MOLECULAR TOOLS FOR DIAGNOSING PLANT HEALTH ISSUES

By

Kevin A. Korus

University of Nebraska

Doctor of Plant Health

Under the Supervision of Professor Gary L. Hein

Lincoln, NE

May 6th, 2016

Abstract

This document discusses the methodologies and gene targets involved in molecular diagnostics for plant associated pests and pathogens. It explores the capabilities and limitations of molecular diagnostics in a routine laboratory as well as the propensity for these tests to be used to identify cryptic or new organisms. A case study is presented that details the methods used to identify an emerging bacterial pathogen of corn in Nebraska. The most successful method for identifying pests and pathogens is through sequence analysis of target genes or gene sequences. The limitations of this method are the result of yet too poorly supplemented databases from which to compare unknown sequences. The success of molecular diagnostics depends on the addition of verified specimens to publicly available DNA sequence databases. Diagnoses made in a routine laboratory may be supplemented by molecular data.

Dedication

This Document is dedicated to my family; Mom, Dad, Gina, Jenny, Jesse, Justin, and Les. Thank you for your influence and unending support of my scholastic pursuits.

Acknowledgements

I would like to give special thanks to the following people who have made this degree possible: Drs. Gary Hein, Gerard Adams, Roger Elmore, and Bob Wright; as well as Nancy Shoemaker, for their support of the Doctor of Plant Health program. Dr. Jim Steadman and the Extension Plant Pathology team for their mentorship and understanding of my role as both a student and professional. The USDA-APHIS-PPQ Beltsville team for their hard work developing insightful and informative diagnostic workshops: Gloria Abad, Mark Nakhla, John Rascoe, Kurt Zeller, John Bienapfl, and Stefano Costanzo. To our collaborators, Dr. Jan Leach and Jillian Lang, (Colorado State University) and Dr. Charles Block (Iowa State) for their support in diagnosing the bacterial corn disease in Nebraska discussed in Chapter II. I would also like to thank Dr. Carolina Camargo for her encouragement, love and support throughout my time as a professional student.

Contents

Index of Figures and Tables

CHAPTER I MOLECULAR TOOLS FOR DIAGNOSING PLANT HEALTH ISSUES

1.1 Introduction

Increased understanding of genomes and genotyping (whole genome sequencing) of microorganisms has led to the development of molecular tools for identification and aids diagnosing plant health issues. With increased specificity, these tools are reducing false positives and negatives associated with identifying causal organisms in a plant health crisis (López et al., 2003). The goal of my internship was to learn molecular methods for pest and pathogen detection to increase the capabilities of the Plant and Pest Diagnostic Clinic (PPDC) at the University of Nebraska – Lincoln. During five workshops (Table 1.1), hands on experience conducting pathogen DNA extraction was practiced to ensure amplifiable target DNA. Experience was gained in evaluation of different polymerase chain reaction (PCR) methodologies that can be employed in diagnosing, depending on the target and end product needed. Sequence analysis was conducted on target amplicons and the limitations of conducting database queries were explored. The workshops also focused on the design and selection of effective primers for organismal and species differentiation between plant pathogens, including viruses, nematodes, bacteria, phytoplasmas, fungi, oomycetes and insects. Finally, I learned fundamental bioinformatics in using nucleotide and amino acid sequence data with genomic databases to explore phylogenetics and population genetics. Previously in the PPDC, when molecular data were needed to support a diagnosis, samples or isolations in culture were sent to another diagnostic laboratory within the National Plant Diagnostic Network (NPDN). Following completion of this internship, I have the competency necessary for performing DNA analysis and resources required to perform the work have been purchased and calibrated for the PPDC under my leadership.

When plant pest, disease or disorder samples are processed by the PPDC, each sample is recorded in an online database called the Plant Diagnostic Information System 2.0. These samples are then uploaded to the NPDN National Repository. When a diagnosis is made, the diagnostician must assign a

confidence level for the genus, species, and subspecies level (if applicable). Confidence levels are selected from the established qualifiers: confirmed, suspected, not detected, and inconclusive. There are many organisms that are easily recognizable to the genus level via identification of morphological characteristics, but they cannot be precisely distinguished to the species level. Fungi, belonging to the genus *Fusarium*, are a great example. It is easy to identify a *Fusarium* culture by morphological characteristics of macroconidia, but identifying it to species by morphology alone is not practically achievable. Often, when *Fusarium* is identified in the clinic, the genus is confirmed, but the species is only approximated based on host index records, symptomology, and disease progress. Molecular tests offer an achievable supplemental means that can now be employed that improve confidence in species identification.

The following chapter discusses the molecular methods available for analyzing DNA and how these methods have been applied for the detection of plant associated insects and the five plant pathogen groups; fungi, bacteria, viruses, nematodes, and phytoplasmas. It is important to note that the field of genomics is advancing rapidly, and these tools and methods are constantly being upgraded or modified, and that they will become of increased value as databases expand with time.

1.2 Molecular Tools and Techniques

DNA is the starting material for all life, and each individual has their own unique DNA "fingerprint". The isolation and characterization of DNA has given us the ability to identify organisms based on the uniqueness of their genetic code. Nucleic-acid hybridization is the process where single stranded (ss) DNA or RNA anneals to its complimentary DNA or RNA strand, respectively. When temperatures are increased, double stranded (ds) DNA and RNA denatures into single complimentary strands. When temperatures cool, ssDNA and ssRNA hybridizes with a complimentary strand (Felsenfeld and Miles, 1967). Nucleic-acid hybridization is the key to DNA replication and transcription and has enabled development of assays like Southern Blots, Northern Blots, PCR, Sanger sequencing and various next generation approaches to sequencing.

Fluorescence *in situ* hybridization (FISH) is a technique that involves using radioactively labeled gene sequences to hybridize to their respective gene on a given chromosome. The binding is very specific, and has allowed researchers to locate a particular gene within the genome (Levsky and Singer, 2003).

Restriction fragment-length polymorphism (RFLP) is a technique that uses restriction enzymes to cleave DNA into fragments. Restriction enzymes (RE) can be developed to detect a 4-6 base pair (bp) region of the genome and make a cut wherever this specific region occurs (Miller and Martin, 1988). Fragmented DNA is then separated by agarose-gel electrophoresis, stained with ethidium bromide, and viewed under UV lighting. When DNA is cut with restriction enzymes, a unique "barcode" develops from the resulting fragments (Miller and Martin, 1988). This barcode can be compared across different species that were cut with the same RE. Restriction fragments differ in both number and size between species because of inserts or deletions between existing RE sites or from mutations that either create or destroy restriction sites (Miller and Martin, 1988; Jenkins et al., 2012). RFLPs of total genomic DNA are too labor intensive, time consuming, and expensive for routine pathogen detection, but they are useful when differentiating between two closely related organisms. However, RFLPs of PCR (RFLP-PCR) amplified products are easily incorporated into a rapid technology for diagnosis.

Amplified fragment-length polymorphism (AFLP) analysis is a similar technique to RFLP analysis. Primers developed for this method target adaptors that are ligated to DNA that has digested with RE producing fragments with compatible ends (Hoy, 2003; Jenkins et al., 2012). The adapter and restriction site sequences become targets for primer annealing and sets of restriction fragments are selectively amplified (Hoy, 2003). The products are then evaluated via electrophoresis. This methodology is expensive and requires considerable preliminary data collection using a range of RE, before direct use for diagnosis with one select RE.

Conventional Polymerase Chain Reaction (PCR) was developed by Kary Mullis in 1983, and it involves the exponential duplication of a single copy of a gene sequence (López et al., 2013; Makkouk and Kumari, 2006). Thermal cycling is used to continually heat and cool the reaction, allowing the DNA

to denature and for enzymatic replication to take place (Makkouk and Kumari, 2006). Target-specific primers (also called oligonucleotides) are used to match complimentary DNA sequences within specific regions of the genome. A heat stable DNA polymerase is used to enable replication of the DNA during the single-stranded condtion. The reaction contains free deoxynucleotides (dNTPs) that are assembled into a new complimentary strand. The newly assembled and replicated strands are then separated by electrophoresis, and their relative sizes are estimated by using a DNA ladder consisting of oligonucleotides of various known lengths. The characteristics of the starting template, cycling parameters (annealing time/temperature), purity, and concentration of dNTPs, buffer composition/stability, and the type of polymerase all affect the sensitivity and efficiency of PCR (López et al., 2003). The ease, rapidity, flexibility, and low expense of conventional PCR make it a well suited methodology for detection and identification of causal agents in plant health diagnosis.

Quantitative Real-Time PCR (qPCR) is a method of DNA replication that allows for the results to be read as the reaction is progressing (Makkouk and Kumari, 2006). DNA probes labelled with fluorescent markers are added to the reaction, and a recordable amount of light is emitted every time there is hybridization at the gene region where the probe annealed (Makkouk and Kumari, 2006). If the target region is present and the primers and probes work correctly, then the fluorescent light will be emitted in an exponential fashion as the reaction progresses (Makkouk and Kumari, 2006). Figure 1.1 shows an example of a positive and a negative reaction. Because DNA replication is done in an exponential fashion, the amount of starting template can be calculated from the amount of product recorded during any part of the reaction (López et al., 2013). Therefore, this method is useful for quantifying plant pathogen titers in host tissue and insect parasitoid numbers inside their insect host (Hoy, 2003; Jenkins et al., 2012). Performing qPCR requires an expensive investment in a specialized thermocycler capable of capturing and recording the fluorescence emitted from the labelled probes. This method is extremely sensitive and can be utilized to process many samples simultaneously with immediate results (Hoy, 2003). Quantitative PCR is best suited for larger laboratories that diagnose high sample volumes as the

initial cost of the equipment and its maintenance is relatively high and reagent kits are a routine additional high cost (Hoy, 2003).

Reverse transcription when coupled with PCR is commonly abbreviated as RT-PCR and is often confused with the abbreviation for real-time PCR. Reverse transcription is required to make complimentary DNA (cDNA) when RNA is the starting template. Oligonucleotide primers on either side of the target gene are extended by a thermostable DNA polymerase in a series of denaturation and extension steps, much like PCR (Webster et al., 2004). RT-PCR is routinely employed for the detection of RNA plant viruses (López et al., 2003; Makkouk and Kumari, 2006; Uyeda and Masuta, 2015; Webster et al., 2004). RT-PCR also is used to make cDNA from messenger RNA, which is then quantified to give an understanding of gene expression in specific tissues or even cells (Hoy, 2003). RT-PCR is complicated by the special handling required for effective RNA isolation and utilization which requires specially treated and handled glassware and plasticware free of the ubiquitous and temperature stable RNases.

Nested PCR is a technique that was developed to reduce the likelihood of primers recognizing incorrect sequences. With nested PCR, two sets of primers are used in two successive runs of PCR. The second set of primers is intended to amplify a secondary target within the product of the first reaction (Hoy, 2003). The first reaction in nested PCR increases the amount of template for the second. This greatly increases the sensitivity and specificity of PCR (Hoy, 2003; Webster, 2004). This process is more specific, but it is also more time consuming and costly since there is a need to run two successive reactions. Nested PCR has proven useful for virus detection when they are present in low titers. It is also useful if DNA polymerase inhibiting compounds are present from plant hosts (Webster et al., 2004). The second round of PCR serves as a dilution to further remove PCR inhibiting compounds (Webster et al., 2004). Nested PCR is subject to false positives due to the ease in which PCR products can cross contaminate test samples in a laboratory environment, and therefore must be employed by particularly careful practioners in rigidly clean labs usually with work areas segregated for particular tasks.

Multiplex-PCR is another commonly used method for detecting unknown pests and pathogens that forgoes the need for two successive reactions. With multiplex-PCR, multiple primer sets are used within a single PCR reaction mixture that target specific gene regions and produce products (amplicons) of different lengths (Jenkins et al., 2012). The amplicons must be different enough in size to be separated via gel electrophoresis. Selecting multiple targets in one reaction saves time, reagents, and money (Hoy, 2003). However, with multiplex PCRs, the possibility of primer-primer interactions is greater and may result in false negatives because the target gene region may not be successfully amplified (Jenkins et al., 2012). The primers used must also have the same temperature and cycling requirements, and the amplicons should be close in size yet discernable via electrophoresis (Hoy, 2003; López et al., 2003). Multiplex-PCR can be used in real-time by attaching different fluorophores to each of the different amplicons being detected (Jenkins et al., 2012). The ability to discern multiple organisms or toxins in a single host is sometimes important in detecting viral, bacterial, and fungal plant pathogens. (López et al., 2003). Multiplex-PCR methods need to be formally proven in replicated tests and the effectiveness reported in literature prior to routine use in a diagnostic lab. These methodologies will likely increase in practical use as a wider selection of proven multiplex reaction mixtures become published for plant diagnosis applications.

Loop-mediated isothermal amplification (LAMP), as the name implies, is a type of DNA replication technique that employs three sets of primers that are active at the same temperature (Notomi et al., 2000). The unique isothermal nature of these primers allows the reaction to occur in a simple heating block, rather than a thermocycler. This not only reduces the cost of the test but allows for molecular detection of pests and pathogens in the field. For example, Agdia Inc. has developed an assay that utilizes human body temperature to accelerate the reaction LAMP reaction. Samples in the field are introduced to the necessary enzymes, primers and reagents required for the reaction and then placed in a device that incubates the sample in your armpit. The products from this reaction are then tested via lateral-flow chromatography for the presence of target amplicons.