IDENTIFYING ISOLATES OF THE FUNGAL GENUS *FUSARIUM* SUSPECTED OF CAUSING VASCULAR WILT DISEASE IN PLANTS BY DNA BARCODING

By

Nader Mostowfi

A capstone project submitted for Graduation with University Honors

May 13, 2021

University Honors
University of California Riverside

APPROVED

Dr. Alexander Putman
Department of Microbiology Plant Pathology

Dr. Richard Cardullo, Howard H Hays Jr. Chair
University Honors

Presentor: Nader Mostowfi, Biochemistry Major
Faculty Mentor: Dr. Alexander Putman, Department of Microbiology and Plant Pathology
Contributing Members:
Lindsey Pedroncelli
Hannah Ayala

Abstract:

Fusarium oxysporum is a soilborne, filamentous fungus and a species complex, or a group of closely related species that are challenging to distinguish. Many members of this complex are host-specific plant pathogens in agriculture that are classified into *forma specialis* (f. sp.) or "special forms" based on their host specificity, which requires greenhouse experiments or DNA sequence analysis. However, *F. oxysporum* can colonize non-host plants. Therefore, it is important to positively identify suspected disease samples to allow growers to better manage the pathogen. For this project I would be identifying a group of isolates that were obtained from diseased plants sent to the Putman Lab and are suspected to belong to the *F. oxysporum* species complex. I would then perform identification by PCR amplifying and sequencing a portion of the translation elongation factor 1-alpha (TEF-1a) gene with a length of 500-600 base pairs in a process known as DNA barcoding. After obtaining the sequences, I would next assemble reads into a consensus sequence for each isolate. Reference sequences from isolates that have been positively identified in reference databases or datasets from published papers would be downloaded. I would then align the sequences of my unknown isolates with the reference sequences and then construct a phylogenetic tree. Unknown isolates would then be identified based on clustering with reference strains with high bootstrap support. All analysis would be conducted on CLC Main Workbench Software. This project taught me how to identify unknown microorganisms to the species level through conducting a proper experiment.

Acknowledgements

For this Capstone project, I am greatly thankful for my mentor Dr. Putman for helping me through the challenges of this Capstone Project. Neither of us were anticipating the circumstances that hit the world last year when we began to plan out this project in late 2019. The consequences of Covid-19 forced us to adapt as the original project planned was no longer possible. Even through this major obstacle, Dr. Putman was very patient with me and made sure I knew the reasoning behind every step of this experiment. Although I was not able to work with him in person for the Capstone, I still enjoyed working with him nonetheless.

I am also very grateful for all the help Lindsey Pedroncelli has provided as well. She was the one to propose the alternative project involving *Fusarium* isolates after it was made clear that undergraduate students could not work in research labs through most of the school year. Lindsey was very helpful and gave me insight whenever it was necessary. She also helped out with the lab work while I could not attend, ordering whatever materials necessary and assisting Hannah with sampling the isolates. While it was certainly a blessing to have worked with Lindsey on this project, it is also bittersweet as this is the last time I will be working alongside her before I graduate.

Finally, I must also thank Hannah Ayala for all her effort in the lab. She was able to perform DNA extractions and PCR for the isolates effortlessly. She was also very elaborate on explaining her tasks behind this project which made it easier for me to note down in my paper. Although I have not been able to meet her in person since she was employed by the Putman Lab after the pandemic, I am still happy that I at least got to work alongside her for the Capstone.

Table of Contents

Introduction

Crops appear simple to understand, with their main purpose being to grow on the fields while bearing food for harvest. It seems that not much is necessary for growing them besides adequate water and sunlight. In reality, plants are more complex. Similar to animals, they can suffer from diseases and illnesses. When plant pathogens and other diseases become apparent, they require research to discover a method of management. Currently, plant pathologists are at this stage with fungus *Fusarium*, a diverse genus containing many species complexes. While this paper focuses on those infecting plants, there are also species of *Fusarium* that are not soilborne. Additionally, there are some that infect humans by weakening their immune system (Nucci, 2007). Identified as early as 1874, there has been no guaranteed prevention method found for soilborne *Fusarium* (Maymon, 2020). It is difficult to pinpoint a specific method to determine which *Fusarium* species are pathogenic and which are not. With its large presence in agriculture through its wide distribution, it is a major goal of farmers to protect their plants from this harmful fungus (Knights, 2004).

Symptoms of *Fusarium* often include wilting and browning of the plant, as well as powdery growth in some cases. The damage is generally noticeable as there is often rotting of the plant (Ma, 2010). This damage can be quite costly to farmers who depend on producing crops. For instance the lettuce industry, which is also vulnerable to *Fusarium*, was worth 1.7 billion dollars in California in 2009 (Scott, 2012). While proper prevention methods would be taken by farmers, there is always the possibility of *Fusarium* infecting the crops. If the crops are infected, it is critical to identify the species to know the specific prevention methods the farmers should take to prevent loss in harvest.

Because of the diverse variety of *Fusarium* fungi present, there are numerous methods

dedicated to preventing the different species of fungi. Mycotoxins are a very common metabolite produced by *Fusarium* which farmers have learned to creatively prevent (Richard, 2007). They found that mulch layers derived from different mustards were able to effectively fight off mycotoxins in wheat, leading to improved yield of wheat (Kadziene, 2019). *Fusarium oxysporum*, another species of *Fusarium*, is eliminated with the use of microwaves (Soriano-Martin, 2005). Other effective strategies include utilizing cover crops, or crops which are not actually meant for harvest, but for covering soil. This prevents weeds and pathogens from growing and allows the harvesting crops to grow successfully (Scott, 2012). Knowing the species of each *Fusarium* strain would help farmers recognize which prevention method would be the best to eliminate it.

The genus *Fusarium* consists of several species complexes, each of which contain numerous species that are closely related and difficult to distinguish. Some are pathogenic while some are actually harmless, or at least nonpathogenic, for plants (R. Mendoza, 2020). With regards to my Capstone project, I would be identifying different species of *Fusarium* to see to which species the isolates belong. More thoroughly, different species of *Fusarium* have *forma speciales* and races, which are more specific classifications of each species based on their pathogenic properties (Edel-Hermann, 2018). Both *forma speciales* and races are not formal taxonomic groups, therefore, their classifications may vary considerably among different scientists. However, scientists follow consistent rules in labeling species, and that pattern applies to *Fusarium* species as well. The phrase *forma speciales* was originally made to help distinguish different strains of the fungi *Puccinia graminis* (Edel-Hermann, 2018).

Our focus would be directed towards *F. oxysporum*, as that is the species complex that is

primarily soilborne and holds much relevance in agriculture, which aligns with the Putman Lab's research on plant pathology. More specifically, I would be examining the *forma specialis*--or special forms--of *F. oxysporum*, which is commonly labeled as f. sp. X and f. sp. Y. The *forma specialis* classification system is used because almost all *F. oxysporum* have a very similar morphological appearance, but differ in which plant species it can cause disease on. I am familiar with the process of sampling them into Petri dishes and extracting DNA from this biology lab class known as Dynamic Genome. Furthermore, I identified a species of a certain organism before using online technology known as BLAST. While I primarily used Qiagen CLC Main Workbench software, BLAST provides a similar experience in using an online program to decipher the species of a certain species by matching the DNA. Having this background knowledge is very helpful as it would make it easier to adapt to the Workbench software.

Materials and Methods

The first step in executing this project would be to isolate the *Fusarium* strains from the plants from which they were present. While this may be clear based on the symptoms the infected plants may be showing, the nonpathogenic *Fusarium* strains must be accounted for as well. When the plants appeared to have begun dying, my lab would have begun to process them. This was done by washing off the soil from the plant and cutting through the crown with a knife. Next, one would take note of the appearance of the crown and record its color and how healthy it appears to be. Then the roots and petioles were cut up and their surfaces sterilized. For sterilization, the roots, petioles and crown pieces were inserted into specimen cups in bleach. The petioles and roots only had to be in bleach for one minute while the crown pieces required two minutes. After being soaked in bleach, all three components needed to have been rinsed in three separate batches of water for one minute each. Once fully rinsed, these portions of the plant

would be dried on a paper towel and transferred onto potato dextrose agar triple plus (PDA+++) petri dishes. PDA+++ is a media amended with three antibiotics that is used to reduce growth of bacteria when isolating from plant tissue. After processing the plants in PDA+++ media, if the morphology of fungal colonies resemble that of a *Fusarium* species, then that suggested a *Fusarium* species is causing the disease observed in the plant. Since there are many *Fusarium* species, there are many different appearances that must be accounted for when examining the petri dishes. For instance, *F. oxysporum* appears as an oblique shade of pink, while mycelium produced by *Fusarium* is pink but much more cloudy. This method provided evidence for the potential presence of *Fusarium* in the plant, which stresses the importance of executing the following steps properly for confirmation.

Following the first step, DNA extraction from the gathered colonies would be necessary to identify the DNA in the next step in confirming its species. DNA extraction with the Qiagen DNeasy Plant Mini Kit is a tedious process that requires enzymes, known as lyases, which break down the RNA and leave only the DNA present. Buffers are necessary as well to provide a barrier between the different pHs in the solution containing the DNA. The centrifuge is utilized to rapidly rotate the DNA solutions to separate the DNA from the other material based on their respective densities. To test for the purity of DNA, the Nanodrop would be used to measure the content of DNA and RNA proteins to study the concentration and purity of DNA present in the sample. The 260/280 ratio would be examined because 260 nm is the wavelength of absorbance maxima for DNA and 280 is the wavelength range for the absorbance maxima of RNA (ThermoScientific, 2006). A ratio of about 1.8 would indicate a pure DNA sample, whereas a ratio of about 2 would indicate more RNA in the sample.

DNA barcoding would be performed on an approximately 700 base pair fragment of the

translation elongation factor 1-alpha gene, which is also referred to as TEF-1a. This gene has been shown in existing research to effectively identify isolates of the *F. oxysporum* species complex (O'Donnell, 2009). Table 1 displays the unknown *F.* oxysporum samples that have been selected for this project.

The 25µL PCR reaction consisted of both the forward EF1 primer and the reverse EF2 primer at a final concentration of 1 µM, ensuring that the TEF-1a region of the DNA molecules would be cloned. Each reaction also had 12.5 µL of GoTaq G2 Green Taq MM and 11 µL of nuclease-free water for each reaction. The cycling parameters for the PCR began with a heating of 95°C for five minutes, and then for another minute. These steps allow for the DNA to denature. The next cycle held the DNA at 57°C for one minute and fifteen seconds to allow for annealing. The following cycle holds the temperature of the DNA at 72°C for one minute to execute the extending stage. These cycles were repeated for each isolate, with the following cycle of 72°C being held for ten minutes for more extending. Finally, the last stage was held at 4°C for an extended period of time for cooling of the samples. These conditions were held for 25 separate PCR reactions, one for each isolate, as well as a few isolates for an unrelated project, and water. Water serves as the negative control in this reaction. This means water is run with the intention of showing no results when examining DNA content. If the water were to have any composition of DNA from this experiment, that would imply an incident occurred in the experiment where the samples were likely contaminated.

After PCR was done, PCR cleanup was performed to eliminate excess primers and nucleotides, which ensures the DNA molecules are pure (Exosap-IT, 2000). The DNA product of the PCR reactions was then run on a gel in gel electrophoresis to measure the size of the DNA strands. The content of the gel was made up of 0.3 g agarose and 30 mL 1 x TBE buffer. 1.5 uL

RedSafe Nucleic Acid Staining Solution was also included in the gel, a chemical used to dye the DNA, making it easier to observe. A 100 base pair ladder was used to reference the size of the DNA present on the gel. Here, the reaction with water showed no bands, meaning the reaction was very unlikely to have any contamination. Once the sequences were amplified and cleaned up, Hannah and Lindsey sent them off to the UCR Genomics Core to have them sequenced.

Sanger Sequencing was also involved in this process. Like PCR, this process involved primers, deoxynucleotides, *Taq* polymerase and a DNA template, with a DNA polynucleotide chain growing off of a primer. Additionally, dideoxynucleotides (ddNTP) were also part of this process, which are nucleotides that lack any hydroxyls on the sugar. Their addition to any growing DNA chain from a primer would cease extension of that DNA, since the following nucleotide has no hydroxyl to bind to on the ddNTP. The four ddNTP (ddATP, ddTTP, ddGTP, ddCTP) are fluorescently labeled, allowing them to be clearly visible once bound to the DNA polynucleotide chain. The ddNTPs could bind to the chains formed by the primers at any time, resulting in DNA chains of numerous lengths (Karki, 2018). These chains were then denatured to become single stranded and separated by size using gel electrophoresis, running on thin capillaries. The shorter strands were able to travel further to the positive side of the gel due to being lighter. The ddNTP of each strand was then identified thanks to their fluorescent lighting. Recording the ddNTP starting from the strand that traveled furthest to the one that traveled the least allows us to obtain the desired sequence.

Qiagen CLC Main Workbench is the next step in analyzing sequences, or small segments of DNA. It is a convenient option as this would allow me to work extensively on my project from home while the university campus is still closed.

After the PCR process, Qiagen Sequencing would be required to observe the digital DNA information obtained and determine the species of the *Fusarium* found in the physical lab process. After the elongation factor region has been sequenced, I would compare the resulting species I have identified to that of a reference sequence, which are sequences from isolates that have been positively identified. These references would be obtained from https://fusarium.mycobank.org/, as well as several journal articles. I would then align the sequences in CLC Workbench to confirm my results and construct a phylogenetic tree in CLC Workbench with my results. This would be an accurate secondary method to help confirm the species of *Fusarium*. The list of the known reference isolates chosen for the analysis, which were carefully chosen through peer review, can be seen in Table 2.

Overall, this process does not have too many procedures that must be worked on in the lab, which is very convenient during this period of social distancing. The true objective of this project depends on the ability to properly examine the DNA correctly and to determine the significance of the discovered data.

Once the isolates were sequenced, I then uploaded the raw sequence reads into the CLC Workbench. Most of the VSP isolates had several disagreements between their forward and reverse strands. This prevented them from forming an assembled sequence. These isolates would require the use of a reference sequence, or a sequence whose base makeup is known, and can be used to locate the unclear TEF-1a region in the primers of the lower quality samples. I utilized the reference sequence FJ985270.1 f sp. *fragaria* for the samples that were unable to assemble on their own. The reference sequence provided the endpoints of the strands stated previously to make locating the desired region possible, which allowed me to perform any necessary trimming.

Once the forward and reverse strands had been assembled together as a sequence and polished, the result was a higher quality consensus sequence. The consensus sequence is the sequence of nucleotides in which both primers agree upon in the assembly. With a high quality consensus sequence obtained for each of the isolates, I can now place them on an alignment, which compares each consensus sequence on a single page and identifies any differences within each TEF-1a region. This allows me to identify any mutations among each consensus sequence, including substitutions in transitions and transversion. Additionally, there may also be insertion and deletion mutations. Studying these mutations allows me to identify the differences among each consensus, and to know where I can replace 'N' bases. For example, if one consensus sequence has an 'N' where all the other ones have a specific nucleotide such as 'G', then I can safely substitute that 'N' base with a 'g', which I have lowercase to indicate that it is a guess. By mitigating these unknowns and eliminating any simple sequence repeats, I would be able to polish up the alignment further to have a more accurate phylogenetic tree.

After learning how to assemble sequences and formulating an alignment, I then looked for reference isolates to compare to the ones my lab has obtained for me. This step is important as it allows me to refer my samples to positively identified isolates to help confirm their species, based on whether or not my isolates match any of the reference ones. I examined scholarly sources that involved studies on *F. oxysporum* isolates that could confirm the genetic composition of those particular isolates. Specifically, I looked at sources that included the nucleotides of the translation elongation factor 1-alpha gene of their isolates, the area of the genome I would be examining to identify the species of my isolates. Assuming the source was officially published and had the necessary nucleotide information, I included it as a reference

which would be added into a table seen in the results section of this paper. I was able to accumulate the genetic code of many *F. oxysporum* species this way, providing me a diverse set of species with which to compare the genetic code of my isolates.

I have made sure to include multiple references for each phylogeny in my spreadsheet; this way I would have genetic variations of a species considered when attempting to identify my isolates. Notable sources of isolates include Kerry O'Donnell's journal article where he and his team constructed a two-locus DNA sequence database (O'Donnell et al., 2009). The research behind this project involved studying the translation elongation factor 1-alpha gene of many *F. oxysporum* species, the same portion of DNA that would be examined in this project. As a result, a diverse range of *F. oxysporum* have been identified at the forma specialis level in among a popset available on genbank. These positively identified sequences have been gathered to form an alignment, which would be critical to study the species of the unknown isolates. Additionally, sequences of several outgroups have been identified as well and included in this alignment. According to my faculty mentor, an outgroup is any taxa that is closely related but not the same as my unknown isolates. Given that these unknown isolates very likely belong to the *F. oxysporum* species, two outgroups have been included from different species belonging to the *Fusarium* genus. These two species are *Fusarium commune* and *Fusarium foetens*. Based on the alignment, the DNA of their translation elongation factor 1-alpha gene is very similar, which would make it less difficult to root and orient the eventual phylogeny tree of the reference alignments.

I then took the reference isolates along with the unknown isolates and formed an alignment of them to make it easier to trim the ends of each sequence. Additionally, forming an

alignment would help locate any mutations among the nucleotides. After aligning them, I trimmed the ends so that only the high quality material was present. Afterwards, I ran a model test, where the CLC Main Workbench applied four models to the alignment I made to determine which one is the most effective for creating a phylogenetic tree. The five models tested were Jukes-Cantor, Felsenstein 81, Kimura 80, Hasegawa Kishino-Yano, and General Time Reversible. There were four tests made for these models, starting with the Hierarchical Likelihood Ratio Test (hLRT). For this test, the Workbench utilized a null hypothesis for each model as well as an alternative hypothesis to calculate which model creates the best tree. It considered the difference between the estimated parameters of the two hypotheses and depending on the p-value, accepted or rejected the null hypothesis. The next test, the Bayesian Information Criterion (BIC) model test, considered the number of parameters and BIC of each model. The lower the BIC number for the model, the more viable the model is. The following test is the Akaike Information Criterion (AIC) Model Test, which predicts the error of each model. According to the Workbench, this test also considered the parameters of each model as well as the ratio between the length of alignment over the number of parameters. Once again, the lower the number for AIC, the model is more competent. Finally, there was the Akaike Information Correction criterion (AICc) Model Test, which used several formulas to determine the most accurate model. Both AIC and AICc are used because while AIC is more applicable for all models in general, while AICc provides less bias (Cavanaugh, 2004). Overall, the Workbench concluded that the best model would be Hasegawa Kishini-Yano (HKY) with topology variation. I proceeded to make a Maximum Likelihood phylogeny tree with the HKY model, which resulted in a phylogeny tree of all of my reference isolates. It helps scientists see how closely

related the isolates are to one another, identifying any homologous, paralogous, and orthologous relationships (Kapli, 2020). It is important to note that this same procedure would be followed with the *Fusarium* isolates I would have received from my lab.

The resulting phylogeny tree determined that *F. commune* and *F. foetens* were the most distinct isolates out of all the samples, which makes sense considering they are the outgroups. The next two isolates most closely related to the outgroups were *F. oxysporum* f. sp. *gladoli* and *F. oxysporum* f. sp. *cubense*. The rest of the samples were much more closely related to each other within several groupings of branches. The organization of this phylogeny tree would make it more convenient in identifying the unknown isolates based on how closely their DNA matches a particular branch on the tree.

When developing the phylogeny tree, it is important to ensure a bootstrap analysis is completed. The bootstrap values are essentially a probability of how likely it is for the branching patterns present to have occurred. These values are accessed by the nucleotide components of each sample, determining how likely it is for one sample to have descended from another. The bootstrap values are displayed as percentages between each descending sample, beginning from the common ancestor. For this project, a bootstrap value of at least 50% is preferred. This would mean that branches are only considered valid if there is at least a 50% possibility of the descending pattern to have occurred between one sample and another. If the bootstrap value is less than 50%, then that branch and the rest of the samples descending from it are no longer considered. After utilizing the bootstrap feature myself, I found that many of the samples that diverged further from the common ancestor had lower bootstrap values, meaning that on this tree, there is less confidence in the descending pattern for the more diverse samples.

After applying the bootstrap data and accessing the results, the next step is to root the tree. The purpose of rooting the tree is to hypothesize the ancestor of the aligned sequences. This step helps visualize the tree and interpret data from it more easily. When rooting the tree, it is important to find where the outgroups diverge from the rest of the samples. This emerging edge on the tree would indicate the region from which the outgroups evolved with respect to the samples I have gathered. This step would make it easier to visualize the tree, as it would then become more organized.

Results

I was able to get a consistent end point for the 3' end of the sequences, being 'GTCACC' the beginning point of the primers was 'GTCGAC'. The quality of the samples is very high. This allowed me to have no issue with assembling the sequences. I was able to trim the forward and reverse strands and ensure a high quality consensus sequence for the alignment.

The alignment I formed depicted the minor differences among the samples, mainly a few substitution mutations in transversions and transitions. I only noticed two indel mutations throughout all the unknown isolates, with an insertion in VSP-83 around nucleotide 620, and a deletion in VSP-57 around nucleotide 600.

Once the tree was developed, I noticed the bootstrap levels were sufficient for most of the samples, with only a few unknown isolates having a poor bootstrap level. This does not mean that their species cannot be identified, but rather the confidence for those specific results according to the tree is not that high. The tree presents all but two of the unknown samples clustered among the reference oxysporum sequences with high bootstrap support of 50%. Unknown isolates VSP-233 and VSP-258 clustered with the outgroup samples, suggesting they

are outgroups of oxysporum as well. While it is difficult to pinpoint the exact forma speciales of each *Fusarium* sample from the phylogeny tree alone, it is still important as they are matched with identified samples.

The phylogeny tree created can be seen in Figure 1. Each unknown sample is closely related to one or two identified reference samples, which significantly narrows down their *forma specialis*. Host species *Fragaria ananassa* was the source of many of the isolates in this project. Those isolates are clustered into two groups on the tree, one on the top and one towards the bottom. The bottom cluster is associated with the reference f. sp. *fragaria,* while the top cluster shows an association with references with f sp. *vanillae* and f. sp. *bouvardiae.*

Isolates derived from the host species *Apium graveolens* also appeared clustered in the same bottom region as the isolates from *Fragaria ananassa,* and nowhere else on the tree. The isolates from the other two host species, *Solanum lycopersicum* and *Capsicum annuum,* are clustered in the top of the tree as well near the reference f. sp. *bouvardiae.*

It appears that with the development of the phylogeny tree, each unknown isolate is closely related to an identified *forma specialis*. It is worth noting that VSP-233 and VSP-258 have been excluded from the tree since they are outgroups. Their inclusion in the tree resulted in a long branch from the outgroups that would make it difficult to view the rest of the isolates. Below is a table which connects each unknown to one or two identified *forma specialis,* based on how closely related they are on the tree seen in Table 3.

Regarding the majority of the reference samples, many of them are deemed nearly identical according to the tree. The tree features several clusters of reference isolates that are lined up on top of one another within each branch. Their nodes are present on a vertical line,

meaning they lack any polymorphic sites between them. The alignment confirms this as well, showing no differences between the TEF-1a genes of these particular references.

Discussion

Perhaps the biggest takeaway from the tree is the clusters of identical references present throughout the tree. This is the first point to target as I continue working on this project. Having many identical references does not provide much significance in analyzing the data since they are exactly the same. If one of those references does not match closely to a VSP isolate, then none of them would, providing almost no useful information. This result proves that it is very important to be aware of gathering references that are not only from reliable sources, but also have enough polymorphisms in their TEF gene to help identify the f. sp. of the unknown isolates..

Most of the unknown isolates are closely related to a reference sample upon the same branch. Some unknown isolates, such as VSP-310 or VSP-293, are considerably further from identified samples. This makes it difficult to identify them to the *forma specialis* level seeing that they are distant from the reference samples on the tree. These unknowns need more time before being more confidently identified, perhaps through developing another tree with another batch of reference samples.

An interesting observation is while most of the unknown isolates were obtained from *Fragaria ananassa*, they all did not cluster together on the tree. There does not appear to be a particular correlation between the host species and the presence of an unknown isolate. Otherwise, all the isolates derived from *Fragaria ananassa* would have most likely accumulated off one particular branch on the tree. This indicates that these *forma specialis* are host specific in

that they target *Fragaria ananassa.* However, they can still grow on other plant species. Since there is not one cluster of isolates from the *Fragaria ananassa* host species, then there either are too few polymorphisms on the TEF gene to distinguish the *forma specialis* of the isolates, or that the unknown isolate was not the source of infection on the plant. There is in fact a possibility that there was a different *Fusarium* fungi present on the plant that caused the infection seen.

It is more difficult to analyze any correlation among the other host species used in this project since there were only one to two isolates gathered from them. Where isolates gathered from *Fragaria ananassa* were grouped into two clusters, the other host species were only seen in one cluster. This may narrow down the VSP isolates from them to more specific f. sp. based on the reference samples. However, there needs to be more research done with a better gathering of references to confirm that idea. Additionally, this would contradict the immense diversity seen in the infection ability of most *Fusarium* species.

The bootstrap levels were a considerable issue when developing the tree. Several of the outer branches had low bootstrap levels which had to have been modified by the bootstrap threshold. The bootstrap threshold value would rearrange the lower quality branches to have an acceptable bootstrap level, which for this tree was 50%. As a result, the outer branches may not be as accurate when depicting the relationship between the unknown and reference samples. This may explain why many of the unknowns on the top of the tree are depicted as being related to *f sp. bouvardiae*. Upon further examination of the alignment, however, the samples listed as being related to f. sp. *bouvardiae* on the tree appear to be very closely related by their nucleotide composition within the TEF-1alpha region. They have little to no variation in comparison to that reference sample. Overall, the phylogeny tree is a major step in identifying unknown isolates, as

this practically confirms they are of the *F. oxysporum* species. While the exact f sp. is unclear, this is a significant step in determining that level of identification as the tree helps narrow it down with reference samples. The next step is to further study the TEF-1alpha regions of the samples to more precisely identify their f. sp..

Examining the VSP samples further on the tree, it is clear that some samples are better identified than others. This judgement is based on the fact that some of the VSP samples are much further from the references on the tree than others. Examples of well identified samples include VSP-219, VSP-296, and VSP-265. These samples appear to be closely related to references on the tree, providing more confidence as to how well identified they are. On the contrary, there are several samples that are rather distant from reference samples on the tree. These samples include VSP-221, VSP-476, and VSP-310. On the alignment, they appear to be similar to existing reference samples. Despite this similarity, they are rather far from those reference samples on the actual tree, raising doubt over their potential f. sp.. This is one major area that requires further research for the future of this project.

This project focused on *Fusarium* isolates engaged heavily on the precision necessary for developing high quality sequences for the alignment. I initially thought my experience with sequences in my biology lab freshman year would make the procedure for this project more clear. In reality, there were many steps that I was required immense training for. While the procedure was not necessarily long, there were many small steps to consider for forming the consensus sequences. If any of these steps were forgotten about or done incorrectly, then the entire alignment would have been affected. I had to reattempt building the alignment numerous times, due to the fact that I would make minor errors when building the consensus sequences.

The most common error I had was with the forward and reverse primers, where I would leave insertions that were present in one of the primers in the consensus sequence. It was very important to identify and remove these insertions, as they would shift the entire sequence, which would compromise the alignment. Another issue I had was depending too much on relying on a reference sequence for assembling the unknown sequences. This mistake led to more insertions in the consensus sequence, which diminished the quality of my unknown samples. Fixing this issue allowed me to obtain a consensus sequence that was less reliant on the data of the reference sequence I used, which meant fewer errors were present.

The fact that there are only two minor mutations among the unknown isolates is certainly favorable regarding the reliability of the results. Given that the TEF-1alpha complex serves the same function in all the isolates, we would expect them to be very similar genetically. If there were significant differences among the unknown isolates on this alignment, that would make the reliability of this alignment questionable.

One method of confirmation I found without relying on the Workbench was assessing the relationships of the unknowns with the references simply based on the tree. VSP-212, for example, appeared to be most closely related to reference f. sp. *batatas* on the tree. Consistently, VSP-212 most closely matched f. sp. *batatas* on the alignment. Another instance includes VSP-296, which closely matches f. sp. *fragariae* and f. sp. *apii* genetically on the alignment. As a result, VSP-296 is also very closely related to these references on the tree too.While this is most likely explained through the CLC Workbench algorithm analyzing the sequences, it is useful to have a way of confirming without having to rely solely on the program. Visually speaking, the organization of the alignment matches very closely to what is apparent on the tree.

Based on the alignment, VSP-233 and VSP-258 had a TEF-1a gene that matched the outgroups of the tree much more closely than any of the other VSP isolates. This supports the results seen on the tree depicting them as not being *oxysporum.* Out of curiosity, I uploaded the sequences of VSP-233 and VSP-258 onto BLAST since they were considered outgroups according to my data. As a result, BLAST also labeled them as *Fusarium* fungi outside of the *oxysporum* species. This helps support the fact that they are outgroups. While BLAST is not entirely reliable, it is ensuring to observe it providing conclusions that are consistent with my results. This provides more confidence in the rest of the results gathered through this project.

The process taken in identifying these unknown samples has great potential with identifying other unknown samples in the greater scientific world. With the possibility of new or unknown organisms being discovered at any point in time, it is very useful to have a practical method of studying their DNA. DNA barcoding is a tool that can be used to identify the species of an organism based on a short portion of its DNA, as opposed to its entire genome. This is done by comparing that segment of DNA to segments of DNA uploaded for positively identified species online. By using this method, it is possible to identify an unknown organism based on its genetic components alone. While tools like BLAST exist online which allow anyone to use DNA barcoding, it is important to go through the process of forming an alignment and phylogeny tree such as in this project. My faculty mentor specified that this is because anyone can upload DNA sequences to the database of genomes, meaning that they may not always be accurate. By having a phylogeny tree and alignment to reference, this grants a source of confirmation to ensure that the results we obtain from a program such as BLAST are actually supported through extensive research. This project was a significant journey in introducing me to the challenges that come

with performing scientific research, specifically in the field of DNA barcoding.

Works Cited

Cavanaugh, Joseph E. "Unifying the Derivations for the Akaike and Corrected Akaike Information Criteria." *Statistics & Probability Letters*, North-Holland, 9 June 1999, www.sciencedirect.com/science/article/pii/S0167715296001289?via%3Dihub.

Cortés-Ciriano, Isidro, et al. "Comprehensive Analysis of Chromothripsis in 2,658 Human Cancers Using Whole-Genome Sequencing." *Nature News*, Nature Publishing Group, 5 Feb. 2020, www.nature.com/articles/s41588-019-0576-7.

Edel-Hermann, V., and C. Lecomte. "Current Status of Fusarium Oxysporum Formae Speciales and Races." *Phytopathology*®, vol. 109, no. 4, 19 Nov. 2018, pp. 512–530., doi:10.1094/phyto-08-18-0320-rvw.

ExoSAP-IT, USB Corporation , 2000.

Ivors, Kelly. "PROTOCOL 06-07.1: DNA Isolation from Phytophthora Mycelium Using the DNeasy Plant Mini Kit (QIAGEN)." *Laboratory Protocols for Phytophthora Species*, 6 Sept. 2016, pp. 1–2., doi:10.1094/9780890544969.06.07.1.

Kadziene, Grazina, et al. "Tillage and Cover Crop Influence on Weed Pressure and Fusarium Infection in Spring Cereals." *Crop Protection*, Elsevier, 20 Sept. 2019, www.sciencedirect.com/science/article/abs/pii/S0261219419303126.

Kapli, Paschalia, et al. "Phylogenetic Tree Building in the Genomic Age." *Nature News*, Nature Publishing Group, 18 May 2020, www.nature.com/articles/s41576-020-0233-0.

Karki, Gaurab. "Sanger's Method of Gene Sequencing." *Online Biology Notes*, 18 Dec. 2018,

www.onlinebiologynotes.com/sangers-method-gene-sequencing/

Knights, E.J. “Fusarium.” *Fusarium - an Overview | ScienceDirect Topics*, Science Direct, 2004, www.sciencedirect.com/topics/food-science/fusarium.

Ma, Li-Jun, et al. “Comparative Genomics Reveals Mobile Pathogenicity Chromosomes in Fusarium.” *Nature News*, Nature Publishing Group, 18 Mar. 2010, www.nature.com/articles/nature08850.

Maymon, Marcel, et al. “The Origin and Current Situation of Fusarium Oxysporum f. Sp. Cubense Tropical Race 4 in Israel and the Middle East.” *Nature News*, Nature Publishing Group, 31 Jan. 2020, www.nature.com/articles/s41598-020-58378-9.

Nucci, Marcio, and Elias Anaissie. “Treatment and Prevention of Fusarium Infection.” *UpToDate*, UptoDate, 20 Apr. 2007, www.uptodate.com/contents/treatment-and-prevention-of-fusarium-infection. R. Alexander,

O' Donnell, Kerry, et al. “A Two-Locus DNA Sequence Database for Typing Plant and Human Pathogens within the Fusarium Oxysporum Species Complex.” *Fungal Genetics and Biology : FG & B*, U.S. National Library of Medicine, 26 Aug. 2009, pubmed.ncbi.nlm.nih.gov/19715767/.

O'Donnell, Kerry, et al. “Fusarium Translation Elongation Factor 1 Alpha Gene, Partial Cds. - PopSet - NCBI.” *National Center for Biotechnology Information*, U.S. National Library of

Medicine, 26 Aug. 2009, www.ncbi.nlm.nih.gov/popset/247892137.

R. Mendoza, et al. "Biological Control of Fusarium Wilt in Crop Plants Using Non-Pathogenic Isolates of Fusarium Species." *Indian Phytopathology*, Springer India, 1 Jan. 1970, link.springer.com/article/10.1007/s42360-020-00202-5.

Raja, Huzefa A, et al. "Fungal Identification Using Molecular Tools: A Primer for the Natural Products Research Community." *Journal of Natural Products*, American Chemical Society and American Society of Pharmacognosy, 24 Mar. 2017, www.ncbi.nlm.nih.gov/pmc/articles/PMC5368684/.

Richard, John L. "Some Major Mycotoxins and Their Mycotoxicoses-An Overview." *International Journal of Food Microbiology*, Elsevier, 20 Oct. 2007, www.sciencedirect.com/science/article/pii/S0168160507003790?casa_token=GoHq-eCM3poAAAAA%3ASapRA6fW44_TG31Dcj9qrSucVejfcAbvljTfLrA5vTqyqage1KXYcIcLG2tnBDbPqLGo-X823H8.

Scott, J, et al. "Archive." *California Agriculture*, University of California, Agriculture and Natural Resources, 1 Jan. 2012,calag.ucanr.edu/archive/?article=ca.v066n01p20.

Soriano-Martín, M.L., et al. "Use of Microwaves in the Prevention of Fusarium Oxysporum f. Sp. Melonis Infection during the Commercial Production of Melon Plantlets." *Crop Protection*, Elsevier, 13 June 2005, www.sciencedirect.com/science/article/abs/pii/S0261219405001055.

Stankovic, S.; Levic, J.; Petrovic, T.; Logrieco, A.; Moretti, A. Pathogenicity and mycotoxin production by Fusarium proliferatum isolated from onion and garlic in

Serbia.Eur. J. Plant Pathol.2007,118, 165–172.

ThermoScientific. *Interpretation of Nucleic Acid 260/280 Ratios*, Thermo Scientific NanoDrop Products , 2006.

Table 1. Suspected VSP *F. oxysporum* isolates obtained from plants exhibiting symptoms of vascular wilt disease.

Isolate Identification Number	Host	City	County
57	*Fragaria ananassa*	Irvine	Orange
71	*Solanum lycopersicum*	Bakersfield	Kern
73	*Fragaria ananassa*	Oxnard	Ventura
83	*Fragaria ananassa*	Irvine	Orange
212	*Fragaria ananassa*	Santa Maria	Santa Barbara
218	*Fragaria ananassa*		Ventura
219	*Fragaria ananassa*		Ventura
221	*Fragaria ananassa*	Nipomo	San Luis Obispo
222	*Apium graveolens*	Thermal	Riverside
225	*Apium graveolens*	Thermal	Riverside
233	*Fragaria ananassa*		Ventura
258	*Apium graveolens*	Mecca	Riverside
265	*Fragaria ananassa*	Nipomo	San Luis Obispo
293	*Capsicum annuum*		Riverside
303	*Fragaria ananassa*		Ventura
320	*Fragaria ananassa*	Oxnard	Ventura
337	*Fragaria ananassa*	Nipomo	San Luis Obispo
371	*Fragaria ananassa*	Ventura	Ventura
476	N/A	N/A	N/A
310	*Fragaria ananassa*	Guadalupe	Santa Barbera
296	*Fragaria ananassa*		Ventura

Table 2. Known reference isolates of the Genus *Fusarium* used for phylogenetic analysis

Species complex	Species Name	f. sp/var identification	Country Discovered:	Year Discovered:	Accession #
oxysporum	*oxysporum*	var. *meniscoideum*	USA	2015	FJ985293.1
oxysporum	*oxysporum*	f. sp *lavandulae*	Italy	2015	FJ985293
oxysporum	*oxysporum*	f. sp. *bouvardiae*	USA	2013	FJ985282.1
oxysporum	*oxysporum*	f. sp. cassiae	USA	2009	FJ985294.1
oxysporum	*oxysporum*	f. sp. albedinis	USA	2006	DQ837688.1
oxysporum	*oxysporum*	f. sp. arctii	USA	2009	FJ985289.1
oxysporum	*oxysporum*	f. sp. allii	USA	2009	FJ985288.1
oxysporum	*oxysporum*	f. sp. apii	USA	2020	MT295485.1
oxysporum	*oxysporum*	f. sp. callistephi	USA	2006	DQ837679.1
oxysporum	*oxysporum*	f. sp. cattleyae	USA	2009	FJ985268.1
oxysporum	*oxysporum*	f. sp. matthiolae	USA	2007	DQ837682.1
oxysporum	*oxysporum*	f. sp. cucurbitacearum	USA	2009	FJ985283.1
oxysporum	*oxysporum*	f. sp. fabae	USA	2006	DQ837684.1
oxysporum	*oxysporum*	f. sp. melongenae	USA	2009	FJ985297.1
oxysporum	*oxysporum*	f. sp. loti	USA	2007	EU313517.1
oxysporum	*oxysporum*	f. sp. rhois	USA	2007	DQ837683.1
oxysporum	*oxysporum*	f. sp. phaseoli	USA	2007	DQ837686.1
oxysporum	*oxysporum*	f. sp. heliotropii	USA	2006	DQ837685.1
oxysporum	*oxysporum*	f. sp. vanillae	USA	2009	FJ985300.1
oxysporum	*oxysporum*	f. sp. elaeidis	USA	2009	FJ985270.1
oxysporum	*oxysporum*	f. sp. lycopersici	Belgium	2009	FJ790393.1
oxysporum	*oxysporum*	f sp. radicis-lycopersici	Belgium	2009	FJ790410.1
oxysporum	*oxysporum*	f sp. fragariae	USA	2019	MN609989.1
oxysporum	*oxysporum*	f. sp. vasinfectum	Spain	2013	KF928931.1
oxysporum	*oxysporum*	f. sp. lilii	Spain	2013	KF928931.1
oxysporum	*oxysporum*	f. sp. lycopersici	Spain	2013	KF928931.1
oxysporum	*oxysporum*	f. sp. radicis-lycopersici	Spain	2013	KF928931.1
oxysporum	*oxysporum*	f. sp. fragariae	Turkey	2014	KJ776747.1
oxysporum	*oxysporum*	f. sp. lactucae	Italy	2019	MK801785.1
oxysporum	*oxysporum*	f. sp. apii	USA	2020	MT295485.1
oxysporum	*oxysporum*	f sp. corianderii	USA	2020	MT295492.1
oxysporum	*oxysporum*	f. sp. lactucae	USA	2018	MH412702.1

oxysporum	*oxysporum*	f sp. fragariae	USA	2016	KX456097.1
oxysporum	*oxysporum*	f sp. fragariae	USA	2016	KX456061.1
oxysporum	*oxysporum*	f sp. batatas	USA	2009	FJ985368.1
oxysporum	*oxysporum*	f sp. dianthi	USA	2009	FJ985284.1
oxysporum	*oxysporum*	f sp. niveum	USA	2009	FJ985410.1
oxysporum	*oxysporum*	f sp. hebes	USA	2009	FJ985423.1
oxysporum	*oxysporum*	f. perniciosum	USA	2009	FJ985413.1
oxysporum	*oxysporum*	f sp. tracheiphilum	USA	2009	FJ985343.1
Outgroup	*foetens*	N/A	USA	2009	FJ985444.1
nisikadoi (Outgroup)	*commune*	N/A	USA	2009	FJ985440.1

Table 3: Identified VSP isolates based on their positioning on the Phylogeny Tree

Group	Number	Host Species	Possibly Related f sp.(s)
VSP	57	*Fragaria ananassa*	f sp. *tulipae*
VSP	71	*Solanum lycopersicum*	Distantly related to f sp. *bouvardiae* and f sp. *vanillae*
VSP	73	*Fragaria ananassa*	Distantly related to f sp. *bouvardiae* and f sp. *vanillae*
VSP	83	*Fragaria ananassa*	Distantly related to f sp. *bouvardiae* and f sp. *vanillae*
VSP	212	*Fragaria ananassa*	f sp. *vanillae*
VSP	218	*Fragaria ananassa*	f sp. *vasinfectum*
VSP	219	*Fragaria ananassa*	f sp. *corianderii*
VSP	221	*Fragaria ananassa*	f sp. *bouvardiae*
VSP	222	*Apium graveolens*	f. *melongenae*
VSP	225	*Apium graveolens*	f. *melongenae* or f *sp. corianderii*
VSP	233	*Fragaria ananassa*	Outgroup
VSP	258	*Apium graveolens*	Outgroup
VSP	265	*Fragaria ananassa*	f sp. *bouvardiae*
VSP	293	*Capsicum annuum*	Distantly related to f sp. *bouvardiae*
VSP	303	*Fragaria ananassa*	f sp. *fragaria*
VSP	320	*Fragaria ananassa*	f sp. *callistephi*
VSP	337	*Fragaria ananassa*	Distantly related to f sp. *bouvardiae*
VSP	371	*Fragaria ananassa*	f sp. *vasinfectum*
VSP	476	N/A	f sp. *conglutinans* or f sp. *cucurbitacearum*
VSP	310	*Fragaria ananassa*	Distantly related to f sp. *bouvardiae* and f sp. *vanillae*
VSP	296	*Fragaria ananassa*	f sp. *fragaria*

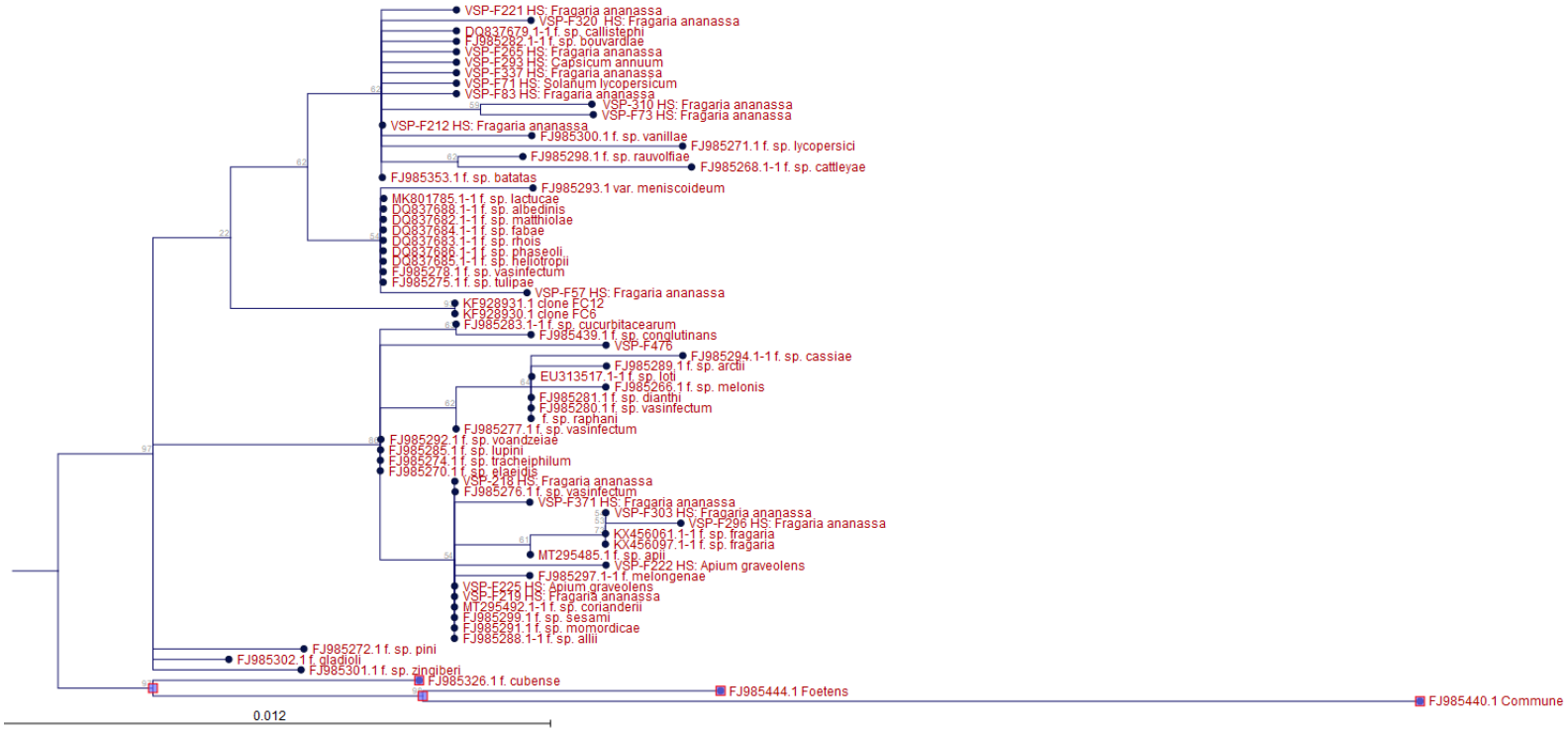

Figure 1. Phylogeny of unknown isolates and reference strains inferred using maximum likelihood.

Note: HS = Host species