8-2020

Identifying Positive Selection in Multiple Subspecies of *Xylella fastidiosa*

Daniel Doroteo Flores
The University of Texas Rio Grande Valley

IDENTIFYING POSITIVE SELECTION IN MULTIPLE SUBSPECIES OF XYLELLA FASTIDIOSA

A Thesis

by

DANIEL DOROTEO FLORES

Submitted to the Graduate College of
The University of Texas Rio Grande Valley
In partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2020

Major Subject: Biology

IDENTIFYING POSITIVE SELECTION IN MULTIPLE SUBSPECIES OF XYLELLA FASTIDIOSA

A Thesis
by
DANIEL DOROTEO FLORES

COMMITTEE MEMBERS

Dr. Erin Schuenzel
Chair of Committee

Dr. Matthew Terry
Committee Member

Dr. Christopher Vitek
Committee Member

August 2020

ABSTRACT

Flores, Daniel Doroteo, Identifying positive selection in multiple subspecies of Xylella fastidiosa. Master of Science (MS), August, 2020, 27 pp., 4 tables, 7 figures, 26 references

For this study, we will be looking to identify positive selection in eight genomes of the bacterial plant pathogen *Xylella fastidiosa*. A previous study used a branching method that identified 2 genes with positive selection along with a site-specific method identifying 34 genes showing positive selection. This current study focused specifically on the site-specific method, resulting in 28 genes (of 1,039 tested) showing positive selection. Of the 28 genes showing positive selection, 12 of them come from the pathogenicity, virulence and cellular structural categories. The remaining genes are found in the biosynthesis, metabolism, macro metabolism, and cellular process categories. Not only does using this type of model help us in understanding the role of selection on a gene in this bacteria, but by using the Selecton website to run the sequences we can also see in which location on the protein the selection is taking place, therefore allowing us to find the specific codon of the gene in which positive selection is taking place. It also allows us to identify other genes that show changes taking place on a specific codon of the gene but may not be showing positive selection as a whole.

DEDICATION

I would like to dedicate this to my family, without their support none of this would be possible. My mother and father have always been supportive of everything I do. Also to my sister, brother-in-law, and my two nieces, whenever I needed a break from things they were the ones who were there for me. The love and patience that all of you have shown and given is the main reason for all of my success.

ACKNOWLEDGMENTS

I would like to acknowledge Dr. Erin Schuenzel, chair of my dissertation committee, for all of her patience and time spent meeting with and assisting me along the way. Every time we met I was always blown away by how much knowledge of the subject you had and shared. Also thanks to my dissertation members: Dr. Matthew Terry and Dr. Christopher Vitek. The advice of all my committee members is and will always be greatly appreciated.

A special thanks to the Student Accessibility Services department of UTRGV for always helping with any barriers I may have faced as a person with a disability.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER I

INTRODUCTION

A long process of evolution has shaped the sequence of amino acids in proteins. Detecting biologically significant sites on a protein, by way of changes in the gene, is important for such areas as drug design and the study of protein function and structure. Sites that are conserved may indicate structurally important sites (Melamed et al. 2004) or ligand binding sites (Gertow et al. 2004). Conserved sites may also be active sites involved in protein-protein interaction (Bridges and Moorhead 2005). Mutations at conserved sites reduce the fitness of the organism, which will be selected against and possibly removed from the population. Highly variable sites may represent sites that have been subjected to positive selection. Variable sites have been regarded as tolerant to functional constraints (Glaser et al. 2003). These sites may be evidence of molecular adaptation, which could mean an evolutionary advantage for the organism (Yang 2005).

In this study, we will look at the genes of the bacterium *Xylella fastidiosa*, a gram-negative gamma subdivision proteobacterium, which lives entirely in the xylem. It causes a variety of diseases to economically and aesthetically important plants. Various strains of the bacteria cause diseases such as Pierce's disease of grapevine (PD), citrus variegated chlorosis (CVC), phony peach disease, pear leaf scorch, almond leaf scorch (ALS), oleander leaf scorch (OLS), and alfalfa dwarf (Simpson et al., 2000). The bacteria essentially clog the xylem, the water-conducting tissue of the plant, and prevent the flow of water and nutrients within the plant,

which results in stress, and eventual death of the plant. In PD, the bacterium eventually kills grapevines or causes them to be unproductive within 2 to 3 years. Southern California went through the destruction of its wine grape industries in the 1880s and suffered a loss of about 40,000 acres of grapes near Anaheim. Currently, Pierce's disease can be found in many of California's wine grape regions as well as other southern states like Arizona and New Mexico, even spreading out into the Georgia and Florida areas. The CVC disease, first recorded in 1987, affected many different commercially important citrus trees especially sweet oranges. The main signs of CVC include chlorotic areas on the upper side of the leaves with gum-like substance on the lower surface of the leaf, foliar wilt (Cordeiro et al. 2014), and small and hardened fruit of no commercial value.

X. fastidiosa is carried from plant to plant by several different species of xylem-feeding insects from the family Cicadellidae called sharpshooters or leafhoppers. A couple of the sharpshooters are the *Homalodisca vitripennis* (Glassy-winged Sharpshooter), *Homalodisca liturata* (Smoke tree Sharpshooter), *Acrogonia terminalis* (this species is one of the two most common sharpshooters found inhabiting citrus in the state of Sao Paulo, Brazil), and the *Graphocephala atropunctata* (blue-green sharpshooter). The glassy-winged sharpshooter (GWSS) is a very good vector for the bacteria because of its ability to move faster and farther, and it has a much wider range of hosts than other sharpshooters (Andersen et al., 2009). According to the California Department of Agriculture (www.cdfa.ca.gov), the GWSS has a host list of more than 110 different plants. Because of its ability to colonize a great variety of host plants that are around vineyards, it can cause more extensive damage. Sharpshooters acquire the bacteria after feeding on an infected plant. The bacterium can survive and form colonies on the inner linings of the mouthparts of the vector. Once the sharpshooter has acquired the bacteria, it

can carry it for the remainder of its life. The vector can lose the bacteria when it molts the inner lining of the inner mouthparts but can reacquire the bacteria after feeding on another infected plant.

Xylem fluid is under a large amount of negative pressure, which means to extract the necessary amount of fluid the sharpshooters must rely on large cibarial muscles to pump out the fluid (Dugravot et al. 2008). The xylem fluid is roughly 95% water that contains a small amount of organic and inorganic molecules. Amino acids and other organic acids are often present in the fluid but found in the very low numbers. The sharpshooter can pump several times its body weight in fluid each day and has highly efficient filter chambers in their alimentary canals that filter out the organic compounds (Andersen et al., 2009). It is during the feeding that the sharpshooters secrete saliva into the plants and then rapidly suck back up leaving behind saliva and fluids from the plant. The bacteria may become dislodged due to the fluid movements in the precibarium. The movement of fluid from the mouth into the plant, from the stylets into the plant cell, allows for the bacteria to enter the plants.

Once inside the plant, the bacteria use a fimbria-like structure and twitching motility to aid in the movement of the bacteria upstream and to form a bacterium-bacterium and plant-bacterium attachment (Meng et al. 2005). The bacteria forms biofilm clumps within the xylem and are usually found in an extracellular matrix composed of extracellular polysaccharides creating a blockage in the xylem. Similar examples of biofilm formation have been found in *Neisseria gonorrhea* and its attachment to human cells (Carbonnelle et al. 2005). The blockage of the xylem has previously been seen in vitro using wood as the surface for the attachment of bacteria using several different strains of *X. fastidiosa* (Marques et al. 2002). The formation of the biofilm in the xylem of host plants is a major factor of the diseases related to his bacteria.

X. fastidiosa also contains proteins on the outer membrane that promote adhesion to the plant and other bacteria. These afimbrial adhesins are a crucial pathogenic factor involved in plant pathogens. They allow the bacteria to bind to other bacteria and the xylem cell walls, which is essential in the early stages of pathogenesis (Hélaine et al. 2005).

Previous Study

The initial study looked at genes from four genomes of the bacterial plant pathogen *Xylella fastidiosa*. There were 1143 genes tested using a branching method and a site-specific method to identify positive selection. For the current study, we examined genes found to contain positive selection having used the sites-specific method. The sites-specific method estimates the ratio of non-synonymous and synonymous mutations for each site within a gene across the phylogeny. The genes used came from 10 different categories, six of which could be grouped into housekeeping and surface genes. The housekeeping genes include the categories of biosynthesis of small molecules, metabolism of small molecules, macro metabolism of DNA, RNA and proteins, and cellular processes. The other four categories included genes with functions that may not have fit into the above categories, may contain mobile genetic elements, and also included two divisions of hypothetical open reading frames (ORF's) that only occur in *X. fastidiosa*, as well as those conserved from other bacterial species.

The site-specific method used the previously identified 34 genes with significant positive selection. To categorize the genes, the online Gene Ontology database (www.geneontology.org) was used. The GO database uses a network of gene data to represent the molecular functions, cellular locations, gene products, and categories for each gene. When looking at the categories of the 34 genes showing positive selection in the original study, 13 are placed into a surface gene category and the other 21 into the housekeeping gene category using the GO database. The

surface genes involve those that interact directly with the environment, which includes pathogenicity/virulence and cellular structure genes. The housekeeping genes involved those categories of biosynthesis, metabolism, macro metabolism, and cellular processes.

Current Study

For this study, we added 4 more genomes to the previous study to determine whether we see the same patterns of selection occurring. By adding 4 more genomes we could identify if there is an increase or decrease in the number of genes that are undergoing positive selection. It will also help to identify under which categories, pathogenic/virulence, or housekeeping genes, they will fall under. By using the Selecton website (www.selecton.tau.ac.il), we can identify which amino acids of the proteins are showing most of the change which helps us identify which part of the proteins is having most of the interaction with the plant defense mechanisms. That in turn helps us identify which codon on the gene is undergoing the most change. By adding more genomes to our study, I expect to see a small decrease in the total number of genes showing positive selection. There may also be some genes showing positive selection that were not on the original study. Because of the high amount of interactions that must take place within the bacteria and the defense mechanisms of the plants, we would expect to see genes involved in pathogenicity/virulence and outer membrane structures to show signs of positive selection. Some examples of these genes would be those responsible for adhesion to one another and the plant's xylem walls as well as outer membrane proteins.

CHAPTER II

MATERIALS & METHODS

The 8 genomes used in this study come from four different strains of *Xylella fastidiosa*. The four subspecies are fastidiosa, multiplex, pauca, and sandyi. The pauca strain can be found in South America with signs of it moving north, and the other three can be found in North America. From these four subspecies, we used eight genomes: 9a5c, M12, M23, Temecula, GB514, Dixon, EB92.1, Ann-1 (Table 1). We obtained the sequences for each gene of the genomes from the National Center for Biotechnology Information website (www.ncbi.nlm.nih.gov). Once the eight sequences were obtained for a particular gene, they were checked and edited appropriately especially when the sequence was in an anti-parallel (inverse complement) form. To align the sequences we used the Clustal Omega website (www.ebi.ac.uk/Tools/msa/clustalo) program which helped in identifying any gaps that may have occurred in one or more of the sequences. The Clustal Omega program clustalo (version 1.1.0) was used for the alignment of the sequences. After all of the editing took place and the sequences were in their proper FASTA format, they were run through the Selecton website to find positive selection. When you are on the Selecton server website you must upload your FASTA format sequence and give your query sequence a name. To receive and save your results, you must also input your email address for the results to be sent to you. Under the evolutionary model portion of the setup, you must select the "Null model: no positive selection (M8a, beta + ω =1)" option along with setting the number of categories for the distribution to 4.

Once that is done you are allowed to submit your sequences and the server will calculate for positive selection.

To detect the level of selection on a given protein, Selecton computes ω, which is the ratio between non-synonymous (Ka) and synonymous (Ks) substitutions. Sites that show ω values significantly higher than one (ω>1) indicate positive selection and sites that show ω values significantly lower than one (ω<1) are indicative of purifying selection (Stern et al., 2007). The null hypothesis, which shows a ω value equal to one (ω=1), is neutral evolution occurring in which mutations occurring at the molecular level are not undergoing natural selection. Rather the allele frequency changes happen because of random fixation of neutral mutations through genetic drift.

The minimal input into the server consists of either a non-aligned file containing homologous coding DNA sequences, or a codon aligned file of coding DNA sequences and a query sequence. The sequences must be FASTA format only. Sequences must be of coding regions only and may not contain internal stop codons. A minimum of three sequences is required, yet it is recommended to work with at least 10 sequences or more to increase the power of the results. Several different evolutionary models can be used on the Selecton software. For this study, we are using two models nested against one another. The first, M8, assumes a ω value greater than 1, which allows for positive selection. It is the default model for the Selecton server. The model is nested against is the M8a model, a variation of the M8 model, which sets the ω value to 1 thus allowing for purifying selection and neutral evolution. By comparing these two models we can statistically test the hypothesis that there is positive selection operating in the protein by comparing it to the hypothesis of the null model that there is no positive selection. The output created by Selecton is the likelihood of each model, using a likelihood ratio test (LRT)

used for comparison when the models are nested. The LRT compares twice the log-likelihood difference of both models to a $\chi 2$ table with one degree of freedom when nested. Because we have one degree of freedom, the target $\chi 2$ value at 95% certainty is 3.84. Any value at that or higher indicates positive selection for that gene.

Biologically significant sites in a gene are identified by comparing the rate of synonymous (Ks) substitutions to non-synonymous (Ka) substitutions. By dividing the Ka substitutions by the Ks substitutions we can create a ratio (ω), for each amino acid on the protein allowing us to identify either positive or purifying (negative) selection. Selecton runs by accepting a set of coding DNA sequences (and a protein databank ID if available), which are then translated to a sequence of amino acids giving you the entire protein. Each position of the protein is given a ω score, which was computed, based on the Selecton algorithms (Doron-Faigenboim et al., 2005).

CHAPTER III

RESULTS

After running the 1,039 genes through the Selecton server, we identified 28 genes showing positive selection (Table 2). Looking at the categories in which these genes come from, we identified that 6 are from the intermediary metabolism category, 3 are from the biosynthesis category, 3 are from the macromolecule metabolism category, 8 are from the cell structure category, 4 are from the cellular processes category, and 4 were from the pathogenicity/virulence category (Table 3). Together the pathogenicity/virulence and cell structure categories make up all surface genes. The biosynthesis, metabolism, cellular process, and macro metabolism categories make up the housekeeping genes. If we look specifically at the surface genes, 227 that were tested, we see that 12 (5.28%) were identified to show positive selection that is barely above the 5% expected by chance. The remaining 812 genes that were tested, housekeeping genes, showed that 16 of those genes (1.97%) showed positive selection. Those 16 genes, while not significant given the sample size, would be good candidates for further testing with a greater sample size subspecies of *X. fastidiosa*. As predicted we identified genes showing positive selection from the cell structure/outer membrane categories as well as the pathogenicity/virulence categories.

The ability to map where positive selection takes place on these proteins is possible by using the Swiss Institute of Bioinformatics website (https://swissmodel.expasy.org/interactive/) to get a hypothetical model (SWISS-MODEL) using the resulting protein sequence from each

genes DNA sequence. It uses the RCSB Protein Data Bank (www.rcsb.org), European Protein Data Bank (www.ebi.ac.uk/pdbe/), as well as the CATH Protein Structure Classification Database (www.cathdb.info) to obtain similar protein sequences as well as their protein structures. Protein structures in the data bank are used as templates to create a hypothetical 3D structure for the protein sequence entered.

We can use genes involved in type IV pili and biofilm productions, which have been identified in past studies to exhibit positive selection. Type IV pili genes have functions, which include the ability for movement using a twitching action, as well as adhering to cell surfaces whether it be other bacteria or the plant cell walls (Li et al., 2007). The two within this category that showed positive selection in the previous study were the pilY1 and pilE genes. If we look at the 3D structure of the pilE gene (Fig. 1) we see a helical structure roughly 25 amino acids long followed by three beta-strands. Using the SWISS-MODEL website, the organism with the closest related sequence is the PilE protein from the *Pseudomonas aeruginosa* bacteria. By looking at our Selecton results, we see that it displays positive selection at the 79^{th} (coil) and 110^{th} (2^{nd} beta-trand) residue of the protein.

Another set of genes that showed positive selections in the previous study was the fimbrial precursor genes. The mrkD gene, a fimbrial adhesin precursor, has shown to be actively evolving under strong positive selection in *E. coli* through the accumulation of mutations, most often happening at the same position in the protein (Stahlhut et al., 2013). It has been found in other closely related pathogens such as Klebsiella, Enterobacter, Serratia, and Citrobactyert. Fimbrial adhesin genes are responsible for the production of biofilms within the xylem causing a lack and blockage of water getting through to the leaves of the plants. They are adhesion organelles that help in attaching to the outside environment as well as each other. The gene

with the highest $\chi 2$ score was the mrkD gene with a score of 34.36. By looking at the Selecton picture result (Fig. 2) of the protein we can see several regions undergoing selection. The residues are shown in yellow and more particularly orange are sites within the gene displaying positive selection.

The mtfA (mannosyltransferase) gene is another that indicated positive selection in the initial *X. fastidiosa* study. It aids in the transfer of mannose to guanosine diphosphate mannose and eventually to a pyrophosphate polyprenol carrier. This is used in the making of exopolysaccharides, which are a fastidian gum of importance to the pathogenicity of Xylella (Alves et al., 2011). The mtfA gene had also been shown to exhibit positive selection on a study using 6 different strains of Saccharomyces. It was suggested that the positive selection going on in the gene was caused mainly by the relaxed purifying selection at both non-synonymous and synonymous sites (Li et al., 2009). For this study, it contained 19 amino acid sites with high amounts of positive selection and a $\chi 2$ score of 15.18.

Along with finding positive selection on a gene as a whole, Selecton shows genes that show positive selection on a particular site of the gene as well as showing us genes undergoing negative selection. The exbD2 (biopolymer transport protein) and rpfA (aconitase) genes, which had shown positive selection in the initial study, were shown to drop below the $\chi 2$ threshold score of 3.84 after adding four more strains for this study. The exbD2 gene, along with tonB, helps in activating the transport of iron through the outer membrane of the cell. Because of the low amount of free iron ions in most host organisms, many gram-negative bacteria have had to develop efficient ways for the bacteria to take in any iron ions available in their environments (Garcia et al., 2007). By looking at the protein sequence created by Selecton for exbD2 (Fig. 3), we can see three sites undergoing change. Another gene that had previously shown positive

selection is rfpA (aconitase). It is part of the rfp (regulation of pathogenicity factors) cluster genes. It has involvement in the levels of extracellular enzymes and extracellular polysaccharide production (Wilson et al., 1998). After adding four more strains in this study, we see that some $\chi 2$ scores drop enough to show negative selection as shown in the Selecton results (Fig. 4).

As in the original study, we also expected positive selection in genes involved in the interaction with the environment of the bacteria as well as pathogenicity/virulence and cellular structure genes. We did continue to see positive selection in 12 of those original 34 genes. As expected when adding more *X. fastidiosa* subspecies to be analyzed with the original, we did see a drop in $\chi 2$ threshold scores of some of those genes. Seven of the original genes that dropped in $\chi 2$ score continued to show positive selection on certain sites of the gene but not enough to show positive according to Selecton analysis. The other 15 genes from the original list showed $\chi 2$ scores below the threshold indicating purifying selection. Three of the genes in the original group of 34 (*pilY1*, *pilE*, and *mrkD*), all part of the cell structure category, had some of the highest $\chi 2$ scores in the original study did have some of the highest $\chi 2$ scores on this study.

CHAPTER IV

DISCUSSION

It is important to note that the original study used CODEML that is part of the Phylogenetic Analysis by Maximum Likelihood (PAML), which is a set of programs used for analyses of DNA or protein sequences using maximum likelihood. The original study used two models, the first which assumes two types of amino acids exist where $\omega_0<1$ (purifying selection) and a second type with $\omega=1$ (neutral). The second model contains a category of sites subject to positive selection where $\omega>1$. CODEML, using two degrees of freedom, then determines if the second model is a better fit and if so then positive selection is determined. The current study used Selecton for all gene sequence analyses that could explain some of the differences in the amount of genes showing positive selection and any new genes showing positive selection that may not have been in the first study. Selecton also uses two models, M8 against M8a, in which M8 assumes ω values come by beta distribution and an extra category where $\omega_s\geq1$ allows for positive selection. It is nested against M8a, in which it's additional category $\omega_s=1$ allowing for purifying or neutral selection. One degree of freedom is used in this current statistical testing and uses a likelihood ratio test and compares twice the log-likelihood difference of the two models on a $\chi 2$ table.

By adding 4 more strains of the *Xylella fastidiosa* bacteria to go along with 4 strains from the original study, we found that 12 of the original 34 genes continued to show positive selection. We then decided to look at the entire genome of the *X. fastidiosa* bacteria and added all of the

other genes in the pathogenicity, cell structure, biosynthesis, metabolism, cellular process, and macro metabolism categories. This resulted in an additional 16 genes showing positive selection, bringing our total to 28. Again by adding more subspecies to 8 analyzed in this study, we would expect the number of 28 to possibly decrease, but statistically, we can have more confidence in the results because we will have increased the power of the likelihood ratios for our genes.

In addition to those genes exhibiting positive selection, there were an additional 16 genes that exhibited positively selected sites (Table 4) within the gene but not enough to identify it as a gene with positive selection as a whole. As we have seen by this study, adding more strains not only decreased the number of genes showing positive selection from the original study but also helped in identifying genes that previously did not display positive selection. By adding another 10 to 12 strains of *X. fastidiosa* to our study we should be able to better narrow down the genes that continue to show positive selection.

Adding more strains of *X. fastidiosa* genes also helps us narrow down which genes and category of genes to target in this and other similar bacteria. Genes that should be targeted for future study should be those responsible for pathogenicity/virulence, especially those genes that are responsible for proteins that are exposed to the outside environment. Many genes have been found to show positive selection because of their exposure to the outside environment of which are involved in the evolutionary arms race between the host and the pathogen itself (Aguileta et al. 2009). Examples of genes from other bacteria that have shown positive selection are the transmembrane proteins PorB of Meningococcus (Urwin et al. 2018). Pseudomonas aeruginosa has a siderophore protein that is secreted by the bacteria and involved in the ferrous ion uptake in its surroundings that has also shown positive selection (Smith et al. 2005). In the bacterial plant pathogen Pseudomonas aeruginosa, the hrpA pilin involved in the anchoring of the bacteria to

the plant cell walls, which is very similar to *X. fastidiosa*, has also shown selection (Guttman et al. 2006). These surface proteins in bacteria have continued to show a departure from neutral or negative selection as compared to other classes and categories of proteins and that is why is important to target this specific type of gene.

Along with the genes on the surface of the bacteria and which have direct interaction with the outside environment, there are more genes worth looking at. A couple of the genes that have the possibility of increasing their $\chi 2$ score with the addition of more strains would be the yxaH, ccmB, rfbB, and vacC/tgt genes. They all have $\chi 2$ values above 2.0. One gene in particular is the metE (methyltransferase) gene with a $\chi 2$ score of 3.42. It is responsible for the synthesis of methionine by a direct transfer of a methyl group to the sulfur atom of homocysteine (Ferrer et al. 2004). Methionine is very important in translation initiation and protein biogenesis as well as being a major component of proteins and peptides.

Another thing to consider is the hypothetical 3D structures of the proteins that we can create. If we look at the structure for the PilY1 protein (Fig. 5), we see an amino acid that is roughly 510 amino acids long when in actuality the length of the protein is 1230 amino acids long. Because of the lack of templates available, it makes it difficult to map all the positively selected sites on the protein when we have only a partial structure of the protein. The 3D hypothetical rope structure of the PilY1 protein (Fig. 6) shows 5 locations on the protein that were available for mapping. We can use the model template alignment of our protein to that of the closest related sequence found in the C-terminal domain of PilY1 structure of *Pseudomonas aeruginosa* (Fig. 7) to find where the location of the amino acids undergoing positive selection. There are approximately 720 amino acids, which do not have a hypothetical structure composed. Continued work on the modeling of proteins is very important identify the sites on the protein

which are changing as well as to better understand that proteins and their interaction with its outside environment.

With the recent advances in genome sequencing, more genomes are now available to be compared and identified which will help establish the role that positive selection plays in not only the host adaptation of the multiple *X. fastidiosa* subspecies, but other species that undergo similar host-pathogen relationships. Genes showing positive selection could be targeted to reduce virulence and pathogenicity of those pathogens in their hosts.

Table. 1 Eight genomes used

	Subsp.	Type	Size(Mb)	GC %	# of genes	Accession #
9a5c	pauca	citrus	2.73	52.6	2905	AE003849
M12	multiplex	recombinant	2.48	51.9	2368	CP000941
M23	multiplex	recombinant	2.57	51.8	2320	CP001011
Temecula	fastidiosa	grape vine	2.52	51.8	2125	AE009442
GB514	fastidiosa	almond	2.52	51.8	2271	CP002165
Dixon	multiplex	almond	2.62	52.0	2408	NZ_AAAL02000001
EB92.1	fastidiosa	nonpathogen	2.48	51.5	2393	NZ_AAAL00000000
Ann-1	sandyi	oleander	2.56	51.9	2375	CP006696

Table.2 Genes showing Positive Selection using Selecton

Gene	Name	$\chi 2$
xf0017	coproporphyrinogen III oxidase, aerobic (hemF)	8.26
xf0032	PilY1 gene product (pilY1)	12.02
xf0033	PilE protein (pilE)	7.46
xf0078	fimbrial adhesin precursor (mrkD)	34.36
xf0083	fimbrial subunit precursor	10.22
xf0366	ribokinase (rbsK)	4.26
xf0423	exodeoxyribonuclease V beta chain (recB/rorA)	9.68
xf0506	virulence-associated protein E (vapE)	4.48
xf0557	electron transfer protein azurin I (az1)	5.488
xf0845	family 3 glycoside hydrolase (xylA)	7.34
xf0851	D-amino acid dehydrogenase subunit (dadA/dadR)	8.24
xf0873	outer membrane protein	4.18
xf0887	mannosyltransferase (mtfA)	15.18
xf1105	dihydrodipicolinate reductase (dapB)	5.62
xf1140	UDP-N-acetylglucosamine pyrophosphorylase (glmU)	5.02
xf1363	soluble lytic murein transglycosylase precursor (slt/sltY)	9.58
xf1390	cytochrome O ubiquinol oxidase, subunit II (cyoA)	5.92
xf1426	ion transporter	9.18
xf1475	ABC transporter ATP-binding protein (ynhD)	4.12
xf1498	NADPH-sulfite reductase, iron-sulfur protein (cysI)	6.74
xf1501	ATP sulfurylase, large subunit(nodQ)	4.06
xf1903	potassium uptake protein (kup/trkD)	4.44
xf2133	ABC transporter ATP-binding protein (yheS)	8.5
xf2225	bifunctional aspartokinase/homoserine dehydrogenase I (thrA/1/2)	9.58
xf2232	catalase/peroxidase (cpeB)	10.42
xf2550	outer membrane hemolysin activator protein (hecB)	4.32
xf2578	two-component system, regulatory protein (actR)	4.016
xf2728	type I restriction-modification system DNA methylase	4.92

Table.3 Number of genes within each Gene Ontology category

Biochemical Functions	# showing Positive Selection
Pathogenicity	4
Cell Structure	8
Biosynthesis	3
Metabolism	6
Cellular Processes	4
Macrometabolism	3

Table. 4 Genes with Positively Selected Sites using Selecton

Gene	Name	$\chi 2$
xf0012	biopolymer transport ExbD2 protein (exbD2)	1.956
xf0134	valyl-tRNA synthetase (valS)	0.88
xf0138	aminopeptidase A/I (pepA/xerB/carP)	0.3
xf0223	queuine tRNA-ribosyltransferase (tgt/vacC)	2.26
xf0255	dTDP-glucose 4,6-dehydratase (rfbB)	2.42
xf0373	fimbrial assembly protein (pilQ)	0.54
xf0445	prolyl-tRNA synthetase (proS/drpA)	0.38
xf0562	sec-independent protein translocase (tatC/mttB)	0.24
xf0608	mannosyltransferase (mtfA)	2.16
xf0651	transport protein (yxaH)	2.34
xf1403	phosphotransferase system HPr enzyme (phbH)	0.154
xf1499	NADPH-sulfite reductase, flavoprotein subunit (cysJ)	1.98
xf1535	citrate synthase (gltA)	1.76
xf2272	5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase (metE)	3.42
xf2456	heme ABC transporter membrane protein (ccmB)	2.66
xf2617	ABC transporter ATP-binding protein (uup)	0.7

Fig. 1 Hypothetical 3D Structure of PilE Protein (2 sites showing positive selection)

Fig.2 Fimbrial Adhesin Precursor (mrkD)

Legend:

The selection scale:

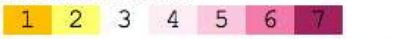


Fig. 3 Biopolymer Transport ExbD2 Protein (exbD2)

Legend:

The selection scale:

Fig. 4 Aconitase (rpfA)

Legend:

The selection scale:

1 2 3 4 5 6 7

Positive selection Purifying selection

Fig. 5 Hypothetical 3D Structure of PilY1 Gene Product

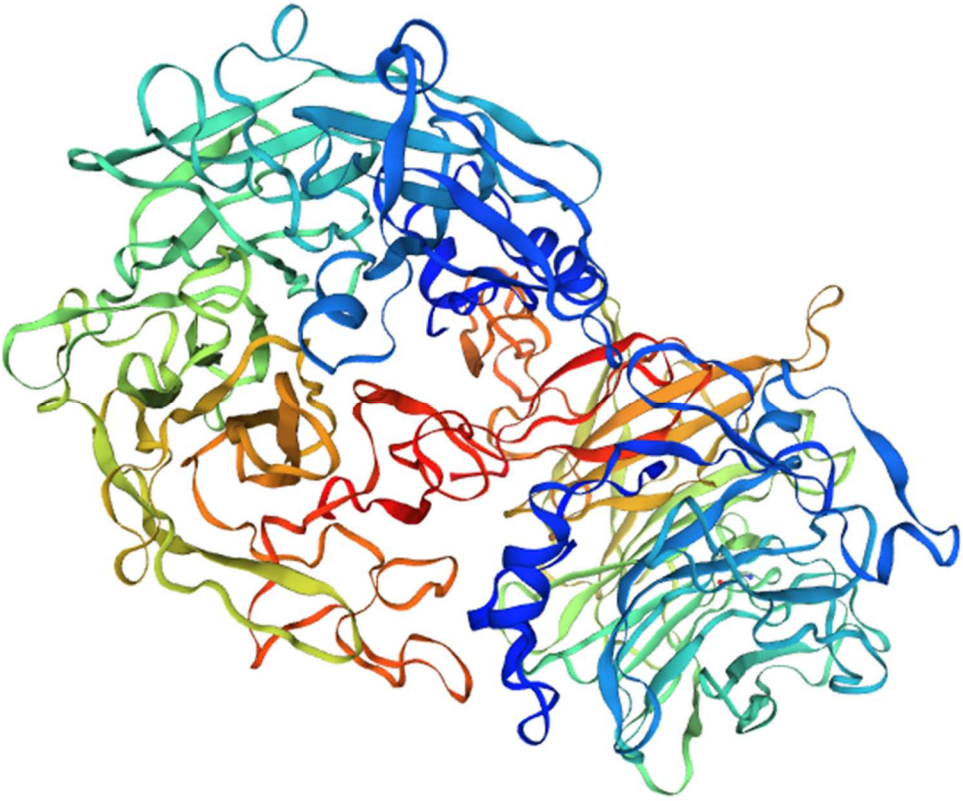


Fig. 6 Hypothetical 3D Rope Structure of PilY1 Gene Product
(same protein shown above) Contains 5 sites showing positive selection

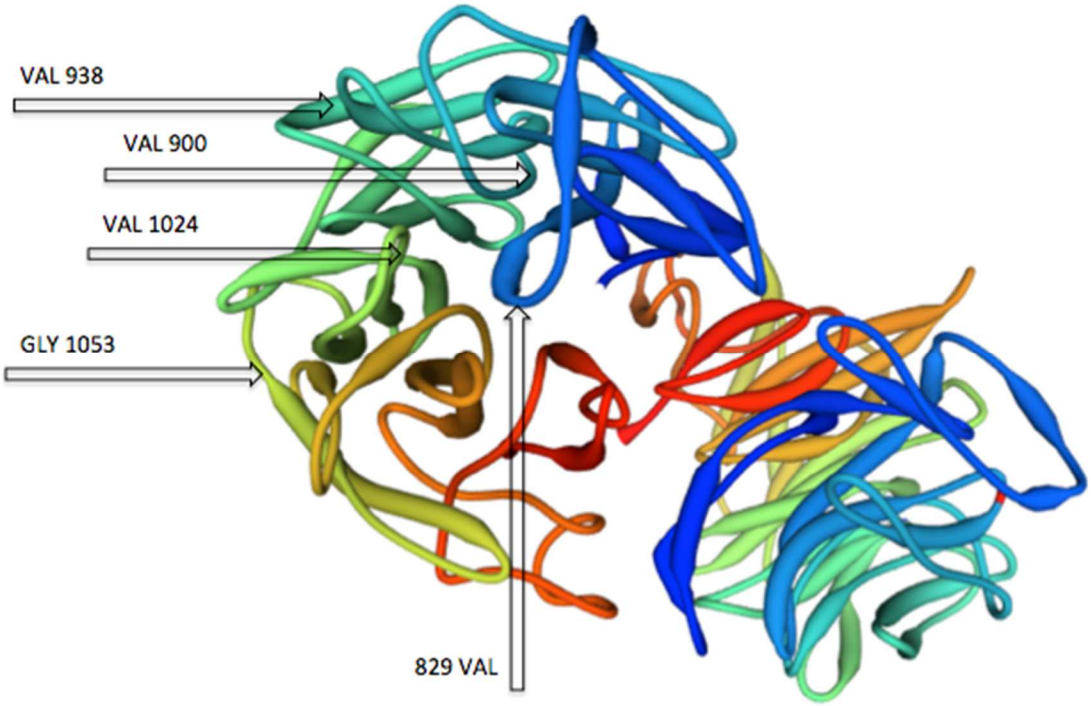
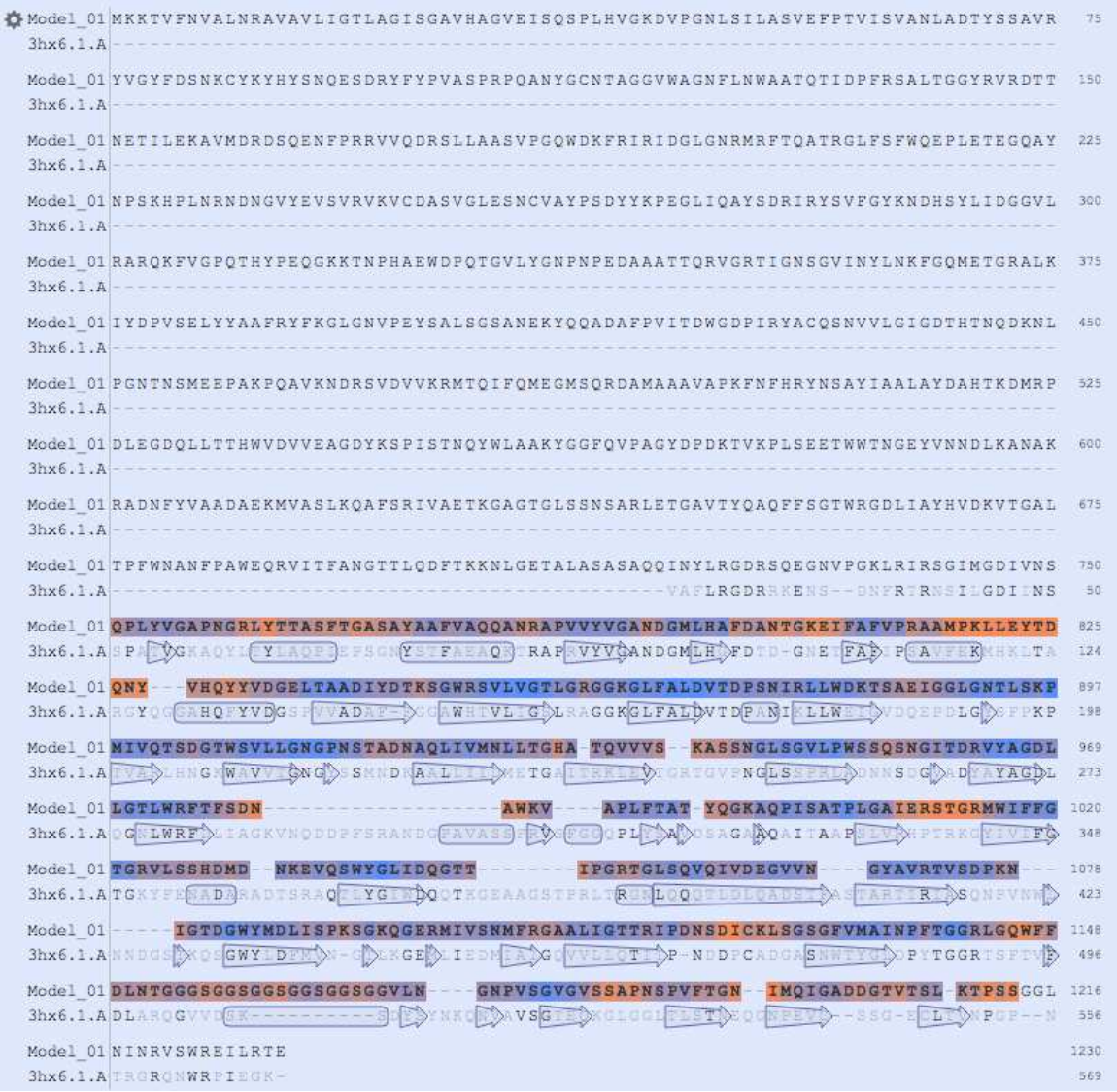


Fig. 7 PilY1 Model Template Alignment



REFERENCES

Aguileta G., Yockteng R., Fournier E., Giraud T., Refre G. 2009. Infection , Genetics and Evolution Rapidly evolving genes in pathogens : Methods for detecting positive selection and examples among fungi , bacteria , viruses and protists. 9:656–670.

Alves C.A., Pedroso M.M., de Moraes M.C., Souza D.H.F., Cass Q.B., Faria R.C. 2011. Real-time investigation of mannosyltransferase function of a Xylella fastidiosa recombinant GumH protein using QCM-D. Biochem. Biophys. Res. Commun. 408:571–575.

Andersen P.C., Brodbeck B. V., Mizell R.F. 2009. Assimilation efficiency of gree and protein amino acids by Homalodisca vitripennis (Hemiptera: Cicadellidae: Cicadellinae) feeding on Citrus sinensis and Vitis vinifera. Florida Entomol. 92:116–122.

Bridges D., Moorhead G.B.G. 2005. 14-3-3 Proteins: A Number of Functions for a Numbered Protein. Sci. STKE. 2005:re10 LP-re10.

Carbonnelle E., Hélaine S., Prouvensier L., Nassif X., Pelicic V. 2005. Type IV pilus biogenesis in Neisseria meningitidis: PilW is involved in a step occurring after pilus assembly, essential for fibre stability and function. Mol. Microbiol. 55:54–64.

Cordeiro A.B., Sugahara V.H., Stein B., Junior R.P.L. 2014. Evaluation by PCR of Xylella fastidiosa subsp . pauca transmission through citrus seeds with special emphasis on lemons (Citrus limon (L .) Burm . f). Crop Prot. 62:86–92.

Doron-Faigenboim A., Stern A., Mayrose I., Bacharach E., Pupko T. 2005. Selection: A server for detecting evolutionary forces at a single amino-acid site. Bioinformatics. 21:2101–2103.

Dugravot S., Backus E.A., Reardon B.J., Miller T.A. 2008. Correlations of cibarial muscle activities of Homalodisca spp. sharpshooters (Hemiptera: Cicadellidae) with EPG ingestion waveform and excretion. J. Insect Physiol. 54:1467–1478.

Ferrer J., Ravanel S., Dumas R. 2004. Crystal Structures of Cobalamin-independent Methionine Synthase Complexed with Zinc, Homocysteine, and Methyltetrahydrofolate. J. Biol. Chem. 2004, 279:44235-44238.

Garcia-Herrero A., Peacock R.S., Howard S.P., Vogel H.J. 2007. The solution structure of the periplasmic domain of the TonB system ExbD protein reveals an unexpected structural homology with siderophore-binding proteins. Mol. Microbiol. 66:872–889.

Gertow K., Bellanda M., Eriksson P., Boquist S., Hamsten A., Sunnerhagen M., Fisher R.M. 2004. Genetic and Structural Evaluation of Fatty Acid Transport Protein-4 in Relation to Markers of the Insulin Resistance Syndrome. J. Clin. Endocrinol. Metab. 89:392–399.

Glaser F., Pupko T., Paz I., Bell R.E., Bechor-Shental D., Martz E., Ben-Tal N. 2003. ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. Bioinformatics. 19:163–164.

Guttman D.S., Gropp S.J., Morgan R.L., Wang P.W. 2006. Diversifying selection drives the evolution of the type III secretion system pilus of Pseudomonas syringae. Mol. Biol. Evol. 23:2342–2354.

Hélaine S., Carbonnelle E., Prouvensier L., Beretti J.L., Nassif X., Pelicic V. 2005. PilX, a pilus-associated protein essential for bacterial aggregation, is a key to pilus-facilitated attachment of Neisseria meningitidis to human cells. Mol. Microbiol. 55:65–77.

Li Y., Hao G., Galvani C.D., Meng Y., De La Fuente L., Hoch H.C., Burr T.J. 2007. Type I and type IV pili of Xylella fastidiosa affect twitching motility, biofilm formation and cell-cell aggregation. Microbiology. 153:719–726.

Li Y.D., Xie Z.Y., Du Y.L., Zhou Z., Mao X.M., Lv L.X., Li Y.Q. 2009. The rapid evolution of signal peptides is mainly caused by relaxed selection on non-synonymous and synonymous sites. Gene. 436:8–11.

Marques L.L.R., Ceri H., Manfio G.P., Reid D.M., Olson M.E. 2002. Characterization of biofilm formation by Xylella fastidiosa in vitro. Plant Dis. 86:633–638.

Melamed D., Mark-Danieli M., Kenan-Eichler M., Kraus O., Castiel A., Laham N., Pupko T., Glaser F., Ben-Tal N., Bacharach E. 2004. The Conserved Carboxy Terminus of the Capsid Domain of Human Immunodeficiency Virus Type 1 Gag Protein Is Important for Virion Assembly and Release. J. Virol. 78:9675–9688.

Meng Y., Li Y., Galvani C.D., Hao G., Turner J.N., Burr T.J., Hoch H.C. 2005. Upstream Migration of Xylella fastidiosa via Pilus-Driven Twitching Motility. J. Bacteriol. 187:5560 LP – 5567.

Simpson A.J.G. et al. 2000. The genome sequence of the plant pathogen Xylella fastidiosa. Nature. 406:151–157.

Smith E.E., Sims E.H., Spencer D.H., Kaul R., Olson M. V. 2005. Evidence for Diversifying Selection at the Pyoverdine Locus of Pseudomonas aeruginosa. 187:2138–2147.

Stahlhut S.G., Chattopadhyay S., Kisiela D.I., Hvidtfeldt K., Clegg S., Struve C., Sokurenko E. V., Krogfelt K.A. 2013. Structural and population characterization of MrkD, the adhesive subunit of type 3 fimbriae. J. Bacteriol. 195:5602–5613.

Stern A., Doron-Faigenboim A., Erez E., Martz E., Bacharach E., Pupko T. 2007. Selecton 2007: Advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic Acids Res. 35:506–511.

Urwin R., Holmes E.C., Fox A.J., Derrick J.P., Maiden M.C.J. 2018. Phylogenetic Evidence for Frequent Positive Selection and Recombination in the Meningococcal Surface Antigen PorB. :1686–1694.

Wilson T.J.G., Bertrand N., Tang J.L., Feng J.X., Pan M.Q., Barber C.E., Dow J.M., Daniels M.J. 1998. The rpfA gene of Xanthomonas campestris pathovar campestris, which is involved in the regulation of pathogenicity factor production, encodes an aconitase. Mol. Microbiol. 28:961–970.

Yang Z. 2005. The power of phylogenetic comparison in revealing protein function. Proc. Natl. Acad. Sci. U. S. A. 102:3179 LP – 3180.

BIOGRAPHICAL SKETCH

Daniel Doroteo Flores received his Masters of Science in Biology at The University of Texas Rio Grande Valley in August of 2020. He had previously received his Bachelor's Degree in Biology with a Minor in Chemistry during the fall semester of 2011 at The University of Texas-Pan American. Daniel has been a teacher at PSJA North Early College High School since 2016. He has taught Advanced Placement Biology, Food Science, and Robotics in his first 4 years. Most of his students are seniors getting ready for post-secondary school life. He also coaches the UIL Science team (Biology, Chemistry, and Physics), robotics team (2 time state qualifiers), and chess team.

During his off time, Daniel is actively involved in sports being the first to compete in the annual Edinburg 10k run by doing it in a hand cycle made for wheelchair users. He is involved with the disabled community by being a starting member of the Spinal Cord Injury support group with the Doctors Hospital Renaissance of Edinburg and as a mentor for patients who have gone through similar injuries. Daniel can be reached by email: bluerooout@hotmail.com.