 23

CHAPTER V. DISCUSSION…………………………………………………………... 25

REFERENCES………………………………………………………………………… 28

BIOGRAPHICAL SKETCH…………………………………………………………… 67

LIST OF TABLES

Page

Table 1: Information regarding the bacterial population study organisms………………….... 31

Table 2: Results of t-tests with unequal variances comparing the population data obtained from plant and vertebrate pathogens……………………………………………………… 32

Table 3: Information regarding the selected pathogens with Type III Secretion Systems…... 33

Table 4: Significant differences in the rate of evolution (one ω vs. free ω) of the Type III Secretion System genes with degrees of freedom………………………………………… 34

Table 5: Significant differences in the rate of evolution (ω vertebrate vs. ω plant) of the Type III Secretion System genes with degrees of freedom……………………………….. 35

Table 6: ω values of ancestral, vertebrate, and plant pathogen branches of the Type III Secretion System genes………………………………………………………………… 36

Table 7: Number of changes within codons ($\omega < 1$ vs. $\omega = 1$) of the Type III Secretion System genes with degrees of freedom………………………………………………… 37

Table 8: Number of changes within codons ($\omega < 1$ vs. $\omega = 1$ vs. $\omega > 1$) of the Type III Secretion System genes with degrees of freedom………………………………………… 38

Table 9: Housekeeping genes selected…………………………………………………. 39

Table 10: Significant differences in the rate of evolution (one ω vs. free ω) of the housekeeping genes with degrees of freedom……………………………………………….. 40

Table 11: Significant differences in the rate of evolution (ω vertebrate vs. ω plant) of the housekeeping genes with degrees of freedom……………………………………………….. 41

Table 12: ω values of ancestral, vertebrate, and plant pathogen branches of the housekeeping genes…………………………………………………………………….. 42

Table 13: Number of changes within codons ($\omega < 1$ vs. $\omega = 1$) of the housekeeping genes with degrees of freedom…………………………………………………………………... 43

Table 14: Number of changes within codons ($\omega < 1$ vs. $\omega = 1$ vs. $\omega > 1$) of the housekeeping genes with degrees of freedom……………………………………………….. 44

LIST OF FIGURES

Page

Figure 1: Rate of evolution of the *YscC* Type III Secretion System gene in plant and vertebrate pathogen lineages.. 45

Figure 2: Rate of evolution of the *YscF* Type III Secretion System gene in plant and vertebrate pathogen lineages.. 46

Figure 3: Rate of evolution of the *YscJ* Type III Secretion System gene in plant and vertebrate pathogen lineages.. 47

Figure 4: Rate of evolution of the *YscL* Type III Secretion System gene in plant and vertebrate pathogen lineages.. 48

Figure 5: Rate of evolution of the *YscN* Type III Secretion System gene in plant and vertebrate pathogen lineages.. 49

Figure 6: Rate of evolution of the *YscQ* Type III Secretion System gene in plant and vertebrate pathogen lineages.. 50

Figure 7: Rate of evolution of the *YscR* Type III Secretion System gene in plant and vertebrate pathogen lineages.. 51

Figure 8: Rate of evolution of the *YscS* Type III Secretion System gene in plant and vertebrate pathogen lineages.. 52

Figure 9: Rate of evolution of the *YscT* Type III Secretion System gene in plant and vertebrate pathogen lineages...…………………………………………………………… 53

Figure 10: Rate of evolution of the *YscU* Type III Secretion System gene in plant and vertebrate pathogen lineages...…………………………………………………………… 54

Figure 11: Rate of evolution of the *YscV* Type III Secretion System gene in plant and vertebrate pathogen lineages...…………………………………………………………… 55

Figure 12: Rate of evolution of the *YscW* Type III Secretion System gene in plant and vertebrate pathogen lineages...…………………………………………………………… 56

Figure 13: Rate of evolution of the *atpC* housekeeping gene in plant and vertebrate pathogen lineages...……………………………………………………………………….. 57

Figure 14: Rate of evolution of the *coaD* housekeeping gene in plant and vertebrate pathogen lineages...……………………………………………………………………….. 58

Figure 15: Rate of evolution of the *folA* housekeeping gene in plant and vertebrate pathogen lineages...……………………………………………………………………….. 59

Figure 16: Rate of evolution of the *greA* housekeeping gene in plant and vertebrate pathogens lineages..……………………………………………………………………….. 60

Figure 17: Rate of evolution of the *nusG* housekeeping gene in plant and vertebrate pathogen lineages...……………………………………………………………………….. 61

Figure 18: Rate of evolution of the *rpmF* housekeeping gene in plant and vertebrate pathogen lineages...……………………………………………………………………….. 62

Figure 19: Rate of evolution of the *rpmG* housekeeping gene in plant and vertebrate pathogen lineages...……………………………………………………………………….. 63

Figure 20: Rate of evolution of the *rpmI* housekeeping gene in plant and vertebrate pathogen lineages...………………………………………………………………………….. 64

Figure 21: Rate of evolution of the *secE* housekeeping gene in plant and vertebrate pathogen lineages...………………………………………………………………………….. 65

Figure 22: Rate of evolution of the *ssb* housekeeping gene in plant and vertebrate pathogen lineages………………………………………………………………………………. 66

CHAPTER I

INTRODUCTION

The relationship between a bacterial pathogen and its host is a constant interaction. As the host evolves defense mechanisms in order to protect itself from the pathogen, the pathogen will, in turn, evolve in order to evade those mechanisms (Anderson *et al*., 2011; Burdon and Thrall, 2009). The bacteria and host essentially participate in an evolutionary arms race leading to possible coevolution. Evolution of the pathogen will depend on the selection pressure exerted by host and the mutation rate of the bacteria.

During this evolutionary process, the genome of a bacterial pathogen can change via point mutations, chromosomal mutations, or horizontal gene transfer. Point mutations cause single base pair mutations in the DNA sequence while chromosomal mutations involve large DNA portions. An example of one such mutation are inversions, where a segment of DNA breaks off a chromosome, inverts its original orientation, and re-inserts itself back into the same chromosome. Horizontal gene transfers (transformation, conjugation, and transduction) result in the acquisition of unrelated genetic information (Treangen and Rocha, 2011), which can enable bacteria to adapt rapidly. A well-known example is how a single bacterium that acquires antibiotic resistance can quickly transfer the resistance gene(s) to other species of bacteria through conjugation, allowing them to become resistant as well (Hawkey and Jones, 2009).

With regards to evolutionary changes that occur in bacteria, point mutations are usually referred to as the raw material that promotes them. Point mutations can affect the amino acid

sequence coded by a specific gene in two ways. If the point mutation is synonymous (silent), such as in a substitution of a single base pair, the alteration of the nucleotide will have no change in the amino acid encoded by the gene. However, if the point mutation is nonsynonymous, this will result in the translation of a different amino acid, and thus, a different protein. Nonsynonymous mutations result in the possible alteration of a single amino acid (missense mutation) or the alteration of an amino acid coding codon for a stop codon (nonsense mutation). Nonsense mutations, base pair insertions, and deletions result in a frameshift of downstream codons that can alter the translation of more than one amino acid resulting in a substantially different protein (Bryant *et al.*, 2012). As the expression of a different protein might lead to a change in phenotype, point mutations play a key role in promoting evolutionary change.

Similar to point mutations, chromosomal mutations can promote evolutionary change as well. An example of such mutations are gene duplications. A gene duplication event occurs when an extra copy of a chromosomal region is synthesized and becomes part of the chromosome. If a bacterium was to undergo binary fission, the resulting bacterium would have the duplicated gene on its genome as well. Serres *et al.* (2009) compared the protein families of various bacterial groups to examine any functional similarities and differences. Their findings led them to conclude that bacterial protein families seem to have arisen by evolution through gene duplication and divergence. Furthermore, gene copies retained over the evolutionary history of bacteria are variants that have led to different bacterial physiologies and taxa.

Aside from point and chromosomal mutations, horizontal gene transfer can also promote the evolution of bacteria. Hawkey and Jones (2009) investigated the rise of multiple drug resistance (MDR) bacteria in recent years. They found that plasmids, which are independent DNA molecules that are generally smaller compared to the chromosomal DNA, are the major

vectors by which MDR bacteria accelerate the dispersal of antibiotic resistance genes. This is due to the fact that, via conjugation, bacteria can transfer the plasmids containing these genes between strains and species. Additionally, Hawkey and Jones (2009) found that some plasmids contain various resistance genes, which facilitate the transfer of multiple antibiotic resistance fairly rapidly within different bacterial groups.

Due to the dramatic and rapid changes possible in the pathogen, the host must also be able to respond. In the host, the defense mechanisms undergoing rapid evolution are typically part of the immune system, which provides an organism with the ability to recognize, respond to, and defend itself against infectious agents (Germain, 2001). Due to this benefit, the appearance of an immune system in plants and animals during their respective evolutionary histories was likely sustained by the process of natural selection. The immune systems of these groups are subdivided into two types: innate and adaptive.

Innate and Adaptive Immune Systems

The innate immune system can be found in both plants and vertebrates. Upon host exposure to pathogens, it initiates an immediate maximal immune response. Additionally, the innate immune system lacks antigen (foreign substance) recognition of pathogens, and exposure results in no immunologic memory, or the ability to recognize a pathogen upon future exposure (Cooper *et al.*, 2009). As a result, the overall response to pathogens is always generic, or non-specific.

The adaptive immune system has several differences to the innate immune system. For example, it is only found in vertebrates. Also, following exposure, a lag time exists prior to the maximal immune response. The adaptive immune system is capable of recognizing pathogens by

their antigens, and exposure results in immunologic memory, which increases overall efficiency upon future exposures (Cooper *et al.*, 2009). By having immunologic memory, the adaptive immune system can change an organism's immunity within its lifespan, while the innate immune system only changes between generations due to inheritance.

Plant Immune System

Although the immune system is subdivided into innate and adaptive components, the hosts' defense mechanisms vary depending on the taxon. For instance, plants have a two-branched innate immune system. The first branch recognizes and responds to molecules common to many classes of microbes, including non-pathogens. The second branch responds to pathogen virulence factors, such as toxins, either directly or through their effects on host targets. The recognition occurs via the activation of disease resistance (R) proteins found within the plant. These proteins are hypothesized to be activated by effector molecules (virulence factors) that are released by the pathogen upon invasion and not through the direct recognition of the pathogens (Jones and Dangl, 2006).

The primary line of defense in plants is chemical and secondary metabolites, which are very successful in dealing with bacteria and fungi. An example is salicylic acid (SA), a hormone that induces the hypersensitive response. This defense mechanism induces cells around an infection site to die and the cell wall to strengthen, forming a barrier that prevents pathogens from invading deeper into the plant (Brodersen *et al.*, 2005).

The hypersensitive response is not always effective however. Harold Flor was the first to demonstrate that, via this immune response, the inheritance of both resistance in the plant and the pathogen's ability to cause disease is controlled by pairs of matching genes, a hypothesis he

referred to as the gene-for-gene relationship. In this relationship, the plant's gene is known as the resistance (*R*) gene, while the pathogen's is called the avirulence (*Avr*) gene. Plants that produce a specific *R* gene product are only resistant to pathogens that produce a corresponding *Avr* gene product (Flor, 1971). If the proteins produced by the pathogen (the *Avr* gene products) are not recognized by the plants' immune system through *R* gene products, the hypersensitive response is not triggered and the plant will eventually succumb to disease.

The gene-for-gene relationship exemplifies the coevolution that exists between plant and pathogen. Jones and Dangl (2006) demonstrate a scenario in which a pathogen that carries an effector gene (*E1*) is recognized by a rare allele (*R1*) in the plant host. The result is higher selection rates for an increased frequency of *R1* in the population. Pathogens in which the effector is mutated become selected for due to their ability to grow in plants that have the *R1* allele. As a result, the effectiveness of the *R1* allele decreases, as well as the fitness and frequencies of the plants that have it. Because the population of the pathogen still contains individuals with *E1*, in the absence of *R1*, both the fitness and frequency of *E1* increases. The result will be the initial selection of *R1* individuals, continuing the coevolution cycle.

Vertebrate Immune System

The vertebrate immune system is a vastly complex network of circulating cells and molecules, as well as large tissues and organs. It has cell-mediated and humoral compartments. In addition, vertebrates possess both innate and adaptive components in their immune system (Zimmerman *et al.*, 2009). A common innate defense in vertebrates is the inflammatory response. Upon exposure to foreign invaders, such as bacteria, the affected area becomes inflamed and nearby blood vessels dilate, becoming permeable for immune cells including

monocytes, macrophages, and neutrophils to exit in order to get rid of the invaders. These three types of cells eliminate the invaders either through direct contact or through phagocytosis (Janeway *et al.*, 2005)—that is, by engulfing them and digesting them inside the cell.

Adaptive aspects of vertebrate immunity include antigen dependency and immunologic memory. B cells and T cells have dissimilar receptors, allowing the recognition of specific antigens; this is particularly important in immunologic memory. Memory cytotoxic T cells are produced following pathogen exposure, which will act faster upon future exposure to the same antigen/pathogen. This increases the efficiency of the adaptive immune system.

The benefits of immunologic memory are usually associated with viral infections. For example, when newly exposed to a specific strain of the flu virus a person was previously exposed to, that person's immune system will generate a more efficient immune response to fight the infection. However, it has also been demonstrated that bacterial pathogens can trigger the activation of immunologic memory after exposure. This is usually exemplified in the effects of vaccines, which contain antigens specific to a bacterial pathogen strain that will elicit an immune response and promote the propagation of memory cells. In a 2012 study by Patel *et al.*, researchers aimed to demonstrate whether memory B cell responses to *Vibrio cholerae* O1, a bacterial pathogen that causes cholera, were associated with protection from secondary exposure. Their findings indicated that the presence of memory B cells specific to this strain in the blood of patients previously exposed led to a 68% decrease in the risk of infection. This suggests that the presence of those memory B cells is an important factor in protection against future exposure to *V. cholerae* O1.

Additional key components of the adaptive immune system are the major histocompatibility (MHC) receptors. These receptors play a key role in the induction of immune

responses via the identification of pathogens (Borghans *et al.*, 2004) and are well known to have elevated rates of evolution (Hughes and Nei, 1988; Hughes and Nei, 1989). Additionally, the diversity of MHC receptors is quite large, with hundreds of different alleles having been identified for some MHC loci. Because adapting pathogens evade identification by the most common MHC alleles, they provide a selective pressure for the host population to have a large variety of rare MHC alleles. If the host population is large enough, a large set of MHC alleles can persist over many generations via host-pathogen coevolution and frequency-dependent selection (Borghans *et al.*, 2004). Furthermore, this suggests that adaptive immunity is adapting to pathogens at these receptors.

This study was undertaken to determine if a correlation exists between different selection pressures exerted from the hosts' immune systems and the rates of evolution for the bacterial pathogens. Furthermore, the absence of immunological memory in the innate immune system suggests that the adaptive immune system exerts a higher selection (coevolutionary) pressure on pathogens, leading to higher mutation rates and higher rates of evolution.

CHAPTER II

ANALYSIS OF THE POINT MUTATION RATE IN CONSERVED GENES WITHIN BACTERIAL PATHOGEN POPULATIONS

The mutation rates of selected bacterial pathogen populations were determined by analyzing their "housekeeping" genes. These genes perform essential functions necessary to maintain cellular functions, such as transcription, translation, and signaling (Butte *et al.*, 2001). Due to their key functions, housekeeping genes are usually conserved within populations.

The gene population data of the pathogens selected for this study was obtained from the online databases Multi Locus Sequence Typing (MLST; http://www.mlst.net/), PubMLST (http://pubmlst.org/), and Plant Associated and Environmental Microbes Database (PAMDB; http://genome.ppws.vt.edu/cgi-bin/MLST/home.pl). These databases utilize a technique known as multilocus sequence typing (MLST), which involves the PCR amplification and sequencing of 5-10 loci of approximately 500 base pairs in length. Each varying sequence is assigned a unique locus variant or allele number, and each combination of locus variants is assigned a sequence type (ST) that is used to indicate a specific set of sequences. The results are used to divide a population and facilitate the comparison of results via the internet. Over the years, the cumulative MLST data that has been gathered includes important information of various bacterial species, including their time and place of isolation, their host or niche, and their clinical or drug resistance profiles (Inouye *et al.*, 2012)

By focusing on analyzing conserved housekeeping genes, if results show differences in the mutation rates of these genes in plant and vertebrate pathogens, this could suggest that differences in selection exerted by the pathogens' respective host immune systems exist. To determine the existence of different mutation rates, point mutations were analyzed. This is due to the fact point mutations are generally referred to as the raw material that promotes evolutionary changes in bacteria, as they can alter the amino acid sequence of a protein, lead to the translation of a different protein, and a change in phenotype. Furthermore, it is expected that the mutation rate for vertebrate pathogens will be higher due to possible higher selection pressure exerted by their respective hosts.

Materials and Methods

Bacterial pathogens with population datasets specific to plant or vertebrate hosts were identified and downloaded from public databases, including MLST, PubMLST, and PAMDB. The gene datasets were saved under FASTA format. The total numbers of bacterial pathogen datasets were seven of plants and 34 of vertebrates; however, only datasets with number of isolates under 500 and within a single species were utilized to control for sampling differences, for a total of four plants pathogens and 11 vertebrates pathogens (Table 1).

The datasets were inputted into DnaSP v. 5.10 (Librado and Rozas, 2009; http://www.ub.edu/dnasp/) for alignment of the alleles and analysis. Data gathered under polymorphic sites included number of sites (length) sampled, variable (polymorphic) sites, total number of mutations, singleton variable sites (two variants), and parsimony informative sites (two variants). For DNA polymorphism, the only values gathered were nucleotide diversity (π). For recombination (R), data gathered included minimum number of recombination events,

estimate of R per gene, and estimate of R between adjacent sites. Lastly, the number of alleles was noted as well. For all values, averages, standard deviations, and variances were calculated.

T-tests were calculated in the website VassarStats (http://www.vassarstats.net/) comparing the data from plant and vertebrate pathogens, including the number of alleles, number of isolates, number of polymorphic sites, number of sites sampled, number of mutations, nucleotide diversity, and number of recombinations. In addition, analyses of covariance (ANCOVAs) were performed between the number of sites sampled and all different data points due to a significant difference observed in plant and vertebrate site sampling.

Results

The P values that resulted from t-tests comparing the significance of the difference between variances were significant; thus, data from t-tests assuming unequal variances were obtained (Table 2). The differences in the means of the number of alleles, number of isolates, nucleotide diversity, and number of recombinations were all not significant ("NS"), as each of their respective P values are > 0.05. Conversely, the differences in the means of number of polymorphic sites, number of sites sampled, and number of mutations are all significant, as all of their P values are < 0.05 (Table 2). Lastly, the results from the ANCOVAS between the number of sites sampled and all different data points indicated no significant covariance was found, indicating the results obtained from the t-tests were not affected due to the sampling difference observed.

Discussion

Analysis of the results suggests that the correlation between the mutation rate of the plant and vertebrate pathogen housekeeping genes with regards to different selection pressures exerted by the pathogens' respective host immune systems is opposite to what was expected. As shown in Table 2, the t-tests comparing polymorphic sites and number of mutations favor plant pathogens, as the mean values for each of those t-tests are higher than those of vertebrate pathogens. This suggests that the mutation rate of plant pathogens is higher than that of vertebrate pathogens, and thus, that the plant hosts are exerting higher selection pressure on the plant pathogens versus vertebrate hosts on the vertebrate pathogens.

Plant pathogens having a higher mutation rate may indicate a difference in the life histories of the plant and vertebrate pathogens. Typically, plant pathogens infect multiple hosts over a wide taxonomic range, as opposed to vertebrate pathogens which usually infect a single host such as humans. For example, *Xyllela fastidiosa*, one of the plant pathogens selected for this study, has a host range of over 150 plant species (Janse and Obradovic, 2010). On the other hand, *Stenotrophomonas maltophilia*, one of the vertebrate pathogens selected, is characterized for being an antibiotic-resistant opportunistic pathogen only in humans (Waters *et al*., 2007). Due to this, a higher selection pressure is exerted on plant pathogens versus vertebrate pathogens, which is indicative by the significantly higher plant pathogen means of both their number of polymorphic sites and mutations as previously stated (Table 2).

To account for this difference in life histories, the mutation rate of more plant pathogens with wide taxonomic ranges, as well as more vertebrate pathogens with narrow host ranges, could be assessed. If the mutation rate of the plant pathogens is also found to be significantly

higher than that of vertebrate pathogens, this might be consistent with the idea that having a wider taxonomic range leads to higher mutation rates in spite of the hosts' defense mechanisms

.

CHAPTER III

ADAPTIVE EVOLUTION OF THE TYPE III SECRETION SYSTEM (T3SS) IN PLANT AND VERTEBRATE PATHOGENS

Pathogenicity genes are generally known to have elevated rates of evolution, possibly due to the pathogen's adaptation in response to high selection pressure from their host. Gram-negative bacteria have a structure known as the Type III Secretion System (T3SS), also known as an injectosome or needle complex, which is an organelle that allows them to inject pathogenic proteins into eukaryotic cells (Galán and Wolf-Watz, 2006; Mueller *et al.*, 2008). This organelle is comprised of a basal body that spans through the inner and outer bacterial membranes, as well as an external needle that protrudes from the bacterial surface. To deliver the bacterial proteins into the host cell, the injectosome exports an additional set of three proteins required, which are known as translocators (Mueller *et al.*, 2008).

The effector proteins secreted by T3SS trigger various host cellular pathways, including rearrangements of the cytoskeleton and defense responses. These proteins have been identified as products of avirulence (*Avr*) genes, and in plants, are responsible for inducing the host defense responses in accordance to the plant's corresponding resistance (*R*) genes. In vertebrates, these bacterial effector proteins are known to suppress host defenses. Among these, *YopJ*, an effector protein found in the vertebrate pathogen *Yersinia pestis*, induces apoptosis in macrophages

(Büttner and Bonas, 2003), whose key role in the vertebrate immune system is to engulf and digest invading pathogens.

The 12 genes of the T3SS that were selected for this study are *YscC*, *YscF*, *YscW*, *YscJ*, *YscR*, *YscS*, *YscT*, *YscU*, *YscV*, *YscN*, *YscQ*, and *YscL*. From these, both *YscF* and *YscO* are part of the needle that protrudes from the T3SS and penetrates the plasma membrane of the host cell. *YscC* and *YscW* are part of the outer rings found in the outer membrane of the bacterial cell. *YscJ*, *YscR*, *YscS*, *YscT*, *YscU*, and *YscV* are all part of the inner rings found in the inner membrane of the bacteria which stabilize the structure of the T3SS. Lastly, *YscN*, *YscL*, and, *YscQ* constitute the ATPase complex that provides energy for the proper functioning of the T3SS (Ogata *et al*., 1999).

The evolution of Type III Secretion System genes is likely dominated by competing forces, including the requirement for conservation of virulence function, the avoidance of host defenses, and possible adaptation to new hosts (Rohmner *et al.*, 2004). Observed differences in the evolutionary rates of these genes in plant and vertebrate pathogens could suggest that differences in selection exerted by the pathogens' respective host immune systems exist.

To determine the evolutionary rates, dN/dS (i.e. ω or Ka/Ks) ratios will be calculated. The ratios are obtained by dividing the number of amino acid changes per site over the number of silent changes per site in each of the genes. If the ratio is less than one, the gene is undergoing negative (purifying) selection, which removes deleterious alleles. If the ratio is equal to one, the gene is undergoing neutral evolution or genetic drift. And if the ratio is greater than one, the gene is undergoing positive selection. Differences in the evolutionary rates of the T3SS between plant and vertebrate pathogens may be exhibited in individual genes of the T3SS undergoing positive selection within a species or in different species undergoing differential selection.

Materials and Methods

12 genes of the Type III Secretion System were selected: *YscF, YscC, YscW, YscJ, YscR, YscS, YscT, YscU, YscV, YscN, YscQ,* and *YscL*. Representative plant pathogens were identified from Mansfield *et al*. (2012). These pathogens represent species that currently have an important impact on agriculture. The pathogens were also matched with vertebrate pathogens of the same subdivision or genus also with known impact on vertebrates. Gene sequences of selected plant and vertebrate pathogens were identified and downloaded using the Kyoto Encyclopedia of Genes and Genomes (KEGG; http://www.genome.jp/kegg/). In total, genes were obtained from five vertebrate pathogens and nine plant pathogens (Table 3).

For alignment of the gene files and construction of phylogenetic trees, the program Clustal Omega was used (Sievers *et al.*, 2011; http://www.clustal.org/omega/). In addition, Codeml in PAML 4 v. 4.7 (Yang, 2007; http://abacus.gene.ucl.ac.uk/software/paml.html) was used to calculate the rate of evolution (ω) for all lineages and determine if positive selection was occurring in a specific lineage (the branch method) or within a specific gene (the sites model). For obtaining the visualized phylograms, the program TreeView X was used (Page, 1996; http://taxonomy.zoology.gla.ac.uk/rod/treeview.html), for which two files were needed. The first was a PAML control file to construct the tree, in which the model was set to 2 and the NS sites to 0. Once run, the output generated the information necessary to generate the tree file. In this tree file, the plant and vertebrate pathogen lineages were numbered as 1 or 2, respectively, in order to generate the correct phylogram on TreeView.

Several models were tested to determine what type of selection is operating and how the rate of evolution varies between the taxa. First, the ω value was forced to be equal for all lineages (i.e. the one ω model). The likelihood value was then compared to the model where ω is

calculated individually for each lineage (i.e. the free ω model). If there was significant heterogeneity in the rates, a third model was run (i.e. the plant ω vs. vertebrate ω model) where each lineage leading to a plant pathogen was given the same estimated ω value while those lineages leading to vertebrate pathogens were given their own estimated ω value. A significantly better fit with this model compared to the one ω value model, which was determined by a likelihood ratio test, would indicate that plant pathogens are evolving at a different rate than the vertebrate pathogens.

If there was no significant heterogeneity from the first comparison, the genes were also tested for positive selection. Three sites models were run; $\omega < 1$ (indicating negative selection), $\omega = 1$ (indicating neutral evolution) and $\omega > 1$ (indicating positive selection). If the positive selection model had significantly better fit than the other two models, then positive selection is operating in that gene.

Results

The results were analyzed to determine if the vertebrate immune system exerts a higher selection pressure on its respective pathogens as opposed to the plant immune system on its pathogens, and if this affects the rates of evolution of the pathogens' Type III Secretion System genes. The branch models were analyzed first. In Table 4, data under "branch" indicates significant differences in the rates of evolution of the specific T3SS genes with degrees of freedom. Non-significant values are denoted by "NS." The values were obtained from comparing the one ω model to the free ω model. They show that all genes had significant differences in the rates of evolution for the different lineages except for *YscF* and *YscW*.

Table 5 indicates the values obtained by comparing the ω values of vertebrate and plant pathogens. In this case, the genes that had significant differences in the rates of evolution are *YscJ*, *YscT*, and *YscV*. Table 6 compares the ω values of ancestral, vertebrate, and plant pathogen branches based on the results of Table 5. Figures 1-12 depict the phylograms that resulted from this analysis. Longer branches on the phylograms indicate higher evolutionary rates of the respective gene in the plant and vertebrate pathogen lineages.

Both Tables 7 and 8 show results of the sites model. In Table 7, data under "sites" indicates the number of changes in codons that have occurred within the various T3SS genes under $\omega < 1$ vs. $\omega = 1$. Table 8 shows the same information, but instead the values were obtained under $\omega < 1$ vs. $\omega = 1$ vs. $\omega > 1$. Both of these tables show that while the sites vary between $\omega < 1$ and $\omega = 1$, only *YscW* was significant for positive selection ($\omega > 1$).

Discussion

The results of the branch method indicate that all genes of the Type III Secretion System except *YscF* and *YscW* are undergoing differential selection. Further analysis of these results demonstrate that, in *YscJ*, *YscT*, and *YscV*, differences in branch ω values among plant, vertebrate, and ancestral lineages exist. As shown in Table 6, the ω value of the ancestral branches in these three genes is higher than the ω values of vertebrate and plant pathogens. This suggests that there was possible adaptive evolution in the ancestral lineages in each of these three genes, a result that was unexpected since we anticipated to observe higher ω values in diverged braches of vertebrate pathogen lineages.

On the other hand, results of the sites method indicate the only T3SS gene undergoing positive selection within a species is *YscW*. This gene is required for the proper functioning of

the *YscC* complex, which is a stable ring-like structure found in the outside portion of the T3SS that serves for both structural support and transport of macromolecules (Bi *et al*., 2009). Additionally, *YscW* is characterized as an outer membrane protein. These type of proteins, some of which are part of the T3SS, are essential for pathogenicity in Gram-negative bacteria by secreting adhesins, toxins, and immunomodulatory compounds (i.e., compounds that affect host immune responses). Once secreted, these compounds allow for binding and invasion to host cells, promote cytotoxicity, and modulate host immune responses (Kuehn and Kesty, 2005). Due to its presence on the exterior of the bacterial cell and its interaction with *YscC*, *YscW* undergoing positive selection makes sense, as this protein is constantly exposed to the environment and under pressure to penetrate host cells.

CHAPTER IV

ANALYSIS OF CONSERVED GENES IN PLANT AND VERTEBRATE PATHOGENS TO DETERMINE DIFFERENCES IN EVOLUTIONARY RATES

The “housekeeping” genes of individual bacterial pathogens were analyzed in order to determine their evolutionary rates. Housekeeping genes perform essential functions necessary to maintain cellular functions, such as transcription, translation, and signaling (Butte *et al.*, 2001). Due to their key functions, housekeeping genes are usually conserved within populations. This is opposite to the expected evolutionary rates of pathogenicity genes, which are generally known to have high rates of evolution, possibly due to their interaction with the host immune system. However, observed differences in the mutation rates of conserved housekeeping genes in plant and vertebrate pathogens could suggest that differences in selection exerted by the pathogens’ respective host immune systems exist.

To determine the evolutionary rates, dN/dS (i.e. ω or Ka/Ks) ratios will be calculated. The ratios are obtained by dividing the number of amino acid changes per site over the number of silent changes per site in each of the genes. If the ratio is less than one, the gene is undergoing negative (purifying) selection, which removes deleterious alleles. If the ratio is equal to one, the gene is undergoing neutral evolution or genetic drift. And if the ratio is greater than one, the gene is undergoing positive selection. In previous studies such as Lieberman *et al*. (2011), it has been demonstrated that pathogenicity genes have a dN/dS ratio greater than one, suggesting the genes

are undergoing positive selection. On the other hand, studies such as Guo *et al*. (2008) have demonstrated that housekeeping genes have dN/dS ratios less than one, which indicates they are undergoing purifying selection.

Observed differences in the evolutionary rates of conserved housekeeping genes in plant and vertebrate pathogens could suggest that differences in selection exerted by the pathogens' respective host immune systems exist. Additionally, the differences in the evolutionary rates of the housekeeping genes between plant and vertebrate pathogens may be exhibited in individual genes undergoing positive selection within a species or in different species undergoing differential selection.

Materials and Methods

The 2004 study performed by Gil *et al.* was used as a guide in selecting bacterial housekeeping genes. The gene sequences of selected plant and vertebrate pathogens were identified and downloaded based on their availability on the National Center for Biotechnology Information (NCBI) online database (http://www.ncbi.nlm.nih.gov/). In total, 10 genes from various housekeeping categories were obtained (Table 9) from the same five vertebrate pathogens and nine plant pathogens used in the previous study (Table 3).

For alignment of the gene files and construction of phylogenetic trees, the program Clustal Omega was used (Sievers *et al.*, 2011; http://www.clustal.org/omega/). In addition, Codeml in PAML 4 v. 4.7 (Yang, 2007; http://abacus.gene.ucl.ac.uk/software/paml.html) was used to calculate the rate of evolution (ω) for all lineages and determine if positive selection was occurring in a specific lineage (the branch method) or within a specific gene (the sites model).

For obtaining the visualized phylograms, the program TreeView X was used (Page, 1996; http://taxonomy.zoology.gla.ac.uk/rod/treeview.html), for which two files were needed. The first was a PAML control file to construct the tree, in which the model was set to 2 and the NS sites to 0. Once run, the output generated the information necessary to generate the tree file. In this tree file, the plant and vertebrate pathogen lineages were numbered as 1 or 2, respectively, in order to generate the correct phylogram on TreeView.

Several models were tested to determine what type of selection is operating and how the rate of evolution varies between the taxa. First, the ω value was forced to be equal for all lineages (i.e. the one ω model). The likelihood value was then compared to the model where ω is calculated individually for each lineage (i.e. the free ω model). If there was significant heterogeneity in the rates, a third model was run (i.e. the plant ω vs. vertebrate ω model) where each lineage leading to a plant pathogen was given the same estimated ω value while those lineages leading to vertebrate pathogens were given their own estimated ω value. A significantly better fit with this model compared to the one ω value model, which was determined by a likelihood ratio test, would indicate that plant pathogens are evolving at a different rate than the vertebrate pathogens.

If there was no significant heterogeneity from the first comparison, the genes were also tested for positive selection. Three sites models were run; $\omega < 1$ (indicating negative selection), $\omega = 1$ (indicating neutral evolution) and $\omega > 1$ (indicating positive selection). If the positive selection model had significantly better fit than the other two models, then positive selection is operating in that gene.

Results

The results were analyzed to determine if the vertebrate immune system exerts a higher selection pressure on its respective pathogens as opposed to the plant immune system on its pathogens, and if this affects the rates of evolution of the pathogens' housekeeping genes. The branch models were analyzed first. In Table 10, data under "branch" indicates significant differences in the rates of evolution of the specific housekeeping genes with degrees of freedom. Non-significant values are indicated by "NS." The values were obtained from comparing the one ω model to the free ω model. They show the genes that had significant differences in the rates of evolution for the different lineages, except for *ssb*, *nusG*, *rpmF*, *rpmI*, and *coaD*.

Table 11 indicates the values obtained by comparing the ω values of vertebrate and plant pathogens. In this case, the only gene that had a significant difference in the rate of evolution was *atpC*. Table 12 compares the ω values of ancestral, vertebrate, and plant pathogen branches based on the results of Table 11. Figures 13-22 depict the phylograms that resulted from this analysis. Longer branches on the phylograms indicate higher evolutionary rates of the respective gene in the plant and vertebrate pathogen lineages.

Tables 13 and 14 show results of the sites model. In Table 13, data under "sites" indicates the number of changes in codons that have occurred within the various housekeeping genes under $\omega < 1$ vs. $\omega = 1$. Table 14 shows the same information, but instead the values were obtained under $\omega < 1$ vs. $\omega = 1$ vs. $\omega > 1$. Both tables shows that while the sites vary between $\omega < 1$ and $\omega = 1$, the genes that were significant for positive selection ($\omega > 1$) are *ssb*, *nusG*, *rpmG*, *rpmI*, *atpC*, and *folA*.

Discussion

Analysis of the results of the branch method indicates that the genes undergoing differential selection are *greA*, *rpmG*, *secE*, *atpC*, and *folA*. However, in *atpC*, differences in branch ω values among plant, vertebrate, and ancestral lineages exist. As shown in Table 12, both ω values of vertebrate and plant pathogens are much higher than that of the ancestral branches. Though this may suggest this gene is undergoing positive selection in both plant and vertebrate pathogen lineages, to determine with certainty whether this is occurring, a sites-branch test would need to be performed. This type of test combines the methodology of both the branch method and the sites method of analyzing the evolutionary rates of specific genes in order to reach more conclusive results.

On the other hand, results of the sites method indicate six house housekeeping genes are undergoing positive selection within a species: *ssb*, *nusG*, *rpmG*, *rpmI*, *folA*, and *atpC*. From these, *folA* codes for the enzyme dihydrofolate reductase. Inhibition of this enzyme is known to lead to a deficiency in thymidylate, which subsequently leads to a disruption of DNA synthesis (Zolli-Juran *et al.*, 2003). Secondly, *ssb* (single-strand binding protein) primarily functions to stabilize and prevent premature re-annealing of single-stranded DNA during replication. Both *rpmG* and *rpmI* are 50S ribosomal genes, which are part of the larger ribosomal subunit in prokaryotes and generally known to be highly conserved. Finally, *nusG* is responsible for various roles during the regulation of translation and antitermination (Steiner *et al.*, 2002).

To determine why these conserved genes are undergoing positive selection, a possible route of further research could be to assess the evolutionary rates of these genes in non-pathogenic bacteria, which are typically not exposed to high selection pressure from a host. For example, both *folA* and *ssb* are involved in DNA processes. It is possible that in pathogenic

bacteria these genes are undergoing positive selection in order to have a beneficial increase in the rate of DNA replication, which would allow them to propagate faster via binary fission. In the case of *rpmG*, *rpmI*, and *nusG*, since these genes play essential roles during the process of protein synthesis, the fact that they are undergoing positive selection could suggest that the bacteria are adapting to synthesize proteins in a faster and more efficient way. Lastly, *atpC* undergoing positive selection could suggest that, being part of the ATP synthase, this gene is evolving possibly to provide more energy for the bacteria. Collectively, all these adaptations would be beneficial to a bacterial pathogen, as they are typically exposed to higher selection pressures than non-pathogenic bacteria.

CHAPTER V

DISCUSSION

Results from the study performed on housekeeping genes at the population level indicated that the mutation rate of plant pathogens is higher compared to that of vertebrate pathogens. Aside from a difference in the life histories of plant and vertebrate pathogens, other confounding factors to host immune system that could be playing a role in the evolution of the selected housekeeping genes would be the degree of virulence of the pathogens, as well as their population structure. For the last two studies, only pathogens with high degrees of virulence were selected; however, for the first study, the degree of virulence of the pathogens chosen could be assessed. This could be performed by comparing the evolutionary rate of these pathogenic bacteria to non-pathogenic bacteria in order to determine whether virulence plays a role in the evolution of housekeeping genes.

With regards to the population structure of the pathogens, Castillo and Greenberg (2007) demonstrated that this factor plays a role in the evolution of housekeeping genes in pathogenic bacteria. Their study revealed that five housekeeping genes of *Ralstonia solanacearum* are undergoing strong purifying selection due to geographic isolation events that have occurred in its population. Further analysis of the results from population study would need to be performed to determine if population dynamics is a possible driver for the evolution of the housekeeping genes selected. This type of analysis could provide results that are consistent with the idea that differential selection, rather than positive selection, drives host adaptation in different

housekeeping genes but more likely so in pathogens with wide host ranges (E. L. Schuenzel, unpublished).

At the species level, results suggest most genes seem to be undergoing differential and/or positive selection. In the Type III Secretion System, the ω values of ancestral lineages are higher that than of plant and vertebrate pathogen lineages in *YscJ*, *YscT*, and *YscV*, suggesting that there was possible adaptive evolution in the ancestral lineages. Among the housekeeping genes, a site-branch test needs to be performed on *atpC* to determine which specific lineages are undergoing positive selection due to the higher ω values observed in vertebrate and plant pathogen lineages as opposed to ancestral lineages. This test was not performed due to a lack of strong statistical power, for which at least 40 gene sequences would be needed and only 14 were assessed.

The genes that are undergoing positive selection within a species were *YscW* for the Type III Secretion System and the housekeeping genes *ssb*, *rpmG*, *rpmI*, *nusG*, *folA*, and *atpC*. From these results, *YscW* undergoing positive selection is consistent with the idea that genes exposed to the outside environment are under high selection pressure from the host immune system, and thus, are constantly evolving. However, for the housekeeping genes, the reasoning behind them undergoing positive selection seems more unclear. It is possible that the bacterial pathogens are adapting to high selection pressures by increasing the rate and efficacy of processes these genes are involved in, including DNA replication, translation, and energy synthesis, but further analyses are needed in order to assess the validity of this.

Finally, the results suggest that while only *YscW* is undergoing positive selection in the T3SS out of 12 genes that were studied, the majority of housekeeping genes of the plant and vertebrate pathogens selected are undergoing positive selection (six out of 10 analyzed). Further research is needed to determine why this is occurring in housekeeping genes that are known to be

conserved as opposed to pathogenicity genes that normally have elevated rates of evolution. For example, the positions of the genes undergoing positive selection could be assessed to determine if they are mutation hot spots, or areas that are susceptible to higher mutation rates compared to the rest of the genome, which would explain their unusually high rates of evolution. Another route of possible further research could be to determine if these housekeeping genes are part of an operon. Since genes in an operon are influenced by a promoter region that acts as an activator, the evolution rate of the promoter or other upstream genes that influence transcriptional rates could have an effect on the evolutionary rates of downstream genes as well. If it is determined that the conserved genes selected are part of an operon, this could explain why they are undergoing positive selection.

REFERENCES

Anderson, J. P., Gleason, C. A., Foley, R. C., Thrall, P. H., Burdon, J. B., and Singh, K. B. 2011. Plants versus pathogens: An evolutionary arms race. Funct. Plant Biol. 37(6): 499-512.

Bi, Y., Du, Z., Yang, H., Guo, Z., Tan, Y., Zhu, Z., and Yang, R. 2009. Reduced apoptosis of mouse macrophages induced by *yscW* mutant of *Yersinia pestis* from the reduced secretion of YopJ and relates to Caspase-3 signal pathway. Scand. J. Immunol. 70: 358-367.

Borghans, J. A. M., Beltman, J. B., and De Boer, R. J. 2004. MHC polymorphism under host-pathogen coevolution. Immunogenetics 55: 732-739.

Brodersen, P., Malinovsky, F. G., Hématy, K., Newman, M., and Mundy, J. 2005. The role of salicylic acid in the induction of cell death in Arabidopsis *acd*11. Plant Physiol. 138: 1037-1045.

Bryant, J., Chewapreecha, C., and Bentley, S. D. 2012. Developing insights into the mechanisms of evolution of bacterial pathogens from whole-genome sequences. Future Microbiol. 7(11): 1283-1296.

Burdon, J. J. and Thrall, P H. 2009. Co-evolution of plants and their pathogens in natural habitats. Science 324(5928): 755-756.

Butte, A. J., Dzau, V. J., and Glueck, S. B. 2001. Further defining housekeeping or “maintenance,” genes: Focus on “A compendium of gene expression in normal human tissues.” Physiol. Genomics 7: 95-96.

Büttner, D. and Bonas, U. 2003. Common infection strategies of plant and animal pathogenic bacteria. Curr. Opin. Plant Biol. 6: 312-319.

Castillo, J. and Greenberg, J. T. 2007. Evolutionary dynamics of *Ralstonia solanacearum*. Appl. Environ. Microbiol. 73(4): 1225-1238.

Cooper, M. A., Elliott, J. M., Keyel, P. A., Yang, L., Carrero, J. A., and Yokoyama, W. M. 2009. Cytokine-induced memory-like natural killer cells. PNAS: 106(6): 1915-1919.

Flor, H. H. 1971. Current status of the gene-for-gene concept. Annu. Rev. Phytopathol. 9: 680-685.

Galán, J. E. and Wolf-Watz, H. 2006. Protein delivery into eukaryotic cells by type III secretion machines. Nature 440(30): 567-573.

Germain, R. N. 2001. The art of the probable: System control in the adaptive immune system. Science 293(5528): 240-245.

Gil, R., Silva, F. J., Peretó, J., and Moya, A. 2004. Determination of the core of a minimal bacterial gene set. Microbiol. Mol. Biol. R. 68(3): 518-537.

Guo, Y., Zheng, W., Rong, X., and Huang, Y. 2008. A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: Use of multilocus sequence analysis for streptomycete systematics. Int. J. Syst. Evol. Micr. 58: 149-159.

Hawkey, P. M. and Jones, A. M. 2009. The changing epidemiology of resistance. J. Antimicrob. Chemoth. 64(1): i3-i-10.

Hughes, A.L. and Nei, M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335: 167-170.

Hughes, A.L. and Nei, M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. PNAS 86: 958-962.

Inouye, M., Conway, T. C., Zobel, J., and Holt, K. E. 2012. Short read sequence typing (SRST): Multi-locus sequence types from short reads. BMC Genomics 13(338): 1-7.

Janeway, C. A., Travers, P., Walport, M., and Shlomchik, M. J. 2005. Immunobiology, 6th ed. Garland Science.

Janse, J. D. and Obradovic, A. 2010. *Xyllela fastidiosa*: Its biology, diagnosis, control and risks. J. Plant Pathol. 92(1): S1.35-S1.48.

Jones, J. D. G. and Dangl, J. L. 2006. The plant immune system. Nature 444: 323-329.

Kuehn, M. J. and Kesty, N. C. 2005. Bacterial outer membrane vesicles and the host-pathogen interaction. Genes Dev. 19: 2645-2655.

Librado, P. and Rozas, J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452.

Mansfield, J., Genin, S., Magori, S., Citovsky, V., Sriariyanum, M., Ronald, P., Dow, M., Verdier, V., Beer, S. V., Machado, M. A., Toth, I., Salmond, G., and Foster, G. D. 2012. Top 10 plant pathogenic bacteria in molecular plant pathology. Mol. Plant Pathol. 13(6): 614-629.

Mueller, C. A., Broz, P., and Cornelis, G. R. 2008. The type III secretion system tip complex and translocon. Mol. Microbiol. 68(5): 1085-1095.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. 1999. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 27: 29-34.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12: 357-358.

Patel, S. M., Rahman, M. A., Mohasin, M., Riyadh, M. A., Leung, D. T., Alam, M. M., Chowdhury, F., Khan, A. I., Weil, A. A., Aktar, A., Nazim, M., LaRocque, R. C., Ryan, E. T., Calderwood, S. B., Qadri, F., and Harris, J. B. 2012. Memory B cell responses to *Vibrio cholerae* O1 lipopolysaccharide are associated with protection against infection from household contacts of patients with cholera in Bangladesh. Clin. Vaccine Immunol. 19(6): 842-848.

Rohmner, L., Guttman, D. S., and Dangl, J. L. 2004. Diverse evolutionary mechanisms shape the type III effector virulence factor repertoire in the plant pathogen *Pseudomonas syringae*. Genetics 167: 1341-1360.

Serres, M. H., Kerr, A. R. W., McCormack, T. J., and Riley, M. 2009. Evolution by leaps: Gene duplication in bacteria. Biol. Direct. 4(46): 1-17.

Sievers, F., Wilm, A., Dineen, D. G., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert. M., Söding, J., Thompson, J. D., and Higgins, D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7: 539.

Steiner, T., Kaiser, J. T., Marinkoviç, S., Huber, R., and Wahl, M. C. 2002. Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. EMBO J. 21(17): 4641-4653.

Lieberman, T. D., Michel, J., Aingaran, M., Potter-Bynoe, G., Roux, D., Davis Jr., M. R., Skurnik, D., Leiby, N., LiPuma, J. J., Goldberg, J. B., McAdam, A. J., Priebe, G. P., and Kishony, R. 2001. Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. Nat. Genet. 43(12): 1275-1280.

Treangen, T. J. and Rocha, E. P. C. 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. PLoS Genet. 7(1): 1-12.

Waters, V. J., Gómez, M. I., Soong, G., Amin, S., Ernst, R. K., and Prince, A. 2007. Immunostimulatory properties of the emerging pathogen *Stenotrophomonas maltophilia*. Infect. Immun. 75(4): 1698-1703.

Yang, Z. 2007. PAML 4: A program package for phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24: 1586-1591.

Zimmerman, L. M., Vogel, L. A., and Bowden, R. M. 2010. Understanding the vertebrate immune system: Insights from the reptilian perspective. J. Exp. Biol. 231: 661-671.

Zolli-Juran, M., Cechetto, J. D., Hartlen, R., Daigle, D. M., and Brown, E. D. 2003. High throughput screening identifies novel inhibitors of *Escherichia coli* dihydrofolate reductase that are competitive with dihydrofolate. Bioorgan. Med. Chem. Lett. 13: 2493-2496.

TABLES

Table 1. Information regarding the bacterial population study organisms

Pathogen Name	Database Obtained From	Important Hosts
Pseudomonas syringae	PAMDB	Over 200 plant species
Ralstonia solanacearum	PAMDB	Potato, tomato, eggplant, banana, ginger, tobacco, sweet pepper, olive
Xanthomonas campestris	(Provided by Dr. E. L. Schuenzel)	Cruciferous plants
Xyllela fastidiosa	PubMLST	Over 150 plant species
Campylobacter fetus	PubMLST	Humans, cattle, sheep
Campylobacter upsaliensis	PubMLST	Humans
Clostridium botulinum	PubMLST	Humans, cows, birds
Corynebacterium diphtheriae	PubMLST	Humans and animals
Haemophilus parasuis	PubMLST	Pigs
Pasteurella multocida	PubMLST	Humans, birds (including poultry), cats, dogs, rabbits, cattle, pigs
Porphyromonas gingivalis	PubMLST	Humans
Stenotrophomonas maltophilia	PubMLST	Humans
Streptococcus agalactiae	PubMLST	Humans, cattle, camels, dogs, cats, fish
Streptococcus dysgalactiae	MLST	Humans and cows
Vibrio vulnificus	PubMLST	Humans

Table 2. Results of t-tests with unequal variances comparing the population data obtained from plant and vertebrate pathogens

Variable	Plant	Vertebrate	P Values
Number of alleles	33.64	30.95	NS
Number of isolates	231.75	143.36	NS
Number of polymorphic sites	73.80	50.47	0.026325
Number of sites sampled	589.28	456.63	<.0001
Number of mutations	85.16	55.72	0.031385
Nucleotide diversity	0.0313	0.0247	NS
Number of recombinations	9.96	4.85	NS

Table 3. Information regarding the selected pathogens with Type III Secretion Systems

Pathogen Name	Important Hosts
Erwinia amylovora ATCC 49946	Apple and pear
Dickeya dadantii Ech586	Pepper, potato, eggplant, tomato, tobacco, broccoli, radishes, celery, carrot, sugar cane, rice, pineapple, asparagus, onions
Pectobacterium carotovorum subsp. *carotovorum*	Potato
Pseudomonas syringae pv. *tomato* DC3000	Tomato
Ralstonia solanacearum GMI1000	Potato, tomato, eggplant, banana, ginger, tobacco, sweet pepper, olive
Xanthomonas campestris pv. *campestris* ATCC 33913	Citrus
Xanthomonas axonopodis pv. *citri* 306	Citrus
Xanthomonas oryzae pv. *oryzae* KACC 10331	Rice
Xanthomonas oryzae pv. *oryzicola*	Rice
Escherichia coli O157 H7 Sakai	Humans and animals
Pseudomonas aeruginosa PAO1	Humans and animals
Salmonella enterica subsp. Enterica serovar Typhi Ty2	Humans, cats, hamsters
Shigella flexneri 301 (serotype 2a)	Humans and primates
Yersinia pestis Angola	Humans and animals

Table 4. Significant differences in the rate of evolution (one ω vs. free ω) of the Type III Secretion System genes with degrees of freedom

Branch Model (One ω vs. Free ω)	
Gene	**Branch**
YscF	NS
YscC	170.2[a], 22[b]
YscW	NS
YscJ	164.3, 24
YscR	87.9, 24
YscS	42.9, 24
YscT	484.7, 24
YscU	130.2, 24
YscV	344.5, 24
YscN	125.8, 24
YscQ	35.9, 22
YscL	56.8, 18

[a] Likelihood ratio statistic

[b] Degrees of freedom

Table 5. Significant differences in the rate of evolution (ω vertebrate vs. ω plant) of the Type III Secretion System genes with degrees of freedom

Branch Model (ω Vertebrate vs. ω Plant)	
Gene	**Branch**
YscF	NS
YscC	NS
YscW	NS
YscJ	136.0[a], 29[b]
YscR	NS
YscS	NS
YscT	410.5, 29
YscU	NS
YscV	198.5, 29
YscN	NS
YscQ	NS
YscL	NS

[a] Likelihood ratio statistic
[b] Degrees of freedom

Table 6. ω values of ancestral, vertebrate, and plant pathogen branches of the Type III Secretion System genes

ω Values			
Gene	**ω Ancestral**	**ω Vertebrate**	**ω Plant**
YscF	0.0008	0.703	0.0591
YscC	0.5092	0.1354	0.1406
YscW	0.4663	2.0727	1.1549
YscJ	**4.0821**	**0.0068**	**0.0342**
YscR	0.0439	0.0035	0.0481
YscS	0.0286	0.1648	0.0358
YscT	**6.3069**	**0.0466**	**0.0254**
YscU	0.103	0.254	0.1246
YscV	**0.6857**	**0.0751**	**0.1067**
YscN	0.1853	0.0449	0.0754
YscQ	0.1385	0.0135	0.1584
YscL	0.1208	0.0024	0.1209

(Bold values denote significant differences)

Table 7. Number of changes within codons (ω < 1 or ω = 1) of the Type III Secretion System genes with degrees of freedom

Sites Model (ω < 1 vs. ω = 1)	
Gene	**Sites**
YscF	38.9[a], 1[b]
YscC	293.5, 1
YscW	10.3, 1
YscJ	88.6, 1
YscR	217.3, 1
YscS	8.5, 1
YscT	1008.7, 1
YscU	293.5, 1
YscV	935.5, 1
YscN	634.4, 1
YscQ	217.3, 1
YscL	6.5, 1

[a] Likelihood ratio statistic
[b] Degrees of freedom

Table 8. Number of changes within codons ($\omega < 1$, $\omega = 1$, $\omega > 1$) of the Type III Secretion System genes with degrees of freedom

Sites Model ($\omega < 1$ vs. $\omega = 1$ vs. $\omega > 1$)	
Gene	**Sites**
YscF	NS
YscC	NS
YscW	13.7[a], 2[b]
YscJ	NS
YscR	NS
YscS	NS
YscT	NS
YscU	NS
YscV	NS
YscN	NS
YscQ	NS
YscL	NS

[a] Likelihood ratio statistic

[b] Degrees of freedom

Table 9. Housekeeping genes selected

Genes used by category
DNA metabolism
ssb (SSB)
RNA metabolism
greA (Transcription elongation factor) *nusG* (Transcription antitermination protein) *rpmF* (50S ribosomal protein L32) *rpmG* (50S ribosomal protein L33) *rpmI* (50S ribosomal protein L35)
Protein processing, folding, and secretion
secE (Membrane-embedded preprotein translocase subunit)
Energy and intermediary metabolism
atpC (ATP synthase ε chain) *coaD* (4'-Phosphopantetheine adenylyltransferase) *folA* (Dihydrofolate reductase)

Table 10. Significant differences in the rate of evolution (one ω vs. free ω) of the housekeeping genes with degrees of freedom

Branch Model (One ω vs. Free ω)	
Gene	**Branch**
ssb	NS
greA	80.3[a], 24[b]
nusG	NS
rpmF	NS
rpmG	53.9, 24
rpmI	NS
secE	44.5, 24
atpC	291.5, 24
coaD	NS
folA	58.1, 24

[a] Likelihood ratio statistic
[b] Degrees of freedom

Table 11. Significant differences in the rate of evolution (ω vertebrate vs. ω plant) of the housekeeping genes with degrees of freedom

Branch Model (ω Vertebrate vs. ω Plant)	
Gene	**Branch**
ssb	NS
greA	NS
nusG	NS
rpmF	NS
rpmG	NS
rpmI	NS
secE	NS
atpC	151.2[a], 29[b]
coaD	NS
folA	NS

[a] Likelihood ratio statistic
[b] Degrees of freedom

Table 12. ω values of ancestral, vertebrate, and plant pathogen branches of the housekeeping genes

ω Values			
Gene	**ω Ancestral**	**ω Vertebrate**	**ω Plant**
ssb	1.6715	1.8596	2.5924
greA	0.1097	0.0506	0.0545
nusG	3.1956	6.3273	5.5617
rpmF	0.1547	0.0608	0.1371
rpmG	0.2744	0.0225	0.0233
rpmI	2.3297	2.5669	5.5533
secE	0.0983	0.0533	0.0259
atpC	**0.6256**	**20.4888**	**12.1701**
coaD	0.0652	0.0418	0.0473
folA	1.50834	3.8185	1.8832

(Bold values denote significant differences)

Table 13. Number of changes within codons ($\omega < 1$ or $\omega = 1$) of the housekeeping genes with degrees of freedom

Sites Model ($\omega < 1$ vs. $\omega = 1$)	
Gene	**Sites**
ssb	75.5[a], 1[b]
greA	148.2, 1
nusG	48.4, 1
rpmF	68.1
rpmG	44.3, 1
rpmI	41.9, 1
secE	19.8, 1
atpC	575.6, 1
coaD	21.7, 1
folA	67.4, 1

[a] Likelihood ratio statistic
[b] Degrees of freedom

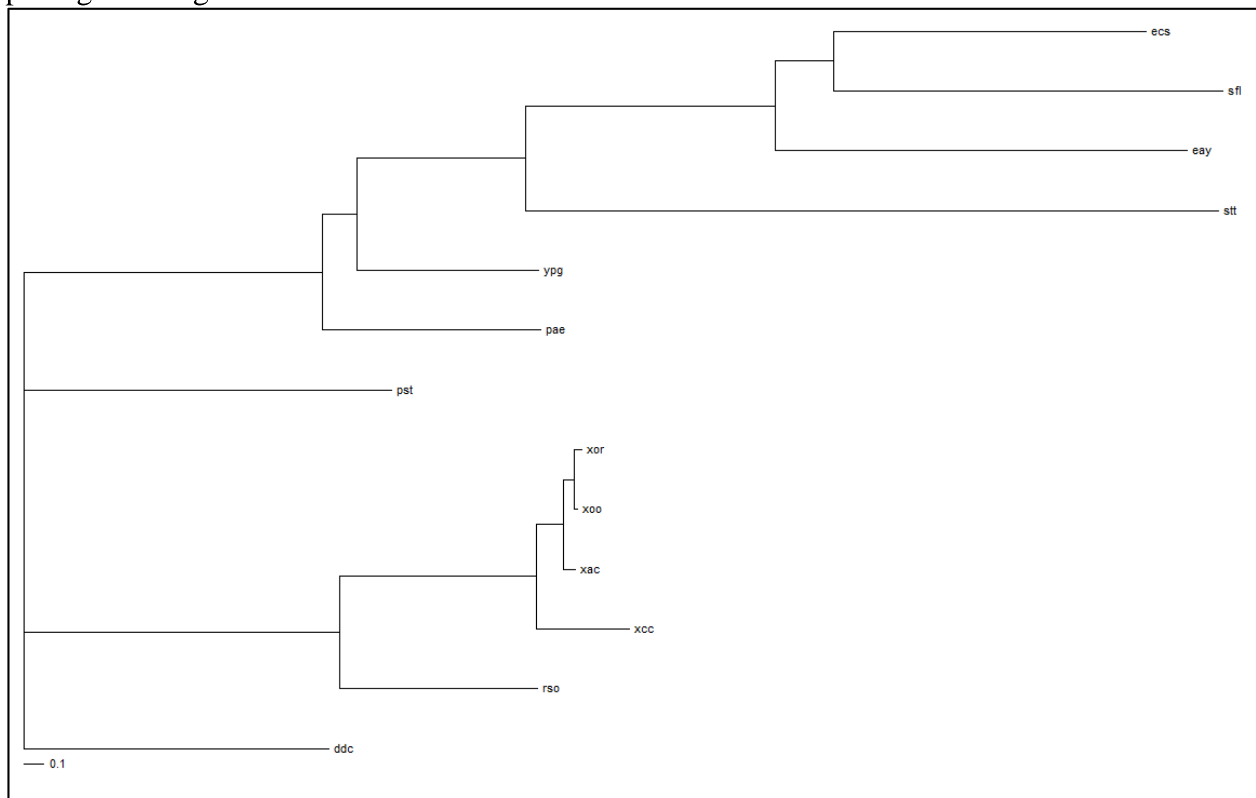
Table 14. Number of changes within codons (ω < 1, ω = 1, ω > 1) of the housekeeping genes with degrees of freedom

Sites Model (ω < 1 vs. ω = 1 vs. ω > 1)	
Gene	**Sites**
ssb	101.3[a], 2[b]
greA	NS
nusG	188.5, 2
rpmF	NS
rpmG	10.4, 2
rpmI	77.1, 2
secE	NS
atpC	505.1, 2
coaD	NS
folA	88.3, 2

[a] Likelihood ratio statistic
[b] Degrees of freedom

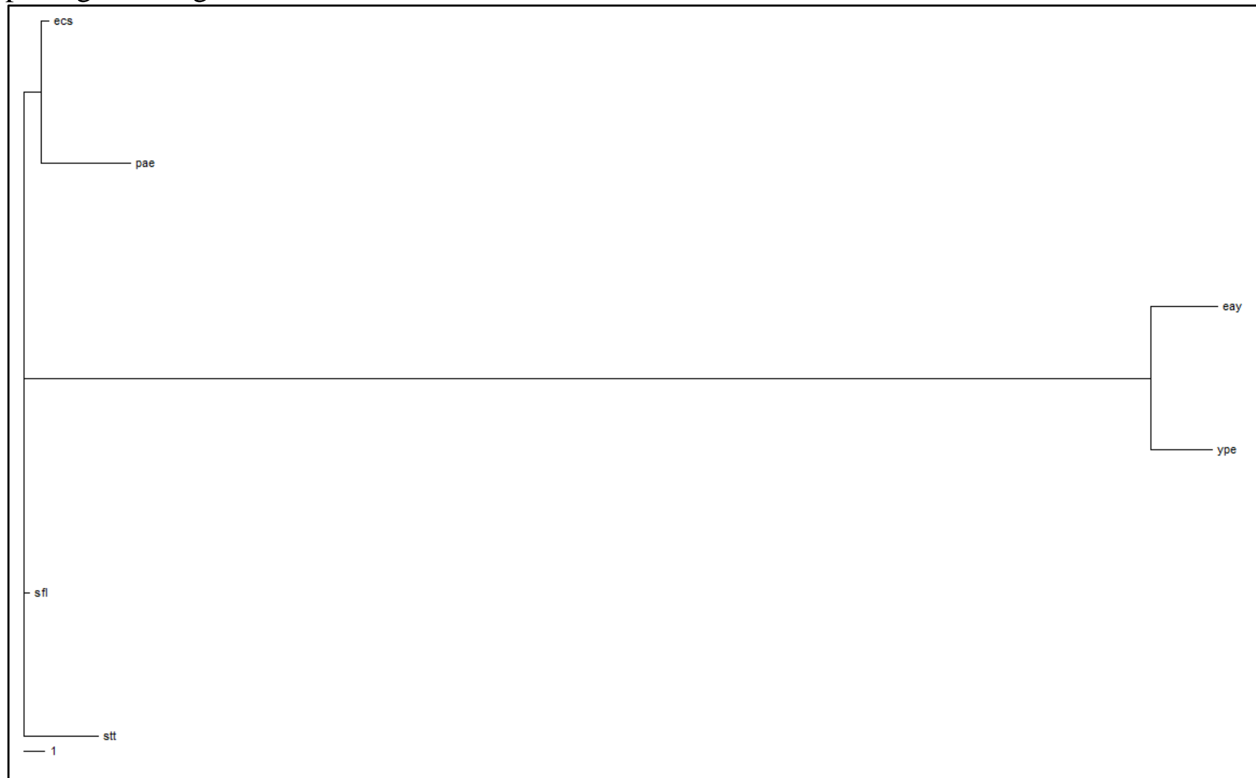
FIGURES

Figure 1. Rate of evolution of the *YscC* Type III Secretion System gene in plant and vertebrate pathogen lineages



- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- Significance was observed for differential selection

Figure 2. Rate of evolution of the *YscF* Type III Secretion System gene in plant and vertebrate pathogen lineages

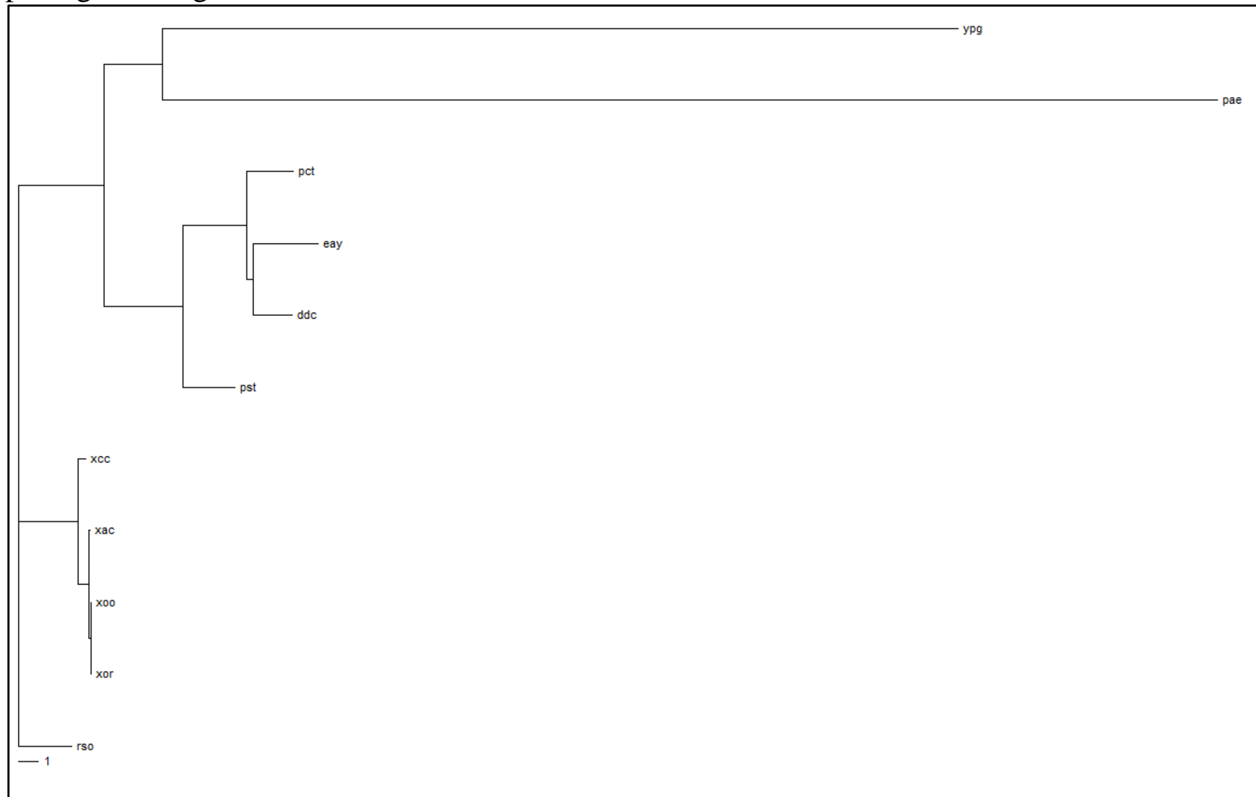


- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- No significance was observed for any type of selection

Figure 3. Rate of evolution of the *YscJ* Type III Secretion System gene in plant and vertebrate pathogen lineages

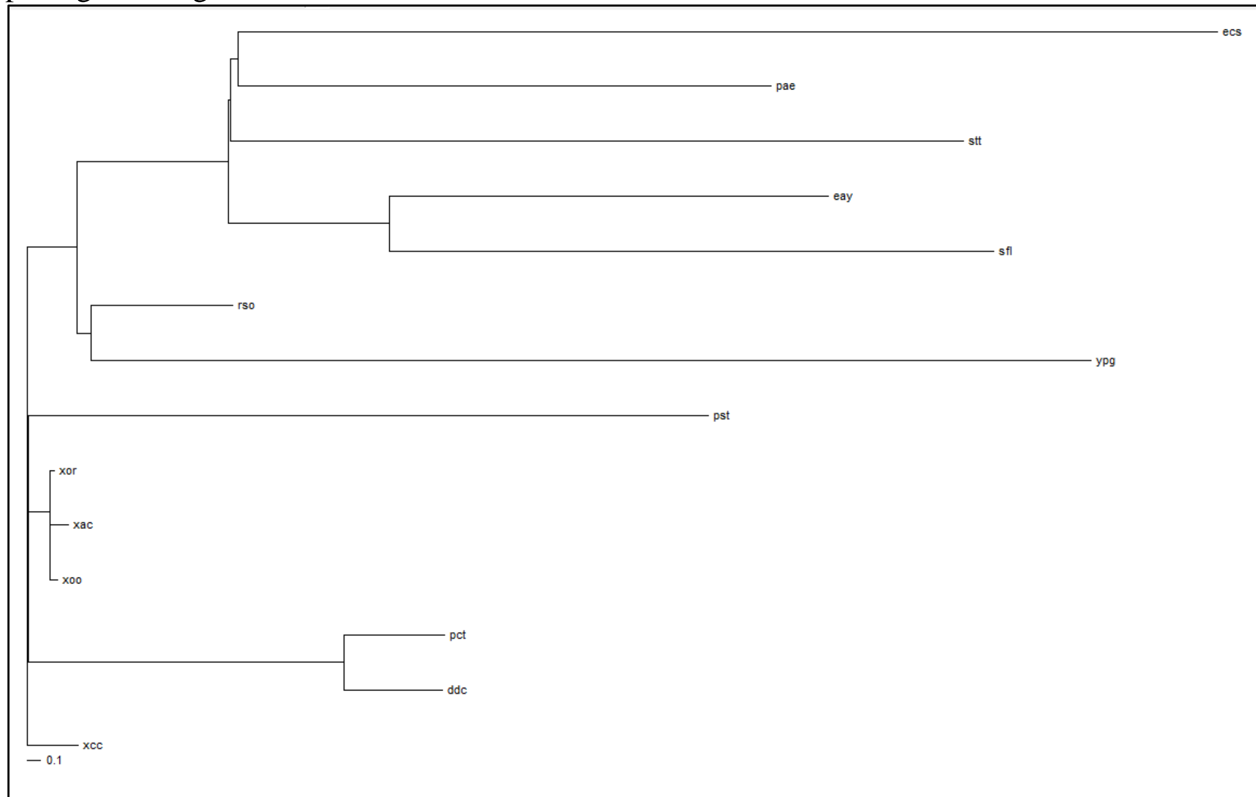
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- Significance was observed for higher differential selection in ancestral lineages

Figure 4. Rate of evolution of the *YscL* Type III Secretion System gene in plant and vertebrate pathogen lineages



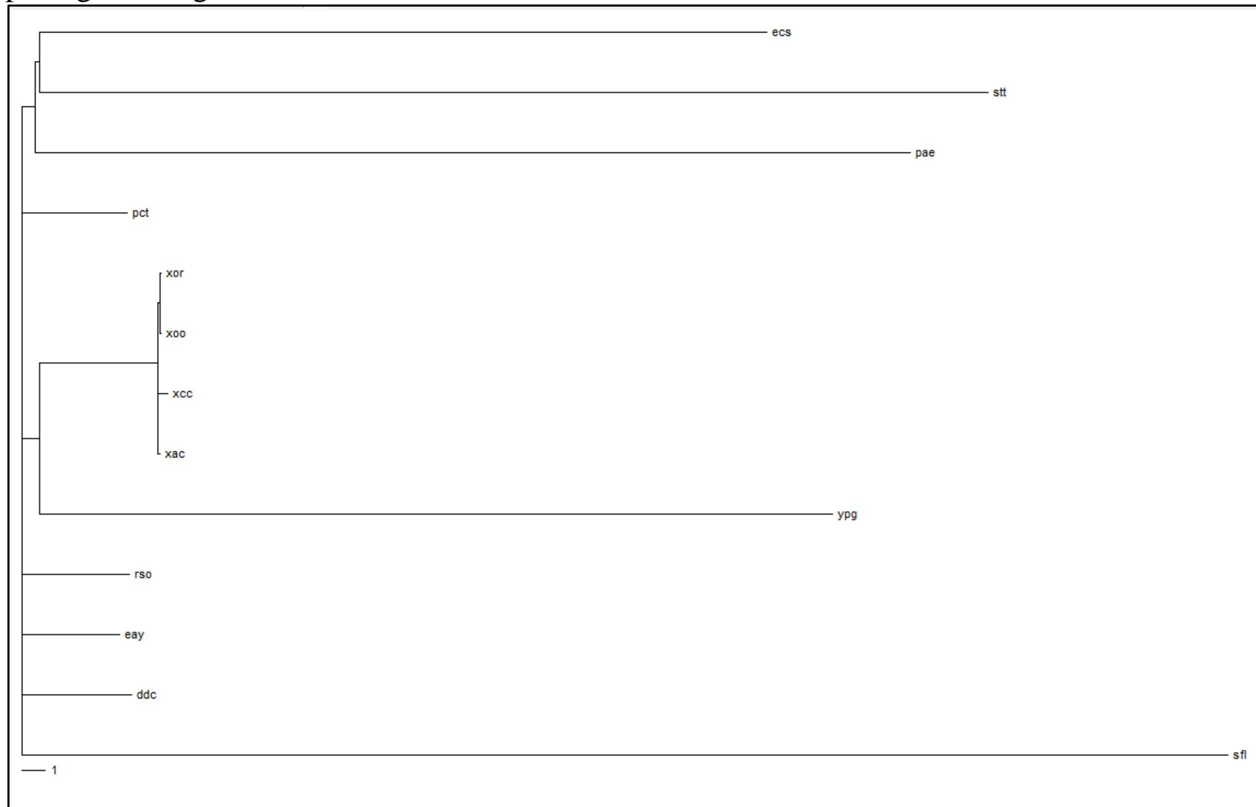
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- Significance was observed for differential selection

Figure 5. Rate of evolution of the *YscN* Type III Secretion System gene in plant and vertebrate pathogen lineages



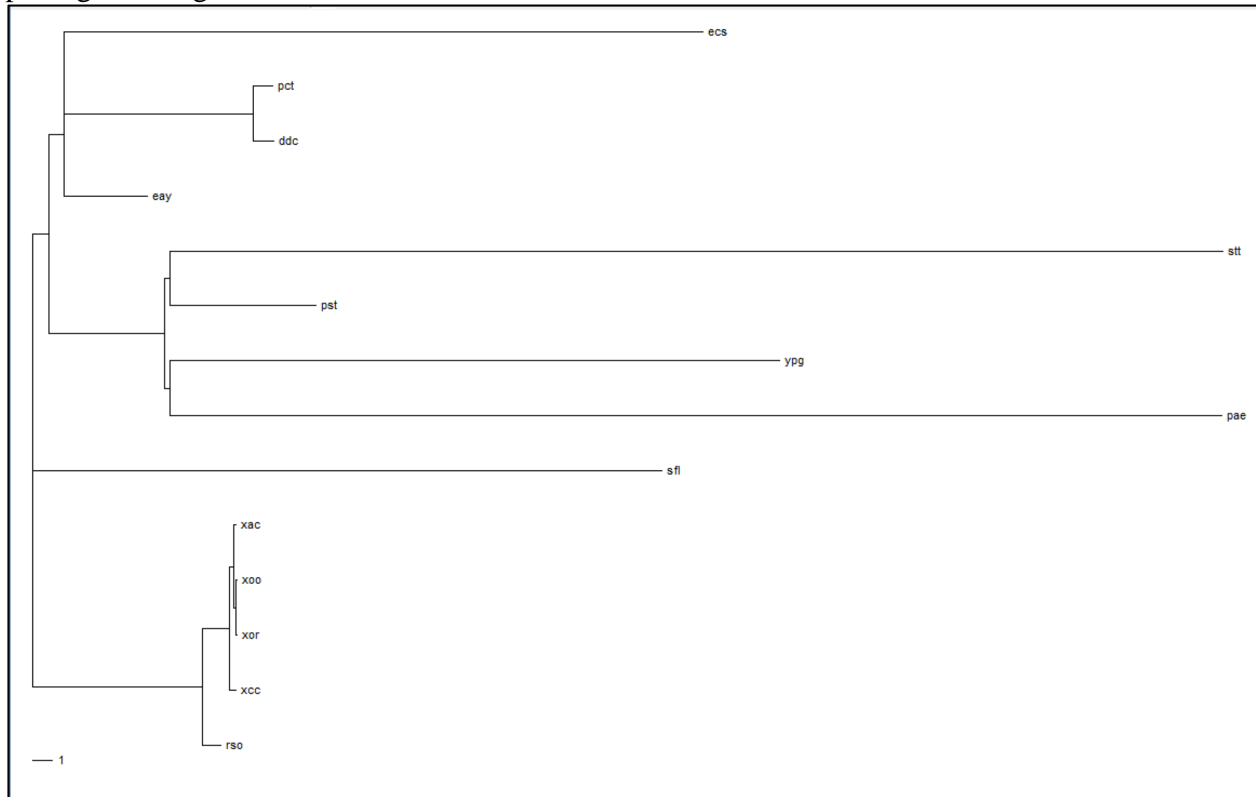
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- Significance was observed for differential selection

Figure 6. Rate of evolution of the *YscQ* Type III Secretion System gene in plant and vertebrate pathogen lineages



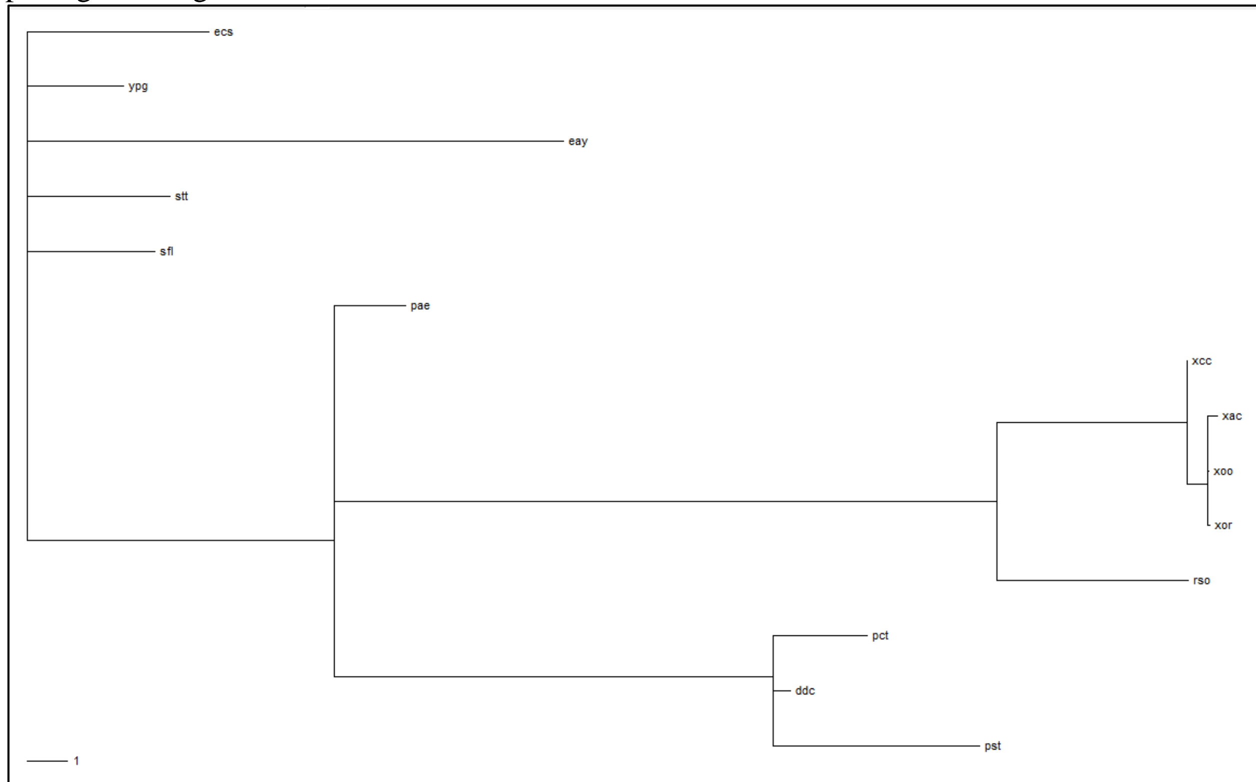
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- Significance was observed for differential selection

Figure 7. Rate of evolution of the *YscR* Type III Secretion System gene in plant and vertebrate pathogen lineages



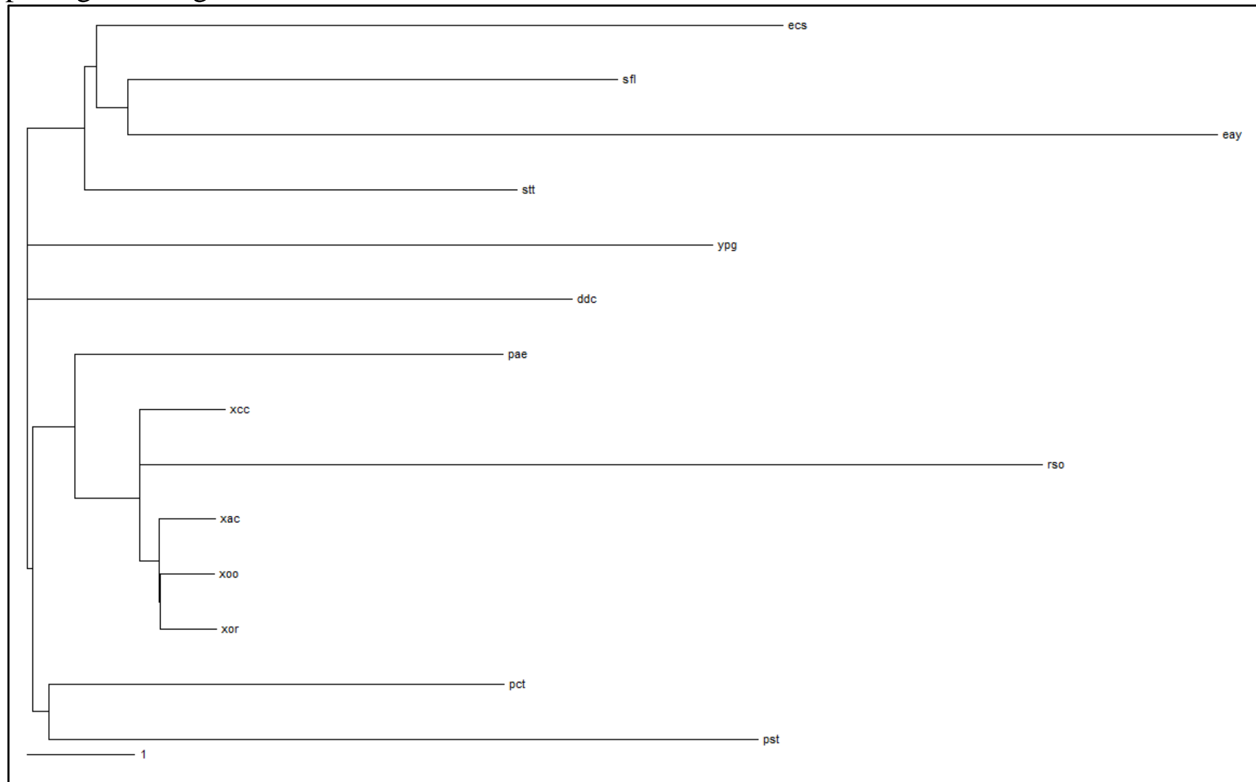
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- Significance was observed for differential selection

Figure 8. Rate of evolution of the *YscS* Type III Secretion System gene in plant and vertebrate pathogen lineages



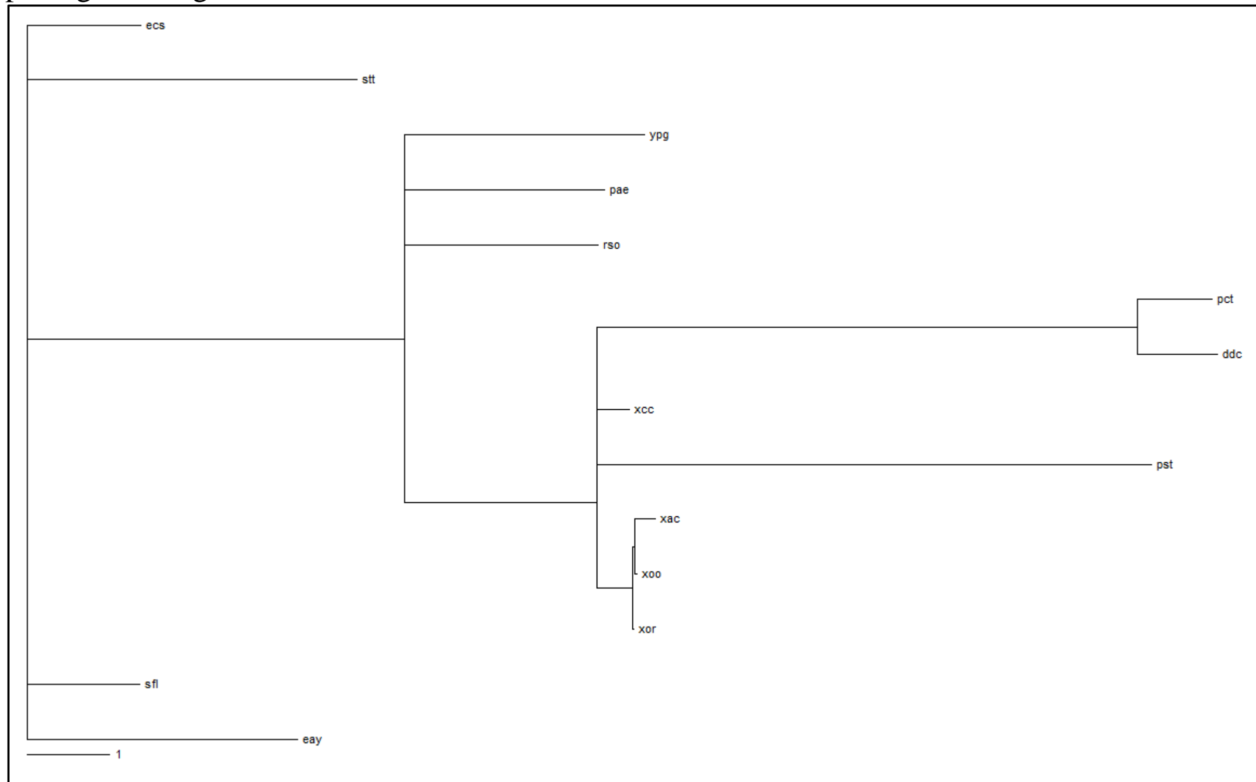
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- Significance was observed for differential selection

Figure 9. Rate of evolution of the *YscT* Type III Secretion System gene in plant and vertebrate pathogen lineages



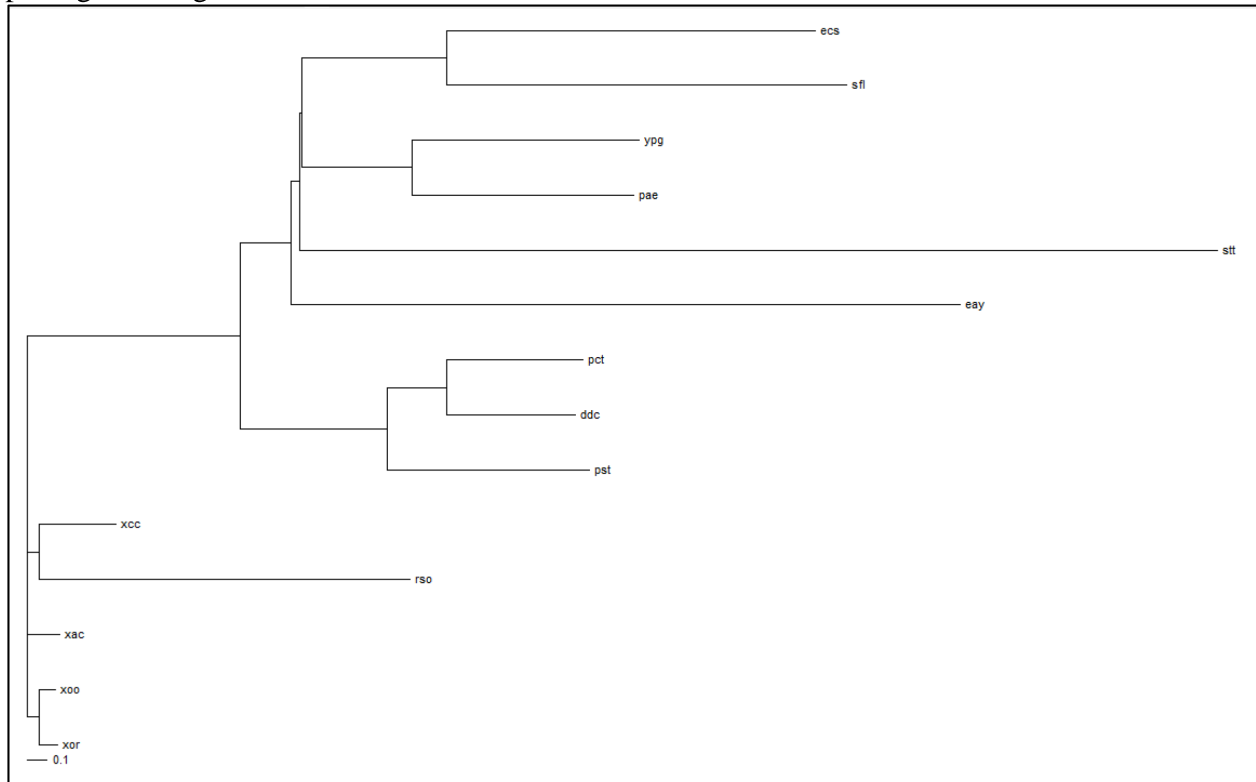
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- Significance was observed for higher differential selection in ancestral lineages

Figure 10. Rate of evolution of the *YscU* Type III Secretion System gene in plant and vertebrate pathogen lineages



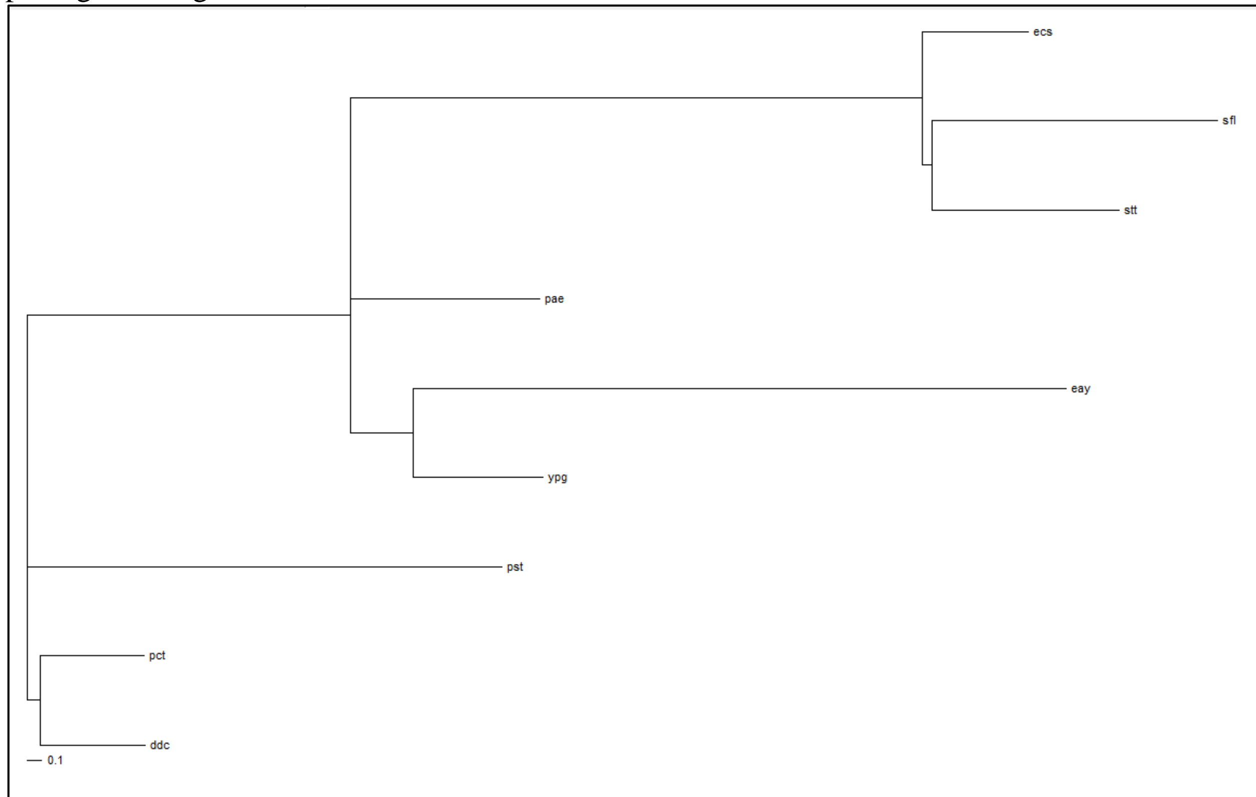
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- Significance was observed for differential selection

Figure 11. Rate of evolution of the *YscV* Type III Secretion System gene in plant and vertebrate pathogen lineages



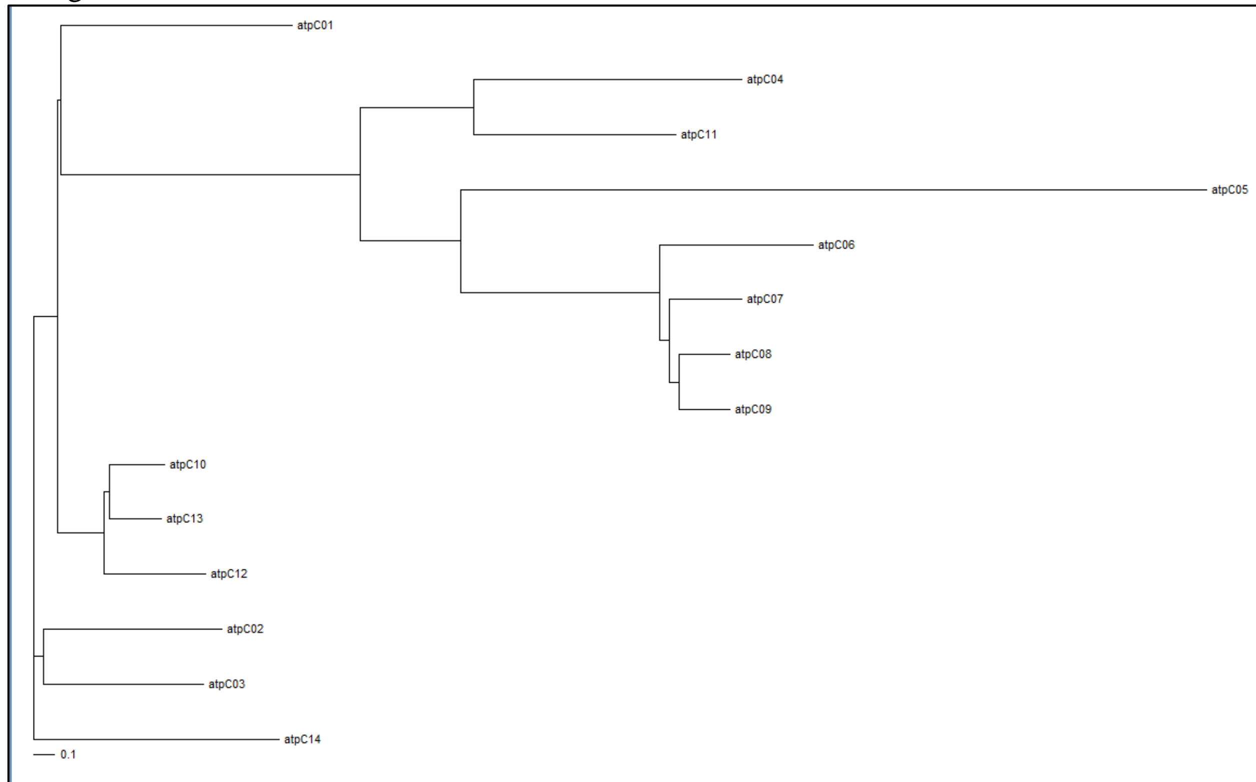
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- Significance was observed higher differential selection in ancestral lineages

Figure 12. Rate of evolution of the *YscW* Type III Secretion System gene in plant and vertebrate pathogen lineages



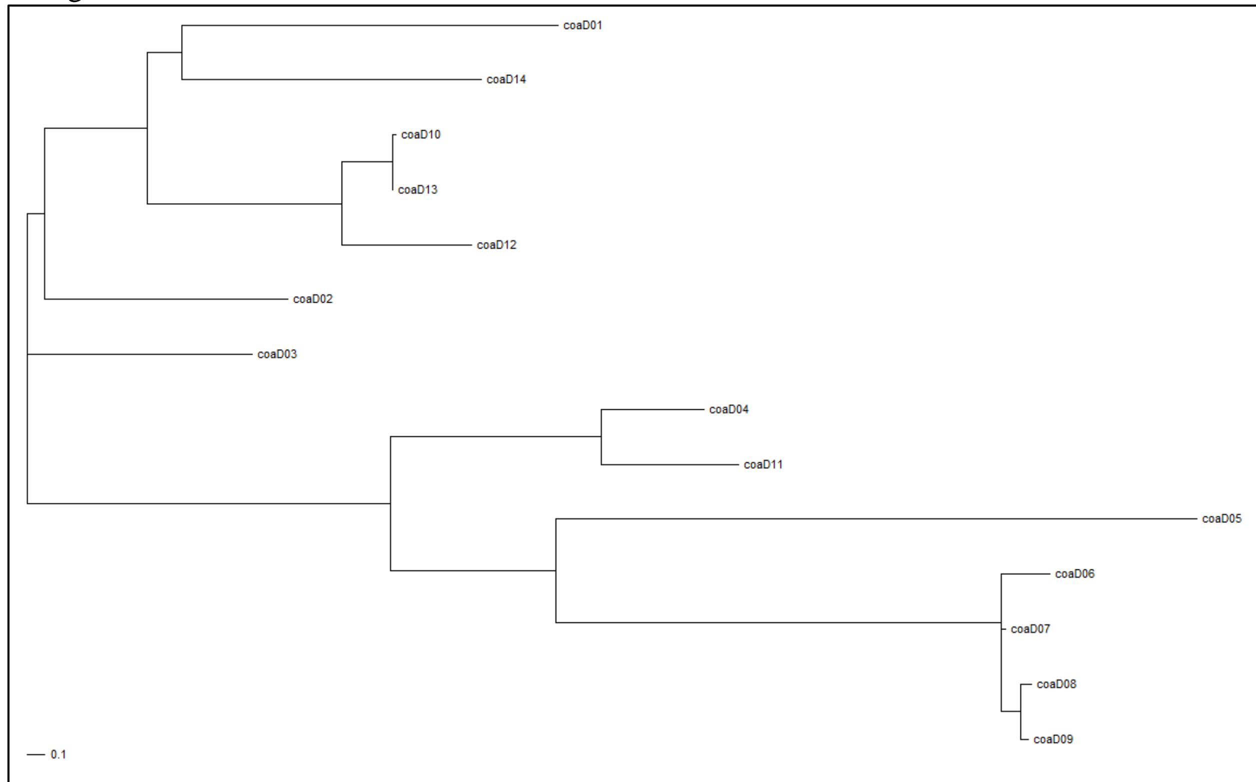
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: pst, xor, xac, rso, eay, pct, ddc, xcc, xoo
- Vertebrate pathogen lineages: ecs, ypg, pae, sfl, stt
- Significance was observed for positive selection

Figure 13. Rate of evolution of the *atpC* housekeeping gene in plant and vertebrate pathogen lineages



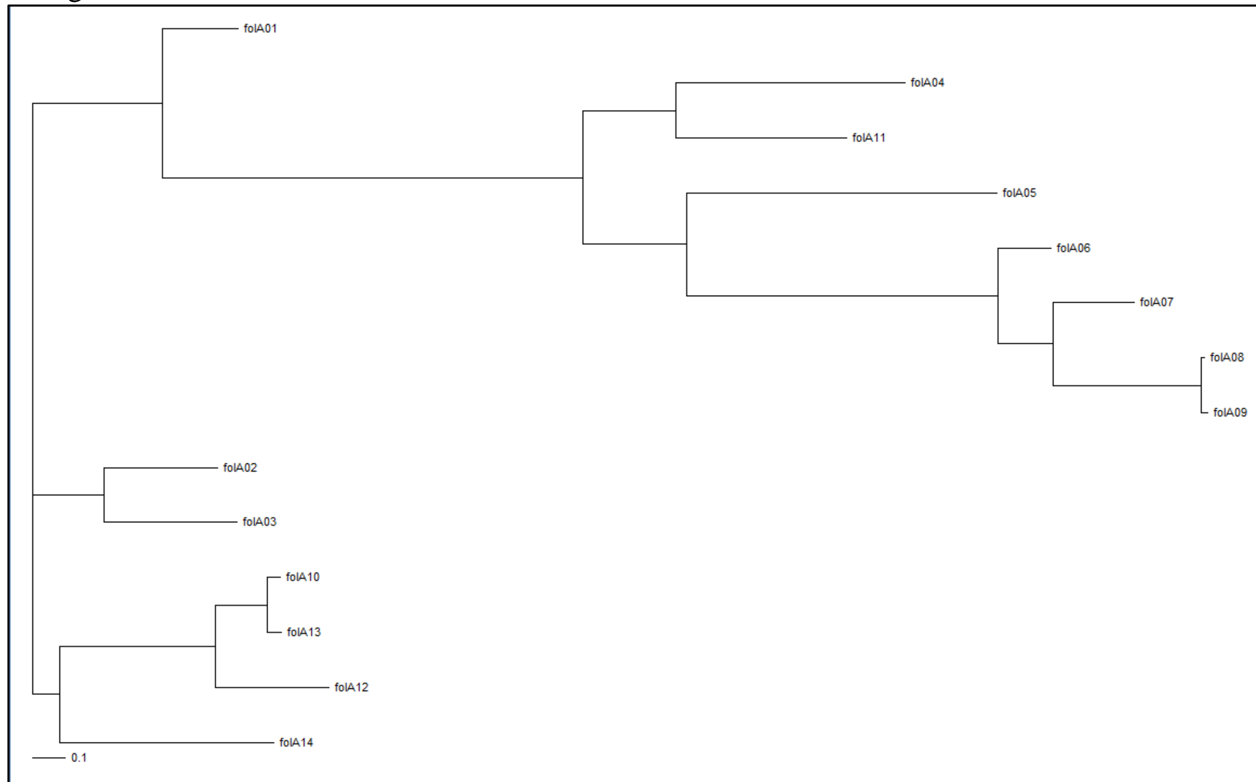
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: 01-09
- Vertebrate pathogen lineages: 10-14
- Significance was observed for positive selection and higher differential selection in vertebrate and plant lineages

Figure 14. Rate of evolution of the *coaD* housekeeping gene in plant and vertebrate pathogen lineages



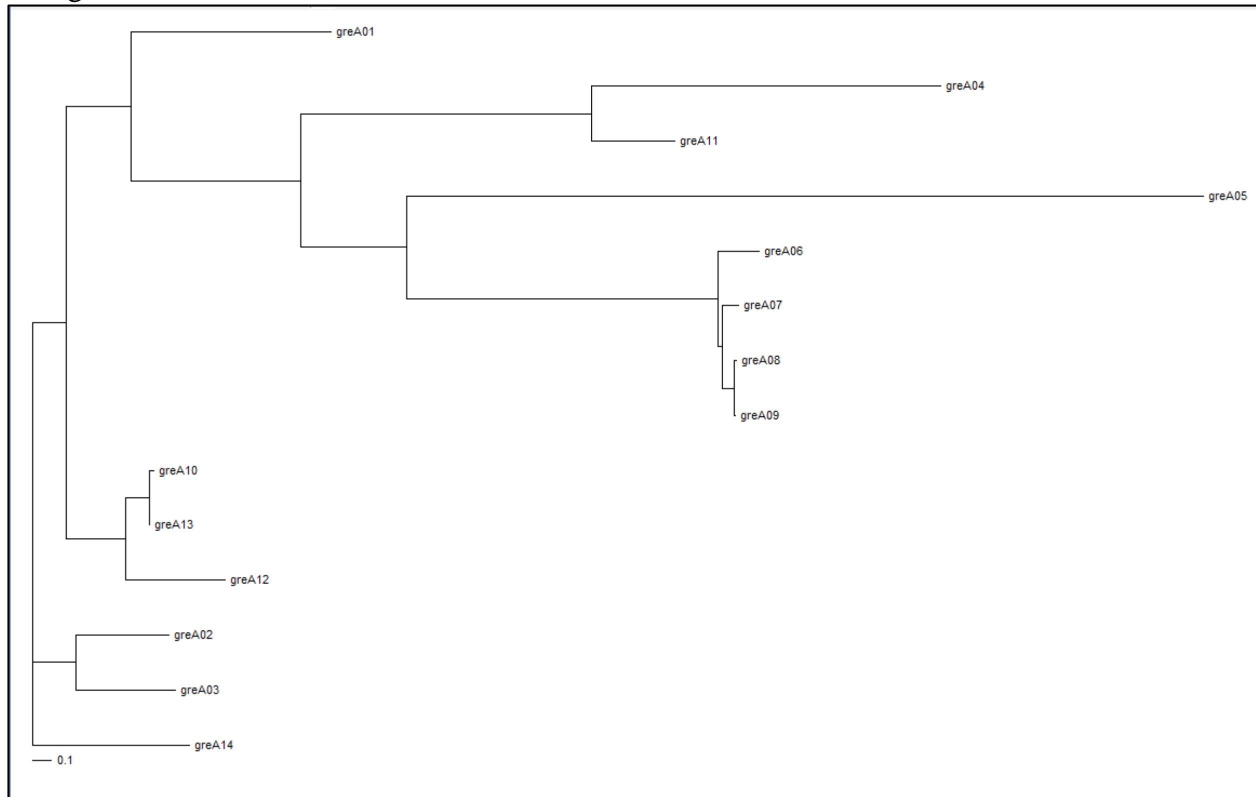
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: 01-09
- Vertebrate pathogen lineages: 10-14
- No significance was observed for any type of selection

Figure 15. Rate of evolution of the *folA* housekeeping gene in plant and vertebrate pathogen lineages



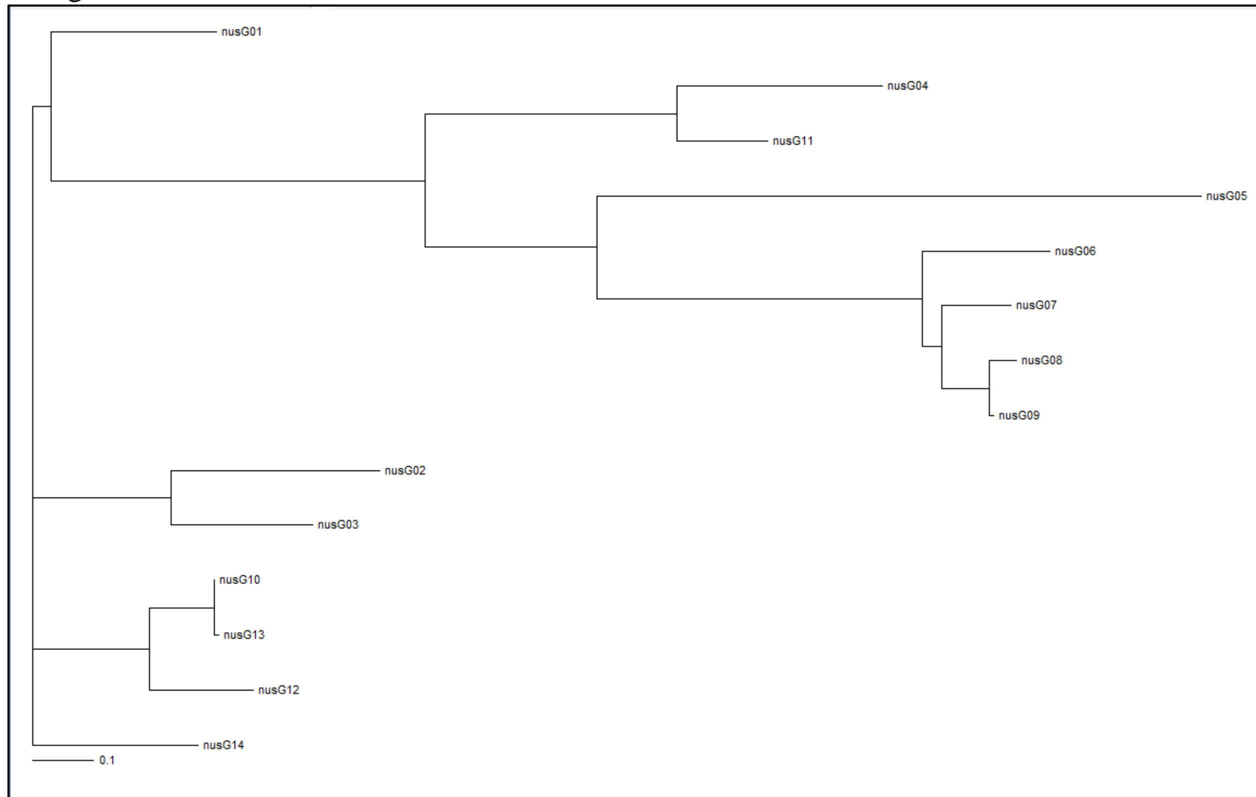
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: 01-09
- Vertebrate pathogen lineages: 10-14
- Significance was observed for positive selection

Figure 16. Rate of evolution of the *greA* housekeeping gene in plant and vertebrate pathogen lineages



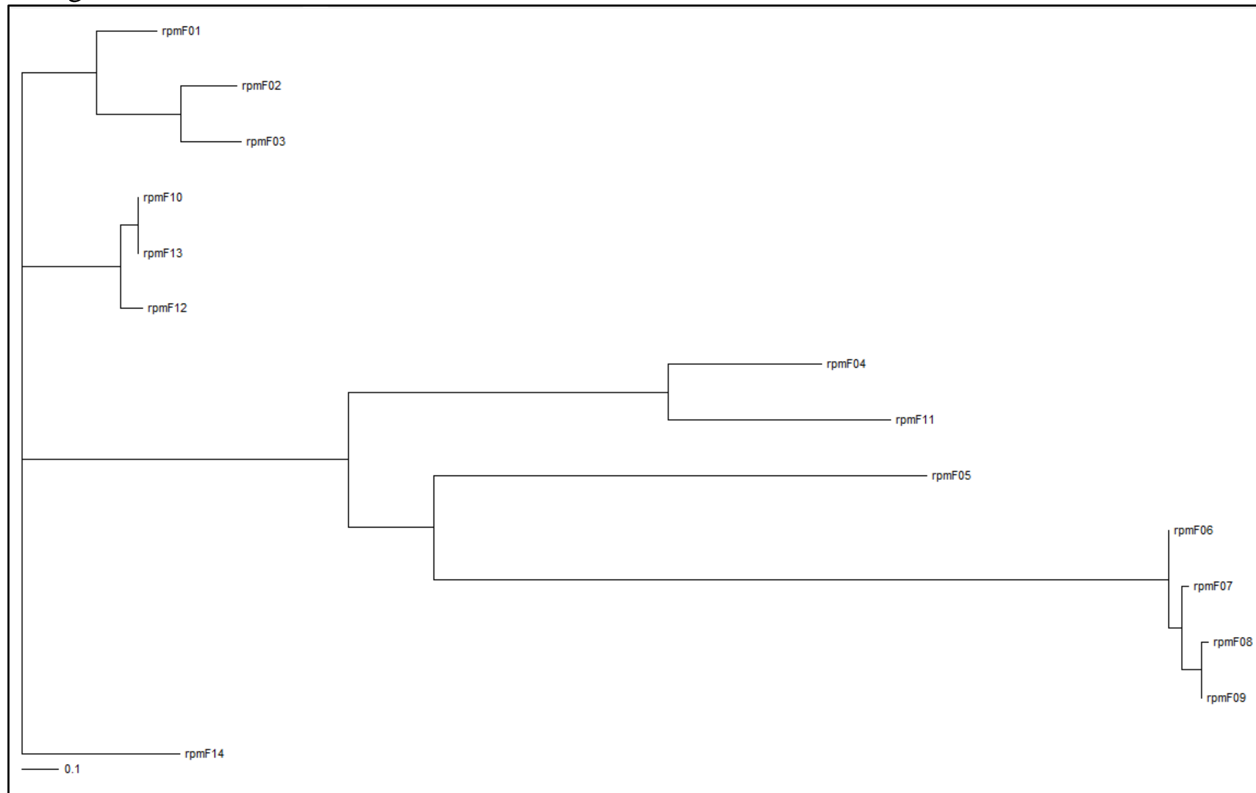
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: 01-09
- Vertebrate pathogen lineages: 10-14
- Significance was observed for differential selection

Figure 17. Rate of evolution of the *nusG* housekeeping gene in plant and vertebrate pathogen lineages



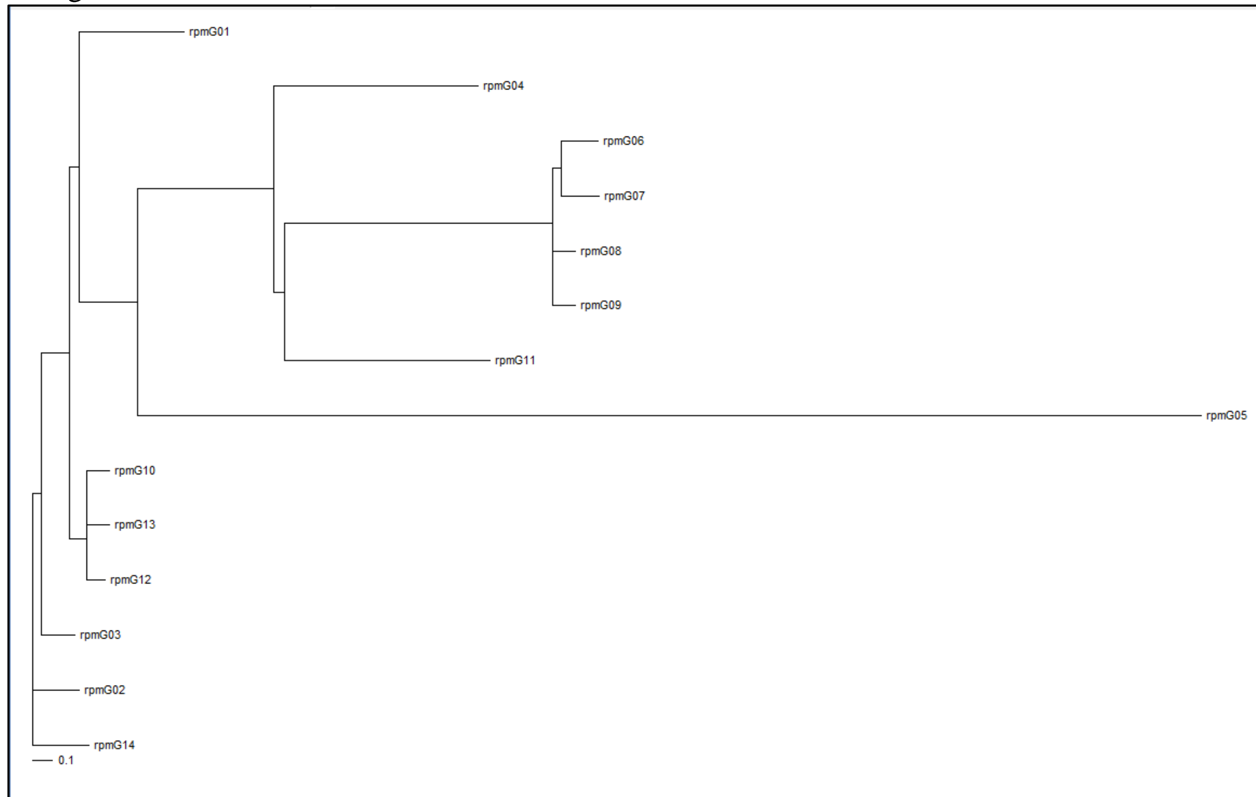
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: 01-09
- Vertebrate pathogen lineages: 10-14
- Significance was observed for positive selection

Figure 18. Rate of evolution of the *rpmF* housekeeping gene in plant and vertebrate pathogen lineages



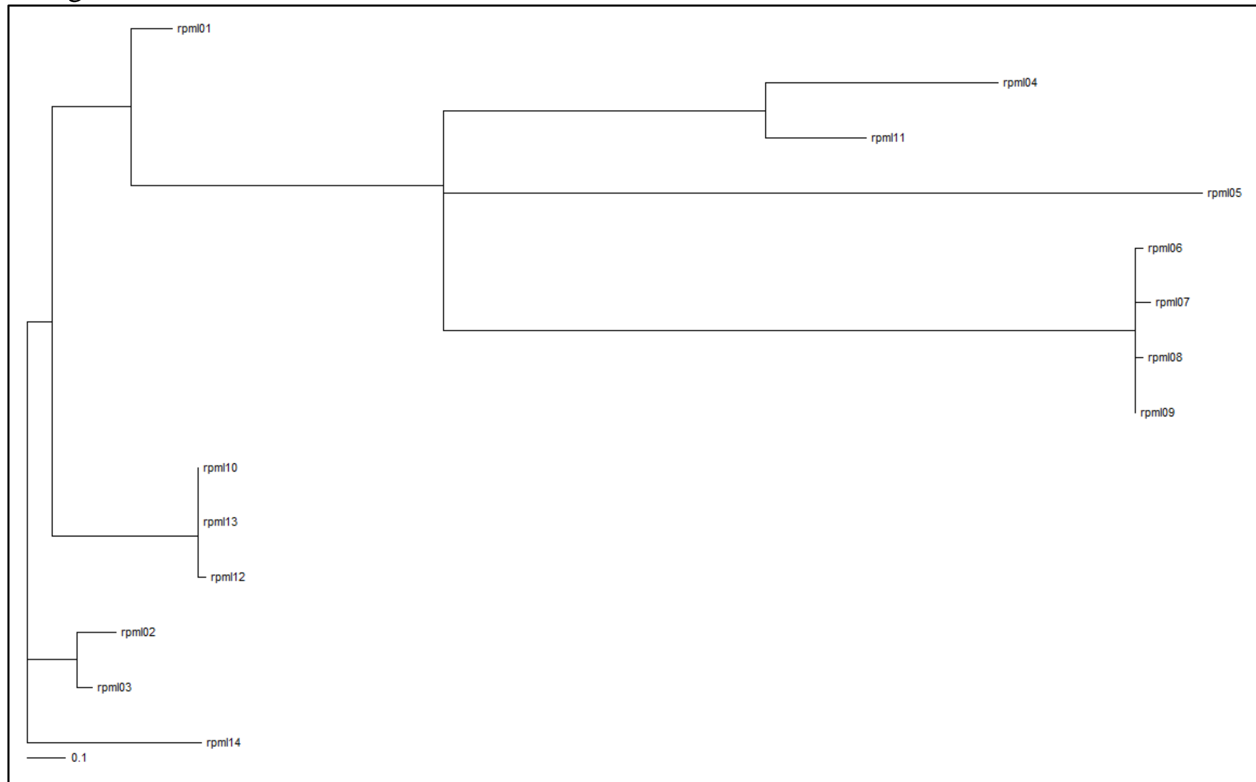
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: 01-09
- Vertebrate pathogen lineages: 10-14
- No significance was observed for any type of selection

Figure 19. Rate of evolution of the *rpmG* housekeeping gene in plant and vertebrate pathogen lineages



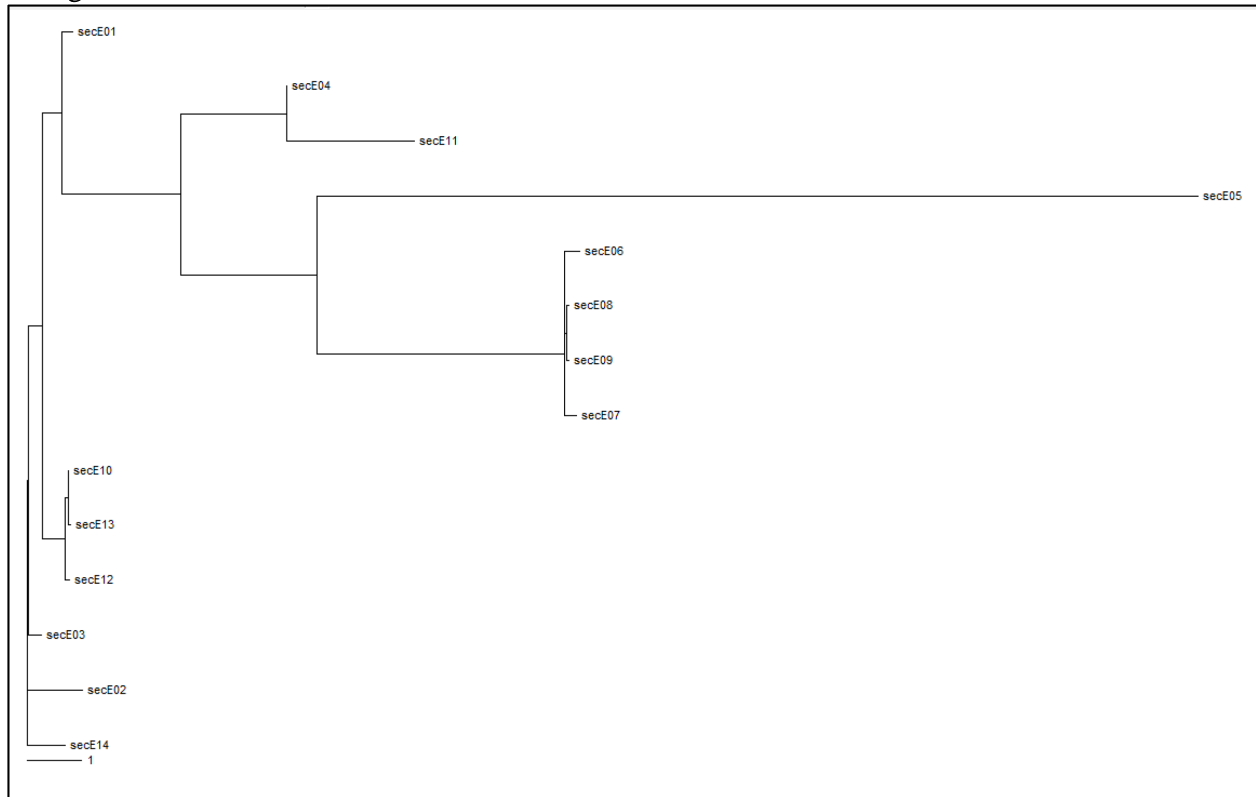
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: 01-09
- Vertebrate pathogen lineages: 10-14
- Significance was observed for positive selection

Figure 20. Rate of evolution of the *rpmI* housekeeping gene in plant and vertebrate pathogen lineages



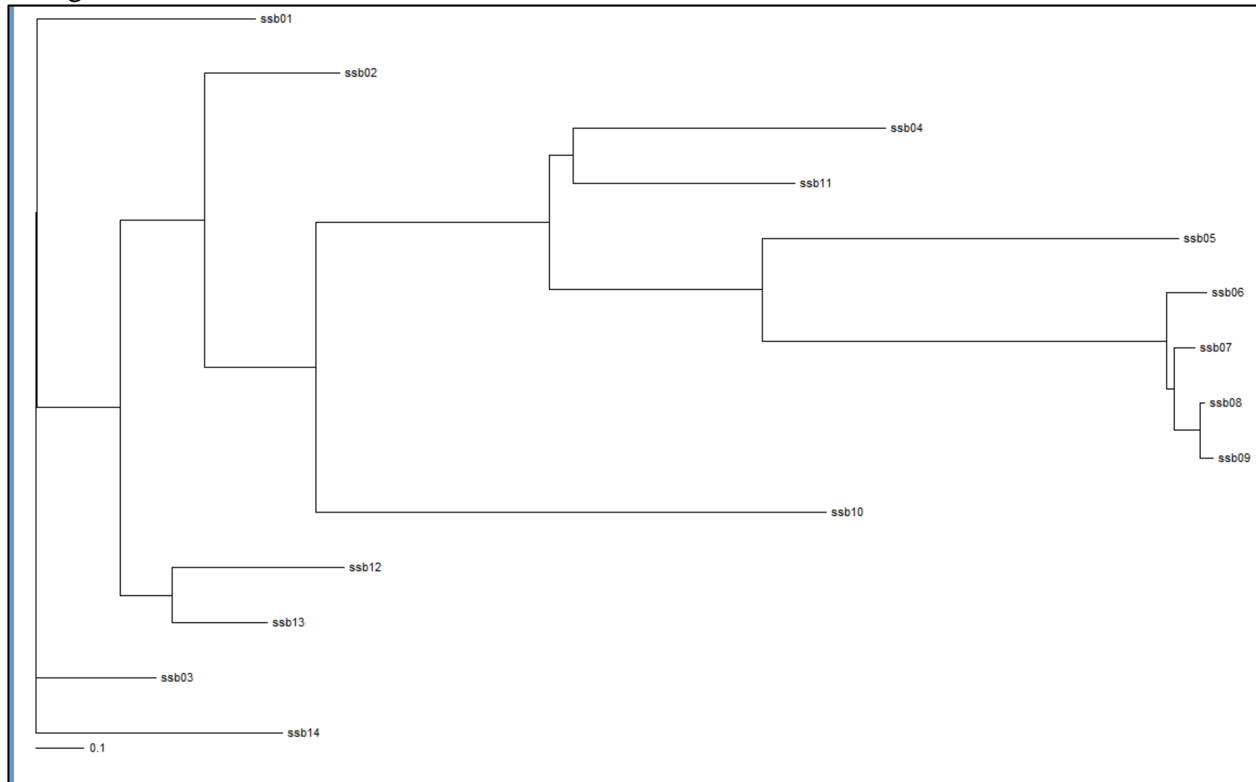
- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: 01-09
- Vertebrate pathogen lineages: 10-14
- Significance was observed for positive selection

Figure 21. Rate of evolution of the *secE* housekeeping gene in plant and vertebrate pathogen lineages



- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: 01-09
- Vertebrate pathogen lineages: 10-14
- Significance was observed for differential selection

Figure 22. Rate of evolution of the *ssb* housekeeping gene in plant and vertebrate pathogen lineages



- Type of phylogram: Unrooted neighbor joining tree
- Longer branches indicate higher evolutionary rates (scale is equivalent to number of substitutions per site)
- Plant pathogen lineages: 01-09
- Vertebrate pathogen lineages: 10-14
- Significance was observed for positive selection

BIOGRAPHICAL SKETCH

Alejandro Cantu is 24 years old and completed his Master of Science in Biology Degree on May, 2014 from the University of Texas-Pan American. In December 2010, he earned his Bachelor of Science in Biology Degree (*summa cum laude*) with a concentration in Education and a minor in Spanish. He currently resides in 1107 E. Valle Vista Ave. in Pharr, Texas.

Alejandro is a high school teacher at PSJA Thomas Jefferson T-STEM Early College High School in Pharr, Texas. He is currently on his third year of teaching. Additionally, he is the school's UIL Academic Coordinator and Biology Coach for UIL Science. Prior to his teaching job, he was a Substitute Teacher also at PSJA ISD from January 2011 to June 2011.