

THE DEVELOPMENT AND CHARACTERIZATION OF
MOLECULAR TOOLS FOR MICROBIAL FORENSICS

By

TEECIE PAIGE BROWN

Bachelor of Science in Biology
Tarleton State University
Stephenville, Texas
2006

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2011

THE DEVELOPMENT AND CHARACTERIZATION OF
MOLECULAR TOOLS FOR MICROBIAL FORENSICS

Dissertation Approved:

Dr. Ulrich K. Melcher

Dissertation Adviser

Dr. Jacqueline Fletcher

Dr. Peter Hoyt

Dr. Patricia Canaan

Dr. Robert Allen

Dr. Mark E. Payton

Dean of the Graduate College

ACKNOWLEDGMENTS

I would sincerely like to thank Dr. Ulrich Melcher for providing the support and encouragement that has influenced my scientific and professional mentality over the past years. Dr. Melcher's obvious love for science and dedication to his students made interacting with him pleasant on a daily basis. I appreciate his sense of humor and his willingness to share a good laugh in times of struggle. Dr. Melcher is a leader and an inspiration and I have thoroughly enjoyed having him as my mentor.

I would also like to thank Dr. Fletcher for her professional guidance and perspective. I appreciated her professional insight and personal advice and enjoyed watching the National Institute for Microbial Forensics & Food and Agricultural Biosecurity (NIMFFAB) grow into the success that it is today.

I am grateful to my committee members, Drs. Peter Hoyt, Patricia Canaan and Robert Allen for their helpful opinions and varied points of view. It has been a pleasure to interact with professionals in diverse fields, and I feel that I gained a unique perspective by working with each of them.

Lastly, I would like to thank my family and friends for their unwavering emotional support. They encouraged me when I was desperate, grounded me when I lost perspective, assisted me when I felt alone and loved me when I was not very loveable. I absolutely would not be the person I am today, without them.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
II. REVIEW OF LITERATURE.....	7
Introduction to microbial and plant pathogen forensics	7
RNA viruses and WSMV as forensic targets	12
Forensic markers and genotyping techniques	19
Current and future challenges associated with microbial forensics	23
III. ASSESSING THE MOLECULAR DIFFERENCES BETWEEN PLANT VIRUS POPULATIONS COLLECTED FROM NATURALLY AND ARTIFICIALLY- INFECTED PLANTS: IMPLICATIONS FOR MICROBIAL FORENSICS	45
Abstract	45
Introduction.....	46
Materials and Methods.....	49
Results.....	54
Discussion.....	59
Acknowledgements.....	63
IV. UTILIZATION OF APYRASE-MEDIATED ALLELE-SPECIFIC PRIMER EXTENSION (AMASE) AND TAG-ARRAY MINISEQUENCING TECHNOLOGIES TO GENETICALLY FINGERPRINT <i>WHEAT STREAK MOSAIC VIRUS</i>	78
Abstract	78
Introduction.....	79
Materials and Methods.....	83
Results.....	89
Discussion.....	94
Acknowledgements.....	97

Chapter	Page
V. CHARACTERIZATION AND EVALUATION OF A UNIVERSAL PLANT VIROCHIP	118
Abstract	118
Introduction	119
Materials and Methods	121
Results	127
Discussion	131
Acknowledgements	134
VI. USE OF FATTY ACID METHYL ESTER PROFILES FOR DISCRIMINATION OF <i>BACILLUS CEREUS</i> T-STRAIN SPORES GROWN ON DIFFERENT MEDIA	150
Chapter Preface	150
Abstract	151
Introduction	152
Materials and Methods	155
Results	160
Discussion	167
Acknowledgements	175
APPENDX	194

LIST OF TABLES

Table	Page
CHAPTER III	
Table 1: Mean similarity scores for nucleic acid and amino acid sequence alignments	71
Table 2: Tests of neutrality for individual and combined data sets	72
Table 3: Estimates of the patterns of nucleotide substitution	73
Table 4: Evidence of recombination within data sets	74
CHAPTER IV	
Table 1: AMASE probe sequences	108
Table 2: Minisequencing primer and tag sequences	109
CHAPTER V	
Table 1: Detection of previously identified plant viruses by the Virochip	141
Table 2: Identification of partially characterized viruses using the Virochip	142
Table 3: Family, genus, species and strain-specific probes for WSMV	143
CHAPTER VI	
Table 1: Sporulation media key	186
Table 2: Fatty acid variables	187
Table 3: Relative abundances of fatty acid markers	188
Table 4: ANOSIM on spore-medium group comparisons	189

LIST OF FIGURES

Figure	Page
CHAPTER II	
Fig. 1. Genome organization of <i>Wheat streak mosaic virus</i>	14
Fig. 2. Image of the wheat curl mite	17
CHAPTER III	
Fig. 1. Agarose gel image of WSMV RT-PCR amplicons after staining with ethidium bromide	75
Fig. 2. Evolutionary relationships of nucleic acid sequences	76
Fig. 3. Plots of the quality of conservation of amino acid alignments.....	77
CHAPTER IV	
Fig. 1. SNP-typing by allele-specific primer extension.....	110
Fig. 2. Solution-based minisequencing	111
Fig. 3. The effect of apyrase concentrations on allele-specific extension reactions	112
Fig. 4. Reproducibility of AMASE.....	113
Fig. 5. Reproducibility of minisequencing	114
Fig. 6. Sensitivity of minisequencing	115
Fig. 7. Effect of primer concentration on extension reactions using amplicons.....	116
Fig. 8. Effect of polymerase concentration on minisequencing.....	117
CHAPTER V	
Fig. 1. Agarose gel image of WSMV RT-PCR amplicons after staining with ethidium bromide.....	144
Fig. 2. Line graph displaying the raw fluorescence intensity profiles for the 50 most intense probes in a WSMV-positive sample (UPVM-P)1_010 and a healthy wheat sample (UPVM-P01_032).....	145

Figure	Page
Fig. 3. Graphical representation of the biological replicates performed using WSMV-infected plant tissue	146
Fig. 4. Representation of viral taxa conserved among the top fifty probes and their relative abundances in all three replicate experiments	147
Fig. 5. Real-Time PCR results for the quantification of WSMV cDNA	148
Fig. 6. Fluorescence intensities of probes using 4.5 pg, 0.45 pg and 0.045 pg of target cDNA.....	149

CHAPTER VI

Fig. 1. Vegetative and spore forms of <i>Bacillus cereus</i> T strain.....	190
Fig. 2. Comparison of spore cultures grown on different medium formulations ...	191
Fig. 3. Two-dimensional nMDS plot of spore FAME profiles grouped by growth medium	192
Fig. 4. Discriminant function analysis of spore FAME profiles.....	193

CHAPTER I

INTRODUCTION

The concern about the potential weaponization and dissemination of microorganisms targeting agriculture become increasingly daunting over the past decade. The accessibility and potential for large-scale dissemination of pathogens makes the United States agricultural sector vulnerable to acts of bioterrorism [1]. Plants are especially vulnerable to such attacks, due to their economic importance, nutritional value and the lack of surveillance associated with wide-scale crop production [2-5].

In the United States alone, plant pathogens have been estimated to cause approximately \$33 billion per year in crop losses [6, 7], of which 65% (\$21 billion) are attributed to exotic pathogens [8]. It is estimated that over 50,000 plant pathogens, including viruses, bacteria, fungi and nematodes, exist in the United States, and pose a significant threat to crop production [3]. Additionally, individual crops, such as wheat, can be attacked by over 200 pathogens [3].

Though constraints on crop production by natural outbreaks of plant diseases are abundant, advancements in technology, increased global movement and the lack of a complete plant pathogen forensic capability have increased the potential for the deliberate release of plant pathogens [9]. A well-developed forensic program that is poised to

respond to the deliberate and malicious dispersal of plant pathogens is essential to deterring such an event, attributing a biocrime and excluding innocent individuals from suspicion following an attack. Thus the fields of microbial forensics and the sub-discipline of plant pathogen forensics have experienced significant growth since the anthrax attacks of 2001 that prompted their conception. Still, several gaps exist within the discipline of plant pathogen forensics, including the lack of trained personnel, infrastructure and the need for validated tools that are capable of genetically characterizing pathogens for use in attribution of a crime [10].

Because forensic investigations are expensive and labor-intensive, the first consideration following a reported act of bioterrorism is the determination of whether such an event has actually occurred. Falsely reporting the release of a biological weapon is a violation of federal law that complicates the immediate response of responsible agencies. The Federal Bureau of Investigation examines many hoaxes involving the release of anthrax each year, and experienced a significant increase in hoaxes following the anthrax attacks of 2001 [11]. The ability to rapidly distinguish between an actual attack and a hoax is critical to the agency's response. Although an attack on plants is likely to elicit a much different response from regulatory agencies, the determination of whether an inoculation event is deliberate or natural is essential to their subsequent actions. Information concerning the differences between a natural outbreak of a plant disease and an intentional introduction of a pathogen is not widely available, and presents a gap in the field of plant pathogen forensics.

Another substantial gap in the field is the lack of validated forensic assays that are capable of accurately and rapidly identifying and genetically fingerprinting plant

pathogens for use in attribution. Current genotyping techniques may be amenable for use as tools for plant pathogen forensics, but would require strenuous validation to withstand the scrutiny imposed in a courtroom [9]. One such technique, microarray, has been recognized for its diagnostic and genotyping capabilities using plant viruses [12-17], as well as its ability to produce reliable single nucleotide polymorphism (SNP) profiles using limited amounts of genetic material for use in human forensics [18]. Since microarrays are capable of detecting an abundance of pathogens in parallel and have been used previously for genotyping purposes, adaptation of this technology for use in plant pathogen forensics is promising.

This thesis describes the exploration of molecular analyses to aid in the discrimination between naturally- and intentionally-infected plants using a model plant virus, and the development and testing of microarray-based technologies for the identification and genetic fingerprinting of plant viruses. Additionally, a microbial forensics method outlining the application of fatty acid profiles from *Bacillus cereus* T-strain spores prepared using different substrates as a method for determining the components used to prepare a sample of *Bacillus* spores, is reported. The objectives for this research were as follows:

- i). Evaluate molecular differences between populations of WSMV in plants collected from a natural outbreak of the virus and plant virus populations derived from mechanically-inoculated plants.
- ii). Assess the use of apyrase-mediated allele-specific primer extension (AMASE) and minisequencing using a microarray slide for generating single nucleotide polymorphism (SNP) profiles for a plant virus.

- iii). Characterize and test a universal plant virus microarray (plant Virochip) using identified and partially characterized plant viruses.
- iv). Examine fatty acid methyl ester (FAME) profiles of *Bacillus cereus* T-strain spores for the presence of molecular markers that are indicative of specific media components used during preparation of the spores*.

*Note: The document entitled “Use of fatty acid methyl ester profiles for discrimination of *Bacillus cereus* T-strain spores grown on different media” was published in 2010 in Applied and Environmental Microbiology. My specific contributions to the project were achieved during a three-month internship at the FBI Counterterrorism and Forensic Science Research Unit (FBI-CFSRU) and are outlined in the chapter preface.

LITERATURE CITED

1. Committee on Biological Threats to Agricultural Plants and Animals, N.R.C., *Countering agricultural bioterrorism*. 2002, Washington, DC: National Academies Press.
2. Casagrande, R., *Biological terrorism targeted at agriculture: The threat to US national security*. *Nonproliferation Review*, 2000. **7**: p. 98-99.
3. Madden, L.V. and M. Wheelis, *The threat of plant pathogens as weapons against U.S. crops*. *Annual Review of Phytopathology*, 2003. **41**(1): p. 155-176.
4. Whitby, S.M., *The potential use of plant pathogens against crops*. *Microbes and Infection*, 2001. **3**(1): p. 73-80.
5. Whitby, S.M., *Biological warfare against crops*. 2002, Basingstoke, U.K.: Palgrave.
6. USBC, *Statistical Abstract of the United States 2001*, U.S.B.o.t. Census, Editor. 2001, U.S. Government Printing Office: Washington, DC.
7. Pimentel, D., *Techniques for reducing pesticides: environmental and economic benefits*. 1997, Chichester, UK: John Wiley and Sons.
8. Pimentel, D., *Habitat factors in new pest invasions*, in *Evolution of insect pests- Patterns of variation*, K.C. Kim and B.A. McPherson, Editors. 1993, John Wiley and Sons: New York. p. 165-181.
9. Budowle, B., et al., *Microbial forensics*, in *Microbial Forensics*, R.G. Breeze, B. Budowle, and S.E. Schutzer, Editors. 2005, Elsevier Academic Press: San Diego

10. Fletcher, J., et al., *Plant pathogen forensics: capabilities, needs, and recommendations*. Microbiology and Molecular Biology Reviews, 2006. **70**(2): p. 450-471.
11. FBI, *FBI warns against anthrax hoaxes*. 2001, FBI National Press Release Office: Washington, D.C.
12. Tiberini, A., et al., *Oligonucleotide microarray-based detection and identification of 10 major tomato viruses*. Journal of Virological Methods. **168**(1-2): p. 133-140.
13. Pasquini, G., et al., *Oligonucleotide microarray-based detection and genotyping of Plum pox virus*. Journal of Virological Methods, 2008. **147**(1): p. 118-126.
14. Grover, V., et al., *Oligonucleotide-based microarray for detection of plant viruses employing sequence-independent amplification of targets*. Journal of Virological Methods. **163**(1): p. 57-67.
15. Bystricka, D., et al., *Oligonucleotide-based microarray: A new improvement in microarray detection of plant viruses*. Journal of Virological Methods, 2005. **128**(1-2): p. 176-182.
16. Nicolaisen, M., *An oligonucleotide-based microarray for detection of plant RNA viruses*. Journal of Virological Methods. **173**(1): p. 137-143.
17. Zhang, Y., et al., *Oligonucleotide microarray with a minimal number of probes for the detection and identification of thirteen genera of plant viruses*. Journal of Virological Methods. **167**(1): p. 53-60.
18. Divne, A.-M. and M. Allen, *A DNA microarray system for forensic SNP analysis*. Forensic Science International, 2005. **154**(2-3): p. 111-121.

CHAPTER II

REVIEW OF LITERATURE

1. Introduction to microbial and plant pathogen forensics

1.1. Microbial forensics, plant pathogen forensics, bioterrorism and biocrimes

The field of microbial forensics was established in response to the anthrax attacks of 2001. Though several definitions exist, microbial forensics can be defined as a scientific discipline that is involved with the collection, characterization and interpretation of evidence collected from the deliberate release, hoax or accidental introduction of microorganisms, for use during the attribution process [1]. Following the historical anthrax letter attacks that resulted in the deaths of five individuals [2], several initiatives were begun to assess the vulnerabilities and gaps associated with the potential for agroterrorism in the United States. Among those recognized by the National Research Council were the lack of a federal response plan for a biological attack on livestock and crop systems, vague and inconsistent criteria for pathogen risk assessment, absence of adequate infrastructure for analyzing high-risk pathogens and insufficient information concerning populations of pathogens [3]. These findings highlighted the vulnerability of plant production, a major contributor to the Gross Domestic Product (GDP), and ultimately the discipline of plant pathogen forensics. This nascent field combines traditional approaches to plant pathology, epidemiology and forensic science with the ultimate goal being the attribution, or assignment of blame, following a deliberate release of a plant pathogen. Though scientific advancements have begun to fill

the gaps in the fields of microbial forensics and plant pathogen forensics, admissibility of newly adapted tools may prove to be challenging.

There are several motives for the intentional release of plant pathogens; however, two main categories can be defined; acts of bioterrorism and biocrimes. The primary difference between the two scenarios is the purpose for the weaponization of the pathogen. While the former is motivated by political, religious, ecological or other ideological purposes with the intent to cause widespread fear and panic [4], the latter is executed to directly harm individuals using a biological weapon, instead of traditional weapons such as guns and knives [5].

1.2. Impacts of plant pathogens on crop production

Plant pathogens pose a significant threat to U.S. crop production. Plants are vulnerable to attack by various species of weeds, viruses, bacteria, fungi, nematodes and arthropods [6]. In the U.S., annual yield losses attributed to plant pathogens have been estimated to be as high as \$33 billion per year [7, 8], of which 65% (\$21 billion) are attributed to exotic pathogens [9]. It is estimated that over 50,000 plant pathogens, including viruses, bacteria, fungi and nematodes exist in the United States, and pose a significant threat to crop production [10]. Additionally, individual crops such as wheat, can be attacked by over 200 pathogens [10]. Although preventative treatments offer some protection for the six major crops produced in the United States (wheat, rice, maize, potatoes, soybeans and cotton), yield losses attributed to plant pathogens remain high (12.5%) [11]. Additionally, non-indigenous plant pathogens represent a major threat to U.S. crop production. An estimated \$21.5 billion is spent each year on damage and control costs associated with these plant pathogens [12]. The detection of exotic pests

also restricts exportation of the commodities, thus further reducing the potential for cost recovery by producers.

Since hosts, vectors and plant pathogens are influenced by weather, global changes in climate may increase disease severity [13]. Furthermore, the rise in global temperature could contribute to the adaptability of host-vector-pathogen systems, alter host resistance and physiology and affect the rates of pathogen development [14]. Though the effects of global climate change are unpredictable, the potential for additional yield losses attributed to plant pathogens in the face of an expanding population is alarming. Fortunately, plant breeding and genetic engineering programs have offered some relief from the devastation of plant pathogens.

1.3. The need for plant pathogen forensics

The health of the U. S. economy is highly dependent upon the success of the agricultural sector, as approximately \$1 trillion is annually generated by the U.S. agricultural industry [15]. In 2008, over \$115 billion was generated by exporting agricultural commodities [16]. The fact that the U. S. is a net exporter of agricultural products highlights the economically important position agriculture plays within the nation, as well as the dependence upon U.S. agriculture production by other countries. The central role that agriculture plays in the health of the U. S. economy, coupled with the ease of deploying biological weapons targeting plants or animals, makes agriculture an attractive target for a bioterrorist attack [17-19]. Therefore, protecting against a bioterrorist attack on crops is among a nation's top priorities [6].

With the available informational resources and the relative ease of access and deployment, the employment of a plant pathogen as a bioweapon may occur in the near

future. In the United States and other first world countries, the intentional introduction of a plant pathogen will reduce crop yield and quality, create panic, undermine the public's confidence in the government and food supply and negatively impact the economy [20]. Given that the global economy is highly dependent upon the status of the U.S. economy, disturbances could directly and indirectly affect other countries' economies, as well. The deliberate introduction of a plant pathogen in a third world country, especially those that are heavily dependent upon a single monoculture for sustenance, would have an even larger effect such as decreased nutrition and governmental decline [20]. Such an attack has been projected to be as devastating as an anthrax attack on a city [21]. Examples of pathogen devastation of crops have occurred throughout history. The Irish potato famine, caused by a population explosion of the endemic fungus *Phytophthora infestans*, caused widespread malnutrition and starvation, ultimately resulting in the death of over 1.5 million people [22]. If such an epidemic were to occur as a result of a deliberate bioterrorist attack on a staple crop, the psychological and humanitarian toll could be massive and the economic damages could undoubtedly impact the global economy.

The unintentional introduction of *Xanthomonas axonopodis* pv. *citri*, the causal agent of citrus canker, into the U. S. in 1995 is a more recent example of the devastation that can occur as a result of the introduction of plant pathogens to economically important crops. Initial plans to eradicate the pathogen cost over \$200 million [23], excluding the millions spent on legal fees and education [24]. The public anxiety associated with the removal of thousands of citrus trees, coupled with the economic damage incurred as a result of the disease, demonstrates the potential impacts that plant pathogens could have if employed as bioweapons.

Rice and wheat are two major crops that supply calories to the population therefore a biological attack on one of these cropping systems could yield serious economic consequences and result in insufficient nutrition of a population [25]. Though the outcome of such an attack in a developing country could affect the health of the population, targeting the same crops in the U. S. would probably have a strong regional economic effect, but would not cause widespread malnutrition [5].

Regardless of whether a plant disease outbreak occurs under natural or human-influenced circumstances, it is important to determine the origin and the causal agent, in order to implement effective management strategies. The difference between traditional plant pathology and epidemiology and plant pathogen forensics is that the latter is concerned with gathering information to be used during the attribution process, or the assignment of blame. This is necessary to insure justice for victims of biocrimes or acts of bioterrorism and to deter potential future attacks using bioweapons [5]. If an intentional inoculation of a plant pathogen is reported, it is the responsibility of the Federal Bureau of Investigation and the USDA-APHIS to determine whether an actual introduction occurred, or if the report is a hoax. Hoaxes can trigger the same expensive, labor and time-intensive responses by these agencies, and may therefore be as serious as an actual attack [26].

1.4. History of Anti-Crop Bioweapons

Anti-crop bioweapons have been explored extensively by several countries over the past century. For example, during World War I, the French explored weaponizing *Phytophthora infestans* [27] (the causal agent of the Irish potato famine) and during the second World War, the Japanese explored the possibility of utilizing a variety of anti-

crop agents to further their cause. Additionally, before being dismantled in 1969, the U.S. developed mass production and dissemination capabilities that were unparalleled [28] and the Soviet Union continued to develop anti-crop programs over the next twenty years [29]. Iraq was also discovered to have been developing anti-crop bioweapons such as *Tilletia laevis* (the causal agent of common bunt of wheat) during the early 1990s [30]. To date, several research initiatives involving defensive biological warfare programs are in existence, and several unconfirmed attacks on valuable crop systems have been unofficially reported. The history of anti-crop biowarfare programs suggests that countries that are serious about utilizing bioweapons usually include anti-crop programs [28]. The rapid detection of pathogens and collection of evidence for use during the attribution process is critical to restoring public confidence following a bioterrorist attack. Nevertheless, biochemical methods capable of detecting and profiling plant pathogens in a forensic capacity are currently insufficient. The development of a forensic tool that is capable of simultaneously diagnosing and genetically typing plant pathogens would, therefore, be of great use to forensic scientists, law enforcement agents and plant pathologists during an investigation involving crop bioterrorism.

2. RNA viruses and WSMV as forensic targets

2.1. RNA viruses as targets for microbial forensics

Analogous to human DNA, plant pathogen genetic material has also been shown to contain large numbers of SNPs, which have been useful for diagnostic [31] and classification [32] purposes. Plant viruses that use RNA as their genetic material, and thus are dependent upon an RNA-dependent RNA polymerase (RdRp) for replication, are

especially polymorphic [33, 34] due to the apparent inability of RdRp to perform proofreading functions [35]. RNA virus genomes are stable enough to permit forensic analysis, but are constantly evolving so that unique genotypes exist within populations [36]. Strong argument has been made for a quasispecies model of population genetics to describe RNA virus populations [37-40]. Recent evidence, however, suggests that the plant pathogen *Wheat streak mosaic virus* (WSMV) [41] and other RNA viruses [42] may not follow this model. Instead, WSMV has been proposed to follow a stochastic model of population variation. Nonconsensus variations in serially passaged WSMV sequences exist at a constant frequency, suggesting that variable sequences that arise during replication by RdRp are usually packaged into virus particles and are not immediately used as a template for replication [43]. Occasionally, the nascent RNA molecules are used as a template for replication, resulting in the establishment of new genotypes within a single plant. Since the field variability of WSMV is complex and the genotypic variability within plants is low [44], an intentional inoculation would presumably result in multiple plants with nearly identical consensus sequences. These observations indicate that RNA viruses may be amenable to forensic applications. However, multiple changes in the SNP profiles of species with rapidly evolving genomes should be compensated for by assessing an adequate number of SNPs.

2.2. *Wheat streak mosaic virus* as a model plant virus

WSMV was chosen as a model pathogen for this study due to its economic importance and prevalence throughout the state of Oklahoma, the abundance of information available for the pathogen and minimal restrictions associated with working with the virus in a laboratory or field setting.

WSMV is a member of the genus *Tritimovirus* within the family *Potyviridae* [45]. The *Potyviridae* family is one of the largest and most economically important families of plant viruses, encompassing over 30% of known plant virus species [46, 47]. Like other species in the *Potyviridae* family, WSMV utilizes a single, positive-sense RNA genome that is 9384 bp long and encodes a polyprotein that is subsequently cleaved by three viral proteases, yielding ten mature proteins that function during replication, infection and transmission [48-50]. The ten mature proteins are: P1, helper component proteinase (HC-Pro), P3, 6K1, cylindrical inclusion (CI), 6K2, viral genome-linked protein-nuclear inclusion protein a (VPg-NIa), proteinase protein of NIa (NIa-Pro), nuclear inclusion protein b (NIb) and the coat protein (CP). Additionally, a newly discovered open reading frame (P3-PIPO) is present within the P3 coding region, but requires a frame shift to be expressed [51]. The genome has a 5'-linked protein (VPg), which mediates cap-independent translation, and a 3' poly A tail [52]. A representation of the viral genome organization of the potyvirus *Turnip mosaic virus* (TuMV) is shown in Figure 1 [53].

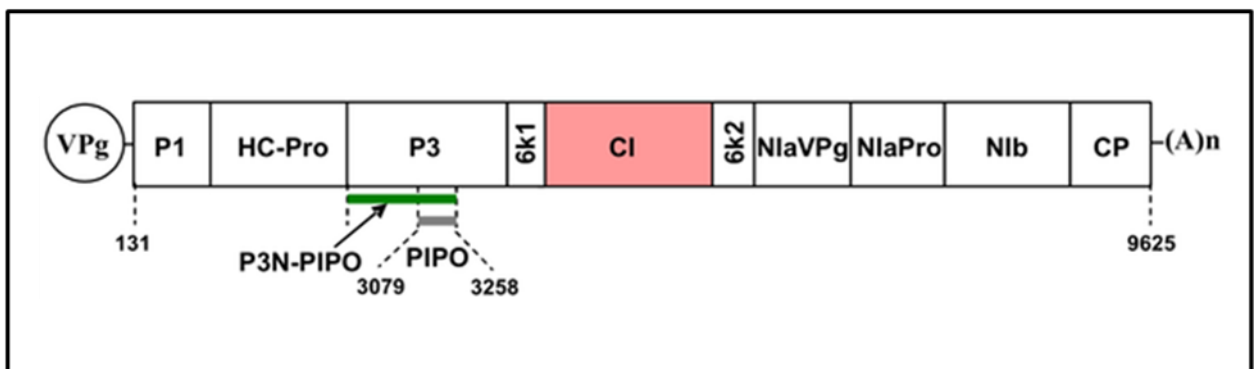


Fig. 1. Genome organization of the potyvirus TuMV (reproduced from reference [53]).

Most of the *Potyviridae* family-encoded proteins are bi- or multi-functional, which is likely a result of the limited genome sizes of plant viruses. The P1 and HC-Pro proteins serve as viral proteinases and are also responsible for suppressing RNA silencing

[55, 56]. Additionally, the P1 protein is involved in viral replication [57] and HC-Pro aids in long-distance movement of potyviruses [58]. The nuclear inclusion protein a (NIa), which serves as the primary proteinase, also displays intrinsic DNase activity and may function to degrade host DNA in the later stages of viral infection [59]. The viral coat protein functions in vector transmission and viral accumulation in plant cells [60]. The coat protein and the cellular inclusion protein are also involved in cell-to-cell and long-distance movement throughout the host [61-65]. The newly-described P3-PIPO protein also is likely to be involved in complex formation at plasmodesmata, and interacts with the CI proteins to facilitate cell-to-cell movement [53].

An abundance of research has been performed on the evolution, mutability and movement of WSMV. Viral replication, evolution and population dynamics were examined and thoughtfully discussed by French and Stenger [66], who argued that WSMV replication is predominantly linear and occurs similarly to the ‘stamping machine model’ proposed by Luria in 1951 [67]. In the same report, they wrote that the number of successful competing genomes in a single cell ranged from 2 to 4, which is low given that a single cell may contain 10^5 to 10^6 virions [67]. Other studies have reported the role of vector transmission and cross-protection bottlenecks in the genetic isolation of WSMV lineages [68]. Together, these data suggest that genetic bottlenecks associated with vector transmission and cell-to-cell movement play an important role in the establishment of new populations of plant viruses within host plants. Furthermore, the fact that WSMV within fields is as variable as the virus collected from different fields in the same region, implies that vector-transmission bottlenecks may contribute to the stochastic establishment of variants from a population of viruses within plants [44]. Another study

[43], in which the occurrence of mutations in WSMV genomes after serial passage of the virus was examined, demonstrated that most of the unique substitutions incurred during replication accumulated and were found at a constant frequency throughout several serial passages. Thus, researchers concluded that nascent genomes are immediately sequestered into virion particles and may not ever be tested for fitness. This study also reported that the largest proportion of mutations occurred in the nuclear inclusion protein a (NIa) (15 allelic shifts at 7 sites) and the coat protein (CP) cistrons (12 allelic shifts at 11 sites) [43].

WSMV is prevalent throughout most wheat growing regions of the United States and other wheat-producing regions including Canada, Mexico, Europe, the Middle East, North Africa, Australia and Asia. WSMV infects wheat, barley, oats, some varieties of maize and millet and other plants in the grass family [69]. Symptoms include yellowing and streaking of the leaves running parallel to the leaf veins, stunting of the plants, curled leaves and occasionally head sterility [70]. Yield losses associated with the pathogen on wheat can be mild or severe and usually range from 5% to 100% per year [71]. The virus may infect its host singly, but is often found in combination with *Triticum mosaic virus* (TriMV) in the Great Plains region of the United States and is capable of inducing cultivar-specific synergism with TriMV [72]. Additionally, WSMV has been found in combination with *High plains virus* in hard red winter wheat in Nebraska [73]. All three viruses are transmitted by the wheat curl mite (WCM), *Aceria tosichella* Keifer [74, 75] (Fig. 2).

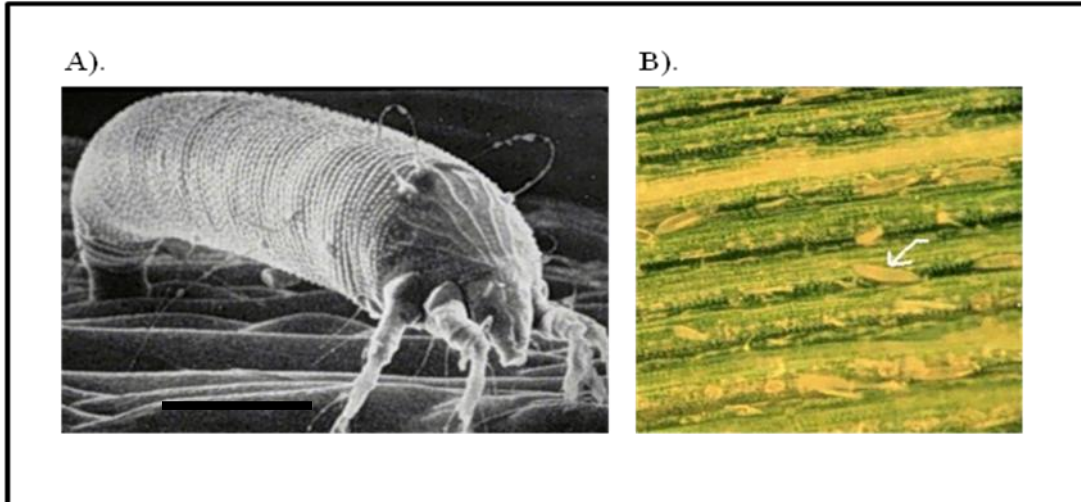


Fig. 2. Images of the wheat curl mite. A). Scanning electron micrograph of the WCM. The bar represents approximately 100 μm . B). Low-resolution photograph of the WCM (reproduced from reference [76]).

Eriophyoid mites, including the WCM, cause little damage and yield losses in the absence of the viruses. However, significant yield losses can result from severe cases of mite infestation [77]. Between seasons, the WCM can over-winter on volunteer wheat, usually produced through weather-related mechanical severance of seeds from heads, and other grass hosts. These alternate hosts act as a ‘green bridge’, and facilitate the survival of both the mites and WSMV. As seedlings develop, mites can transfer to the newly emerging plants and infect the wheat. At the end of the season, mites that have serendipitously arrived at alternate hosts will complete the cycle by reproducing and infecting the next-season’s wheat. Feeding by the WCM causes the leaves to curl inward toward the midvein, and this curling is often used as a marker for the presence of the wheat curl mite and WSMV [70].

Though eriophyoid mites have two pairs of small legs at the front of their bodies, they largely rely on wind dispersal for movement across large areas, such as the movement from the initial source of inoculum to a wheat host. Surprisingly, mites can efficiently translocate by clinging onto other insects or birds that subsequently contact

susceptible host plants [78]. Temperature and light also play a role in the dispersal of the wheat curl mite. For example, one study showed that under controlled environmental conditions, eight-fold more mites were mobile at 24°C than at 12°C and more mites were dispersed during periods of light than in darkness [79]. With the exception of eggs, the WCM is capable of transmitting the virus at all stages of development; however, transmission efficiency is affected by the developmental stage of the mite. One study showed that adults transmitted WSMV approximately 52% of the time, while nymphs and mites in the quiescent stage transmitted the virus with 83% and 94% efficiency, respectively [80]. Interestingly, the same study showed that WCM reproduction increases when the mites are feeding on WSMV-infected tissue, as compared to other cereal virus-infected plants [80].

Dual-purpose wheat production, in which the crop is grazed during the vegetative state and subsequently harvested as a grain crop, is gaining popularity among growers due to its potential benefits to livestock gains [81]. Some concerns regarding the transmission of WSMV during grazing of dual-purpose wheat grazing and crop production infected with the virus prompted researchers to assess mechanical transmission by salivary transfer by animals. One such study tested for the presence of the virus in sheep saliva, the ability of the saliva to serve as a source of inoculum and the ability of sheep to transfer the virus after feeding on heavily-infected wheat tissue. These studies showed that WSMV cannot be amplified from saliva 30 minutes after feeding, and that sheep are incapable of mechanically transferring WSMV to healthy wheat [82]. Since the wheat curl mite is the only known vector for WSMV, considerable efforts have been devoted to the production of wheat cultivars that are resistant to the mite and the

characterization of resistant cultivars for genetic engineering purposes [83-86]. PCR-RFLP analyses on wheat curl mite mitochondrial and ribosomal DNA collected from fields in Nebraska, Kansas and Montana showed that two biotypes of the wheat curl mite exist: type 1 (NE biotype) and type 2 (all others) [80]. Additionally, PCR-RFLP studies on the nuclear and mitochondrial DNA from mites collected in Australia also demonstrate the existence of two distinct lineages [87]. Semi-persistent transmission of WSMV has been shown to be dependent upon the amino terminus of the HC-Pro gene for WSMV [88].

3. Forensic markers and genotyping techniques

3.1. SNPs as genetic markers for fingerprinting viruses

Size-based forensic techniques, including restriction fragment length polymorphism (RFLP) [89], amplified fragment length polymorphism (AFLP) [90], random amplified polymorphic DNA (RAPD) [91] and simple sequence repeat (SSR) analyses [92], have been used historically to discover human DNA genotypes. Despite the discriminatory power provided by these methods, forensic science is currently moving towards using more informative genetic profiling methods, such as SNP-typing [93]. SNPs are attractive forensic markers for human identification for two important reasons. First, most genetic variation throughout the human genome is attributed to SNPs. It is estimated that a SNP occurs every 100-300 bases throughout the human genome, resulting in a total of 10 to 30 million SNPs [94]. Human SNPs are often associated with the onset of disease. Therefore, exhaustive efforts are being dedicated to the identification of human SNPs for treatment and diagnosis. Databases such as

HapMap, which contains a plethora of human genetic information [95] including data for over 3.7 million human SNPs (<http://www.hapmap.org/>), are being developed in an effort to advance knowledge about diseases and pharmacological responses associated with individual SNPs. Although the databases are not necessarily being constructed for forensic purposes, useful forensic data is being uncovered as a result of such efforts. For example, SNPs have been shown to yield information about an individual's skin pigmentation [96] or eye color [97], which may prove useful in creating physical profiles for use in identifying potential suspects. The second reason SNPs are useful for forensic profiling is that SNP-typing can be performed on samples that have been severely degraded, which is often the case for forensic samples. Because most SNP assays are based upon primer extension or hybridization, a much shorter intact sequence is required for analysis. Multiple SNPs can be used to generate the same quality of information obtained from classical forensic approaches when intact DNA is not available. During the first major microbial forensics investigation involving the anthrax attacks of 2001, researchers used SNPs as one of the genetic markers to tie anthrax spores isolated from the letters to the source [98], demonstrating their usefulness as forensic targets. SNPs exist within all organisms and are thus universal markers that can be targeted by forensic techniques.

3.2. Bacterial biomarkers

Bacteria present unique genetic markers that can also be targeted by forensic assays. Some molecular techniques that have been applied to bacteria, usually for epidemiological, identification or characterization purposes, include amplified fragment length polymorphism (AFLP) [99-102], multilocus sequence typing (MLST) [103-105],

direct sequencing methods [106-108], random amplified polymorphic DNA (RAPD) [109-111], pulsed-field gel electrophoresis (PFGE) [112, 113] and other genome-based techniques. Though these genotyping techniques offer high-resolution discrimination of bacteria at various taxonomic levels, nongenetic signatures may offer additional information that would be useful during a forensic investigation. Some of the previously applied nongenetic assays examined the elemental signatures of *Bacillus subtilis* spores prepared in different media [114], the characterization of stable isotopic signatures of *Bacillus globigii* and *Erwinia agglomerans* through growth on different media [115], ratios of stable isotopes in *Bacillus subtilis* cells and spores [116] and analyses of the fatty acid compositions of microbial communities by fatty acid methyl ester (FAME) analyses. Because phospholipids are synthesized during microbial growth and are degraded quickly during cell death, they provide an immediate assessment of the status of a microbial community. FAME profiles are unique to individual bacterial species [117] and strains [118, 119] and have therefore been useful in bacterial classification [120-122] and studies involving the characterization of microbial communities [123, 124]. Furthermore, fatty acid compositions of *Bacillus* spp. reflect the composition of the media used to prepare the bacterial cells [125-127]. These data indicated that fatty acid profiles produced from *Bacillus* spp. may provide insight into the components of the media used to prepare the cells prior to sporulation, information that would prove useful during the processes of evidence collection, exclusion of innocent individuals and the attribution of a biocrime similar to the anthrax attacks of 2001. This assertion was tested and confirmed using *Bacillus cereus* T-strain spores prepared on ten different media [128].

3.3. Microarray as a tool for identifying and forensically genotyping plant viruses

Over the past decade, numerous microarray platforms have been reported for use in identifying plant viruses [129-137], phytoacteria [138-140], plant-pathogenic fungi [141] oomycetes and nematodes [142]. By combining the discriminatory power of the genetic markers with the miniaturization of microarray, a single sample can be screened for the presence of virtually an unlimited number of target sequences. Multiple variations of the microarray platform exist. Though some of the earliest microarrays utilized cDNA spots to capture fluorescently-labeled targets, oligonucleotide-based platforms have largely replaced the use of cDNA arrays due to the simplicity of designing oligonucleotides *in silico* and subsequently printing slides using a robotic instrument.

In addition to its use in diagnostics, microarrays have also been developed for use in a forensic context [143-147]. Previous studies have been developed using solution-based minisequencing targeting both nuclear and mitochondrial SNPs [148, 149]. Panels of Y chromosome SNPs have also been genotyped through the use of microarrays [150]. More recently, a microarray platform was developed and validated for 124 loci of combined human nuclear, mitochondrial and Y chromosome SNP-typing. Researchers estimated the accuracy of the system to be 99.83% with a call rate of 99.66% [151]. Some attention has also been given to the optimization of printing, slide chemistries, and SNP-typing strategies (single base extension-tag arrays vs. the use of allele-specific oligonucleotides) for the utilization of microarrays in forensics [152]. Microarrays have also been developed for use in microbial forensics [153, 154]; however, the application of microarrays specifically for use in plant pathogen forensics has not yet been reported in the literature.

4. Current and future challenges associated with microbial forensics

Though considerable effort has been dedicated to the development of tools for use in microbial forensics, the validation and admissibility of traditional tools in microbial forensic context may not be straightforward. Some cases involving HIV have previously tested the limitations of microbial forensics in a courtroom, and set the precedence for future cases involving the weaponization of pathogens. One such case involved the intentional injection of a woman with a cocktail of the hepatitis C and human immunodeficiency viruses [155] by her ex-lover who was a gastroenterologist and therefore had access to samples containing the two pathogens. Epidemiology and phylogenetic analyses were used to trace the HIV back to patients that the doctor treated prior to the suspected inoculation event. The admissibility of the phylogenetic analyses were challenged but ultimately the analyses were deemed admissible [156]. Using the phylogenetic analyses, in combination with a traditional epidemiological approach, prosecutors proved that the HIV was derived from two patients of the gastroenterologist. The doctor was tried and convicted of attempted second degree murder.

Challenges associated with the proper collection and storage of pathogens for future analyses may also prove problematic for time-dependent signatures such as fatty acid profiling, and long-term storage of microbial evidence. Additionally, the physical collection of infectious materials for testing is more dangerous than that of traditional forensic investigations [157], though this is not necessarily a major concern for plant pathogen forensics, as most plant pathogens are not a threat to human or animal health. However, toxins produced by some plant pathogens may force forensic specialists to

exercise caution when approaching an intentionally-inoculated field. The availability of reference strains for high-threat pathogens and the development of microbial forensic databases would enhance plant pathogen forensic capabilities, and will likely provide weight to forensic results that is necessary for admissibility of evidence in the courtroom [158].

LITERATURE CITED

1. Budowle, B., et al., *Building microbial forensics as a response to bioterrorism*. Science, 2003. **301**(5641): p. 1852-1853.
2. *Famous cases & criminals*. Amerithrax or Anthrax investigation [cited 2011 April 20, 2011]; Available from: <http://www.fbi.gov/about-us/history/famous-cases/anthrax-amerithrax/amerithrax-investigation>.
3. Committee on Biological Threats to Agricultural Plants and Animals, N.R.C., *Countering agricultural bioterrorism*. 2002, Washington, DC: National Academies Press.
4. Carus, W.S., *Bioterrorism and biocrimes: The illicit use of biological agents since 1900*. 2002, Amsterdam, The Netherlands: Fredonia Books.
5. Budowle, B., et al., *Microbial forensics*, in *Microbial Forensics*, R.G. Breeze, B. Budowle, and S.E. Schutzer, Editors. 2005, Elsevier Academic Press: San Diego
6. Madden, L.V. and M. Wheelis, *The threat of plant pathogens against U.S. crops*. Annual Review of Phytopathology, 2003. **41**: p. 155-176.
7. USBC, *Statistical Abstract of the United States 2001*. 2001, U.S. Government Printing Office: Washington, DC.
8. Pimentel, D., *Techniques for reducing pesticides: environmental and economic benefits*. 1997, Chichester, UK: John Wiley and Sons.
9. Pimentel, D., *Habitat factors in new pest invasions*, in *Evolution of insect pests- Patterns of variation*, K.C. Kim and B.A. McPherson, Editors. 1993, John Wiley and Sons: New York. p. 165-181.

10. Madden, L.V. and M. Wheelis, *The threat of plant pathogens as weapons against U.S. crops*. Annual Review of Phytopathology, 2003. **41**(1): p. 155-176.
11. Oerke, E., *Crop losses to animal pests, plant pathogens, and weeds*, in *Encyclopedia of Pest Management*, D. Pimentel, Editor. 2007, CRC Press: Boca Raton.
12. Pimentel, D., et al., *Environmental and economic costs of nonindigenous species in the United States*. Bioscience, 2000. **50**(1).
13. Chakraborty, S., A.V. Tiedemann, and P.S. Teng, *Climate change: potential impact on plant diseases*. Environmental Pollution, 2000. **108**(3): p. 317-326.
14. Coakley, S.M., H. Scherm, and S. Chakraborty, *Climate change and plant disease management*. Annual Review of Phytopathology, 1999. **37**(1): p. 399-426.
15. Parker, H.S., *Agricultural bioterrorism: A federal strategy to meet the threat*. 2002, U.S. Government Printing Office: Washington, D.C.
16. USDA. *Foreign Agricultural Trade of the United States (FATUS)*. 2008 [cited 2008 December 19]; Available from: <http://www.ers.usda.gov/Data/FATUS/>.
17. Frazier, T.W. and D.C. Richardson, eds. *Food and agricultural security: guarding against natural threats and terrorist attacks affecting health, national food supplies, and agricultural economics*. 1999, New York Academy of Sciences: New York.
18. Horn, F.P. and R.G. Breeze, *Agriculture and food security*, in *Food and Agricultural Security: Guarding against Natrual Threats and Terrorist Attacks Affecting Health, National Food Supplies, and Agricultural Economics*, T.W.

- Frazier and D.C. Richardson, Editors. 1999, New York Academy of Sciences: New York.
19. Casagrande, R., *Biological terrorism targeted at agriculture: The threat to US national security*. *Nonproliferation Review*, 2000. **7**: p. 98-99.
 20. Fletcher, J., et al., *Plant pathogen forensics: capabilities, needs, and recommendations*. *Microbiology and Molecular Biology Reviews*, 2006. **70**(2): p. 450-471.
 21. Paul Rogers, S.W.a.M.D., *Biological Warfare Against Crops*. *Scientific American*, 1999. **280**(6): p. 6.
 22. Bourke, P.M.A., *Emergence of potato blight, 1843-46*. *Nature*, 1964. **203**(4947): p. 805-808.
 23. APSnet, *Citrus Canker: The pathogen and its impact*, M.L. Elliott, Editor. 2002.
 24. Brown, K., *Florida fights to stop Citrus Canker*. *Science*, 2001. **292**(5525): p. 2275-2276.
 25. Ryan, J.R. and J.F. Glarum, *Biosecurity and bioterrorism: containing and preventing biological threats*. 2008, London, England: Elsevier, Inc.
 26. Fletcher, J., et al., *Microbial forensics and plant pathogens: attribution of agricultural crime*, in *Wiley Handbook of Science and Technology for Homeland Security*, J. Voeller, Editor. 2008, John Wiley & Sons, Inc. p. 1880-1894.
 27. Geissler, E. and J.E.V.C. Moon, *French activities related to biological warfare*, in *Biological and Toxin Weapons: Research, Development and Use from the Middle Ages to 1945 (Sipri Chemical & Biological Warfare Studies, No. 18.)*. 1999, Oxford University Press 1999. p. 70-90.

28. Whitby, S., *Biological warfare against crops*. 2002, Hampshire, UK: Palgrave.
29. Alibek, K., *The Soviet Union's anti-agricultural biological weapons*, in *Food and Agricultural Security: Guarding Against Natural Threats and Terrorist Attacks Affecting Health, National Food Supplies, and Agricultural Economics*, T.W. Frazier and D.C. Richardson, Editors. 1999, NY Acad. Sci.: New York. p. 18-19.
30. Whitby, S., *Anti-crop biological warfare - implications of the Iraq and US programs*. Defense & security analysis, 1997. **13**(3): p. 303.
31. Lievens, B., et al., *Detecting single nucleotide polymorphisms using DNA arrays for plant pathogen diagnosis*. FEMS Microbiology Letters, 2005. **255**(1): p. 129-138.
32. USDA-ARS. *Epidemiology and management of Xylella fastidiosa (Xf) and other exotic and invasive diseases and insect pests 2007 Annual Report 2007*;
Available from:
http://www.ars.usda.gov/research/projects/projects.htm?ACCN_NO=411381&fy=2007.
33. Drake, J., *Rates of spontaneous mutation among RNA viruses*. Proceedings of the National Academy of Sciences, 1993. **90**(9): p. 4171-4175.
34. Drake, J.W. and J.J. Holland, *Mutation rates among RNA viruses*. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(24): p. 13910-13913.
35. Steinhauer, D.A., E. Domingo, and J.J. Holland, *Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase*. Gene, 1992. **122**(2): p. 281-288.

36. Stenger, D.C., D.L. Seifers, and R. French, *Patterns of polymorphism in Wheat streak mosaic virus: sequence space explored by a clade of closely related viral genotypes rivals that between the most divergent strains*. *Virology Journal*, 2002. **302**(1): p. 58-70.
37. Domingo, E., *Quasispecies theory in virology*. *Journal of Virology*, 2002. **76**(1): p. 463-465.
38. Tagariello, G., et al., *Hepatitis C virus quasispecies in the natural course of HCV-related disease in patients with haemophilia*. *Haemophilia*, 2004. **10**(1): p. 81-86.
39. Naraghi-Arani, P., S. Daubert, and A. Rowhani, *Quasispecies nature of the genome of Grapevine fanleaf virus*. *Journal of General Virology*, 2001. **82**(7): p. 1791-1795.
40. Domingo, E., et al., *The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance -- a review*. *Gene*, 1985. **40**(1): p. 1-8.
41. French, R. and D.C. Stenger, *Population structure within lineages of Wheat streak mosaic virus derived from a common founding event exhibits stochastic variation inconsistent with the deterministic quasi-species model*. *Virology*, 2005. **343**(2): p. 179-189.
42. Jenkins, G.M., et al., *Evidence for the non-quasispecies evolution of RNA viruses*. *Molecular Biology and Evolution*, 2001. **18**(6): p. 987-994.
43. Hall, J.S., et al., *Structure and temporal dynamics of populations within Wheat streak mosaic virus isolates*. *Journal of Virology*, 2001. **75**(21): p. 10231-10243.

44. McNeil, J.E., et al., *Characterization of genetic variability among natural populations of Wheat streak mosaic virus*. *Phytopathology*, 1996. **86**(11): p. 1222-1227.
45. Stenger, D.C., et al., *Phylogenetic relationships within the family Potyviridae: Wheat streak mosaic virus and Brome streak mosaic virus are not members of the genus Rymovirus*. *Phytopathology*, 1998. **88**(8): p. 782-787.
46. Riechmann, J.L., S. Lain, and J.A. Garcia, *Highlights and prospects of potyvirus molecular biology*. *Journal of General Virology*, 1992. **73**(1): p. 1-16.
47. Fauquet, C.M., et al., eds. *Virus taxonomy: Eighth report of the International Committee on the Taxonomy of Viruses 2005*, Elsevier Academic: San Diego, CA. 819-841.
48. Choi, I.-R., et al., *A plant virus vector for systemic expression of foreign genes in cereals*. *The Plant Journal*, 2000. **23**(4): p. 547-555.
49. Choi, I.-R., et al., *Mapping of the PI proteinase cleavage site in the polyprotein of Wheat streak mosaic virus (genus Tritimovirus)*. *Journal of General Virology*, 2002. **83**(2): p. 443-450.
50. Choi, I.-R., D.C. Stenger, and R. French, *Multiple interactions among proteins encoded by the mite-transmitted Wheat Streak Mosaic Tritimovirus*. *Virology*, 2000. **267**(2): p. 185-198.
51. Chung, B.Y.-W., et al., *An overlapping essential gene in the Potyviridae*. *Proceedings of the National Academy of Sciences*, 2008. **105**(15): p. 5897-5902.
52. Kneller, E.P., A.M. Rakotondrafara, and W.A. Miller, *Cap-independent translation of plant viral RNAs*. *Virus Research*, 2006. **119**(1): p. 63-75.

53. Wei, T., et al., *Formation of complexes at plasmodesmata for Potyvirus intercellular movement is mediated by the viral protein P3N-PIPO*. PLoS Pathog, 2010. **6**(6): p. e1000962.
54. ExPASy. *ViralZone: Tritimovirus*. April 28, 2011]; Available from: http://expasy.org/viralzone/all_by_species/619.html.
55. Valli, A., et al., *RNA silencing suppression by a second copy of the P1 serine protease of Cucumber vein yellowing Ipomovirus, a member of the family Potyviridae that lacks the cysteine protease HCPro*. Journal of Virology, 2006. **80**(20): p. 10055-10063.
56. Kasschau, K.D., et al., *P1/HC-Pro, a viral suppressor of RNA silencing, interferes with Arabidopsis development and miRNA function*. Developmental cell, 2003. **4**(2): p. 205-217.
57. Verchot, J. and J. Carrington, *Evidence that the potyvirus P1 proteinase functions in trans as an accessory factor for genome amplification*. Journal of Virology, 1995. **69**(6): p. 3668-3674.
58. Saenz, P., et al., *Host-specific involvement of the HC protein in the long-distance movement of Potyviruses*. Journal of Virology, 2002. **76**(4): p. 1922-1931.
59. Anindya, R. and H.S. Savithri, *Potyviral NIa proteinase, a proteinase with novel deoxyribonuclease activity*. Journal of Biological Chemistry, 2004. **279**(31): p. 32159-32169.
60. Andrejeva, J., et al., *Potyvirus helper component-proteinase and coat protein (CP) have coordinated functions in virus--host interactions and the same CP*

- motif affects virus transmission and accumulation. Journal of General Virology, 1999. 80(5): p. 1133-1139.*
61. Tatineni, S., D.H. Van Winkle, and R. French, *The N-terminal region of Wheat streak mosaic virus coat protein is a host- and strain-specific long-distance transport factor. Journal of Virology, 2010: p. 1718-1731.*
 62. Carrington, J.C., P.E. Jensen, and M.C. Schaad, *Genetic evidence for an essential role for potyvirus CI protein in cell-to-cell movement. The Plant Journal, 1998. 14(4): p. 393-400.*
 63. Dolja, V.V., et al., *Capsid protein determinants involved in cell-to-cell and long distance movement of Tobacco etch Potyvirus. Virology, 1995. 206(2): p. 1007-1016.*
 64. Rodríguez-Cerezo, E., et al., *The coat and cylindrical inclusion proteins of a Potyvirus are associated with connections between plant cells. Virology, 1997. 236(2): p. 296-306.*
 65. Roberts, I.M., et al., *Ultrastructural and temporal observations of the Potyvirus cylindrical inclusions (CIs) show that the CI protein acts transiently in aiding virus movement. Virology, 1998. 245(1): p. 173-181.*
 66. French, R. and D.C. Stenger, *Evolution of Wheat streak mosaic virus: dynamics of population growth within plants may explain limited variation*. Annual Review of Phytopathology, 2003. 41(1): p. 199-214.*
 67. Luria, S.E., *The frequency distribution of spontaneous bacteriophage mutants as evidence for the exponential rate of phage reproduction. Cold Spring Harbor symposia on quantitative biology, 1951. 16: p. 463-70.*

68. Hall, J.S., et al., *Three distinct mechanisms facilitate genetic isolation of sympatric Wheat streak mosaic virus lineages*. *Virology*, 2001. **282**(2): p. 230-236.
69. Brakke, M.K., *Wheat streak mosaic virus*. CMI/AAB Descriptions of Plant Viruses, 1971. **48**.
70. Nyvall, R.F., *Field crop diseases*. Vol. 3. 1999, New York: Iowa State University Press.
71. Christian, M.L. and W.G. Willis, *Survival of Wheat streak mosaic virus in grass hosts in Kansas from wheat harvest to fall wheat emergence*. *Plant Disease*, 1993. **77**: p. 239-242.
72. Tatineni, S., et al., *Wheat cultivar-specific disease synergism and alteration of virus accumulation during co-infection with Wheat streak mosaic virus and Triticum mosaic virus*. *Phytopathology*, 2010. **100**(3): p. 230-238.
73. Mahmood, T., G. Hein, and S. Jenson, *Mixed infection of hard red winter wheat with High plains virus and Wheat streak mosaic virus from wheat curl mites in Nebraska*. *Plant Disease*, 1998. **82**(3).
74. Slykhuis, J.T., *Aceria tulipae Keifer (Acarina: Eriophyidae) in relation to the spread of wheat streak mosaic*. *Phytopathology*, 1955. **45**: p. 116-128.
75. Seifers, D.L. and T.J. Martin, *Identification of the wheat curl mite as the vector of Triticum mosaic virus*. *Plant Disease*, 2009. **93**(1).
76. Townsend, L., D. Johnson, and D. Hershman. *Wheat streak mosaic virus and the wheat curl mite*. 2010; Available from:
<http://www.ca.uky.edu/entomology/entfacts/ef117.asp>.

77. Harvey, T.L., T.J. Martin, and D.L. Seifers, *Wheat yield reduction due to wheat curl mite (Acari: Eriophyidae) infestations*. Journal of Agricultural and Urban Entomology, 2002. **19**: p. 9-13.
78. Jeppson, L.R., H.H. Keifer, and E.W. Baker, *Mites injurious to economic plants*. 1975, London, England: University of California Press, Ltd. .
79. Nault, L.R. and W.E. Styer, *The dispersal of Aceria tulipae and three other grass-infesting Eriophyid mites in Ohio*. Annals of the Entomological Society of America, 1969. **62**: p. 1446-1455.
80. Siritwetiwat, B., *Interactions between the wheat curl mite, Aceria tosichella Keifer (Eriophyidae), and Wheat streak mosaic virus and distribution of wheat curl mite biotypes in the field*. 2006, [Lincoln, Neb. : University of Nebraska-Lincoln.
81. Torell, R., et al., *Wheat pasture grazing: Agronomic, cultural and livestock management*, in *Fact Sheet 99-39*. 1999, University of Nevada Cooperative Extension Service. p. 4.
82. Fahim, M., et al., *Does grazing of infected wheat by sheep result in salivary transmission of Wheat streak mosaic virus?* Crop and Pasture Science. **61**(3): p. 247-254.
83. Chen, Q., R. Conner, and A. Laroche, *Molecular characterization of Haynaldia villosa chromatin in wheat lines carrying resistance to wheat curl mite colonization*. TAG Theoretical and Applied Genetics, 1996. **93**(5): p. 679-684.

84. Malik, R., et al., *Assessment of Aegilops tauschii for resistance to biotypes of wheat curl mite (Acari: Eriophyidae)*. Journal of Economic Entomology, 2003. **96**(4).
85. *A spontaneous translocation that transfers wheat curl mite resistance from decaploid Agropyron elongatum to common wheat*. Genome, 1988. **30**(3): p. 289-292.
86. Chen, Q., et al., *Molecular characterization of the genome composition of partial amphiploids derived from Triticum aestivum×Thinopyrum ponticum and T. aestivum×Th. intermedium as sources of resistance to wheat streak mosaic virus and its vector, Aceria tosichella*. TAG Theoretical and Applied Genetics, 1998. **97**(1): p. 1-8.
87. Carew, M., et al., *Molecular markers indicate that the wheat curl mite, Aceria tosichella Keifer, may represent a species complex in Australia*. Bulletin of Entomological Research, 2009. **99**(05): p. 479-486.
88. French, R.C., et al. *Amino acid substitutions of cysteine residues near the amino terminus of Wheat streak mosaic virus HC-Pro abolishes virus transmission by the wheat curl mite*. in American Society for Virology, 26th Annual Meeting. 2007. Corvallis, OR: Unpublished.
89. Botstein, D., et al., *Construction of a genetic linkage map in man using restriction fragment length polymorphisms*. The American Journal of Human Genetics, 1980. **32**(314-331).
90. Vos, P., et al., *AFLP: a new technique for DNA fingerprinting*. Nucleic Acids Research, 1995. **23**(21): p. 4407-4414.

91. Williams, J.G.K., et al., *DNA polymorphisms amplified by arbitrary primers are useful as genetic markers*. Nucleic Acids Research, 1990. **18**(22): p. 6531-6535.
92. Tautz, D., *Hypervariability of simple sequences as a general source for polymorphic DNA markers*. Nucleic Acids Research, 1989. **17**(16): p. 6463-6471.
93. Kidd, K.K., et al., *Developing a SNP panel for forensic identification of individuals*. Forensic Science International, 2006. **164**: p. 20-32.
94. United States Department of Energy Office of Science. *Human Genome Project Information*. 2008 September 19, 2008; Available from:
http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml.
95. International HapMap Consortium, *A haplotype map of the human genome*. Nature, 2005. **437**: p. 1299-1920.
96. Anno, S., T. Abe, and T. Yamamoto, *Interactions between SNP alleles at multiple loci contribute to skin color differences between caucasoid and mongoloid subjects*. Int J Biol Sci, 2008. **4**(2): p. 81-86.
97. Sturm, R.A., et al., *A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color*. The American Journal of Human Genetics, 2008. **82**(2): p. 424-431.
98. Enserink, M., *Anthrax investigation: Full-genome sequencing paved the way from spores to a suspect*. Science, 2008. **321**(5891): p. 898-899.
99. Lin, J.-J., J. Kuo, and J. Ma, *A PCR-based DNA fingerprinting technique: AFLP for molecular typing of bacteria*. Nucleic Acids Research, 1996. **24**(18): p. 3649-3650.

100. Duim, B., et al., *High-resolution genotyping of Campylobacter strains isolated from poultry and humans with amplified fragment length polymorphism fingerprinting*. Applied and Environmental Microbiology, 1999. **65**(6): p. 2369-2375.
101. Willems, R.J.L., et al., *Host specificity of Vancomycin-resistant Enterococcus faecium*. Journal of Infectious Diseases, 2000. **182**(3): p. 816-823.
102. Keim, P., et al., *Molecular evolution and diversity in Bacillus anthracis as detected by amplified fragment length polymorphism markers*. Journal of Bacteriology, 1997. **179**(3): p. 818-824.
103. Daffonchio, D., et al., *Strategy for identification of Bacillus cereus and Bacillus thuringiensis strains closely related to Bacillus anthracis*. Appl. Environ. Microbiol., 2006. **72**(2): p. 1295-1301.
104. Calmin, G., F. Lefort, and L. Belbahri, *Multi-loci sequence typing (MLST) for two Lacto-Acid Bacteria (LAB) species: Pediococcus parvulus and P. damnosus*. Molecular Biotechnology, 2008. **40**(2): p. 170-179.
105. Margos, G., et al., *MLST of housekeeping genes captures geographic population structure and suggests a European origin of Borrelia burgdorferi*. Proceedings of the National Academy of Sciences, 2008. **105**(25): p. 8730-8735.
106. Shigenobu, S., et al., *Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.* APS. Nature, 2000. **407**(6800): p. 81-86.
107. Clarridge, J.E., III, *Impact of 16S rRNA gene sequence analysis for identification of Bacteria on clinical microbiology and infectious diseases*. Clinical Microbiology Reviews, 2004. **17**(4): p. 840-862.

108. Comas, I.a., et al., *Genotyping of genetically monomorphic bacteria: DNA sequencing in Mycobacterium tuberculosis highlights the limitations of current methodologies*. PLoS ONE, 2009. **4**(11): p. e7815.
109. Rossetti, L. and G. Giraffa, *Rapid identification of dairy lactic acid bacteria by M13-generated, RAPD-PCR fingerprint databases*. Journal of Microbiological Methods, 2005. **63**(2): p. 135-144.
110. Seppola, M., et al., *Random amplification of polymorphic DNA (RAPD) typing of carnobacteria isolated from hindgut chamber and large intestine of Atlantic cod (Gadus morhua l.)*. Systematic and Applied Microbiology, 2006. **29**(2): p. 131-137.
111. Shangkuan, Y.H., et al., *Molecular characterization of Bacillus anthracis using multiplex PCR, ERIC-PCR and RAPD*. Letters in Applied Microbiology, 2001. **32**(3): p. 139-145.
112. Zhong, W., et al., *Differentiation of Bacillus anthracis, B. cereus, and B. thuringiensis by using pulsed-field gel electrophoresis*. Applied and Environmental Microbiology, 2007. **73**(10): p. 3446-3449.
113. Gautom, R., *Rapid pulsed-field gel electrophoresis protocol for typing of Escherichia coli O157:H7 and other gram-negative organisms in 1 day*. Journal Clinical Microbiology, 1997. **35**(11): p. 2977-2980.
114. Cliff, J.B., et al., *Differentiation of spores of Bacillus subtilis grown in different media by elemental characterization using time-of-flight secondary ion mass spectrometry*. Applied and Environmental Microbiology, 2005. **71**(11): p. 6524-6530.

115. Horita, J. and A. Vass, *Stable-isotope fingerprints of biological agents as forensic tools*. Journal of Forensic Sciences, 2003. **48**(1): p. 122-126.
116. Kreuzer-Martin, H.W., et al., *Microbe forensics: Oxygen and hydrogen stable isotope ratios in Bacillus subtilis cells and spores*. Proceedings of the National Academy of Sciences, 2003. **100**(3): p. 815-819.
117. Vandamme, P., et al., *Polyphasic taxonomy, a consensus approach to bacterial systematics*. Microbiological Reviews, 1996. **60**(2): p. 407-438.
118. Leonard, R., et al., *Comparison of MIDI Sherlock system and pulsed-field gel electrophoresis in characterizing strains of methicillin-resistant Staphylococcus aureus from a recent hospital outbreak*. Journal of Clinical Microbiology, 1995. **33**(10): p. 2723-2727.
119. Wintzingerode, F.v., et al., *Identification of environmental strains of Bacillus mycoides by fatty acid analysis and species-specific 16S rDNA oligonucleotide probe*. FEMS Microbiology Ecology, 1997. **24**(3): p. 201-209.
120. Welch, D.F., *Applications of cellular fatty acid analysis*. Clinical Microbiology Reviews, 1991. **4**(4): p. 422-438.
121. Lu, Y. and P. Harrington, *Classification of bacteria by simultaneous methylation–solid phase microextraction and gas chromatography/mass spectrometry analysis of fatty acid methyl esters*. Analytical and Bioanalytical Chemistry. **397**(7): p. 2959-2966.
122. Xu, M., F. Basile, and K.J. Voorhees, *Differentiation and classification of user-specified bacterial groups by in situ thermal hydrolysis and methylation of whole*

- bacterial cells with tert-butyl bromide chemical ionization ion-trap mass spectrometry*. Analytical Chimica Acta, 2000. **418**(2): p. 119-128.
123. Merrill, L. and L. Halverson, *Seasonal variation in microbial communities and organic malodor indicator compound concentrations in various types of swine manure storage systems*. Journal of Environmental Quality, 2002. **6**: p. 2074-2085.
124. Siciliano, S., et al., *Differences in the microbial communities associated with the roots of different cultivars of canola and wheat*. Canadian Journal of Microbiology, 1998. **44**: p. 844-851.
125. Kaneda, T., *Factors affecting the relative ratio of fatty acids in Bacillus cereus*. Canadian Journal of Microbiology, 1971. **17**(2): p. 269-275.
126. Kaneda, T., *Fatty acids in the genus Bacillus I. Iso- and Anteiso-fatty acids as characteristic constituents of lipids in 10 species*. Journal of Bacteriology, 1967. **93**(3): p. 894-903.
127. Weerkamp, A. and W. Heinen, *The effect of nutrients and precursors on the fatty acid composition of two thermophilic bacteria*. Archives of Microbiology, 1972. **81**(4): p. 350-360.
128. Ehrhardt, C.J., et al., *Discrimination of Bacillus cereus T-strain spores grown on different media using Fatty Acid Methyl Ester (FAME) profiles*. Applied and Environmental Microbiology, 2010: p. AEM.02443-09.
129. Deyong, Z., et al., *Differentiation of Cucumber mosaic virus isolates by hybridization to oligonucleotides in a microarray format*. Journal of Virological Methods, 2005. **123**(1): p. 101-108.

130. Tiberini, A., et al., *Oligonucleotide microarray-based detection and identification of 10 major tomato viruses*. Journal of Virological Methods. **168**(1-2): p. 133-140.
131. Engel, E.A., et al., *A diagnostic oligonucleotide microarray for simultaneous detection of grapevine viruses*. Journal of Virological Methods. **163**(2): p. 445-451.
132. Abdullahi, I. and M. Rott, *Microarray immunoassay for the detection of grapevine and tree fruit viruses*. Journal of Virological Methods, 2009. **160**(1-2): p. 90-100.
133. Pasquini, G., et al., *Oligonucleotide microarray-based detection and genotyping of Plum pox virus*. Journal of Virological Methods, 2008. **147**(1): p. 118-126.
134. Lenz, O., K. Petrzik, and J. Spak, *Investigating the sensitivity of a fluorescence-based microarray for the detection of fruit-tree viruses*. Journal of Virological Methods, 2008. **148**(1-2): p. 96-105.
135. Boonham, N., et al., *Detection of potato viruses using microarray technology: towards a generic method for plant viral disease diagnosis*. Journal of Virological Methods, 2003. **108**(2): p. 181-187.
136. Bystricka, D., et al., *Oligonucleotide-based microarray: A new improvement in microarray detection of plant viruses*. Journal of Virological Methods, 2005. **128**(1-2): p. 176-182.
137. Lee, G.P., et al., *Plant virus cDNA chip hybridization for detection and differentiation of four cucurbit-infecting Tobamoviruses*. Journal of Virological Methods, 2003. **110**(1): p. 19-24.

138. Pelludat, C., B. Duffy, and J. Frey, *Design and development of a DNA microarray for rapid identification of multiple European quarantine phytopathogenic bacteria*. European Journal of Plant Pathology, 2009. **125**(3): p. 413-423.
139. Costa de Oliveira, R., et al., *Competitive hybridization on spotted microarrays as a tool to conduct comparative genomic analyses of Xylella fastidiosa strains*. FEMS Microbiology Letters, 2002. **216**(1): p. 15-21.
140. Cho, J.-C. and J.M. Tiedje, *Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays*. Applied and Environmental Microbiology, 2001. **67**(8): p. 3677-3682.
141. Lezar, S. and E. Barros, *Oligonucleotide microarray for the identification of potential mycotoxigenic fungi*. BMC Microbiology, 2010. **10**(87).
142. Szemes, M., et al., *Diagnostic application of padlock probes-multiplex detection of plant pathogens using universal microarrays*. Nucleic Acids Research, 2005. **33**(8): p. e70.
143. Teletchea, F., et al., *Molecular identification of vertebrate species by oligonucleotide microarray in food and forensic samples*. Journal of Applied Ecology, 2008. **45**(3): p. 967-975.
144. Sigurdsson, S., et al., *A microarray system for genotyping 150 single nucleotide polymorphisms in the coding region of human mitochondrial DNA*. Genomics, 2006. **87**(4): p. 534-542.
145. Vallone, P., J. Jakupciak, and M. Coble, *Forensic application of the Affymetrix human mitochondrial resequencing array*. Forensic Science International. Genetics, 2007. **1**(2): p. 196-198.

146. Balogh, M.K., K. Bender, and P.M. Schneider, *Application of Nanogen microarray technology for forensic SNP analysis*. International Congress Series, 2006. **1288**: p. 43-45.
147. Bogus, M., et al., *Rapid microarray-based typing of forensic SNPs*. International Congress Series, 2006. **1288**: p. 37-39.
148. Divne, A.-M. and M. Allen, *A DNA microarray system for forensic SNP analysis*. Forensic Science International, 2005. **154**(2-3): p. 111-121.
149. Divne, A.M., et al., *A novel DNA microarray system for analysis of limited forensic evidence material*. International Congress Series, 2003. **1239**: p. 9-10.
150. Raitio, M., et al., *Y-chromosomal SNPs in Finno-Ugric-speaking populations analyzed by minisequencing on microarrays*. Genome Research, 2001. **11**(3): p. 471-482.
151. Krjutskov, K., et al., *Evaluation of the 124-plex SNP typing microarray for forensic testing*. Forensic Science International: Genetics, 2009. **4**(1): p. 43-48.
152. Lareu, M., et al., *Typing Y-chromosome single nucleotide polymorphisms with DNA microarray technology*. International Congress Series, 2003. **1239**: p. 21-25.
153. Willse, A., et al., *Quantitative oligonucleotide microarray fingerprinting of Salmonella enterica isolates*. Nucleic Acids Research, 2004. **32**(5): p. 1848-1856.
154. Chandler, D.P., et al., *Diagnostic oligonucleotide microarray fingerprinting of Bacillus isolates*. J. Clin. Microbiol., 2006. **44**(1): p. 244-250.
155. Metzker, M.L., et al., *Molecular evidence of HIV-1 transmission in a criminal case*. Proceedings of the National Academy of Sciences, 2002. **99**(22): p. 14292-14297.

156. Budowle, B. and R. Harmon, *HIV legal precedent useful for microbial forensics*. Croatian Medical Journal, 2005. **46**(4): p. 515-521.
157. Keim, P., *Microbial forensics: A scientific assessment*. 2003, American Academy of Microbiology: Washington, DC.
158. Butler, J., *Fundamentals of forensic DNA typing*. 2010, San Diego: Elseiver, Inc. .

CHAPTER III

ASSESSING THE MOLECULAR DIFFERENCES BETWEEN PLANT VIRUS POPULATIONS COLLECTED FROM NATURALLY AND ARTIFICIALLY-INFECTED PLANTS: IMPLICATIONS FOR MICROBIAL FORENSICS

Abstract

Plant pathogens can affect crop production by reducing the yield and nutritional value of foodstuffs. The utilization of plant pathogens as bioweapons could disturb national and global markets. Protecting against biocrimes involving plant pathogens includes the development of a complete plant pathogen forensic capability. Prior to launching a full-scale forensic investigation, responsible agencies must determine whether an outbreak is likely to be naturally or malevolently caused. Currently, few resources are available to aid in this determination. Here, we report the molecular evaluation of two populations of plant viruses with the goal of identifying molecular differences between a natural and an intentional inoculation event. One population was derived from two natural outbreaks of *Wheat streak mosaic virus* (WSMV) and the other population consisted of two groups of wheat that were mechanically inoculated with WSMV. Lower frequencies of recombination were observed in intentionally inoculated plants, as compared to field-derived plants. Additionally, the number of unique

haplotypes, haplotype diversities and average pair-wise nucleotide diversities were higher in the field-derived sequences than sequences derived from mechanically-inoculated plants. To our knowledge, this is the first report of the application of molecular analyses to identify differences between natural and intentional inoculation events for use in plant pathogen forensics. These findings may provide additional support for forensic scientists and government agencies charged with the task of determining if a plant virus was deliberately introduced to a field.

Introduction

The potential weaponization of plant pathogens for the intentional dissemination upon a crop system is a growing concern for producers and consumers worldwide [1-3]. The prospective use of plant pathogens as bioweapons has prompted much discussion over the past decade [4-9], demonstrating an increasing awareness of the possibilities associated with their weaponization. The deliberate introduction of pathogens into a cropping system could have severe consequences for the producer, surrounding crops, the crop system as a whole, and potentially other downstream processes that are dependent upon the commodity. For example, the United States is the world's largest producer of corn, generating over ten billion bushels of corn each year [10], most of which is used to feed domestic and over-seas livestock [11]. Corn is also being explored as an alternative fuel source, thus a successful biological attack on corn production could deleteriously affect livestock production and increase the cost of fuel. Such an attack could also reduce the crop yield and nutritional value of products, potentially harm the environment, negatively affect trade and damage the U.S. economy [8].

The ability to attribute a biocrime to an individual or a group of individuals after a biocrime or bioterrorist event has occurred is essential to maintaining a balanced biosecurity plan and deterring biological attacks on plants [12]. A complete microbial forensic capability involves the determination of whether an outbreak occurred by means of a natural or deliberate inoculation, identification of the source of an outbreak pathogen and the ultimate attribution of the crime [13]. Since forensic investigations are labor-intensive, time-consuming and expensive, the ability to discriminate between an intentional introduction of a plant pathogen and a natural outbreak is desirable. One mechanism of rapidly assessing the likelihood that an intentional inoculation event has occurred is by examining the disease patterns via global positioning systems (GPS) and remote sensing technologies [14]. Currently, tools that are capable of discriminating specifically between likely natural and intentional introductions are limited. One such tool, recently developed using *Wheat streak mosaic virus* as a model pathogen, presents weighted factors that identify deviations from observations of a natural inoculation event [15]. This tool includes some molecular information about the virus. However additional information regarding the patterns of inheritance and genetic relatedness of pathogens at a molecular level would enhance its capability.

Previous research has demonstrated that plant viruses exist within host plants as mixed populations of sequence variants [16]. The diversity within these populations of closely-related genotypes arise mostly through errors by the RNA-dependent RNA polymerase (RDRP) during replication, which serves as a mechanism for the rapid adaptation of viruses to changing selection pressures [17]. Various processes, including selection, mutation, genetic drift, genetic bottlenecks and recombination, can lead to the

production of new virus sequences or the establishment of minor sequences within plants. The population diversity of plant viruses within hosts can be substantial and is dependent upon host-virus interactions [18]. Variability among viral isolates collected from different fields is quite complex [19, 20]. Within-field variability of WSMV is comparable to that of WSMV isolates collected in different fields within a small geographic region [21]. Thus, during a natural outbreak of a plant virus disease, populations are expected to be variable among plants within individual fields. In contrast, a deliberate introduction of plant viruses into a field is expected to result in a relatively homogeneous distribution of viral genotypes throughout the field, assuming that a single source of inoculum is used to infect all plants. The expectation of homogeneity would provide a molecular basis to aid in the assessment of whether a viral introduction resulted from an artificial or natural event. To test the hypothesis that plant virus natural disease outbreaks differ molecularly from plant viruses isolated from intentionally inoculated plants, nucleic acid and translated amino acid sequences from *Wheat streak mosaic virus* collected from two fields in Oklahoma were compared to viral sequences derived from two sets of mechanically-inoculated plants.

To assess the degree of molecular difference between the two types of inoculation events, viral sequences from the two inoculation events were subjected to a series of molecular tests. Molecular variance was assessed using an analysis of molecular variance (AMOVA) for the entire population of sequences, groups of sequences (the two types of inoculation events) and sub-populations within the groups (two sets of manually-inoculated plants and plants collected from two naturally-occurring outbreaks of WSMV). Pair-wise sequence heterogeneity was also assessed between the two types of

inoculation events and for sub-populations within the two groups using F_{ST} . The number of haplotypes and the haplotype and nucleotide diversities were also compared between sequences in each type of inoculation event.

To determine whether the two populations of plant viruses were subject to different evolutionary pressures, three tests of neutrality were performed and Ka/Ks ratios were determined for the individual data sets. Individual data sets were also examined for their transition/transversion preferences and for evidence of recombination. We report molecular differences between populations of WSMV nucleic acid sequences derived from a natural setting and those from mechanically-inoculated plants. These data may provide the forensic community with useful information to assist in determining whether a disease outbreak is natural or deliberate.

Materials and Methods

Sample collection from naturally-infected fields

Fifteen plants were collected from two western Oklahoma wheat fields that showed typical symptoms of WSMV and were assumed to be naturally infected with the virus. Fields were sampled by collecting twenty-five single wheat tillers at three-meter intervals, moving in a direction perpendicular to the suspected linear sources of inocula for the two outbreaks. The two fields were located in Texas County, OK and were separated by approximately three kilometers. Plants in field one were mildly symptomatic, presenting streaking and yellowing that was observed throughout a presumed infection gradient that extended approximately 50 meters into the field. Plants

in field two displayed severe streaking and stunting symptoms characteristic of early infection by WSMV.

Growth chamber plant growth and inoculation

To propagate WSMV for use as a source of inoculum, twenty-five *Triticum aestivum* cv. Chisholm wheat seeds were planted (5 seeds in each pot, each pot approximately 5” tall and 4” in diameter) in Sunshine Redi-earth plug and seedling mix [Sun Gro Horticulture, Inc., Bellevue, WA, USA] and grown in a plant growth chamber [Percival Scientific, Inc., Perry, IA, USA] using a 15 hour-day cycle with minimum and maximum temperatures of 19°C and 25°C, respectively. Seven-10 day old seedlings were mechanically inoculated with a laboratory-derived strain of WSMV (‘OK-LAB’), which was collected in a wheat field in southern Oklahoma, propagated in healthy wheat plants and stored at -80°C until being used as a source of inoculum. Mechanical inoculation was achieved by grinding 100 mg of WSMV-infected frozen leaf tissue in 500 µl of water using a mortar and pestle, combining with Celite[®] [Manville Products, Inc., Denver, CO, USA] and rubbing leaf tissue (one leaf/plant) until the leaf appeared to be water soaked. Tissue was harvested 14-21 days post-inoculation by cutting leaf tissue in approximately 1” long sections, thoroughly mixing clippings and storing at -80°C until being used as a source of inoculum.

Growth chamber-grown individual tillers that were used in this study were planted and grown using the previously described conditions. Two batches of *Triticum aestivum* cv. Chisholm plants (25 plants per batch) were mechanically inoculated with the mixture of clippings, as previously described, 7-10 days after planting. Whole wheat

tillers were harvested 14-21 days post-inoculation, packaged individually and stored at -80°C, prior to RNA extraction and sequencing.

RNA extraction and RT-PCR

RNA was extracted from 100mg of tissue from each of the individual tillers using the RNeasy Plant Minikit (#74904) [Qiagen, Inc., Germantown, MD, USA]. Total RNA was eluted in 50 µL of nuclease-free water and used immediately as a template for reverse transcription.

Reverse transcription was performed in 20 µL reactions containing 200 U of Moloney murine leukemia virus reverse transcriptase (#28025-013) [Invitrogen, Inc., Carlsbad, CA, USA], 2 pmoles of primer C1 (5'-TAC TTG ACT GGG ACC CGA A-3') [22], approximately 2 µg of total RNA, 2 mM of each dNTP, 0.1 mM DTT, first-strand buffer and nuclease-free water, according to the manufacturer's instructions. Reverse transcription reaction tubes were incubated in a 37°C water bath for one hour prior to inactivation of the enzyme at 65°C for 15 minutes. Reverse transcription products were used as templates in polymerase chain reactions containing 1 U of platinum *Taq* DNA polymerase high fidelity (# 11304-102) [Invitrogen, Inc., Carlsbad, CA, USA], 0.2 mM each dNTP, 0.4 mM MgSO₄, 0.2 µM of primer C1 and 0.2 µM of primer XC1 (5'-AAC CCA CAC ATA GCT ACC AAG-3') [22], 1x high fidelity PCR buffer and water in a total reaction volume of 50 µL. Tubes were placed in a programmable thermal controller (PTC-100) [MJ Research, Inc., St. Bruno (Quebec), Canada] and initially denatured by heating at 94°C for 1 minute before 35 cycles of 1 minute at 94°C, 45 seconds at 50°C and 1 minute at 68°C were performed, followed by a final extension step at 68°C for 10 minutes.

To remove residual primers and dNTPs, RT-PCR products from WSMV-infected plant samples were subjected to the QIAquick[®] PCR Purification Kit (#28104) [Qiagen, Inc., Germantown, MD, USA] and the amplicons were purified according to the manufacturer's instructions.

DNA sequencing and clean-up

Purified RT-PCR products were sequenced using the BigDye[®] Terminator v1.1 Cycle Sequencing Kit (#4337450) [Life Technologies, Inc., Carlsbad, CA, USA] in reactions containing 2 μ L of the BigDye reaction mix, 1 μ M of the C1 or XC1 primer, 1x buffer, 1 μ L of template and water, in a final volume of 10 μ L. Cycling reactions were performed in a T-Personal Thermocycler [Biometra, Göttingen, Germany]. An initial denaturation step of 96°C for 2 minutes was followed by sixty cycles of 96°C for 30 seconds, 50°C for 30 seconds and 60°C for 2 minutes.

Sequencing products were precipitated by adding 12 μ L of nuclease-free water, 5 μ L of 3 M ammonium acetate and 57 μ L of 100% ethanol to the sequencing reactions, mixing and centrifuging at 1400 RCFs for 30 minutes. Pellets were subsequently washed in 70% ethanol and centrifuged at 1400 RCFs for 15 minutes. Residual ethanol was removed from the tubes and the pellets were incubated at room temperature for 10 minutes prior to re-suspension in nuclease-free water. Sequencing products were subjected to capillary electrophoresis using an ABI Model 3730 DNA Analyzer (#3730S) [Life Technologies, Inc., Carlsbad, CA, USA].

Replicate sequences were aligned using the EMBOSS [23] alignment program *emma*, which is EMBOSS' interface to Clustal W [24], and consensus sequences were generated using the EMBOSS *cons* program. Consensus sequences were subsequently

aligned using Clustal W in MEGA5 [25]. A standard IUB weight matrix was applied to nucleic acid sequences and a Gonnet matrix was used for amino acid alignments. Trace files were examined for variable and absent nucleotides and dubious sites were assigned manually, based on peak height. If a base was not clearly defined at a locus, the position was assigned an 'N'. Consensus sequences were inspected manually and variable ends were trimmed to remove erroneous positions, leaving 898 nucleotide-long sequences corresponding to positions 8297-9195 of WSMV Sidney 81 (AF057533). The trimmed sequences were used in all of the analyses. After the final sequences were determined, translated amino acid sequence alignments from each of the four data sets were submitted to EMBOSS' plotcon software to assess the quality of sequence conservation, using a standard window size of 4 (Fig. 3).

Molecular analyses

To determine if WSMV populations that were collected from the panhandle differentiated from WSMV that was originally collected from southern Oklahoma, pair-wise F_{ST} comparisons and analyses of the molecular variance (AMOVA) were conducted using Arlequin v. 3.11 [26]. Mean similarity scores were determined for individual data sets using a standard IUB weight matrix for nucleic acid and a Gonnet weight matrix for amino acid sequences, using the Clustal W program [24]. For combined data sets, percent similarities were generated for each individual data set and combined data sets, and used to assess the over-all mean similarity scores of the data sets. Haplotype and pair-wise nucleotide diversities were calculated, tests for neutrality were performed and K_a/K_s ratios were determined using the DNASP [27] software package. MEGA5 was

used to determine average substitution rates for transitions and transversions [25] and RDP was used to assess recombination within groups of data sets [28].

To assess the relatedness of the WSMV amplicons collected from the two fields and the manually-inoculated sequences, a UPGMA phylogenetic tree was generated for nucleic acid sequences using the software MEGA5 [25]. Evolutionary distances were calculated for nucleic acid sequences using the Kimura 2-paramater method [29] and 1000 replications were used to construct a consensus bootstrap tree [30].

Results

WSMV-infected tissue was collected from two fields located in the Oklahoma panhandle (Texas County) during a single growing season. Additionally, two sets of *Triticum aestivum* cv. Chisholm (25 individual plants per set) were inoculated from a single source of viral inoculum and total RNA was used as a template in RT-PCR prior to sequencing. To determine if the range of viral genome sequence variability would differ between naturally-infected wheat and mechanically-infected plants, WSMV-derived nucleic acid sequences obtained from infected plants in the two fields were compared to those from mechanically-inoculated, growth-chamber grown wheat using a variety of molecular analyses.

The infection rates for the two sets of mechanically-inoculated plants were 58.8% and 45.8% for ‘Laboratory 1’ and ‘Laboratory 2’, respectively. 100% of the plants collected from both fields were positive for WSMV, as determined by the presence of a 1,254 bp amplicon corresponding to nucleotides 8117-9371 of WSMV Sidney 81 (AF057533) (Fig. 1).

Field-derived groups of virus sequences and mechanically-inoculated virus sequences were all closely related and did not cluster separately in the phylogenetic tree, as indicated by the low bootstrap values for most of the nodes (Fig. 2). All field and laboratory-derived sequences were more closely related to the two American WSMV strains (Sidney 81 and Type) than to the divergent El Batán strain of WSMV that occurs in central Mexico [31].

Mean similarity scores were calculated for each individual sample set and for combined field and lab sample sets (Table 1). Nucleotide sequences were also converted to amino acid sequences and pair-wise amino acid similarity scores were generated for individual and combined data sets. Mean similarity scores for nucleic acid alignments of individual data sets ranged from 96.62% to 97.69%. Average similarity scores for amino acid alignments from individual data sets were higher than their nucleic acid counterparts, with the exception of Field 2, and ranged from 96.22% to 98.31%. The average similarity scores for field and laboratory-derived WSMV nucleic acid sequences were comparable to similarity scores between the two North American strains of WSMV, Sidney 81 (AF057533) and Type (AF285169) (97% similar), using the same query position. The similarity scores for the amino acid sequences were all higher than the WSMV Type and Sidney 81 mean amino acid similarity score (96%). The conservation of amino acid alignments for each group revealed distinct patterns of invariable residues (Fig. 3).

Structure of field and laboratory-derived WSMV sequence populations

To characterize the sequence populations, the four data sets were subjected to an analysis of molecular variance (AMOVA). The total contribution to the total variance

was distributed among sub-populations within groups (Lab 1, Lab 2, Field 1 and Field 2), between groups (combined Lab sequence sets and combined Field sequence sets) and within the over-all population (all sequences combined). Only a fraction of the total variance (0.36%) was observed among sub-populations within groups. The variance observed among laboratory-derived and field-derived sequences was somewhat higher (5.02%) than the contribution of variance among groups. However, the vast majority of the total variance (94.62%) was observed within the over-all population of sequences.

The genetic differentiation of WSMV coat protein sequences was estimated by determining pair-wise F_{ST} values, using a 95% confidence interval. Pair-wise comparisons of the two sub-populations in each group (Lab 1 vs. Lab 2) yielded a non-significant F_{ST} value (0.006), as did the two field-derived (Field 1 vs. Field 2) data sets ($F_{ST}=0.003$). However, a significant F_{ST} value (0.052) was observed when the combined field-derived sequences were compared to the combined laboratory-derived sequences.

Genetic diversity

Genetic diversity was measured by determining the number of unique haplotypes in each sequence set (Lab and Field), calculating the haplotype diversity and comparing the average pair-wise nucleotide diversities of sequences in each of the two types of inoculation events.

The number of unique haplotypes observed in the field-derived sequence set (20/20) was greater than that of the mechanically-inoculated plants (16/19). Likewise, the haplotype and pair-wise nucleotide diversities were slightly higher in the field-collected sequence sets (1.00 and 0.023) than the laboratory-inoculated sequence sets (0.98 and 0.018), respectively.

Tests of neutrality

To determine if selection pressures acted upon viruses in the two types of data sets differently, individual and combined data sets were subjected to three tests of neutrality (Table 2). All three tests are based on the null hypothesis that all mutations are selectively neutral [32]. Tajima's D statistic compares the differences between the number of segregating sites (S) and the average number of nucleotide differences (k) [33], Fu and Li's D^* statistic is based on the number of singletons (mutations occurring only once throughout the data set) and the total number of mutations and the F^* statistic is based on differences between the number of singletons and the average number of differences (between) pairs of sequences (k) [34].

The three tests of neutrality were not significant at $\alpha=0.05$ for any of the individual data sets, indicating that the sequences are selectively neutral. Combining the two laboratory data sets and the two field-derived nucleic acid sequence sets resulted in a significant test statistic for Tajima's D ($P<0.05$), suggesting that the sequences are not selectively neutral, but not for Fu and Li's F^* or D^* statistics.

Mean pair-wise diversity estimates (π) [35] among all four groups of individual data sets and the two combined data sets were similar, and ranged from 1.4%-2.7%. These estimates of pair-wise diversities are consistent with mean pair-wise nucleotide diversity estimations among whole genome sequences of closely related WSMV isolates (2% of 9384 nucleotides) [36].

The Watterson's estimator ($\hat{\theta}_w$), which is used to estimate the mutation rate of a population, was slightly lower for the combined lab-derived sequences than each of the

individual groups, separately. The Ka/Ks ratios for all four individual and the two combined data sets were below 1.

Transition and transversion bias

Transition (purine/purine and pyrimidine/pyrimidine) and transversion (purine/pyrimidine) substitution rates were determined for each data set by the Maximum Composite Likelihood (MCL) estimate of the pattern of nucleotide substitution [37] (Table 3). Because the parent nucleotide at any given position cannot be determined with absolute certainty, the directionality of nucleotide changes was ignored.

A clear bias for transitions was observed in all four individual data sets. This result is consistent with previously reported transition/transversion ratios for *Wheat streak mosaic virus* coat protein sequences from closely related U.S., Canadian and Turkish isolates [36]. C↔U transitions were consistently higher than A↔G transitions in all data sets, while the transversions (A↔C, A↔U, G↔C, G↔U) occurred at a much lower, but constant frequency. The patterns of substitutions were somewhat variable for the laboratory-derived WSMV nucleic acid sequences, but consistent within the two field data sets. All four data sets displayed high *R* values, indicative of a strong bias for transitions over transversions.

Recombination

Nucleic acid sequences from each data set were tested for evidence of recombination using recombination detection programs in RDP3 [28]. Recombination events that were supported by at least two different statistical tests ($\alpha=0.05$) are reported in Table 4. Three recombination events were detected in the plant samples collected

from the two fields (15%); however, no evidence of recombination was observed in any of the mechanically-inoculated plants.

Discussion

The ability to discriminate between a naturally occurring plant virus disease outbreak and one due to the deliberate introduction of plant viruses is essential to microbial forensics. We report molecular differences between the two events discerned by examining WSMV amplicons obtained from wheat plants in a natural outbreak, and those produced from the deliberate inoculation of wheat using WSMV. To our knowledge, this is the first study that assesses the molecular differences between naturally-occurring and intentionally-inoculated plants using a plant virus, for use in microbial forensics.

An abundance of nucleotide substitutions was observed in the examined viral genome regions (252 observed nucleotide substitutions). These polymorphisms appeared to occur randomly across the examined region of the coat protein, but were slightly more concentrated at both ends of the sequences than in the interior regions. No obvious patterns of substitution were observed except for the strong preference for synonymous substitutions and the transition bias.

Previous findings suggested that random drift is the major contributor to the evolution of the *Wheat streak mosaic virus* genome, that negative selection contributes to the observed variability in WSMV and that the two processes can simultaneously affect different regions of the WSMV genome [31]. Our findings were consistent with these reports. The patterns of conservation of amino acids at specific positions throughout the

coat protein amplicons (Fig. 3) suggests that mutations at some amino acid positions are under strong negative selection, while others are somewhat plastic and tolerant to change.

The highest contribution of molecular variance observed using AMOVA was found when all of the sequences were combined (94.62%). The laboratory and field-derived sequence sets (lab and field) contributed 5.02% and the sub-populations in each group (Lab 1, Lab 2, Field 1 and Field 2) contributed only 0.36% to the total variance. In the absence of genetic differentiation, the variance is expected to be distributed among groups within the hierarchy. Since this was not the case, pair-wise F_{ST} values were calculated for the two groups (Field vs. Lab), and were found to be significant between the two groups, but not within individual groups (Field vs. Field and Lab vs. Lab). These data indicate that WSMV sequences derived from the two sources of inocula are differentiated, which may be due to spatial or temporal effects or the type of inoculation event. Because the reason for the observed differentiation cannot be determined with confidence, sequences in each of the two data sets were not directly compared. Instead, measurements of diversity were determined separately for each type of inoculation event and subsequently compared. The number of unique haplotypes, the haplotype diversities and pair-wise nucleotide diversities were all higher in the field-derived sequence set than in the mechanically-inoculated sequence set. These data indicate that differences in genetic diversity exist between the two types of inoculation events and may be useful in discriminating between an intentional introduction event and a natural virus outbreak.

Recombination events were observed in consensus sequences of WSMV isolated from the two fields; however, there was no evidence of any recombination in consensus sequences derived from mechanically-inoculated plants. One explanation for this

observation is that the relatively shorter infection duration of the inoculated plants likely limited the amount of recombination within the manually-inoculated plants. Plants growing in the field could have been inoculated up to three months before sampling occurred. Still, some evidence of recombination is expected even for short incubation periods. The complete absence of recombination therefore suggests that an extremely low frequency or the complete absence of recombination could serve as a marker for intentionally inoculated plants; however, additional research is necessary to confirm these ideas. If the time difference between the two types of inoculation events is solely responsible for the lack of recombination observed within the mechanically-inoculated plants, it may be possible to estimate an infection time based on the number of observed recombination events.

The average ratio of non-synonymous (K_a) to synonymous (K_s) substitutions is an indicator of the amount and type of selective pressure acting upon a gene. A K_a/K_s ratio >1 indicates positive selection, a K_a/K_s ratio <1 implies that negative selection is acting upon the coding sequence and a ratio equal to one indicates that the gene is evolving neutrally. The low K_a/K_s ratios observed for all four individual data sets implies that negative selection is influencing the viral evolution at some nucleic acid positions in plants from both the field and manually-inoculated WSMV coat protein sequences. This interpretation is further supported by the obvious patterns of conserved amino acids (and variable amino acids) for the four data sets (Fig. 2). While Fu and Li and Tajima's tests indicated that the sequences were evolving neutrally, these tests are based on the number of segregating sites and the average number of nucleotide differences, without regard for the position of the substitutions. A significant test statistic

was observed for Tajima's D in the combined field data sets. This can probably be explained by the detection of recombination within both field data sets, since Tajima's D and other tests for neutrality assume that recombination is absent within the queried data set [38].

There are some considerations regarding the applicability of this research to an actual introduction event. Since the manually-inoculated wheat used as a surrogate in this study was grown and maintained in a growth chamber, environmental effects that could have influenced the viral evolution of populations in the field were largely absent. Since the two populations (lab and field) were shown to be evolving stochastically with some sites being selective, these differences should not have affected the rate of viral evolution. Still, the extent to which environmental factors shape populations of viruses is unclear, and further studies involving a field-inoculation with WSMV should be conducted to confirm the observed differences between the two scenarios. Additionally, the suspected shorter incubation time of the laboratory-inoculated samples, as compared to the samples collected from the field, could have resulted directly in the reduced variability and the lack of recombination within the manually-inoculated plants. If this were true, molecular examination of samples collected from the targeted field would need to be tested soon after an attack was suspected to have occurred.

The molecular analyses performed during this study were based on the assumption that a bioterrorist or a biocriminal would utilize a relatively homogeneous source of inoculum. If this assumption were violated, the molecular differences observed in this study may not hold true. For this, and other unforeseen reasons, a multi-pronged approach to plant pathogen forensics, including a combination of molecular testing and

traditional forensic investigation, is necessary to develop a complete plant pathogen forensic capability.

The application of molecular-based viral population analyses during a forensic investigation involving the weaponization of plant viruses may complement currently available forensic capabilities and aid decision makers, investigators and prosecutors charged with protecting the public from biocrimes and acts of bioterrorism. The molecular analyses described here would thus be one of many tools that would be used during a plant pathogen forensic investigation.

Acknowledgements

The authors would like to thank Dr. Carla Garzon for her assistance with the viral population analyses and data interpretation; Dr. Stephanie Rogers for her assistance with wheat growth and inoculation; Ken Thomas and Janet Rogers for their assistance with sequencing. This research was funded by USDA-NIFA and the Oklahoma Agricultural Experiment Station.

LITERATURE CITED

1. Suffert, F., É. Latxague, and I. Sache, *Plant pathogens as agroterrorist weapons: assessment of the threat for European agriculture and forestry*. Food Security, 2009. **1**(2): p. 221-232.
2. Lutwick, S.M., et al., *Plant pathogens as biological weapons against agriculture*, in *Beyond Anthrax*. 2009, Humana Press. p. 1-29.
3. Gullino, M.L., et al., *Crop biosecurity: definitions and role in food safety and food security*, in *Crop Biosecurity*. 2008, Springer Netherlands. p. 1-10.
4. Young, J.M., et al., *Plant-pathogenic bacteria as biological weapons - real threats?* Phytopathology, 2008. **98**(10): p. 1060-1065.
5. Madden, L.V. and M. Wheelis, *The threat of plant pathogens as weapons against U.S. crops*. Annual Review of Phytopathology, 2003. **41**(1): p. 155-176.
6. NRC, *Countering agricultural bioterrorism*. 2002, Washington, DC: NRC, National Academies Press.
7. Strange, R.N., et al., *Bioterrorism: a threat to plant biosecurity?*, in *The Role of Plant Pathology in Food Safety and Food Security*, Springer Netherlands. p. 115-132.
8. Fletcher, J., et al., *Plant pathogen forensics: capabilities, needs, and recommendations*. Microbiology and Molecular Biology Reviews, 2006. **70**(2): p. 450-471.
9. Ancona, V., D.N. Appel, and P. de Figueiredo, *Xylella fastidiosa: a model for analyzing agricultural biosecurity*. Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science. **8**(2): p. 171-182.

10. Applegate, T., et al. *Ag 101*. Major crops grown in the United States 2009 September 10, 2009 [cited 2011 March 29].
11. *Livestock*. 2010 2008 [cited 2011 March 30, 2011]; Available from: <http://ncga.com/livestock>.
12. Fletcher, J., et al., *Microbial forensics and plant pathogens: attribution of agricultural crime*. Wiley Handbook of Science and Technology for Homeland Security. 2008: John Wiley & Sons, Inc.
13. Budowle, B., et al., *Microbial forensics*, in *Microbial Forensics*, R.G. Breeze, B. Budowle, and S.E. Schutzer, Editors. 2005, Elsevier Academic Press: San Diego
14. Oerke, E.-C., et al., *Spatial and temporal dynamics of plant pathogens*, in *Precision Crop Protection - the Challenge and Use of Heterogeneity*, Springer Netherlands. p. 27-50.
15. Rogers, S., et al. *CBIAT: Crop Bioagent Introduction Intent Assessment Tool*. 2011 [cited 2011 March 30]; Available from: <http://bioinfosu.okstate.edu/nimffab/>.
16. Hall, J.S., et al., *Structure and temporal dynamics of populations within Wheat streak mosaic virus isolates*. Journal of Virology, 2001. **75**(21): p. 10231-10243.
17. Holland, J., et al., *Rapid evolution of RNA genomes*. Science, 1982. **215**(4540): p. 1577-1585.
18. Schneider, W.L. and M.J. Roossinck, *Genetic Diversity in RNA Virus Quasispecies Is Controlled by Host-Virus Interactions*. Journal of Virology, 2001. **75**(14): p. 6566-6571.

19. Iglesias, N., et al., *Population structure of Citrus tristeza virus from field Argentinean isolates*. *Virus Genes*, 2008. **36**(1): p. 199-207.
20. Kurath, G., J.A. Heick, and J.A. Dodds, *RNase protection analyses show high genetic diversity among field isolates of Satellite tobacco mosaic virus*. *Virology Journal*, 1993. **194**(1): p. 414-418.
21. McNeil, J.E., et al., *Characterization of genetic variability among natural populations of Wheat streak mosaic virus*. *Phytopathology*, 1996. **86**(11): p. 1222-1227.
22. Carver, J.D., *Molecular typing of Wheat streak mosaic virus for forensic applications*. 2007.
23. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: The European Molecular Biology Open Software Suite*. *Trends in Genetics*, 2000. **16**(6): p. 276-277.
24. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. *Nucleic Acids Research*, 1994. **22**(22): p. 4673-4680.
25. Tamura, K., et al., *MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0*. *Molecular Biology and Evolution*, 2007. **24**(8): p. 1596-1599.
26. Excoffier, L., G. Laval, and S. Schneider, *Arlequin (version 3.0): An integrated software package for population genetics data analysis*. *Evolutionary Bioinformatics*, 2007. **2005**(EBO-1-Excoffier(Sc)): p. 0.

27. Librado, P. and J. Rozas, *DnaSP v5: a software for comprehensive analysis of DNA polymorphism data*. *Bioinformatics*, 2009. **25**(11): p. 1451-1452.
28. Martin, D.P., et al., *RDP3: a flexible and fast computer program for analyzing recombination*. *Bioinformatics*. **26**(19): p. 2462-2463.
29. Kimura, M., *A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences*. *Journal of Molecular Evolution*, 1980. **16**: p. 111-120.
30. Felsenstein, J., *Confidence limits on phylogenies: an approach using the bootstrap*. *Evolution*, 1985. **39**(4): p. 783-791.
31. Choi, I.R., et al., *Contributions of genetic drift and negative selection on the evolution of three strains of Wheat streak mosaic tritimovirus*. *Archives Of Virology*, 2001. **146**(3): p. 619-628.
32. Kimura, M., *The neutral theory of molecular evolution*. 1983, Cambridge, Massachusetts: Cambridge University Press.
33. Tajima, F., *Evolutionary relationship of DNA sequences in finite populations*. *Genetics*, 1983. **105**(2): p. 437-460.
34. Fu, Y.X. and W.H. Li, *Statistical Tests of Neutrality of Mutations*. *Genetics*, 1993. **133**(3): p. 693-709.
35. Nei, M., *Molecular Evolutionary Genetics*. 1987, New York: Columbia University Press.
36. Stenger, D.C., D.L. Seifers, and R. French, *Patterns of polymorphism in Wheat streak mosaic virus: sequence space explored by a clade of closely related viral*

- genotypes rivals that between the most divergent strains*. Virology Journal, 2002. **302**(1): p. 58-70.
37. Tamura, K., M. Nei, and S. Kumar, *Prospects for inferring very large phylogenies by using the neighbor-joining method*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(30): p. 11030-11035.
38. Lemey, P., M. Salemi, and A.-M. Vandamme, eds. *The Phylogenetic Handbook*. Vol. 2. 2009, Cambridge University Press.
39. Cann, R., M. Stoneking, and A. Wilson, *Mitochondrial DNA and human evolution*. Nature, 1997. **325**(1).
40. Nei, M. and W.H. Li, *Mathematical model for studying genetic variation in terms of restriction endonucleases*. Proceedings of the National Academy of Sciences, 1979. **76**(10): p. 5269-5273.
41. Tajima, F., *Statistical method for testing the neutral mutation hypothesis by DNA polymorphism*. Genetics, 1989. **123**(3): p. 585-595.
42. Sneath, P. and R. Sokal, *Numerical taxonomy: the principles and practice of numerical classification*. 1973, San Francisco: Freeman.

Figure Captions

Fig. 1. Agarose gel image of WSMV RT-PCR amplicons after staining with ethidium bromide. Lanes 3-7 contain RT-PCR amplicons of a 1,254 bp region encompassing the *Wheat streak mosaic virus* coat protein. The sample in lane 2 was negative for the virus, and positive and negative controls were included in lanes 7 and 8, respectively.

Amplicons were generated from total nucleic acids extracted from plant samples collected from field 2 and used as templates in sequencing reactions. A low mass ladder [Invitrogen, Inc., Carlsbad, CA, USA] was included in the gel electrophoresis for estimation of amplicon size and concentration (lane 1).

Fig. 2. Evolutionary relationships of nucleic acid sequences. Evolutionary histories were inferred by constructing a UPGMA tree [42]. The bootstrap consensus tree inferred from 1000 replicates [30] is taken to represent the evolutionary history of the taxa analyzed [30]. Branches corresponding to partitions reproduced in less than 70% bootstrap replicates are collapsed. The tree is drawn to scale, with branch lengths in the same unit as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method [29] and are in the units of the number of base substitutions per site. The analysis involved 40 nucleotide sequences. All positions containing gaps and missing data were eliminated and evolutionary analyses were conducted in MEGA5 [25].

Fig. 3. Plots of the quality of conservation of amino acid alignments. Translated amino acid sequences derived from the two mechanically-inoculated groups of plants, (A) and

(B) (Lab 1 and Lab 2, respectively), and the WSMV-infected plants (C) and (D) (Field 1 and Field 2, respectively) were aligned using the EMBOSS program emma. Alignment files were submitted to the program plotcon in the EMBOSS package to generate plots of the quality of conservation of the alignments. Higher similarity scores presented on the Y-axis represent areas that are conserved, while lower similarity scores represent variability within the alignment. The X-axis corresponds to the amino acid position within the alignment.

Table 1. Mean similarity scores for nucleic acid and amino acid sequence alignments.

Individual data sets			
Alignment type ^a	Sample	Mean ^b	Standard Deviation
N.A.	Lab 1	97.00	0.93
	Lab 2	97.69	0.67
	Field 1	96.62	0.81
	Field 2	96.84	0.73
A.A.	Lab 1	98.31	0.95
	Lab 2	98.06	1.39
	Field 1	97.42	1.39
	Field 2	96.22	1.61
Combined data sets			
Alignment type ^a	Sample	Mean ^b	Standard Deviation
N.A.	Lab	97.31	0.89
	Field	96.73	0.78
A.A.	Lab	98.20	1.17
	Field	96.82	1.61

^aN.A. refers to nucleic acid sequences and A.A. refers to translated amino acid sequences.

^bTen consensus sequences were used for comparison for each data set, except lab 2, which consisted of 9 consensus sequences.

Table 2. Tests of neutrality for individual and combined data sets.

Individual data sets												
	Statistical Parameter					Neutrality Tests						
	m^a	S^b	π^c	$\hat{\theta}_W^d$	k^e	K_a/K_s^f	D^g	$p(D)$	F^{*h}	$p(F^*)$	D^{*i}	$p(D^*)$
Lab 1	10	65	0.024	0.027	21.156	0.035	-0.590	>0.10	-0.390	>0.10	-0.264	>0.10
Lab 2	9	37	0.014	0.015	11.889	0.046	-0.640	>0.10	-0.505	>0.10	-0.389	>0.10
Field 1	10	84	0.027	0.035	23.267	0.053	-1.246	>0.10	-1.342	>0.10	-1.168	>0.10
Field 2	10	66	0.020	0.028	17.178	0.068	-1.456	>0.10	-1.742	>0.10	-1.571	>0.10
Combined data sets												
	Statistical Parameter					Neutrality Tests						
	m^a	S^b	π^c	$\hat{\theta}_W^d$	k^e	K_a/K_s^f	D^g	$p(D)$	F^{*h}	$p(F^*)$	D^{*i}	$p(D^*)$
Lab	19	74	0.018	0.026	15.520	0.041	-1.293	>0.10	-1.176	>0.10	-0.890	>0.10
Field	20	110	0.023	0.039	18.516	0.053	-1.827	<0.05	-2.432	0.10>P> 0.05	-2.194	0.10>P>0.05

Data compiled using the program DNASP [27].

Neutrality tests were performed using $\alpha=0.05$. Bolded statistics are significant.

^aNumber of sequences included in the analysis.

^bNumber of polymorphic (segregating) sites [39].

^cThe direct estimate of per-site heterozygosity derived from the average number of pair-wise sequence differences in the sample [40].

^d $\hat{\theta}_W = 2Ne\mu$, where N is the effective population size, and μ is the mutation rate per nucleotide (or per sequence) and per generation [35].

^eAverage number of nucleotide differences [33].

^fAverage ratio of non-synonymous/synonymous mutations [35].

^gTajima's D statistic [41].

^hFu and Li's F* statistic [34].

ⁱFu and Li's D* statistic [34].

Table 3. Estimates of the patterns of nucleotide substitution.

	Transitions		Transversions				<i>R</i>
	A↔G	C↔U	A↔C	A↔U	C↔G	G↔U	
Lab 1	37.66	46.23	4.38	4.14	3.92	3.68	5.22
Lab 2	31.57	42.11	7.18	6.81	6.34	5.97	2.79
Field 1	31.63	45.07	6.31	5.99	5.66	5.34	3.26
Field 2	34.67	41.27	6.56	6.23	5.80	5.47	3.16

Each entry shows the probability of substitution (*r*) [37] in either direction. For simplicity, the sum of *r* values is made equal to 100. The overall transition/transversion bias is *R*, where $R = [A * G * k_1 + T * C * k_2] / [(A + G) * (T + C)]$. Codon positions included were 1st+2nd+3rd. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA5 [25].

Table 4. Evidence of recombination within data sets.

Field 1				
Sample	Beginning Breakpoint	Ending Breakpoint	Possible Parent	Method ^a
1	114	691	6	MC, Ch, SS, 3S
2	89	859	1	BS, MC, Ch, 3S
Field 2				
Sample	Beginning Breakpoint	Ending Breakpoint	Possible Parent	Method ^a
5	32	677	16	SS, 3S

^aBS= BootScan, MC= MaxChi, Ch= Chimaera, SS= SiScan and 3S= 3Seq.

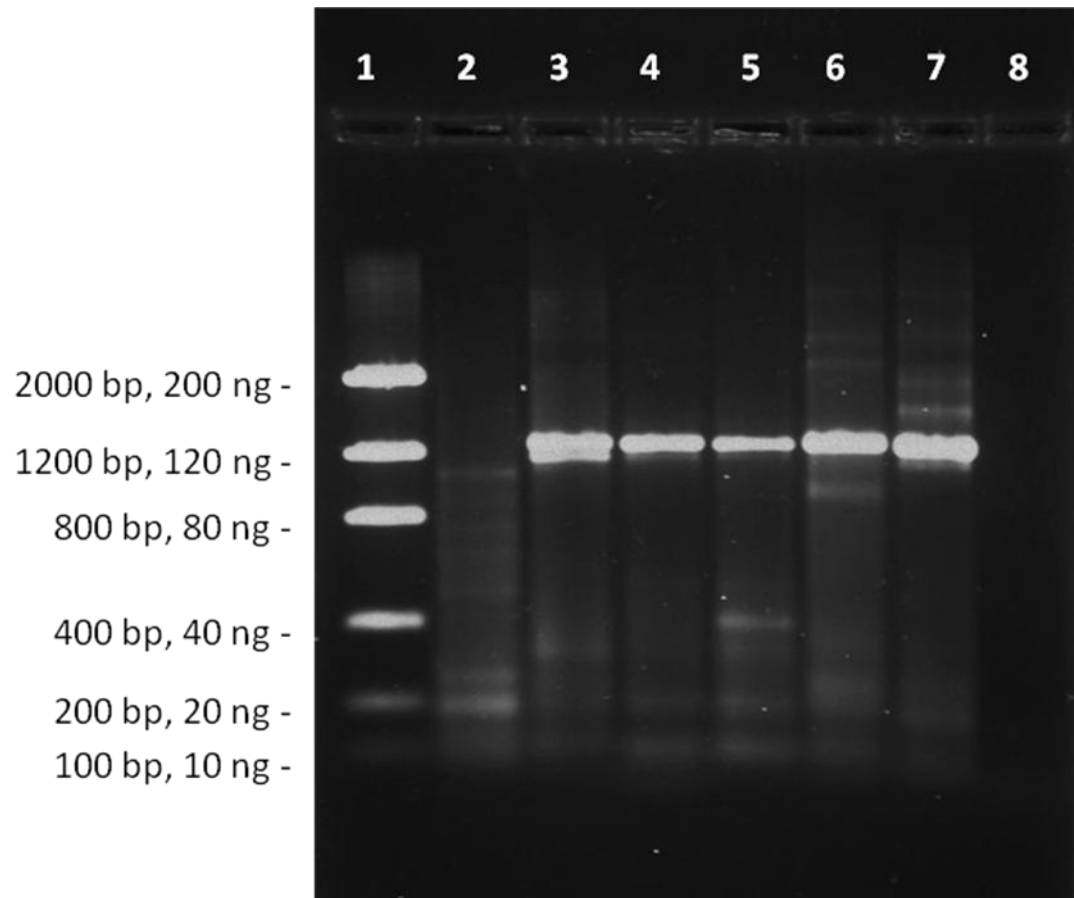


Fig. 1.

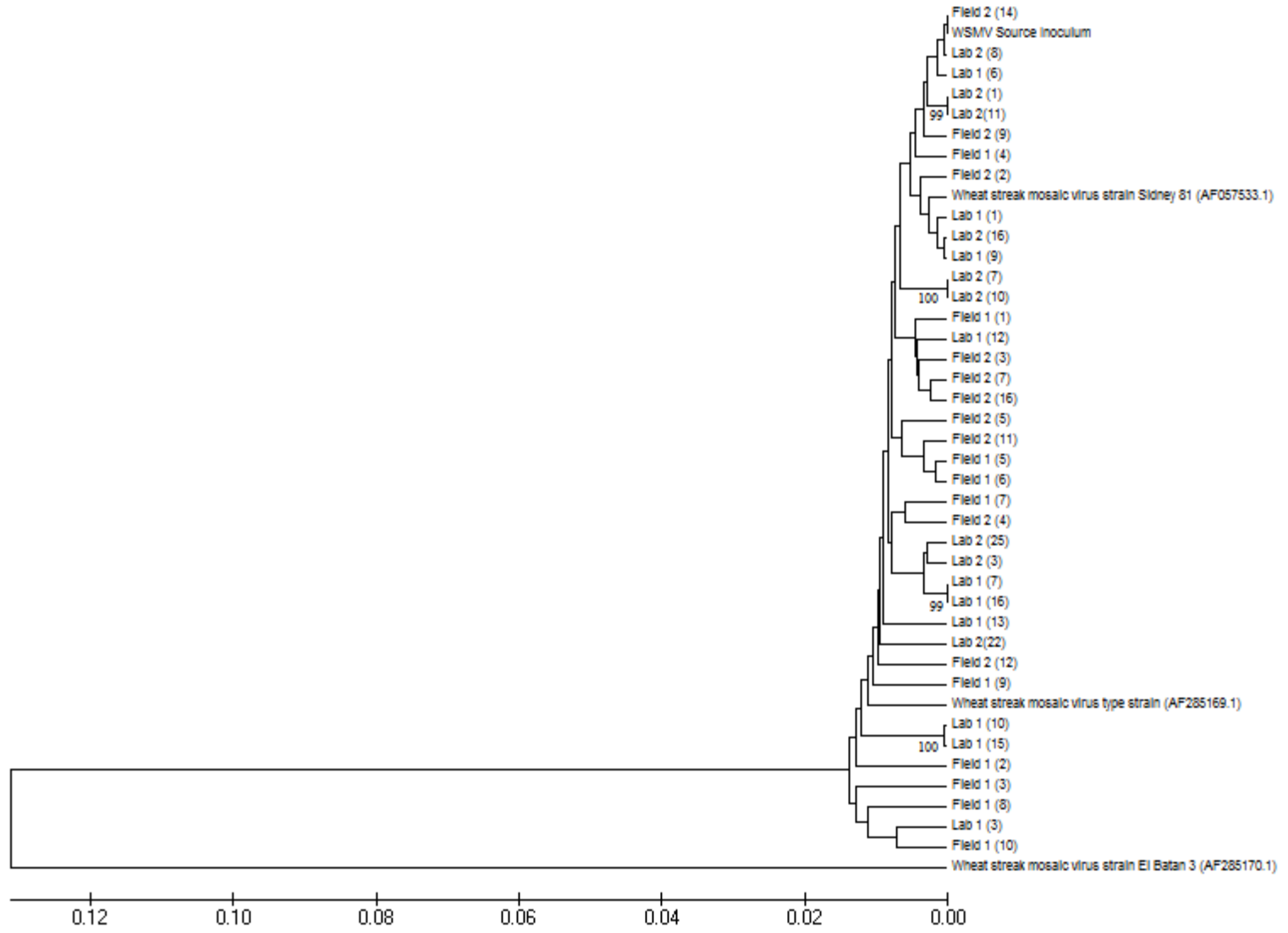


Fig. 2.

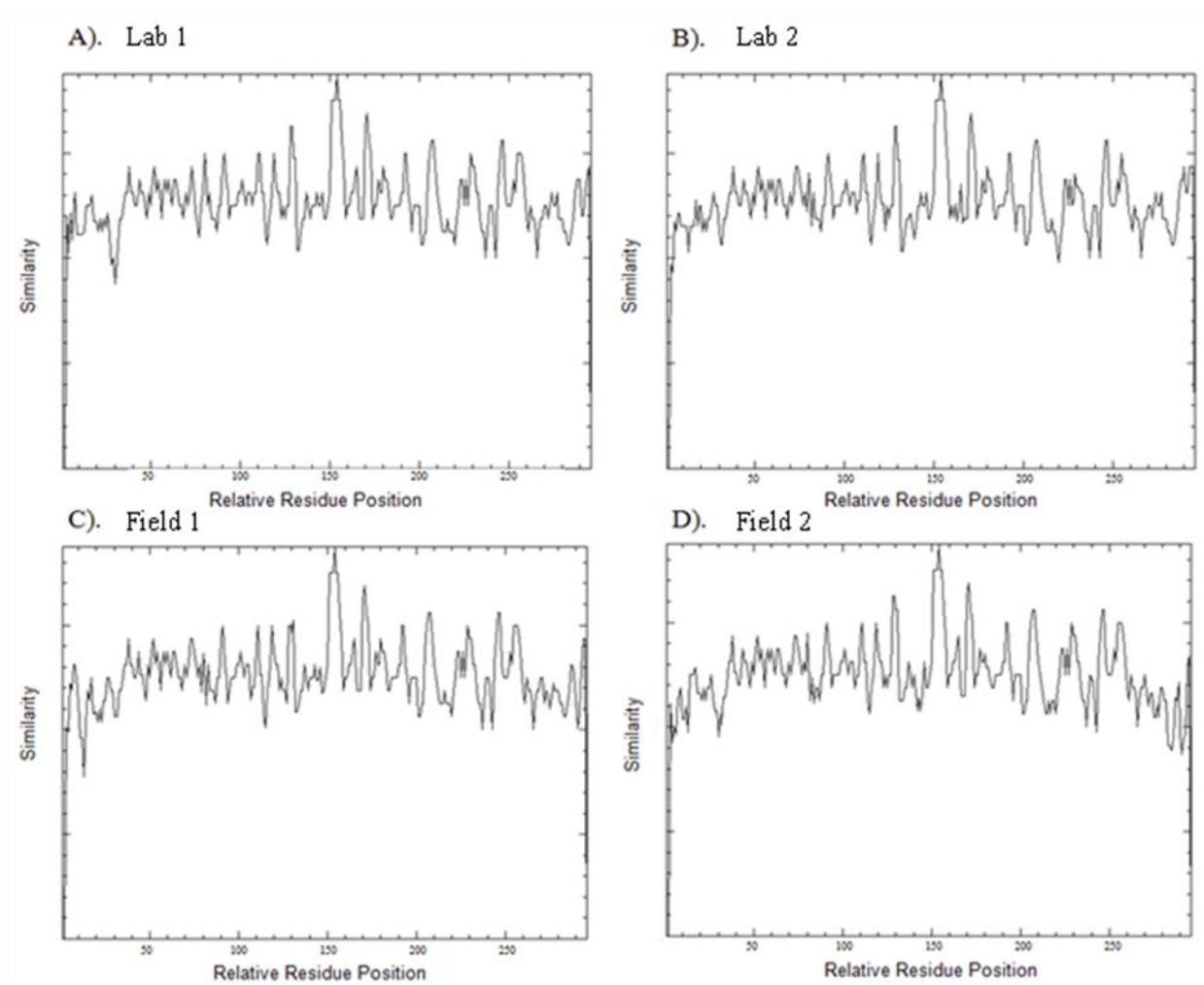


Fig. 3.

CHAPTER IV

UTILIZATION OF APYRASE-MEDIATED ALLELE-SPECIFIC PRIMER EXTENSION (AMASE) AND TAG-ARRAY MINISEQUENCING TECHNOLOGIES TO GENETICALLY FINGERPRINT *WHEAT STREAK MOSAIC VIRUS*

Abstract

The U.S. Department of Homeland Security has identified several gaps in the security of the agricultural sector, indicating that one of the nation's largest economic sectors is vulnerable to a biological attack. Currently, validated forensic tools that could be applied during a plant pathogen forensic investigation are needed for a balanced U.S. forensic capability. The goal of this project was to develop tools that are capable of simultaneously identifying and forensically profiling plant pathogens for use during the attribution of a biocrime. Apyrase-Mediated Allele-Specific primer Extension (AMASE) and a tag-array minisequencing system were employed to profile single nucleotide polymorphisms (SNPs) from the model pathogen *Wheat streak mosaic virus* (WSMV). Using synthetic targets, the amount of non-specific extension was negligible for both systems. Although AMASE proved to be difficult to replicate and thus not suitable for forensic applications, the mean values between technical replicates (n=3) using a single target concentration for the minisequencing assay were indistinguishable using one-way

ANOVA ($\alpha=0.05$). At increasing target concentrations, linear regression yielded an average R^2 value of 0.90 for the three extension primers. A theoretical limit of detection (4.3 fmol of target/reaction) was extrapolated from observed detection limits using synthetic targets and the amount of base misincorporation by the polymerase was found to be negligible (<1000 RFUs) for all types of misincorporation events. The universality of the tag-array minisequencing tool will provide a foundation for the development of similar systems for other plant and human pathogens. This tool also will be useful for rapid diagnostics and epidemiology studies during natural plant disease outbreaks.

Introduction

In 2002, the National Research Council identified the United States agricultural sector as being vulnerable to acts of bioterrorism, and highlighted the need for preparedness [1]. The central role that agriculture plays in the United States economy, coupled with the ease with which an attack could be successfully launched, makes it possible that the agricultural sector will be targeted for such an attack in the future.

Plant pathogens could be utilized to launch an attack on crop systems and are attractive bioweapons for several reasons. First, plant pathogens are easily attainable, and are thus sometimes referred to as ‘the poor man’s weapon of mass destruction’ [2]. Secondly, the vast acreage utilized for crop production in the United States makes it impossible to constantly monitor human activity within fields. The moral threshold of intentionally inoculating plants is lower than that for infecting animals and humans and the legal ramifications of a successful attack have not yet been established. In a well-developed country, such as the United States, a successful attack would have detrimental

effects on the economy and consumer confidence, resulting in higher costs of commodities and fear associated with the food supply [3]. In underdeveloped countries that are highly dependent upon imported crops and monocultures for sustenance, the consequences of a successful attack would be devastating, and could lead to mass starvation and death.

Validated tools that are suitable for genetically ‘fingerprinting’ plant pathogens for forensic purposes are currently nonexistent. However, several genotyping systems based on Single Nucleotide Polymorphisms (SNPs) are currently being utilized for research and diagnostic purposes. SNPs are attractive forensic markers because they occur frequently, and are informative and easy to detect, and can be recovered from highly degraded samples that are often the only source of genetic evidence recovered from crime scenes. Because of their value as forensic markers, SNPs were used as key discriminatory elements during the investigation of the anthrax attacks of 2001 [4]. Existing SNP-based genotyping techniques could easily be adapted for use during a microbial forensics investigation. Current technologies include direct sequencing, followed by bioinformatic techniques [5-8], denaturing high performance liquid chromatography (DHPLC) [9-11], single base extension of primers (SBE) [12-14], polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) [15, 16], real-time PCR [17-19], microarray [20-25] and nanotechnology-based techniques [26, 27]. Although all of these systems can be used to generate SNP profiles, most of the current technologies require extensive sequence knowledge prior to genotyping. Microarray and direct sequencing techniques allow for the simultaneous diagnosis and genotyping of pathogens, a highly desirable capability during a forensic investigation.

Although direct sequencing of pathogens provides ample sequence data for comparing samples, a profile ‘fingerprint’ that is representative of the target sequence would simplify the data analysis and interpretation procedures. Microarray is a good choice for a genotyping platform because arrays can house hundreds of thousands of oligonucleotides, and can therefore serve to diagnose and genotype pathogen targets simultaneously. Microarrays are portable, require limited machinery, and are becoming increasingly cost-effective. Here, we describe the adaptation of two microarray-based platforms, apyrase-mediated allele-specific primer extension and a tag-array minisequencing system for the simultaneous diagnosis and SNP-typing of the model pathogen *Wheat streak mosaic virus* (WSMV), and discuss their potential for validation.

For a forensic tool to be useful, researchers must first establish confidence in its utility by demonstrating its range of use under defined conditions. Budowle et. al. [28] defined validation as a process that assesses the ability of a procedure to perform reliably under specified conditions, explicitly states assay conditions, determines inherent limitations associated with the assay, identifies potential contributors to variability so that they may be controlled, and provide a basic framework for data interpretation. They described three levels of validation; i) developmental validation, ii) internal validation and iii) preliminary validation.

Apyrase-mediated allele-specific primer extension (AMASE), a microarray-based method, is capable of determining the identity of a single base at a specific location within a sequence [29]. AMASE utilizes apyrase, an ATP diphosphohydrolase, to facilitate competition between extension reactions performed on a slide using perfectly matched primer/target complexes and primer/target complexes with a mismatch at the 3'

end (SNP-typing position) (Fig. 1). It has been shown that the reaction kinetics of DNA polymerase are slower when a mismatch occurs at the 3' end of a primer than when the bases are perfectly complementary [30]. AMASE exploits the differences in reaction kinetics between the matched and mismatched primer extension reactions by removing nucleoside triphosphates from the pool of available substrates, thus creating competition for the nucleoside triphosphates by the polymerase for incorporation into the extending nucleic acids. As a result of the differences in reaction kinetics between DNA polymerase bound to perfectly matched and that bound to mismatched primers, the former are capable of integrating fluorescently labeled nucleotides into longer extension products, which yield intense fluorescent spots, while the latter yield little or no fluorescence. SNP-typing is possible when each base is positioned at the 3' end of the primers and AMASE is performed. In the absence of apyrase, some types of mismatches have been shown to drastically reduce PCR product yield, while other types of mismatches have little effect on the PCR reactions [31-34]. AMASE has already proven useful for re-sequencing [35] and SNP-typing [29, 36] human genes. The high-throughput and specific nature of the technique could allow for its adaptation to plant viruses and microbial forensic applications.

Minisequencing on a microarray slide, also referred to as arrayed primer extension (APEX), is also appropriate for large-scale SNP-typing applications [37]. Since minisequencing has a 40:1 signal to noise ratio, there is a high probability that the correct SNPs can be assigned at each position [38], making this technique ideal for use in a forensic context. Similar to AMASE, minisequencing utilizes DNA polymerase in an extension reaction performed either in solution or directly on a microarray slide, to

genotype a target sequence. The major difference between the two methods is that AMASE is performed using fluorescently-labeled deoxynucleoside triphosphates and thus extends a microarray-attached probe by several nucleotides, while minisequencing reactions incorporate a single, fluorescently-labeled dideoxynucleoside monophosphate into each of the extension products (Fig. 2). Minisequencing has recently been used to detect SNPs throughout the human genome [39] and solution-based minisequencing reactions have been used for forensic applications [40, 41]. Single base extension (SBE)-tag arrays have proven to accurately identify 99% of queried SNPs [42]. Because of its high fidelity, simplicity and high-throughput nature, this method was explored as an alternative to the AMASE procedure.

Materials and Methods

Identification of Single Nucleotide Polymorphisms for *Wheat streak mosaic virus*

To identify informative SNPs, a multiple sequence alignment was generated using ClustalW and whole genome sequences of WSMV listed in GenBank with the following accession numbers: EU914918.1, AF454454.1, FJE48358.1, AF511643.2, AF511630.2, FJ348359.1, AF511618.2, AF285169.1, AF511615.2, NC_001886.1, AF057533.1, AF511619.2, AF454455.1, AF511614.2, EU914917.1 and AF285170.1). Single nucleotide polymorphisms were chosen based on their immediate proximity to highly conserved (<10% polymorphic positions) stretches, spanning at least 15 nucleotides for AMASE, and 17 nucleotides for minisequencing. Potential primers were queried for their ability to form intermolecular structures using the program Mfold. For AMASE, these primers were designed to have predicted melting temperatures between 50°C and

56°C and targeted the plus and minus strands of WSMV (Table 1). Minisequencing primers were designed to have predicted melting temperatures ranging from 52°C and 67°C, 40%-60% GC content and to target the plus and minus strands of the model virus. Minisequencing primers were also designed to have an ‘anti-tag’ sequence at the 5’ end of the primer for capturing the fluorescently-labeled minisequencing products by the universal ‘tags’ printed on the microarray (Table 2).

Oligonucleotide probes for AMASE were synthesized [Integrated DNA Technologies, Coralville, Iowa, USA] with an amine linker group and a six-carbon spacer at the 5’ end of the sequence for coupling to the microarray slide, and elevating the oligos off of the surface of the slide, respectively. Universal sequences from a Universal Tag Sequences Library [Affymetrix, Santa Clara, California, USA] were assessed for their ability to form intermolecular structures, and their ability to cross-hybridize with wheat, by subjecting the sequences to Mfold and a NCBI BLAST search, respectively (Table 2). Universal tags for capturing minisequencing products were synthesized with amine groups on the 3’ end. A Cy3-labeled oligonucleotide was also obtained and printed onto the microarray slide for orientation of the microarray grid after scanning. This oligo consisted of a 5’ amine group, followed by a 6-carbon spacer, an 18nt random sequence and terminated in a Cy3 fluorophore (5’- /5AmMC6/AATGGACGGCAACACTCG/3Cy3Sp/-3’). All amine-labeled oligos were dissolved in nuclease-free water at 100 µM and stored at -20°C prior to printing.

Fabrication of custom microarrays

Slides were printed at the Oklahoma State University Biochemistry and Molecular Biology Array and Bioinformatics Core Facility using an Omnigrid 100 microarray

printer [Genomic Solutions, Ann Arbor, MI, USA]. Oligonucleotide probes for AMASE, universal 'tag' sequences for minisequencing and Cy3-labeled molecules for both projects were diluted to a final concentration of 20 μM in 150 μM sodium phosphate buffer (pH 8) and printed in quadruplicates onto CodeLink (NHS ester) activated microarray slides [SurModics In Vitro Diagnostic Products, Eden Prairie, Minnesota, USA] under 50% humidity. Four duplicate spots were printed per array, and twelve replicate arrays were printed on each slide, and spaced to fit in the center of the wells of either 3x8 or 2x8 hybridization cassettes [ArrayIt, Sunnyvale, California, USA], which allow the slide to be divided into individual hybridization chambers. After printing the microarray slides, the humidity was raised to 75% for over-night incubation. Prior to blocking and washing, the array positions were marked on the back of the slides for positioning of the hybridization cassettes.

Post-coupling processing

Slides were placed in a slide rack and reactive groups were blocked using a pre-heated blocking solution [0.1 M Tris, 50 mM ethanolamine (pH 9.0)] for 30 minutes at 50°C, followed by a brief rinsing in deionized water. Slides were subsequently washed with 4x SSC, 0.1% SDS (pre-warmed to 50°C) for 30 minutes, rinsed briefly with deionized water and dried by spinning in a Microarray High-Speed Centrifuge (MHC) [ArrayIt, Sunnyvale, California, USA].

Hybridization of targets

Microarray hybridization cassettes [ArrayIt, Sunnyvale, California, USA] were positioned over the microarray slides, so that the individual arrays were in the center of the wells. Targets were combined with 30 μL of the hybridization buffer (4x SSC, 0.1%

SDS) and applied to the individual wells. The individual chambers, created by the hybridization cassette, were sealed on the top with transparent tape and allowed to incubate at 50°C and 300 rpms for one hour.

Washing, scanning and data processing

Slides were removed from the hybridization chamber, quickly rinsed with 4x SSC, and washed twice with 2x SSC, 0.1% SDS at 50°C for 5 minutes, once with 0.2x SSC at room temperature for 1 minute, and once with 0.1x SSC at room temperature for 1 minute. Slides were dried by centrifuging at top speed in a MHC.

Scanning was performed using an Axon GenePix 4400A 4-laser confocal scanner [Molecular Devices, Inc, Sunnyvale, CA, USA] at wavelengths of 532 nm (for orientation of the gene array list) and 635 nm. Local background values were subtracted from each of the spots individually, and values of quadruplicate spots in each array were averaged.

Control Preparation

Control targets ~300 bp long were prepared for AMASE by performing RT-PCR using *Wheat streak mosaic virus* as a template, and the forward primer F2 (5'-GAGGAAAGGAAAAAATGCTG-3') and reverse primer 6+ (5'-AAAATGTTGTACCGGCGCTT-3'). RT-PCR products were subsequently cloned into a pSC-A-amp/kan plasmid vector (240205) [Agilent Technologies, Inc., Santa Clara, CA, USA] and cloned cells (pWSMV) were stored in 20% glycerol at -80°. To produce control targets for hybridization to the array, the ~300 bp insert was amplified by PCR using M13 Reverse (5'-GGAAACAGCTATGACCAT-3') and M13 -20 (5'-TGACCGGCAGCAAATG-3') primers. These amplicons were subsequently dried in a

VacuFuge plus Vacuum Concentrator [Eppendorf North America, Inc., Westbury, New York, USA], re-suspended in 30 μ L of hybridization buffer and boiled at 100°C for one minute before being applied to the array.

Synthetic targets were designed for optimization of the minisequencing protocol. These sequences were designed to be complementary to the minisequencing primers, but with variable nucleotides at the SNP-typing position. Synthetic targets were synthesized by Integrated DNA Technologies [Coralville, Iowa, USA].

Apyrase-mediated allele-specific primer extension reactions

After control targets were hybridized to the covalently-linked SNP-typing probes on the slide surface, 60 μ L of the extension reaction containing 8 mU of apyrase (M0393S) [New England Biolabs, Inc., Ipswich, Massachusetts, USA], 20 U of Klenow (3'-5' exo-) DNA polymerase and buffer (M0212M) [New England Biolabs, Inc., Ipswich, Massachusetts, USA], and water was added to the individual wells of the hybridization cassette. A 40 μ L mixture of the three unlabeled dNTPs (2.5 μ M each), and the fourth Cy5-labeled dNTP (2.5 μ M) was subsequently added to each of the wells to initiate the competitive extension reaction.

Multiplexed minisequencing reactions

Solution-based minisequencing reactions were performed in a 15 μ L reaction mix containing 0.1 μ M of three unlabeled dideoxynucleoside triphosphates (ddNTPs), 0.1 μ M of the fourth Cy5-labeled ddNTP, 6.67 mM DTT, 10 nM of each minisequencing primer, 10 nM minisequencing target, 8 U Thermo Sequenase DNA polymerase and Thermo Sequenase buffer (E79000Y) [GE Healthcare, Piscataway, NJ, USA]. Cyclic minisequencing was performed in a programmable thermal controller (PTC-100) [MJ

Research, Inc., St. Bruno (Quebec), Canada] by performing 35 cycles of 95°C for 30 seconds and 53°C for 30 seconds.

RT-PCR amplification of WSMV

For the optimization of primer and polymerase concentrations for minisequencing, total RNA was extracted from healthy and WSMV-infected wheat tissue using the RNeasy Plant Mini Kit (74904) [Qiagen, Inc., Valencia, CA, USA].

A ~1200 bp region of the WSMV coat protein was amplified by RT-PCR using the primers C1 (5'-TACTTGACTGGGACCCGAA-3') and XC1 (5'-AACCCACACATAGCTACCAAG-3') [43]. Reverse transcription was performed using 200 U of Moloney murine leukemia virus reverse transcriptase (M-MLV RT) (28025-013) [Invitrogen, Inc., Carlsbad, CA, USA] in 20 µL master mix containing 2 pmole of primer C1, 1 µg of total RNA, 0.5 mM dNTPs, 10 mM DTT and First-Strand buffer, according to the manufacturer.

PCR was performed in a programmable thermal controller (PTC-100) [MJ Research, Inc., St. Bruno (Quebec), Canada] in reactions containing 10 U of *Taq* DNA polymerase (10342053) [Invitrogen, Inc.] in 50 µL master mix containing PCR buffer, 1.5 mM MgCl₂, 0.2 mM dNTPs, 0.2 µM forward and reverse primers, and approximately 1 µg cDNA. Reactions were subjected to an initial denaturation step of 94°C for 1 minute, followed by 35 cycles of 94°C, 30 seconds, 53°C, 45 seconds, 72°C, 1 minute and a final extension step of 72°C for 5 minutes.

Following RT-PCR, un-reacted dNTPs and primers were removed by performing a shrimp alkaline phosphatase/exonuclease I treatment on PCR amplicons. A 10.5 µL master mix containing 6.80 µL of RT-PCR products, 0.05 M Tris HCl, 7.62 mM MgCl₂,

6 U of exonuclease I (M0293S) [New England Biolabs, Inc., Ipswich, MA, USA] and 10.5 U shrimp alkaline phosphatase (EF0511) [Fermentas, Inc., Glen Burnie, MD, USA] was incubated at 37°C for one hour. Shrimp alkaline phosphatase and exonuclease I enzymes were subsequently inactivated at 65°C for 15 minutes.

Results

Effect of apyrase concentration on AMASE reactions

Perfectly complementary synthetic targets ((7+)R, (7+)R Mini, (8+)R, (8+)R Mini and WSMV (6+)) or targets that had a single mismatch corresponding to the base at the 3' end of the array-linked probe were mixed and hybridized to a microarray slide. A hybridization cassette was placed over the individual arrays to partition the slide into individual reaction chambers. Extension reactions containing increasing amounts of apyrase were subsequently added to each well. The ability of the enzyme to decrease non-specific extension was assessed, based on the amount of fluorescence displayed by non-specific probes (Fig. 3).

The addition of apyrase significantly reduced non-specific extension by non-target probes. When 24 mU of apyrase was added to the reactions, non-specific extension was completely eliminated. A slight reduction in fluorescence intensity was observed at the highest concentration of apyrase (40 mU); however, this reduction did not significantly alter the fluorescence intensities for the positive controls. The type of correct base-pair did not appear to significantly alter the extension reactions, as probe/target pairs with A:T and G:C at the SNP-typing position displayed similar

fluorescence intensities. The type of mismatch also did not affect the ability of apyrase to diminish extension.

Reproducibility of AMASE

To assess the reproducibility of the AMASE procedure, duplicate elongation reactions were performed on two separate arrays using positive and negative control synthetic targets. Probes WSMV (6+), (8+)R and (8+)R Mini were perfectly complementary to the positive control synthetic targets. The positive control target oligonucleotides were engineered to have each of the four bases at the position corresponding to one nucleotide downstream from the 3' end of the probe. Negative control synthetic targets were either engineered to have a mismatch at the 3' base of the probe ((6-)a, (6-)b, (6-)c, (8-)Ra, (8-)Rb and (8-)Rc)) or were not complementary to any of the probes. The synthetic targets were mixed and hybridized. Subsequently, a hybridization cassette was placed over the slide and used to sequester individual arrays into separate reaction chambers. Each elongation reaction (in separate wells) contained a different fluorescently-labeled nucleotide, as well as the other three un-labeled nucleotides, and 24 mU of apyrase was used for each of the reactions.

Although the fluorescence intensities were relatively low, probe WSMV (6+) was extended in the presence of apyrase and all four of the nucleotides yielded comparable fluorescence values in the first experiment (Fig. 4, A). In the duplicate experiment (B), however, this probe was elevated above the background fluorescence intensities in only one of the four reactions (Cy5-dCTP). Probes (8+)R and (8+)R mini were also extended in the first experiment, but in only half of the reactions (Cy5-dGTP and Cy5-dATP), and to a much lesser extent than in the reaction that incorporated Cy5-dATP in the replicate.

Non-specific extension was much lower in the first replicate than in the second, indicating that the apyrase was not as effective in the latter. Since a large amount of variation was observed between the two replicate experiments (A) and (B), and between extension reactions with the two data sets, it is not possible to assess the amount of base-preference by the polymerase.

Reproducibility of minisequencing

The lack of reproducibility of the AMASE procedure prompted the assessment of the reproducibility of the minisequencing technique. A synthetic target that was complementary to the Anti-Tag8-Mini 1 minisequencing primer was used as a template during the thirty-six separate minisequencing reactions, in which a single Cy5-ddATP was incorporated into the single base-extension products. A hybridization cassette was placed over the array to sequester arrays into individual hybridization wells. To assess between-well and between-slide variability, fluorescently-labeled minisequencing products were allowed to hybridize to tag sequences in each of 12 individual hybridization chambers.

The standard deviation of fluorescence intensities for quadruplicate spots in a single array averaged 732.44 among the three datasets (Fig. 5). Inter-slide variability was assessed by calculating the standard deviation for the average intensities of quadruplicate spots among arrays on a single slide. The average standard deviation for the three replicates was 1771.21. Intra-array variability (reproducibility) was assessed by subjecting the background-subtracted quadruplicate averages for each array on each of the three slides (12 array spot averages/slide) to a one-way analysis of variance (ANOVA) statistical test using a 0.05 level of significance (Fig. 4). Since the $F_{\text{obs}}=1.04$

and the $F_{\text{critical}}=3.89$, the mean values for each replicate were found to be statistically indistinguishable.

Quantitation and sensitivity of minisequencing

Three concentrations of synthetic targets (10 fmol/reaction, 20 fmol/reaction and 30 fmol/reaction) were spiked into extension reactions containing three multiplexed minisequencing primers (Anti-Tag8-Mini 1, Anti-Tag12-Mini 3 and Anti-Tag13-Mini C), and cDNA from healthy wheat tissue and a ddNTP mix including fluorescently-labeled ddATP (Fig. 6).

In the absence of the synthetic targets (healthy wheat cDNA only), all three primers displayed very low fluorescence intensities ($\mu=1801.69$, $\sigma=913.59$), indicating that the three minisequencing primers did not cross-hybridize with host sequences, form internal structures, or interact with each other during extension (data not shown). The average of the negative controls (fluorescence $\mu=1801.69$) plus two standard deviations ($\sigma=913.59$) was used to determine the minimum threshold value of 3628.87 fluorescence units for positive extension reactions. The average R^2 value for the linear regression for the three extension primers was 0.90. The theoretical limit of detection was calculated for the three positive control datasets by extrapolating the regression lines to the minimum threshold value and averaging the three target concentrations. Based on these data, the average theoretical limit of detection is 4.3 fmol target per reaction.

Specificity of minisequencing

To assess the degree of misincorporation by the polymerase during minisequencing reactions, synthetic targets were combined with mixtures of fluorescently-labeled ddNTPs. Each of the 12 types of misincorporation events (A/T,

A/G, A/C, T/A, T/G, T/C, G/A, G/T, G/C, C/A, C/T and C/G) were queried using synthetic targets with each of the four nucleotides in the SNP-typing position in combination with different fluorescently-labeled/unlabeled ddNTPs. After hybridizing minisequencing reactions from each combination of fluorescently-labeled ddNTP/SNPs, fluorescence intensity values for each type of misincorporation event were found to be lower than 1000 relative fluorescence units.

Effect of primer concentration on minisequencing

The effect of primer concentration was assessed by duplexing increasing amounts of the two primers into extension reactions containing WSMV amplicons, produced from the pWSMV clones, and cDNA from a healthy wheat plant. The extended fluorescent products were added to individual wells of a hybridization cassette and captured by the covalently-linked tags corresponding to the anti-tag sequences at the 5' ends of the minisequencing primers (Fig. 7).

The primer concentration did not significantly affect the fluorescence intensities produced with either of the two primers. The average fluorescence intensities for primer Anti- Tag42- Mini 56/* and Anti- Tag53- Mini 55/* were 27,599.20 ($\sigma=2478.46$) and 11124.80 ($\sigma=1936.47$), respectively. Both primers yielded fluorescence intensities that were significantly higher than the average background values plus two standard deviations (252.50).

Effect of polymerase concentration on minisequencing

Enzyme concentration effects were assessed by increasing the amount of polymerase in each of three extension reactions. The concentrations of enzyme were 0.064, 0.128 and 0.192 U/ μ L in reactions containing duplexed primers (10 nM each) and

amplicons from the pWSMV clones in the healthy wheat (cDNA) background.

Reactions containing variable amounts of enzyme were added to individual wells of a hybridization cassette on a single microarray slide (Fig. 8).

Doubling the amount of polymerase resulted in a significant increase in the fluorescence intensities for both primers (Anti- Tag42- Mini 56/* (+34.42%) and Anti- Tag53- Mini 55/* (+22.20%)). Primer Anti-Tag42- Mini 56/* did not show a significant increase in fluorescence intensity when the amount of polymerase was tripled (+0.4%), however primer Anti- Tag42- Mini 56/* increased by 13.08%. Increasing the enzyme concentration did not appear to affect the background noise, as all negative control values were consistently below 1%.

Discussion

Over the past decade, microarrays have been increasingly recognized for their diagnostic and genetic fingerprinting potential. Although whole genome sequencing is becoming gradually more affordable and provides ample amounts of sequence information, microarrays offer researchers the ability to probe for previously identified discriminatory features, such as SNPs, that are useful during a microbial forensics investigation, thus microarrays are currently a more cost-effective approach than direct sequencing for genetic ‘fingerprinting’ [44].

Enzyme-based microarrays are inherently more specific than microarrays alone, allowing for the potential use of such methods for forensic applications. Here, we explored the use of AMASE and solution-based minisequencing, followed by capture on a universal tag-array, for genotyping WSMV.

Single base extension in the presence of moderate concentrations of apyrase (≥ 24 mU/reaction) strongly reduced the amount of non-specific extension (Fig. 3); however, the AMASE results proved to be difficult to replicate (Fig. 4). Since the extension reactions were performed in individual wells directly on the surface of the microarray, the inability to quickly and thoroughly mix the reactions probably resulted in the observed inconsistencies between replicates. Inconsistent or incomplete mixing of the reactions causes fluctuations in the concentration of fluorescently-labeled substrates by the polymerase or apyrase, resulting in the variability in over-all fluorescence values observed in these experiments. Because the principle of AMASE is highly dependent upon the competition between the polymerase and the apyrase, the ability to consistently mix the components of the extension reactions and initiate the competitive reaction is essential. Thus, performing AMASE directly on the surface of a microarray slide produces results that are too variable to be validated and used in a forensic setting.

Under the specified conditions and using a single target concentration, the minisequencing assay proved to be reproducible, based on a one-way ANOVA test with a 0.05 level of significance (Fig. 5). Linear regression on the three minisequencing extension products in the presence of increasing amounts of target nucleic acids was variable (Fig. 6), displaying an average R^2 value of 0.90. Since the variability between slides was assessed by examining the averages at a single target concentration, this indicates that the minisequencing procedure is more variable when target concentrations fluctuate and thus cannot be used to quantify the amount of SNPs in a sample. However the low background observed in all of the negative controls allows for the fingerprinting of SNPs. The procedure demonstrated specificity of all of the primers, as seen by the

minimal background noise and non-specific extension observed in the absence of target molecule, implying that the primer design strategy employed here was sufficient to eliminate most cross-hybridization during extension (Fig. 7 and data not shown). The data shown here indicate that the minisequencing method can detect at least 10 fmol of targets per reaction, and is estimated to have a limit of detection of approximately 4.3 fmol of targets per reaction (Fig. 6). While the range of primer concentrations explored (10 nM-30 nM) did not appear to affect the SNP-typing ability of the assay (Fig. 7), increasing the polymerase concentration from 0.064 U/ μ L to 0.128 U/ μ L resulted in a significant increase in the fluorescence intensities of the positive controls (Fig. 8). At the highest concentration of the enzyme (0.192 U/ μ L) the fluorescence intensities of positive controls were not significantly increased, therefore 0.128 U/ μ L was found to be the optimum concentration of polymerase under the specified conditions, based on cost effectiveness and discrimination capability. Probe sequence appeared to influence the fluorescence intensity in all three of the titration experiments (Figs. 6-8). However, it is unclear as to whether this effect was due to the hybridization of the minisequencing primer to its target during the extension reaction, or to the hybridization of the anti-tag sequence to its complementary tag on the microarray.

These data indicate that solution-based minisequencing may be suitable for validation and use as a forensic tool. However, further experimentation is required before this tool would be suitable for validation. Technical difficulties associated with microarray equipment delayed the progress and prevented the further development of this technique.

One limitation of this method is the inability of the platform to quantify SNPs. Because plant viruses exist as genetically heterogeneous populations within their hosts, the ability to determine the ratio of SNPs at specific locations within the genome could provide another dimension to a forensic profile. Furthermore, this information could be used to explore virus evolution and virus population dynamics. Additional research involving the forensic comparison of genetic profiles from plant viruses is also necessary. Although the major goal of this research was to explore the use of minisequencing for forensic purposes, minisequencing can also be used as a tool for classification, epidemiology and evolutionary studies.

Acknowledgements

The authors would like to thank Dr. Peter Hoyt for his technical assistance and troubleshooting guidance and Dr. Hoyt and Praveen Yerramsetti for their assistance with microarray printing. This research was funded by USDA-NIFA and the Oklahoma Agricultural Experiment Station.

LITERATURE CITED

1. Committee on Biological Threats to Agricultural Plants and Animals, N.R.C., *Countering agricultural bioterrorism*. 2002, Washington, DC: National Academies Press.
2. Madden, L.V. and M. Wheelis, *The threat of plant pathogens as weapons against U.S. crops*. Annual Review of Phytopathology, 2003. **41**(1): p. 155-176.
3. Fletcher, J., et al., *Plant pathogen forensics: capabilities, needs, and recommendations*. Microbiology and Molecular Biology Reviews, 2006. **70**(2): p. 450-471.
4. Van Ert, M.N., et al., *Strain-specific single-nucleotide polymorphism assays for the Bacillus anthracis Ames strain*. J. Clin. Microbiol., 2007. **45**(1): p. 47-53.
5. Sauerbrei, A., et al., *Genetic profile of an Oka varicella vaccine virus variant isolated from an infant with Zoster*. Journal of Clinical Microbiology, 2004. **42**(12): p. 5604-5608.
6. Zitoudi, K., et al., *Genetic variation in Myzus persicae populations associated with host-plant and life cycle category*. Entomologia Experimentalis et Applicata, 2001. **99**: p. 303 - 311.
7. Wu, X., et al., *SNP discovery by high-throughput sequencing in soybean*. BMC Genomics. **11**(1): p. 469.
8. Branicki, W., et al., *Beyond HV1 and HV2--identification of valuable mitochondrial DNA single nucleotide polymorphisms*. International Congress Series, 2004. **1261**: p. 100-102.

9. Ribas, G., M.J. Neville, and R.D. Campbell, *Single-nucleotide polymorphism detection by denaturing high-performance liquid chromatography and direct sequencing in genes in the MHC class III region encoding novel cell surface molecules*. Immunogenetics, 2001. **53**(5): p. 369-381.
10. Hou, Y.P., et al., *Single nucleotide polymorphisms detected by Temperature-modulated high-performance liquid chromatography*. International Congress Series, 2003. **1239**(Supplement 1): p. S9-S14.
11. Cole, D.E.C., et al., *Calcium-sensing receptor mutations and denaturing high performance liquid chromatography*. Journal of Molecular Endocrinology, 2009. **42**(4): p. 331-339.
12. Deshpande, A., Y. Valdez, and J.P. Nolan, *Multiplexed SNP genotyping using primer Single-Base Extension (SBE) and microsphere arrays*. Current Protocols in Cytometry. 2001: John Wiley & Sons, Inc.
13. Tomas, C., et al., *Typing of 48 autosomal SNPs and amelogenin with GenPlex SNP genotyping system in forensic genetics*. Forensic Science International: Genetics, 2008. **3**(1): p. 1-6.
14. Endicott, P., et al., *Multiplexed SNP typing of ancient DNA clarifies the origin of andaman mtDNA haplogroups amongst south Asian tribal populations*. PLoS ONE, 2006. **1**(1): p. e81.
15. Ota, M., et al., *Single nucleotide polymorphism detection by polymerase chain reaction-restriction fragment length polymorphism*. Nature Protocols, 2007. **2**(11): p. 2857-2864.

16. Chew, C.H., et al., *PCR-RFLP genotyping of C1q mutations and single nucleotide polymorphisms in Malaysian patients with Systemic Lupus Erythematosus*. *Human Biology*, 2008. **80**(1): p. 83-93.
17. Rudi, K. and A.L. Holck, *Real-time closed tube single nucleotide polymorphism (SNP) quantification in pooled samples by quencher extension (QEXT)*. *Nucleic Acids Research*, 2003. **31**(19): p. e117.
18. Pont-Kingdon, G. and E. Lyon, *Direct molecular haplotyping by melting curve analysis of hybridization probes: beta 2-adrenergic receptor haplotypes as an example*. *Nucleic Acids Research*. **33**(10): p. e89.
19. Waku-Kouomou, D., et al., *Genotyping Measles Virus by Real-Time Amplification Refractory Mutation System PCR Represents a Rapid Approach for Measles Outbreak Investigations*. *Journal of Clinical Microbiology*, 2006. **44**(2): p. 487-494.
20. Iwasaki, H., et al., *Accuracy of genotyping for single nucleotide polymorphisms by a microarray-based single nucleotide polymorphism typing method involving hybridization of short allele-specific oligonucleotides*. *DNA Research*, 2002. **9**(2): p. 59-62.
21. Mao, X., B.D. Young, and Y.J. Lu, *The application of single nucleotide polymorphism microarrays in cancer research*. *Current Genomics*, 2007. **8**: p. 219-228.
22. Teh, M.-T., et al., *Genomewide single nucleotide polymorphism microarray mapping in basal cell carcinomas unveils uniparental disomy as a key somatic event*. *Cancer Research*, 2005. **65**(19): p. 8597-8603.

23. Zhang, G.Q., et al., *Multiplex PCR and oligonucleotide microarray for detection of single-nucleotide polymorphisms associated with Plasmodium falciparum drug resistance*. Journal of Clinical Microbiology, 2008. **46**(7): p. 2167-2174.
24. Rickert, A.M., et al., *Quantitative genotyping of single-nucleotide polymorphisms by allele-specific oligonucleotide hybridization on DNA microarrays*. Biotechnology and Applied Biochemistry, 2005. **42**(1): p. 93-96.
25. Cherkasova, E., et al., *Microarray analysis of evolution of RNA viruses: Evidence of circulation of virulent highly divergent vaccine-derived polioviruses*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(16): p. 9398-9403.
26. Lee, K.-J., et al., *Single nucleotide polymorphism detection using Au-decorated single-walled carbon nanotube field effect transistors*. Journal of Nanomaterials, 2011. **2011**.
27. Wang, Y. and B. Liu, *Label-free single-nucleotide polymorphism detection using a cationic tetrahedralfluorene and silica nanoparticles*. Analytical Chemistry, 2007. **79**(18): p. 7214-7220.
28. Budowle, B., et al., *Criteria for validation of methods in microbial forensics*. Applied and Environmental Microbiology, 2008. **74**(18): p. 5599-5607.
29. O'Meara, D., et al., *SNP typing by apyrase-mediated allele-specific primer extension on DNA microarrays*. Nucleic Acids Research, 2002. **30**(15): p. e75-.
30. Ahmadian, A., et al., *Genotyping by apyrase-mediated allele-specific extension*. Nucleic Acids Research, 2001. **29**(24): p. e121-.

31. Wu, D.Y., et al., *Allele-specific enzymatic amplification of beta -globin genomic DNA for diagnosis of sickle cell anemia*. Proceedings of the National Academy of Sciences of the United States of America ; Vol/Issue: 86:8, 1989: p. Pages: 2757-2760.
32. Kwok, S., et al., *Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency virus type 1 model studies*. Nucleic Acids Research, 1990. **18**(4): p. 999-1005.
33. Newton, C.R., et al., *Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS)*. Nucleic Acids Research, 1989. **17**: p. 2503-2516.
34. Ayyadevara, S., J.J. Thaden, and R.J.S. Reis, *Discrimination of primer 3'-nucleotide mismatch by taq DNA polymerase during polymerase chain reaction* Analytical Biochemistry, 2000. **284**(1): p. 11-18.
35. Ericsson, O., et al., *Microarray-based resequencing by apyrase-mediated allele-specific extension*. Electrophoresis, 2003. **24**(19-20): p. 3330-3338.
36. Shamma, C., et al., *ThalassoChip, an array mutation and single nucleotide polymorphism detection tool for the diagnosis of \hat{P}^2 -thalassaemia*. Clinical Chemistry and Laboratory Medicine. **48**(12): p. 1713-1718.
37. Lindroos, K., et al., *Minisequencing on oligonucleotide microarrays: comparison of immobilisation chemistries*. Nucleic Acids Research, 2001. **29**(13): p. e69-.
38. Kurg, A., et al., *Arrayed Primer Extension: solid-phase four-color DNA resequencing and mutation detection technology*. Genetic Testing, 2000. **4**(1): p. 1-7.

39. Ho-Pun-Cheung, A., et al., *Detection of Single-Nucleotide Polymorphisms in cancer-related genes by minisequencing on a microelectronic DNA chip*, in *Microarrays*, J.B. Rampil, Editor. 2007, Humana Press.
40. Ferri, G., et al., *Minisequencing-based genotyping of Duffy and ABO blood groups for forensic purposes*. *Journal of Forensic Sciences*, 2006. **51**(2): p. 357-360.
41. Morley, J.M., et al., *Validation of mitochondrial DNA minisequencing for forensic casework*. *International Journal of Legal Medicine*, 1999. **112**(4): p. 241-248.
42. Hirschhorn, J.N., et al., *SBE-TAGS: An array-based method for efficient single-nucleotide polymorphism genotyping*. *Proceedings of the National Academy of Sciences of the United States of America*, 2000. **97**(22): p. 12164-12169.
43. Carver, J.D., *Molecular typing of Wheat streak mosaic virus for forensic applications*, in *Department of Forensic Science*. 2007, Oklahoma State University.
44. Budowle, B., et al., *Genetic analysis and attribution of microbial forensics evidence*. *Critical Reviews in Microbiology*, 2005. **31**(4): p. 233-254.

Figure captions

Fig. 1. SNP-typing by allele-specific primer extension. The colored blocks represent each of the four bases. The primer is covalently linked to the microarray slide so that the incoming targets can hybridize to the complementary probes directly on the slide surface. After adding the DNA polymerase, probes with a mismatch at the 3' end (SNP-typing position) (A) will not be elongated, while those with a perfectly matched 3' end (B) will be elongated and thus incorporate fluorescently-labeled dNTPs into the elongation product. When the nucleotide-hydrolyzing enzyme apyrase is added, the method is termed apyrase-mediated allele-specific primer extension (AMASE).

Fig. 2. Solution-based minisequencing. A). A minisequencing primer is designed to anneal directly upstream from the queried SNP on the target molecule. Tag sequences, complementary to the anti-tag sequence located at the 5' end of the minisequencing primer, are covalently-linked to the microarray. During solution-based extension, either a fluorescently-labeled ddNTP or an unlabeled ddNTP is incorporated at the end of the minisequencing primer and terminates the reaction. The incorporated ddNTPs are complementary to the queried SNP. B). After the minisequencing reactions are complete, fluorescently-labeled products are 'captured' by tag sequences corresponding to specific minisequencing primers. The minisequencing primers and the fluorescent bases are determined by scanning the microarray slides. A dual-color approach can be achieved by performing extension reactions in separate tubes, incorporating a different fluorescent ddNTP in each reaction (using a red dye for positive samples and a green dye and subsequently hybridizing reactions in separate wells on a single microarray slide.

Fig. 3. The effect of apyrase concentrations on allele-specific extension reactions.

Synthetic targets that were either perfectly complementary to the AMASE probes ((7+)R, (7+)R mini, (8+)R, (8+)R mini and WSMV (6+)) or targets with a mismatch at the SNP-typing position (WSMV (6-)a, WSMV (6-)b, WSMV (6-)c, (7-)Ra, (7-)Rb, (7-)Rc, (8-)Ra, (8-)Rb and (8-)Rc) were hybridized to the probes on the microarray slide.

Subsequently, the hybridization cassette was placed over the slide surface to sequester arrays into individual hybridization chambers, and elongation reactions containing increasing concentrations of apyrase were added to the wells. Base pairs (matches=A:T, G:C; mismatches=T:T, T:G, G:G, C:T, A:G) at the SNP-typing position are bracketed at the end of each probe.

Fig. 4. Reproducibility of AMASE. Two replicate experiments were performed (A) and (B), where synthetic targets, perfectly complementary to probes WSMV (6+), (8+)R and (8+)R Mini, were allowed to hybridize to the array. Targets with each of the four bases at the SNP-typing position were included, and all four fluorescently-labeled nucleotides were added to separate reactions to assess the base preference of the polymerase.

Fig. 5. Reproducibility of minisequencing. A synthetic target that was complementary to the Anti- Tag8- Mini 1 minisequencing primer served as the template during the minisequencing reaction. A single, fluorescently-labeled nucleotide (Cy5-ddATP) was incorporated into the minisequencing product in all three replicates. Average standard deviation values for quadruplicate spots in an array are depicted by the error bars. Quadruplicate spot averages in each array (wells) are presented as individual bars. These

averages (12/slide) were used to assess intra-slide variability and reproducibility using a one-way ANOVA.

Fig. 6. Sensitivity of minisequencing. Synthetic targets that were complementary to the Anti-Tag8-Mini 1, Anti- Tag12-Mini 3 or the Anti- Tag13- Mini C primers were spiked into extension reactions in increasing concentrations. To test the ability of the assay parameters in a host background, synthetic targets were spiked into extension reactions containing cDNA from healthy wheat tissue that was reverse-transcribed using M-MLV RT. Primer/target combinations were chosen so that fluorescently-labeled ddATP was added to the 3' end of all three primers. Cy5-ddATP was added to extension reactions containing synthetic targets (positive controls) and Cy3-ddATP was added to extension reactions without synthetic targets (negative controls).

Fig. 7. Effect of primer concentration on extension reactions using amplicons. WSMV amplicons were used as targets in the minisequencing reactions. Increasing amounts of two minisequencing primers (Anti- Tag42- Mini 56/* and Anti- Tag53- Mini 55/*) were used to determine the optimum primer concentration per reaction. Healthy wheat cDNA was included in separate minisequencing reactions and treated as negative controls. All four fluorescently-labeled ddNTPs were included in the minisequencing reactions (Cy5-ddNTPs were used for the positive controls and Cy3-ddNTPs were used for the negative controls). 0.064 U/ μ L of the DNA polymerase (Sequenase) used in all minisequencing reactions. Error bars indicate the standard deviation of quadruplicate spots in each array.

Fig. 8. Effect of polymerase concentration on minisequencing. WSMV amplicons were used as targets in the minisequencing reactions. Increasing amounts of the DNA polymerase (Sequenase) were titrated into the minisequencing reactions containing 10 nM of the specified primer (either Anti- Tag42- Mini 56/* or Anti- Tag53- Mini 55/*). Enzyme concentrations are bracketed. All four fluorescently-labeled ddNTPs were included in the minisequencing reactions (Cy5-ddNTPs were added to positive control reactions and Cy3-ddNTPs were added to negative control reactions). 65,000 relative fluorescent units was considered to be the maximum (100%) fluorescence intensity, and used to calculate the % total fluorescence for the positive and negative control reactions. Error bars indicate the standard deviation of the percent total fluorescence for quadruplicate spots in the arrays.

Table 1. AMASE probe sequences.

Oligo ^a	Sequence ^b (5'→3')	Length (nt)	% GC Content	T _m (°C)
WSMV (6+)	/5AmMC6/GGCCATGTTGTAAA <u>A</u>	15	40	53.9
WSMV (6-) a	/5AmMC6/GGCCATGTTGTAA <u>A</u> T	15	40	53.6
WSMV (6-) b	/5AmMC6/GGCCATGTTGTAA <u>A</u> G	15	46.7	54.2
WSMV (6-) c	/5AmMC6/GGCCATGTTGTAA <u>A</u> C	15	46.7	54.6
WSMV (11+)	/5AmMC6/GGCCTATGAGTATA <u>A</u> C	16	43.8	52.3
WSMV (11-) a	/5AmMC6/GGCCTATGAGTATA <u>A</u> T	16	37.5	51.3
WSMV (11-) b	/5AmMC6/GGCCTATGAGTATA <u>A</u> G	16	43.8	51.9
WSMV (11-) c	/5AmMC6/GGCCTATGAGTATA <u>A</u> A	16	37.5	51.6
(7+) F	/5AmMC6/TTTACGACTCACATAT <u>T</u> C	18	33.3	54.7
(7-) Fa	/5AmMC6/TTTACGACTCACATAT <u>T</u> T	18	27.8	54.5
(7-) Fb	/5AmMC6/TTTACGACTCACATAT <u>T</u> G	18	33.3	55.0
(7-) Fc	/5AmMC6/TTTACGACTCACATAT <u>T</u> A	18	27.8	53.6
(7+) R	/5AmMC6/GAATATGTGAGTCGTA <u>A</u> A	18	33.3	54.7
(7-) Ra	/5AmMC6/GAATATGTGAGTCGTA <u>A</u> T	18	33.3	54.5
(7-) Rb	/5AmMC6/GAATATGTGAGTCGTA <u>A</u> G	18	38.9	54.8
(7-) Rc	/5AmMC6/GAATATGTGAGTCGTA <u>A</u> C	18	38.9	55.1
(8+) R	/5AmMC6/GCCATGTTGTAAAAT <u>A</u> C	17	35.3	53.8
(8-) Ra	/5AmMC6/GCCATGTTGTAAAAT <u>A</u> T	17	29.4	53.0
(8-) Rb	/5AmMC6/GCCATGTTGTAAAAT <u>A</u> G	17	35.3	53.5
(8-) Rc	/5AmMC6/GCCATGTTGTAAAAT <u>A</u> A	17	29.4	53.3
(8+) R Mini	/5AmMC6/GCCATGTTGTAAAAT <u>A</u>	16	31.3	51.9
(7+) F Mini	/5AmMC6/TTTACGACTCACATAT <u>T</u>	17	29.4	53.3
(7+) R Mini	/5AmMC6/GAATATGTGAGTCGTA <u>A</u>	17	35.3	53.5

^aUnderlined nucleotides indicate the SNP-typing position. Probes were synthesized with a 5' amine group to facilitate binding to the microarray surface.

^bOligos with a '+' in the name are complementary to the genomic WSMV sequences at the SNP-typing position, based on a multiple sequence alignment.

Table 2. Minisequencing primer and tag sequences.

Minisequencing Primers				
Primer	Sequence ^a (5'→3')	Primer Length (nt)	% GC Content	T _m (°C)
Anti-Tag8-Mini 1	<i>GCGACGCCCAAATCTGTAAAGGAAGC</i> ATTTTCTGGTCAAA	20	40.0	61.9
Anti-Tag12-Mini 3	<i>CCGTGGATCAAATCAGTCAAGT</i> GAAAGATCTTGCTCGA	18	44.4	59.7
Anti-Tag13-Mini C	<i>CGCTTGCTCAAAGGGAAGAAAA</i> ACTTTCATGGAACCG	17	41.2	57.9
Anti-Tag53-Mini 55/*	<i>GTGCCAGAGACAACGATCCAT</i> CCTCATTAGTACCCGCACTCA	22	50.0	67.0
Anti-Tag42-Mini 56/*	<i>CCTATTGAAGAAGATCCGAC</i> ACAAATCCACGAGTACCAGG	20	50.0	64.1
Oligonucleotide Tags				
Tag	Sequence ^b (5'→3')			
Tag 8	TTTACAGATTTGGGCGTCGCTTTTTTTTTTTTTTT/3AmM/			
Tag 12	TTGACTGATTTGATCCACGGTTTTTTTTTTTTTTTT/3AmM/			
Tag 13	TTCTCCCTTTGAGCAAGCGTTTTTTTTTTTTTTTT/3AmM/			
Tag 53	TGGATCGTTGTCTCTGGCACTTTTTTTTTTTTTTTTT/3AmM/			
Tag 42	GTCGGATCTTCTTCAATAGGTTTTTTTTTTTTTTTT/3AmM/			

^aPrimer sequences consisted of an anti-tag sequence at the 5' end (italicized) and a WSMV-specific primer sequence at the 3' end. Anti-tag sequences were excluded from the % GC content and T_m calculations.

^bTag sequences were synthesized with a 3' amine group to facilitate binding to the microarray surface.

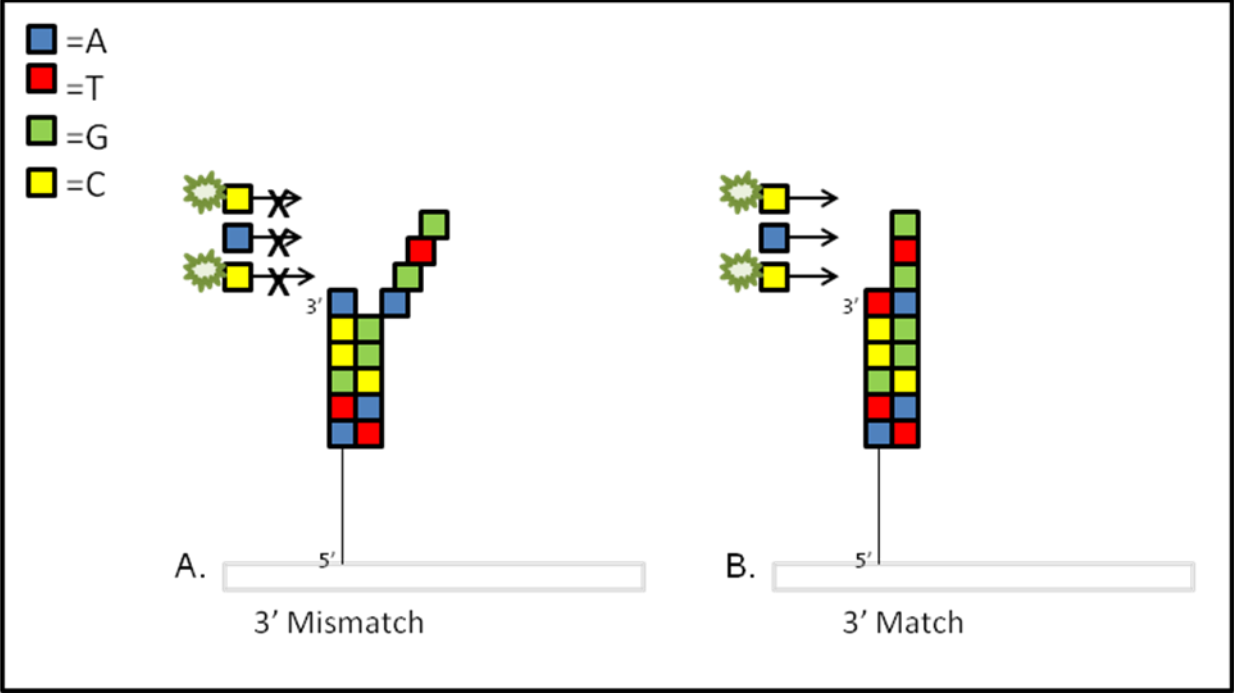


Fig. 1.

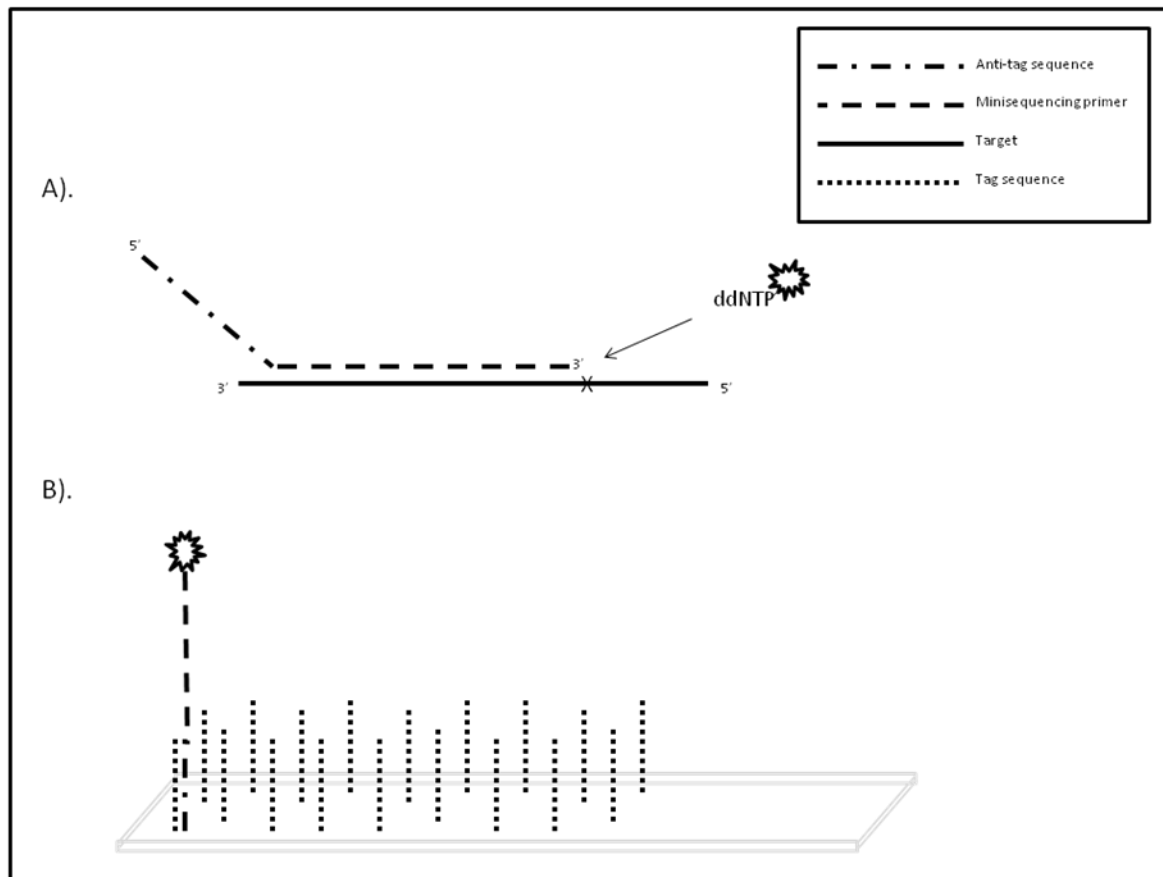


Fig. 2.

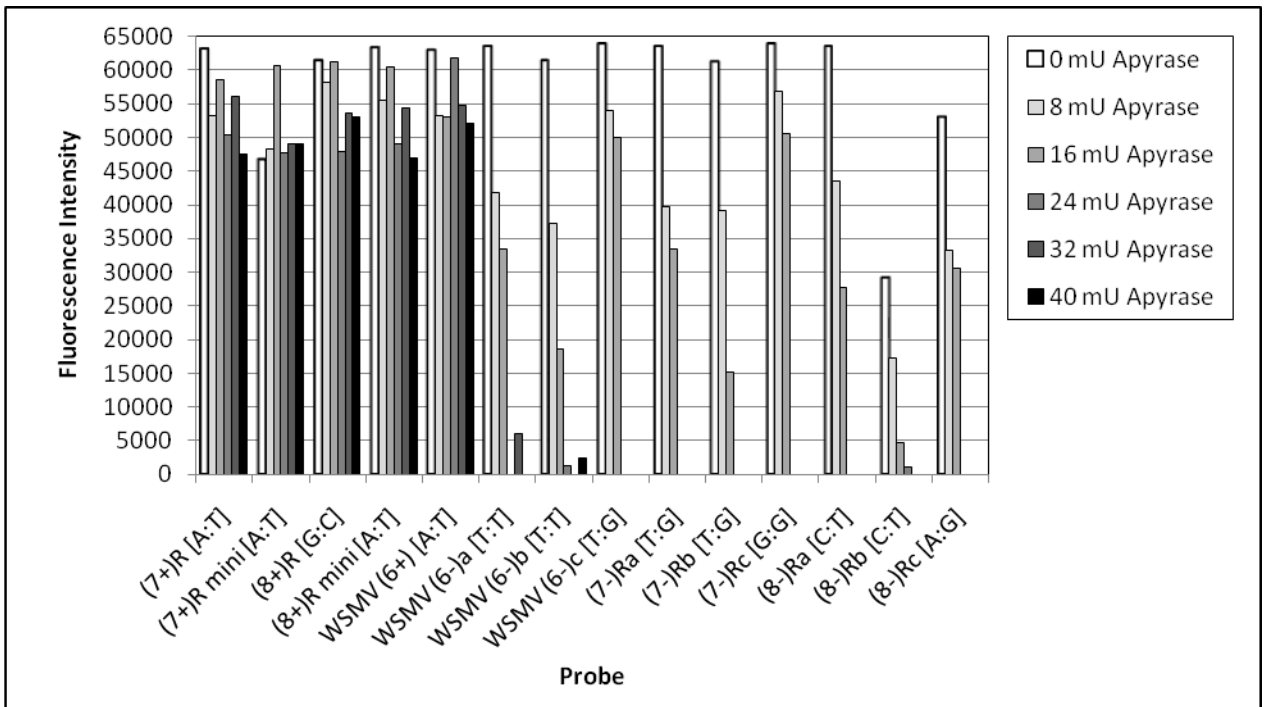
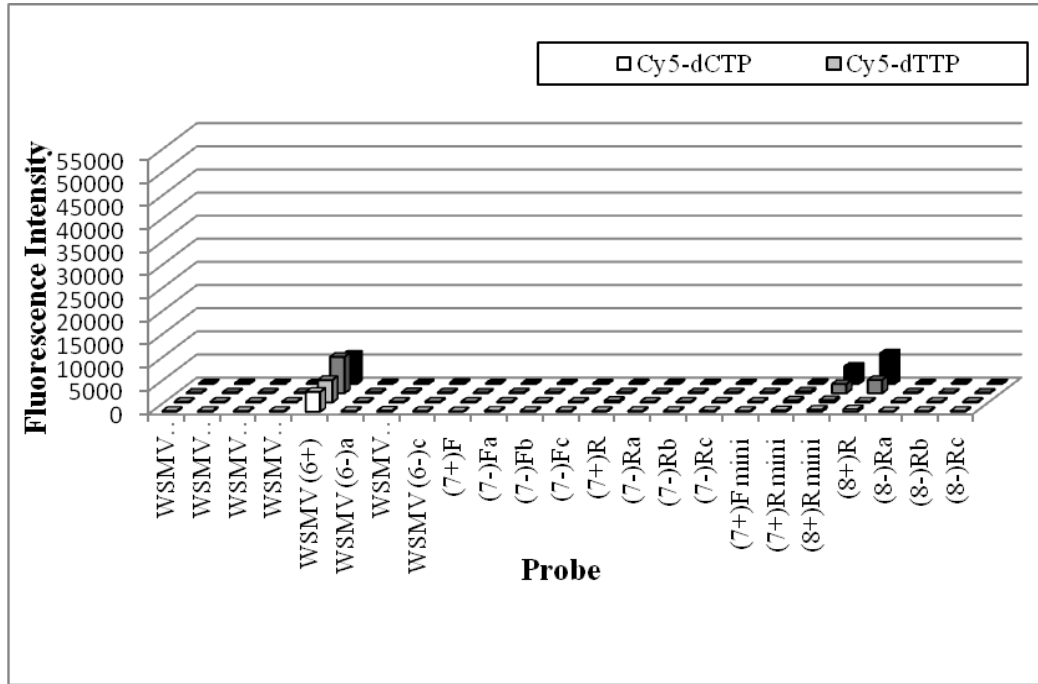


Fig. 3.

A.



B.

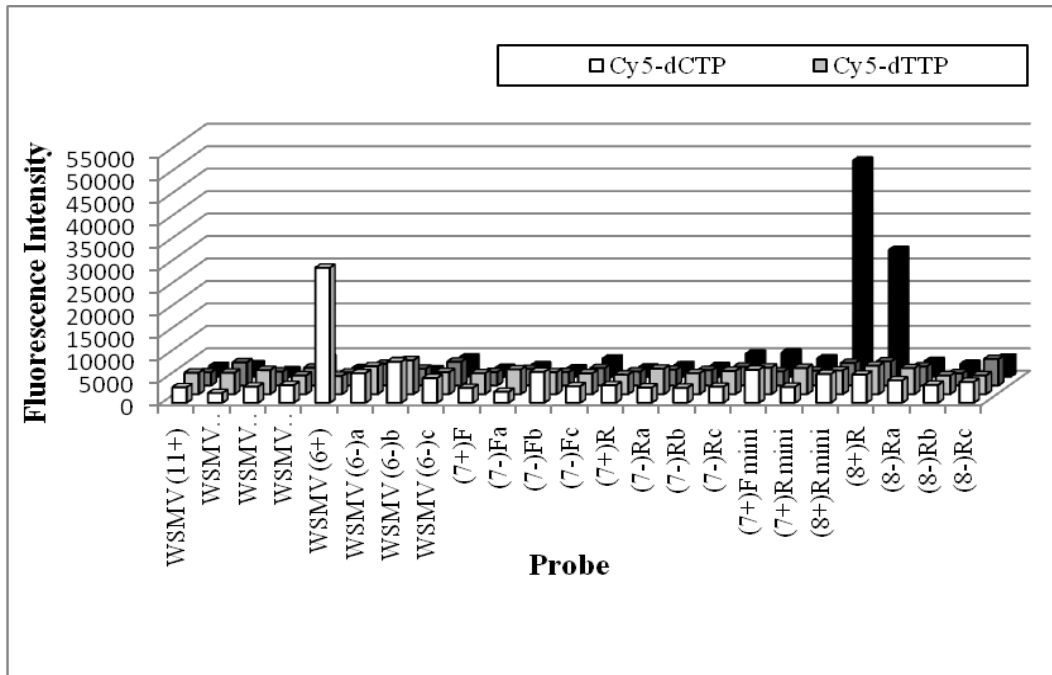


Fig. 4.

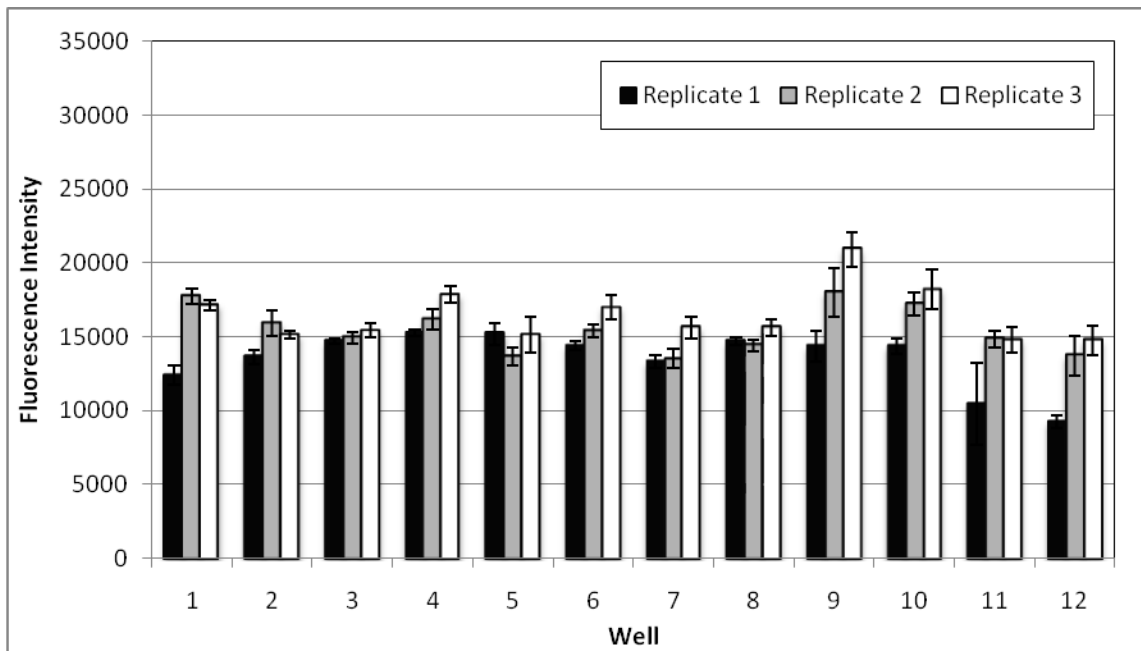


Fig. 5.

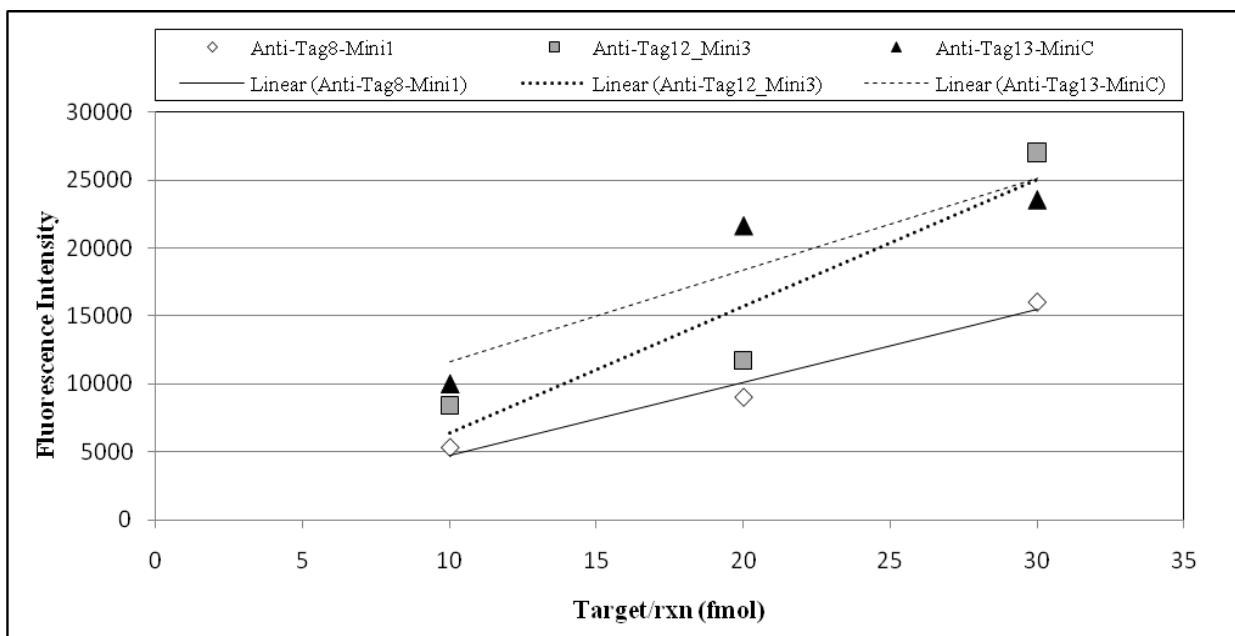


Fig. 6.

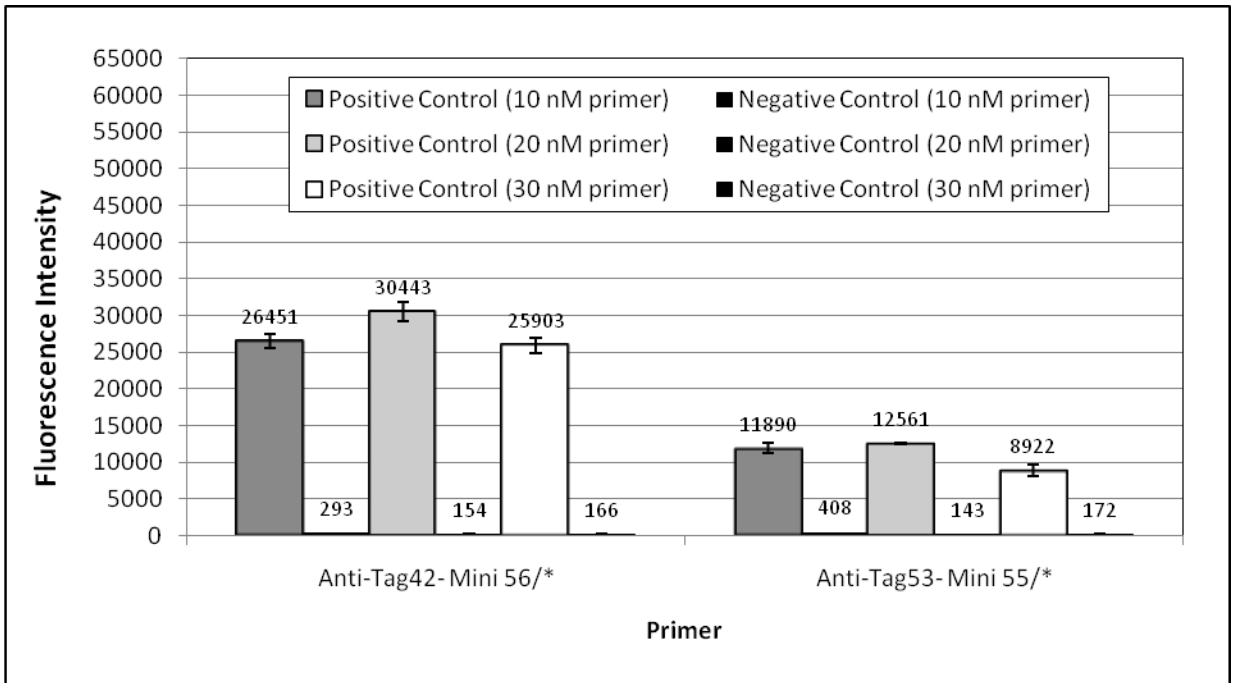


Fig. 7.

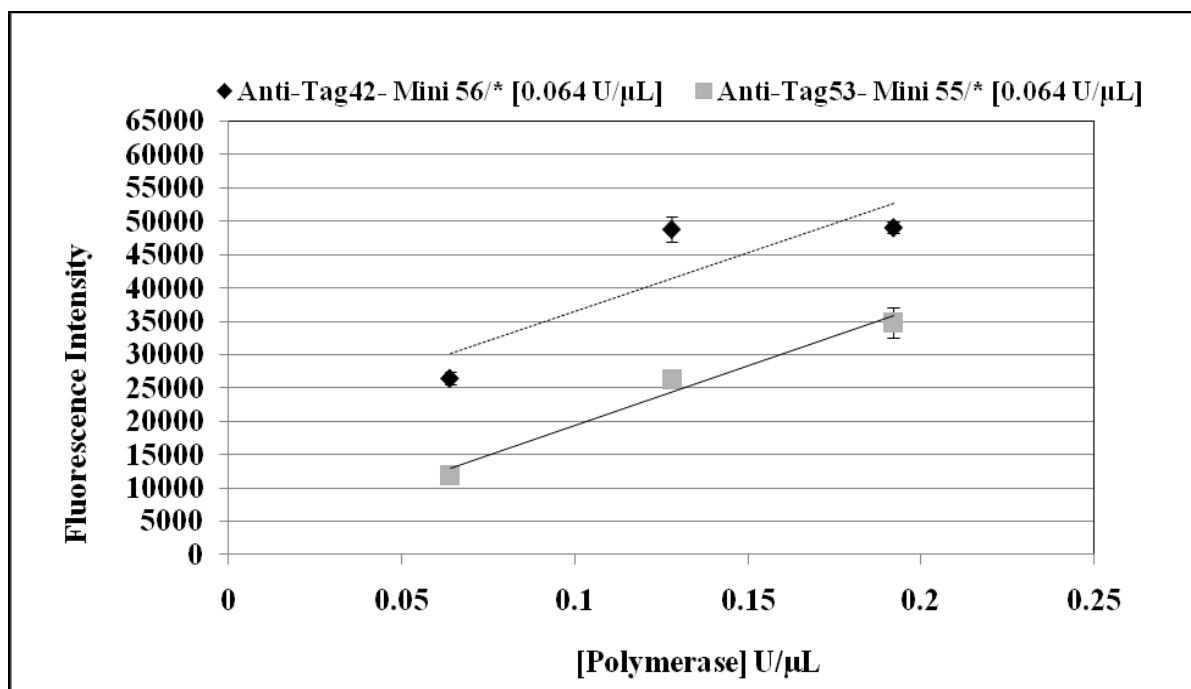


Fig. 8.

CHAPTER V

CHARACTERIZATION AND EVALUATION OF A UNIVERSAL PLANT VIROCHIP

Abstract

Over the past decade, microarray technology has gained popularity in diagnostic applications, due to its ability to screen simultaneously for the presence of multiple pathogens. The advent of a human virus-specific virochip and its utility during the 2003 SARS epidemic demonstrated the ability of a single array to identify both previously characterized and novel viruses. In an effort to produce a similar product for the identification of plant viruses, a plant virochip housing over 9,300 genus and species-specific probes is described. The array includes probes from every recognized viral and viroid genome listed in GenBank as of December of 2009. Using *Wheat streak mosaic virus* as a model pathogen, I examined the reproducibility of the array using biological replicates and queried the chip using different target concentrations. Previously characterized and partially characterized viruses were also hybridized to the array to determine its current detection capabilities.

Introduction

The ability to rapidly and accurately identify plant viruses within various types of plant tissues is important to plant pathologists, epidemiologists and regulatory and quarantine agencies such as the USDA animal and plant health inspection service. Microarray technology is a powerful system in which multiple probes specific for a number of different viruses can be combined, and is thus growing in popularity among researchers and diagnosticians. Over the past decade, numerous microarrays have been developed for the identification and differentiation among viral species and genotypes within plants [1-9]. Although various types of arrays have been developed to differentiate among viruses at different taxonomic levels, a universal plant virus microarray that is capable of identifying all plant viruses has yet to be reported. Such a platform would aid researchers during a disease outbreak or the introduction of a non-indigenous plant pathogen by allowing for the rapid and accurate identification of the pathogen and by providing taxonomic information as a reference. The value in such a system was demonstrated in 2003 when a universal microarray system was used to identify a novel coronavirus, the causal agent of the SARS epidemic [10, 11], and a universal Zip Code array was subsequently developed to identify and determine the genotypes of the viral nucleic acids at variant positions [12].

The vast number of probes that can be housed on a single microarray, coupled with the tendency of probes to cross-hybridize, complicates data interpretation and can lead to the misidentification of pathogens. Therefore, efforts to apply computational algorithms and statistical methods to identify viral species within a sample are currently underway [13-16]. One such algorithm, E-predict, reported in 2005, was used in

combination with the previously described human Virochip to test a variety of clinical samples for the presence of viruses [10, 17]. The E-predict strategy compares observed hybridization patterns from a microarray experiment to theoretical hybridization energy profiles obtained for each set of probes.

The development and testing of the plant Virochip is a collaborative effort by several research groups from various institutes, including the USDA agricultural research service (USDA-ARS), Danforth Center, Washington University, the University of Utah, Cornell University and Oklahoma State University. Our role in the project is to characterize the array in its current state and to test the ability of the Virochip to accurately identify previously characterized plant viruses.

The Virochip houses over 9,300 probes specific to approximately 1,300 viral and viroid species. Family, genus and species-level probes were included for all recognized plant viruses and viroids listed in GenBank as of December of 2009 and positive control probes from conserved host sequences also were incorporated. The array design should lend to the detection of previously characterized viral species, previously uncharacterized isolates and novel plant viruses belonging to characterized genera [John Hammond, USDA-ARS, personal communication].

In this initial assessment of the array, biological replicates and dilution series experiments, using the model virus *Wheat streak mosaic virus* (WSMV), were used to assess the reproducibility of the extraction, labeling and hybridization scheme from replicate experiments using various target concentrations. The array was tested for its ability to identify previously characterized viruses collected and diagnosed by the Plant Disease and Insect Diagnostic Laboratory (PDIDL) at Oklahoma State University. The

Virochip also was used to screen plant samples containing partially characterized plant viruses collected from the Tallgrass Prairie Preserve of northeastern Oklahoma. Though the raw intensity values for the probes were used in the current analyses, an algorithm similar to that of E-predict (T-predict) will ultimately be applied to the hybridization signatures and used for viral identification.

These results were used to assess the current state of the Virochip and help discriminate between useful and uninformative probes. Additional hybridizations are required to fully validate the platform for use in diagnostics and to further refine the T-predict algorithm.

Materials and Methods

Probe design

Oligonucleotide probes were designed by Dr. Kael Fisher at the University of Utah School of Medicine. Plant virus genera were identified using the international committee on taxonomy of viruses' host range annotations and via personal communication with Drs. John Hammond and Claude Fauquet (Danforth Center). All sequences for the plant virus genera were obtained from GenBank (2010-01-15). The sequence records were filtered and redundant subsequences were removed using iterative MegaBLAST. The filtered sequences were considered as densely overlapping 60mer tiles. Each tile was aligned to the complete unfiltered set of plant virus sequence records. Predicted binding energies were calculated (AOS energies) and each tile was assigned to an informative taxon, based on the taxonomic specificity found in the pattern of alignments. Energies of <-30 kcal/mol were considered significant. Taxa with

significant predicted interaction with a tile were identified and linked to the appropriate taxa. Probes were chosen from the list of sequence tiles using a probe selection algorithm.

Sample collection

Plant samples used in this study were derived from three sources. First, healthy wheat tissue was prepared by planting five seeds of *Triticum aestivum* cv. 'Chisholm' in 4" diameter pots containing Sunshine Redi-earth plug and seedling mix [Sun Gro Horticulture, Inc., Bellevue, WA, USA]. Seeded pots were then placed in a plant growth chamber [Percival Scientific, Inc., Perry, IA, USA] and plants were grown for 7-10 days using a 15 hour-day cycle with minimum and maximum temperatures of 19°C and 25°C, respectively. Wheat plants were either left to grow or they were manually inoculated with WSMV 'OK-LAB', which was originally collected in a field located near Lawton in southern Oklahoma and propagated by infecting healthy wheat plants in the laboratory. Mechanical inoculation was achieved by grinding 100 mg of WSMV-infected frozen leaf tissue in 500 µl of water using a mortar and pestle, combining with Celite[®] [Manville Products, Inc., Denver, CO, USA] and rubbing leaf tissue (one leaf/plant) until the leaf appeared to be water soaked. Tissue was harvested 14-21 days post-inoculation by cutting leaf tissue in approximately 1" long sections, thoroughly mixing clippings and storing at -80°C until being used as a source of inoculum. The propagated clippings were subsequently used to infect 7-15 day-old healthy wheat seedlings, and the seedlings were allowed to grow until 7-10 days post-inoculation. Healthy and WSMV-infected tissue was harvested and stored at -80°C until being used as a template for reverse transcription and hybridization.

The second source of virus-infected plant tissue was the Tallgrass Prairie Preserve in northeastern Oklahoma. Over a three-year period from 2005 through 2008, various species of plants were collected by various individuals and tested by sequencing [19-21] and microarray analyses [22] for the presence of viruses in a natural habitat through the Plant Virus Biodiversity and Ecology project [18].

Lastly, plant samples were collected by the Plant Disease and Insect Diagnostic Laboratory (PDIDL) at Oklahoma State University. Upon receipt of plant tissues, the PDIDL tests for the presence of pathogens using molecular and serological diagnostic analyses, including diagnostic PCR and ELISA. All plant samples were stored in a non-frost-free freezer at -80°C until processed.

Printing oligonucleotide microarrays

Oligonucleotide probes were printed onto the microarray slide using a Linear Servo Arrayer at Washington University by Drs. D. Wang and B. Gagewadi. A single spot was printed for each oligonucleotide. In total, 54 plant housekeeping gene control probes, 4,581 forward and 4,518 reverse complement probes were printed on the arrays. A hybridization control probe (probe 70) spiked into each of the oligonucleotides prior to printing, allowed for the proper orientation of the array grid.

Total nucleic acid extraction

The extraction of total nucleic acids (TNA) from plant tissues was carried out similarly to previously reported methods [23, 24]. Briefly, plant leaves were macerated by adding 100 mg of tissue to a 2 ml screw-cap tube containing 12, 2 mm glass beads, adding 1 ml of a solution consisting of 7 ml buffer f2 (2% CTAB, 100 mM Tris-HCl pH 8.0, 2% PVP40, 20 mM EDTA pH 8.0 and 2 M NaCl), 3 ml 5 M potassium acetate and

100 μ L β -mercaptoethanol, chilling the tube at -20°C for 7 minutes, placing the samples in a Mini-Beadbeater-1 cell disrupter (#3110BX)[Biospec Products, Inc., Bartlesville, OK, USA] and performing two, thirty-second rounds of 4500 oscillations per minute. Tubes were subsequently chilled 65°C for 8 minutes then centrifuged at $16000\times g$ for 5 minutes. A volume of 700 μ l of the supernatant was combined with an equal volume of chloroform/isoamyl alcohol (24:1, v/v) and the mixture was vortexed for 15 seconds. Tubes were then centrifuged at $15000\times g$ for 15 minutes and the upper aqueous phase was added to a new tube. Twenty μ l of 3M potassium acetate (pH 5.2) was added to the supernatant and vortexed before the mixture was transferred to an EconoSpin column (#1920-250) [Epoch Life Science, Inc., Missouri City, TX, USA] and the column was centrifuged at $>10000\times g$ for 15 seconds. The column was washed twice with 500 μ l of a 50 ml solution containing 10 μ M Tris-HCl (pH 7.4), 100 mM NaCl, 35 ml of 100% EtOH and 13.5 ml H_2O and the TNA was eluted in 65 μ L of nuclease-free H_2O by spinning at $>10000\times g$ for 1 minute.

cDNA labeling, hydrolysis and clean-up procedures

The cDNA was indirectly labeled by incorporating amino-allyl dUTPs into cDNA, prior to dye coupling. The priming reaction consisted of a 2 $\mu\text{g}/\mu\text{l}$ oligo dT/random hexamer mixture, 0.24 $\mu\text{g}/\mu\text{l}$ TNA and water in a 15.5 μl volume. Tubes containing priming reactions were incubated at 65°C for 5 minutes before being transferred to ice for 5 minutes. cDNA synthesis reactions containing 20 mM DTT, 3 μl of a 12.5 mM dNTP mix (1 mM of dATP, 1 mM dGTP, 1 mM dCTP, 0.4 mM dTTP and 0.6 mM aa-dUTP), RT buffer and 300 U Superscript II Reverse Transcriptase (#18064-014) [Invitrogen, Inc., Carlsbad, CA, USA] and H_2O in a volume of 14.5 μl were mixed

and added to each priming reaction. Tubes of reverse transcription reactions were incubated at 42°C for 1 h.

After the first-strand reactions were complete, a 100 mM NaOH, 10 mM EDTA solution was added and the samples were allowed to incubate in a 65°C water bath for 15 minutes. A DNA clean & concentrator kit (#D4033) [Zymo Research, Inc., Irvine, CA, USA] was used to prepare labeled cDNA products by adding 1 ml of Zymo DNA binding buffer to the reaction products, loading the mixture into a Zymo-Spin Column and centrifuging at 14000xg for 30 seconds. Columns were washed with 200 µl of Zymo Wash Buffer and centrifuged again at 14000xg for 30 seconds before adding 10 µl nuclease-free H₂O and centrifuging the column at 14000xg for 30 seconds.

Amino-allyl dye coupling

One µl of 1 M NaHCO₃ (pH 9.0) and 1 µl of Cy3 reactive dye (8,000 pmole/µl suspended in DMSO) (#RPN5661) [GE Healthcare, Inc., Piscataway, NJ, USA] were added to the labeled cDNA products to initiate the coupling reaction. One µl (100 ng) of probe 70 was combined with 1 µl of NaHCO₃ (pH 9.0), 8 µl H₂O and 1 µl of Cy5 reactive dye (8,000 pmole/µl suspended in DMSO). Dye-coupling reactions were allowed to incubate at room temperature in the dark for 1 h. The removal of uncoupled dyes was achieved using the DNA clean & concentrator kit [Zymo Research, Inc., Irvine, CA, USA] by adding 200 µl of Zymo DNA binding buffer to each sample, mixing well and loading the solution into a Zymo-Spin column. Samples were centrifuged at 14000xg for 30 seconds and washed twice with 200 µl of Zymo wash buffer, spinning again at 14000xg for 30 seconds each. Samples and tubes containing probe 70 were eluted in 24 µl and 10 µl of H₂O, respectively, by spinning at 14000xg for 30 seconds.

Array hybridization, washing and scanning

Hybridization mixtures were prepared by adding 10 μ l of Cy5-labeled probe 70, 6 μ l of 20x SSC and 1 μ l of HEPES (total volume of 17 μ l) to a mixture containing 24 μ l of Cy3-labeled cDNA and 1 μ l of 10% SDS for a total volume of 42 μ l. Tubes containing dye-labeled products were boiled in a 100°C water bath for 2 minutes and cooled at room temperature for 5 minutes before being applied to the surface of the array. To facilitate even hybridization of the labeled products, 25 x 40 mm LifterSlips (#25x40I247723) [Erie Scientific, Inc., Portsmouth, NH, USA] were placed over the surface of the array prior to the addition of the products. Slides were sealed in a hybridization cassette [Corning Life Sciences, Lowell, MA, USA] containing 10 μ l of 3.75x SSC at each end of the array and allowed to incubate in a 65°C water bath for 4 h. Slides were subsequently washed in two wash solutions. First, slides were submerged in wash solution I containing 340 ml H₂O, 10 ml 20x SSC and 1 ml 10% SDS, and plunged up and down 10-20 times. The LifterSlip was dislodged during this initial wash. Next, slides were washed for 30 seconds in wash solution II containing 350 ml H₂O and 1 ml of 20x SSC. Slides were dried by centrifuging for 1 minute in a Microarray High-Speed Centrifuge [ArrayIt, Sunnyvale, CA, USA] and stored in a vacuum-sealed microscope slide box at room temperature prior to scanning.

Slides were scanned using an Axon GenePix 4400A 4-laser confocal scanner [Molecular Devices, Inc., Sunnyvale, CA, USA], scanning at 532 nm and 635 nm for detection of virus-labeled cDNA and orientation of the grid, respectively. Image acquisition and signal analyses were performed using GenePix Pro 4.0 software and raw

intensity values were used to assess the reproducibility, sensitivity and hybridization of virus-containing RNA from tissue.

Real-Time PCR

Prior to performing real-time PCR, positive control plasmids were prepared by amplifying a 110 bp region of the WSMV CI gene using the forward primer WSMV_F2 (5'-GCAAGGAAAATCGGAAATGA-3') and the reverse primer WSMV_R2 (5'-TTCTGCAACACTGGAGCAAC-3') [Mohammad Arif, Oklahoma State University, personal communication] and subsequently cloning the amplicon into a pCR2.1[®] TOPO[®] TA vector using the TOPO[®] TA Cloning kit (#K4550-01) [Invitrogen, Inc., Carlsbad, CA, USA]. Plasmid vectors were isolated from overnight cultures using a Qiaprep Spin Miniprep Kit (#27104) [Qiagen, Inc., Germantown, MD, USA]. Real-time PCR was performed using a Rotor-Gene 6000 Real-Time Rotary analyzer [Corbett Life Sciences, Inc., Concorde, NSW] and the target was quantified using a ten-fold serial dilution of the positive-control plasmid clone, the SYBR green master mix kit [Invitrogen, Inc., Carlsbad, CA, USA] and the same WSMV-specific primers that were used to produce the 110 bp amplicons for cloning.

Results

Reproducibility of the array

Total nucleic acids from wheat tissue (*Triticum aestivum* cv. Chisholm) that was previously inoculated with WSMV was confirmed virus-infected using the forward (5'-TACTTGACTGGGACCCGAA-3') and reverse (5'-AACCCACACATAGCTACCAAG-3') primers C1 (nt 9351-9371 of WSMV Sidney 81 genome NCBI accession number

AF057533) and XC1 (nt 8117-8135 of WSMV Sidney 81 genome NCBI accession number AF057533), respectively [25] (Fig. 1), were used for three replicate experiments. The sequenced amplicons were 97% similar to the above mentioned WSMV Sidney 81 strain. Total nucleic acid was also extracted from healthy wheat tissue from the same cultivar for use as a negative control. A graph displaying the top fifty highest intensity values from the hybridization reactions using WSMV-infected and healthy wheat tissue is shown in Fig. 2. Although the virus-infected tissue produced slightly higher fluorescence intensities than the healthy tissues, the fluorescence intensity profiles for the two datasets did not differ significantly.

The top fifty probes having the highest fluorescence intensity values (excluding plant-specific positive control probes) were compared to assess the reproducibility among replicates. When the fifty most intense probes from the healthy wheat tissue were compared to the top fifty probes in all three replicates, only 33% of the probes were different. The probes that differed between the two data sets (healthy wheat vs. WSMV-infected wheat) were those that specifically targeted WSMV and the closely related *Triticum mosaic virus*, and three others that targeted viruses related to WSMV. All three of the WSMV-targeting probes that yielded intensity values in the top 50 were among the highest for each replicate (Fig. 3). The remaining five probes produced intensity values similar to those produced by random associations with unrelated viruses. Since two of the WSMV-specific probes were designed to detect the divergent WSMV El Batán strain, and did not display significant alignment to the WSMV Sidney 81 strain that was used for hybridization, only six probes should have produced significant fluorescent intensities for WSMV (Table 3 and Appendix). Fifty percent (3/6) of the probes that were designed

to detect WSMV at each of the taxonomic levels detected in the 50 most fluorescent probes; one targeting the family (*Potyviridae*), genus (*Tritimovirus*) and species (WSMV) (Fig. 4).

Only 48% of the probes of the list consistently detected virus among all three virus-infected replicate experiments. This percentage was calculated by dividing the number of total probes in the top 50 most fluorescent positions that were conserved in all three replicates (24) by the total number of probes queried (50) and multiplying by 100%. Probes corresponding to viruses in nine different families were observed in the list. No obvious patterns of fluorescence intensities were observed among the three replicates or the positive controls.

Effects of target concentration on hybridization efficiency

To determine the degree to which the target concentration restricts the detection of the viral target by the Virochip, the concentration of the amino-allyl labeled WSMV cDNA targets was determined by Real-Time PCR (Fig. 5). Labeled cDNA was subsequently diluted in a ten-fold series, and solutions containing 4.5 pg, 0.45 pg and 0.045 pg of the cDNA targets were subjected to the dye coupling, hybridization and scanning procedures (Fig. 6). As expected, the relative intensities of top fifty highest fluorescing probes decreased with the target concentration. In total, 46% of the probes consistently produced high fluorescence values in all three samples, which is only 2% lower than that of the three biological replicates. The same three WSMV-specific probes produced high fluorescence intensities. Linear regression analyses on the three dilutions of the three probes (without normalization) were 0.96, 0.92 and 0.99 for the family, species and genus-level probes, respectively.

Detection of previously diagnosed viruses and unknown viruses using the Virochip

To test the detection capabilities of the Virochip, a collection of plants infected with various plant viruses was subjected to the extraction, labeling and hybridization procedures (Table 1). These infected plant samples were collected by the Oklahoma State University Plant Disease and Insect Diagnostic Laboratory (PDIDL) and diagnosed using diagnostic PCR and ELISA. Because the family and genus-specific probes were designed to detect specific subgroups of species within the taxonomic groups, not all species should be identified by all family and genus-specific probes. Therefore, only the family and genus-specific probes that were designed using the queried species are displayed in the table.

The fifty most intensely fluorescing probes were examined for each hybridization reaction. Most of the target-specific probes failed to produce elevated spot intensities and none of the samples were identified by 100% of the virus-specific probes. Correctly performing probes did not appear to fall into any taxonomic categories.

Four samples infected with viruses that were classified to a specific taxonomic level (family or genus) by sequencing were also tested against the array (Table 2). These samples, collected from the Tallgrass Prairie Preserve of northeastern Oklahoma, were infected with viruses from the *Caulimoviridae*, *Partitiviridae* and *Reoviridae* families. Probes targeting these families and the associated genera (if known) were present on the array. Probes that were consistent with the previous taxonomic assignment of those viruses displayed high fluorescence intensities in half of the samples (2/4). Both samples infected with viruses from the *Partitiviridae* family were detected by species-specific

probes; however, probes for two different viral genera were fluorescent in the hybridization reaction involving the array ‘UPVM-P01_150’.

Discussion

Over the past decade, microarray technology has been recognized for its potential to function as a diagnostic tool for researchers working in human, animal and plant pathology. The ability to screen a single plant for thousands of plant viruses while simultaneously identifying novel viruses would be useful to regulatory agencies, plant pathologists, ecologists, epidemiologists and researchers, alike. We tested a previously fabricated universal plant Virochip that was designed to detect all plant viruses at the family, genus and species-levels.

Several approaches can be used to analyze microarray data. Using an algorithm, such as E-predict [17], to compare the observed patterns of hybridization to patterns of theoretical energy profiles would yield a distinct output. However, the ability of such an algorithm to identify plant viruses requires that the family, genus and species-specific probes housed on the array hybridize to their intended targets. Furthermore, cross-hybridization of probes to host nucleic acids may complicate pattern recognition, thereby impeding the ability of the algorithm to accurately identify target viruses. The identification of cross-hybridizing probes and probes that fail to identify their specified targets, or that identify them inconsistently, prior to the application of such an algorithm is necessary. Hence the fifty probes having the highest fluorescent values in each dataset were examined in this study. This approach allowed for the assessment of the target-

specific probes, and for the identification of non-specific interactions, and the potential for strand-bias and simplified between-dataset comparisons.

Much variability was observed among the fifty highest intensity probes for the three biological replicates, as well as those for the three dilution hybridizations. When the fifty highest intensity-producing probes for healthy wheat were compared to the probes that were common among the three biological replicates, only a third were different and a number of the variable probes were those that were specific to the target virus WSMV and the closely-related *Triticum mosaic virus*. These data suggest that the non-specific fluorescent probes displayed high intensity values as a result of cross-hybridization with host nucleic acids.

Target-specific probes for most of the previously diagnosed viruses did not display fluorescence values within the top fifty highest intensities. However WSMV was properly identified by three probes. One possible explanation for this observation is that most of the plant tissue containing the previously identified viruses was over five years old thus the virus particles could have degraded over time, as compared to the relatively fresh tissue infected with WSMV. Because plant-specific positive control probes displayed high fluorescence values, this is not the likely explanation. Another possible explanation is that the high degree of cross-reactivity by non-specific probes contributed to the masking of the true positive probes. Yet another explanation is that the probe-design strategy did not adequately address the factors that may influence target binding. Interestingly, all three of the probes that detected WSMV targeted the antisense strand of the virus, while their reverse complements failed to produce intense fluorescence values. Though this result is unexpected, as a higher concentration of sense-strand nucleic acids

are produced during replication than anti-sense, the phenomenon has been reported previously in the literature [22]. The molecular basis for the observed strand bias is unclear; however, the potential for strand bias should be considered in microarray probe design and the development of pattern detection programs such as E-predict (and T-predict).

Two of the partially characterized samples collected from the Tallgrass Prairie Preserve were detected by family-specific probes (slide IDs: UPVM-P01_126 and UPVM-P01_150). Probes for two genera in the *Partitiviridae* generated fluorescence intensities in the 50 highest fluorescent positions in one of the reactions (slide ID: UPVM-P01_150). Since the virus was previously classified only to the family level, the possibility of the correct identification of the two viral species within the same plant cannot be rejected. However, since one probe corresponded to a fungal-infecting virus (*Partitivirus* genus) [26], and the other identified a plant-infecting virus (*Alphacryptovirus* genus) [26], cross-hybridization is most likely responsible for the conflicting results.

Previous work has demonstrated that multiple factors influence the ability of a microarray probe to bind its intended target with high sensitivity and specificity. Some of these factors include steric hindrances [27] inter-probe interactions [28] and secondary structure within RNA target molecules [29]. Furthermore, variation of probe-generated signal intensities is highly dependent upon the probe sequence [30, 31]. Therefore it is not surprising that off-target effects were observed within the vast number of probes included on the described Virochip and that some target probes yielded higher fluorescence intensities than others. Since the virus-infected plant samples reported here

were the first to be tested by the Virochip, it is likely that nonspecific probes will be re-designed or removed from the array to reduce the amount of cross-hybridization between virus-specific probes and host-derived sequences. Additionally, the application of the E-predict-like algorithm (T-predict) to apply a statistical method for diagnosis and virus discovery will further enhance the ability of the array to identify queried viruses.

Further research is necessary to refine and fully assess the capabilities of the described Virochip, including the removal of non-specific and cross-hybridizing probes, the development of the T-predict algorithm and further testing and validation of the array using additional plant viruses. The continuous discovery of novel viruses by researchers will require that the array is updated periodically to include new viruses, if it is to remain relevant. The potential for novel virus discovery by the array will likely contribute to the pool of knowledge for plant virus evolution. The rapid screening ability of the procedure would also benefit inspection agencies, such as the DHS-CBP, that are charged with monitoring incoming commodities at the United States borders.

Acknowledgements

The authors would like to thank Jen Olson and Dr. Stephanie Rogers for providing virus-infected plant tissues and Dr. Mohammad Arif for his assistance with Real-Time PCR. This research was supported by USDA-NIFA and the Oklahoma Agricultural Experiment Station.

LITERATURE CITED

1. Deyong, Z., et al., *Differentiation of Cucumber mosaic virus isolates by hybridization to oligonucleotides in a microarray format*. Journal of Virological Methods, 2005. **123**(1): p. 101-108.
2. Tiberini, A., et al., *Oligonucleotide microarray-based detection and identification of 10 major tomato viruses*. Journal of Virological Methods. **168**(1-2): p. 133-140.
3. Engel, E.A., et al., *A diagnostic oligonucleotide microarray for simultaneous detection of grapevine viruses*. Journal of Virological Methods. **163**(2): p. 445-451.
4. Abdullahi, I. and M. Rott, *Microarray immunoassay for the detection of grapevine and tree fruit viruses*. Journal of Virological Methods, 2009. **160**(1-2): p. 90-100.
5. Pasquini, G., et al., *Oligonucleotide microarray-based detection and genotyping of Plum pox virus*. Journal of Virological Methods, 2008. **147**(1): p. 118-126.
6. Lenz, O., K. Petrzik, and J. Spak, *Investigating the sensitivity of a fluorescence-based microarray for the detection of fruit-tree viruses*. Journal of Virological Methods, 2008. **148**(1-2): p. 96-105.
7. Boonham, N., et al., *Detection of potato viruses using microarray technology: towards a generic method for plant viral disease diagnosis*. Journal of Virological Methods, 2003. **108**(2): p. 181-187.
8. Bystricka, D., et al., *Oligonucleotide-based microarray: A new improvement in microarray detection of plant viruses*. Journal of Virological Methods, 2005. **128**(1-2): p. 176-182.

9. Lee, G.P., et al., *Plant virus cDNA chip hybridization for detection and differentiation of four cucurbit-infecting Tobamoviruses*. Journal of Virological Methods, 2003. **110**(1): p. 19-24.
10. Wang, D., et al., *Viral discovery and sequence recovery using DNA microarrays*. PLoS Biology, 2003. **1**(2): p. e2.
11. Ksiazek, T.G., et al., *A novel Coronavirus associated with Severe Acute Respiratory Syndrome*. New England Journal of Medicine, 2003. **348**(20): p. 1953-1966.
12. Long, W.-H., et al., *A universal microarray for detection of SARS coronavirus*. Journal of Virological Methods, 2004. **121**(1): p. 57-63.
13. Wong, C., et al., *Optimization and clinical validation of a pathogen detection microarray*. Genome Biology, 2007. **8**(5): p. R93.
14. Watson, M., et al., *DetectiV: visualization, normalization and significance testing for pathogen-detection microarray data*. Genome Biology, 2007. **8**(9): p. R190.
15. Rehrauer, H., et al., *PhyloDetect: a likelihood-based strategy for detecting microorganisms with diagnostic microarrays*. Bioinformatics, 2008. **24**(16): p. i83-i89.
16. Allred, A., et al., *VIPR: A probabilistic algorithm for analysis of microbial detection microarrays*. BMC Bioinformatics. **11**(1): p. 384.
17. Urisman, A., et al., *E-Predict: a computational strategy for species identification based on observed DNA microarray hybridization patterns*. Genome Biology, 2005. **6**(9): p. R78.

18. Wren, J.D., et al., *Plant virus biodiversity and ecology*. PLoS Biology, 2006. **4**(3): p. e80.
19. Melcher, U., et al., *Evidence for novel viruses by analysis of nucleic acids in virus-like particle fractions from Ambrosia psilostachya*. Journal of Virological Methods, 2008. **152**(1-2): p. 49-55.
20. Muthukumar, V., et al., *Non-cultivated plants of the Tallgrass Prairie Preserve of northeastern Oklahoma frequently contain virus-like sequences in particulate fractions*. Virus Research, 2009. **141**(2): p. 169-173.
21. Roossinck, M.J., et al., *Ecogenomics: using massively parallel pyrosequencing to understand virus ecology*. Molecular Ecology. **19**: p. 81-88.
22. Grover, V., et al., *Oligonucleotide-based microarray for detection of plant viruses employing sequence-independent amplification of targets*. Journal of Virological Methods. **163**(1): p. 57-67.
23. Echevarría-Machado, I., et al., *A simple and efficient method for isolation of DNA in high mucilaginous plant tissues*. Molecular Biotechnology, 2005. **31**(2): p. 129-135.
24. Li, R., et al., *A reliable and inexpensive method of nucleic acid extraction for the PCR-based detection of diverse plant pathogens*. Journal of Virological Methods, 2008. **154**(1-2): p. 48-55.
25. Carver, J.D., *Molecular typing of Wheat streak mosaic virus for forensic applications*, in *Department of Forensic Science*. 2007, Oklahoma State University.

26. Mahy, B. and M.v. Regenmortel, eds. *Partitiviruses of fungi*. Desk encyclopedia of plant and fungal virology. 2010, Elsevier, Ltd.: San Diego.
27. Shchepinov, M.S., S.C. Case-Green, and E.M. Southern, *Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays*. Nucleic Acids Research, 1997. **25**(6): p. 1155-1161.
28. Forman Jonathan, E., et al., *Thermodynamics of Duplex Formation and Mismatch Discrimination on Photolithographically Synthesized Oligonucleotide Arrays*, in *Molecular Modeling of Nucleic Acids*. 1997, American Chemical Society. p. 206-228.
29. Ratushna, V., J. Weller, and C. Gibas, *Secondary structure in the target as a confounding factor in synthetic oligomer microarray design*. BMC Genomics, 2005. **6**(1): p. 31.
30. Zhang, L., M.F. Miles, and K.D. Aldape, *A model of molecular interactions on short oligonucleotide microarrays*. Nat Biotech, 2003. **21**(7): p. 818-821.
31. Royce, T.E., J.S. Rozowsky, and M.B. Gerstein, *Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification*. Nucleic Acids Research, 2007. **35**(15): p. e99.

Figure Captions

Fig. 1. Agarose gel image of WSMV RT-PCR amplicons after staining with ethidium bromide. Lane 2 contains a negative control sample and lanes 3 and 4 contain RT-PCR amplicons corresponding to a 1,254 bp region encompassing the *Wheat streak mosaic virus* coat protein. A low mass ladder [Invitrogen, Inc., Carlsbad, CA, USA] was included in the gel electrophoresis for estimation of amplicon size and concentration (lane 1).

Fig. 2. Line graph displaying the raw fluorescence intensity profiles for the 50 most intense probes in a WSMV-positive sample (UPVM-P01_010) and a healthy wheat sample (UPVM-P01_032). Total nucleic acids were extracted from WSMV-infected tissue and healthy wheat tissue and used as a template for labeling via incorporation of aa-dUTPs during reverse transcription. After the dye coupling reaction, samples were hybridized to separate arrays. Both samples were scanned using 100% laser power and 650 PMT gain. The 50 most intense fluorescent spots in each dataset, excluding positive controls, are presented.

Fig. 3. Graphical representation of the biological replicates performed using WSMV-infected plant tissue. Probes that displayed the highest fluorescence values (top 50) and were consistent among all three replicates are presented. For reference purposes, three positive control probes are also displayed.

Fig. 4. Representation of viral taxa conserved among the top fifty probes and their relative abundances in all three replicate experiments. Probes were derived from nine viral and one unknown family (large chart) and the majority of the viral genera included probes from the *Tymoviridae* and *Potyviridae* families. Within the *Potyviridae* family, five probes corresponded to the *Tritimovirus* genus (small chart), of which three were specific for WSMV.

Fig. 5. Real-Time PCR results for the quantification of WSMV cDNA. Total nucleic acids were extracted from WSMV-infected plant tissue and used as a template for RT-PCR, incorporating aa-dUTPs. The amino-allyl labeled cDNA was subsequently used as a template for qPCR (A). Plasmid clones containing a 110 bp fragment of the WSMV CI protein were used to quantify the cDNA target. 10 ng, 1 ng and 0.1 ng of the plasmids were used in duplicate to produce a standard curve (B). Healthy wheat and a no template control reactions were also included in the qPCR procedure. The initial target concentration was found to be 469 fg/ μ l but was diluted in water to 450 fg/ μ l for dye coupling and array hybridization.

Fig. 6. Fluorescence intensities of probes using 4.5 pg, 0.45 pg and 0.045 pg of target cDNA. (A) The 50 highest fluorescing probes for each group and (B) eight WSMV-specific probes the clustered column graph for each of the target amounts. The three highest columns correspond to family, genus and species-specific probes for WSMV listed in Table 1.

Table 1. Detection of previously identified plant viruses by the Virochip.

Chip ID	Genome	Previously Identified Viruses			Observed Probe
		Family	Genus	Species	
UPVM-P01_010	(+) ssRNA	<i>Potyviridae</i>	<i>Tritimovirus</i>	<i>Wheat streak mosaic virus</i>	<i>Potyviridae</i> : 13241967_nt00101.60 <i>Tritimovirus</i> :13241967_nt07976.60 <i>Wheat streak mosaic virus</i> :13241967_nt03001.60
UPVM-P01_011	(+) ssRNA	<i>Potyviridae</i>	<i>Bymovirus</i>	<i>Wheat spindle streak mosaic virus</i>	None
UPVM-P01_019	(+) ssRNA	<i>Virgaviridae</i>	<i>Tobamovirus</i>	<i>Tobacco mosaic virus</i>	None
UPVM-P01_035	(+) ssRNA	<i>Alphaflexiviridae</i>	<i>Potexvirus</i>	<i>Hosta virus X</i>	None
UPVM-P01_043	(-) ssRNA Tripartite	<i>Bunyaviridae</i>	<i>Tospovirus</i>	<i>Impatiens necrotic spot virus</i>	<i>Tospovirus</i> : 641812_nt3701.60_rc
UPVM-P01_053	(+) ssRNA	<i>Luteoviridae</i>	<i>Luteovirus</i>	<i>Barley yellow dwarf virus-PAV</i> <i>Triticum mosaic virus</i>	<i>Barley yellow dwarf virus-SGV</i> : 533388_nt0600.60_rc <i>Barley yellow dwarf virus-SGV</i> : 533388_nt0600.60 <i>BYDV-PAV::Luteovirus</i> : 165909482_nt3469.60_rc None
UPVM-P01_054	(+) ssRNA	<i>Potyviridae</i>	<i>Tritimovirus</i>	<i>Wheat streak mosaic virus</i>	<i>Potyviridae</i> : 22164035_nt0986.60 <i>Tritimovirus</i> : 13241967_nt0976.60
UPVM-P01_069	(+) ssRNA Bipartite	<i>Secoviridae</i>	<i>Nepovirus</i>	<i>Tobacco ringspot virus</i>	None
UPVM-P01_079	(+) ssRNA	<i>Potyviridae</i>	<i>Tritimovirus</i>	<i>Oat necrotic mottle virus</i>	None
UPVM-P01_085	(+) ssRNA	<i>Luteoviridae</i>	<i>Luteovirus</i>	<i>Barley yellow dwarf virus</i>	None

Table 2. Identification of partially characterized viruses using the Virochip.

Unclassified Viruses						
Chip ID	Genome	Family	Genus	Species	PVBE ID	Observed Probe
UPVM-P01_106	dsDNA Circular	<i>Caulimoviridae</i>	<i>Caulimovirus</i>	Unknown	05TGP00107	None
UPVM-P01_126	dsRNA	<i>Partitiviridae</i>	Unknown	Unknown	05TGP00147	<i>Partitiviridae::Partivirus::Ophiostoma partivirus 1 segment RNA1::Viruses: 109804205_nt1101.60_rc</i>
UPVM-P01_150	dsRNA	<i>Partitiviridae</i>	Unknown	Unknown	06TGP01221	<i>Partitiviridae::Partivirus::Ophiostoma quercus partivirus::Ophiostoma quercus partivirus:125716535_nt340.60_rc Partitiviridae::Alphacryptovirus::Vicia cryptic virus M::Vicia cryptic virus M: 166203315_nt2051.60_rc</i>
UPVM-P01_162	dsRNA	<i>Reoviridae</i>	<i>Oryzavirus</i>	Unknown	05TGP00040	None

Table 3. Family, genus, species and strain-specific probes for WSMV.

Probes			BLAST Results*			
Taxonomic group	Probe ID	Sequence (5' → 3')	Accession	Description	E value	Max ident
Family	13241967_nt00101.60	TCATTGTGAGCTCTCGCATAGAGATAAGCAATGGCAACAGCGAATGTTTGCTCGGTGAC	EU914918.1	Wheat streak mosaic virus isolate Saadat-Shahr, complete genome	2e-22	100%
Family	13241967_nt00101.60_rc	GTCACCGAGCAAACAATTCGCTGTTGCCATTGCTTATCTCTATGCGAGAGCTCACAAATGA	EU914918.1	Wheat streak mosaic virus isolate Saadat-Shahr, complete genome	2e-22	100%
Genus	13241967_nt07976.60	GCTATTCAGAGCGCTATTATTGCAGCATAACGTTGGAAGCTTTCGGTTATGATGAATTCACG	EU914917.1	Wheat streak mosaic virus isolate Naghadeh, complete genome	2e-22	100%
Genus	13241967_nt07976.60_rc	CGTGAATTCATCATAACCGAAAGCTTCCACGTATGCTGCAATAATAGCGCTCTGAATAGC	EU914917.1	Wheat streak mosaic virus isolate Naghadeh, complete genome	2e-22	100%
Species	13241967_nt03001.60	TACTCACCTTTCGCGAAACGCTTACAGGTGGGTATTTACAACATGGCCGCGAACGTCTT	FJ348359.1	Wheat streak mosaic virus isolate Arg2 polyprotein gene, complete cds	2e-22	100%
Species	13241967_nt03001.60_rc	AAGACGTTTCGCGGCATGTTGTAAAATACCCACCTGTAAGCGTTTCCGCAAAGGTGAGTA	FJ348359.1	Wheat streak mosaic virus isolate Arg2 polyprotein gene, complete cds	2e-22	100%
Species (Strain)	11066855_nt5898.60	ATAAGCGTGGTTTGCCAGTAGGGTATGCCGACCACAGAGGTGAATGGCGCCAGACGCAAC	AF285170.1	Wheat streak mosaic virus strain El Batan 3 complete genome	2e-22	100%
Species (Strain)	11066855_nt5898.60_rc	GTTGCGTCTGGCGCATTACCTCTGTGTGTCGCATACCTACTGGCAAACCACGCTTAT	AF285170.1	Wheat streak mosaic virus strain El Batan 3 complete genome	2e-22	100%

*The top result for each query is listed. A complete list of matches for each query can be found in the Appendix.

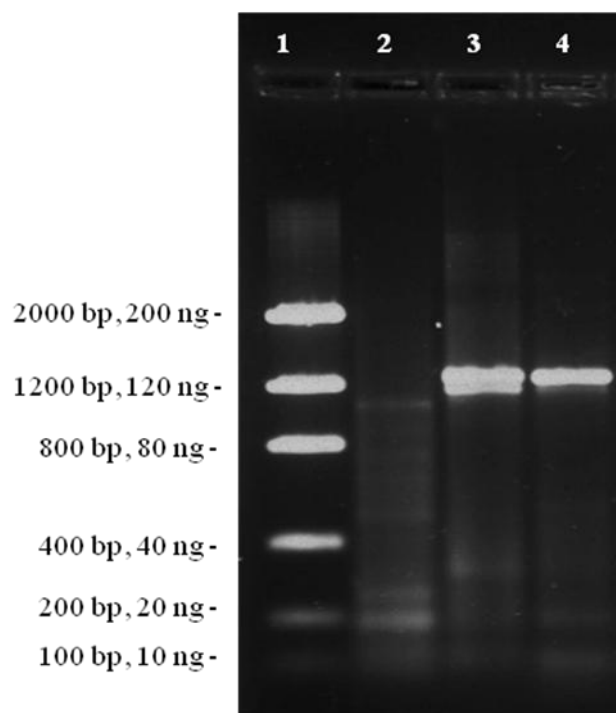


Fig. 1.

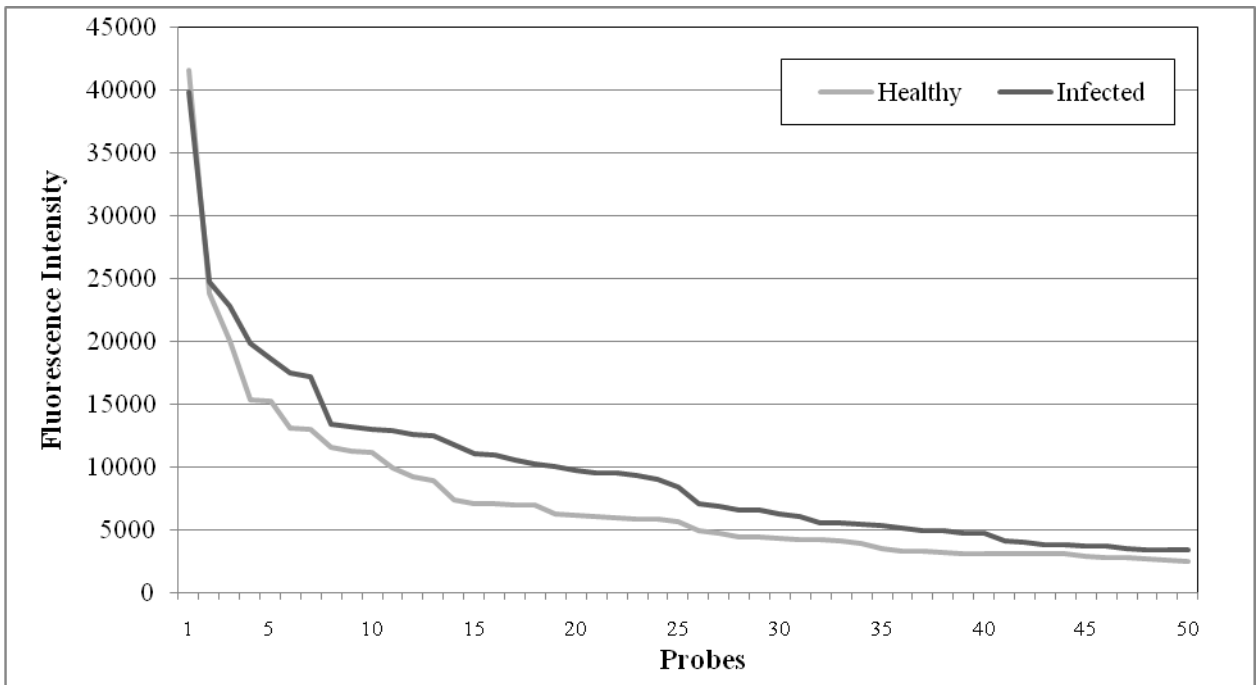


Fig. 2.

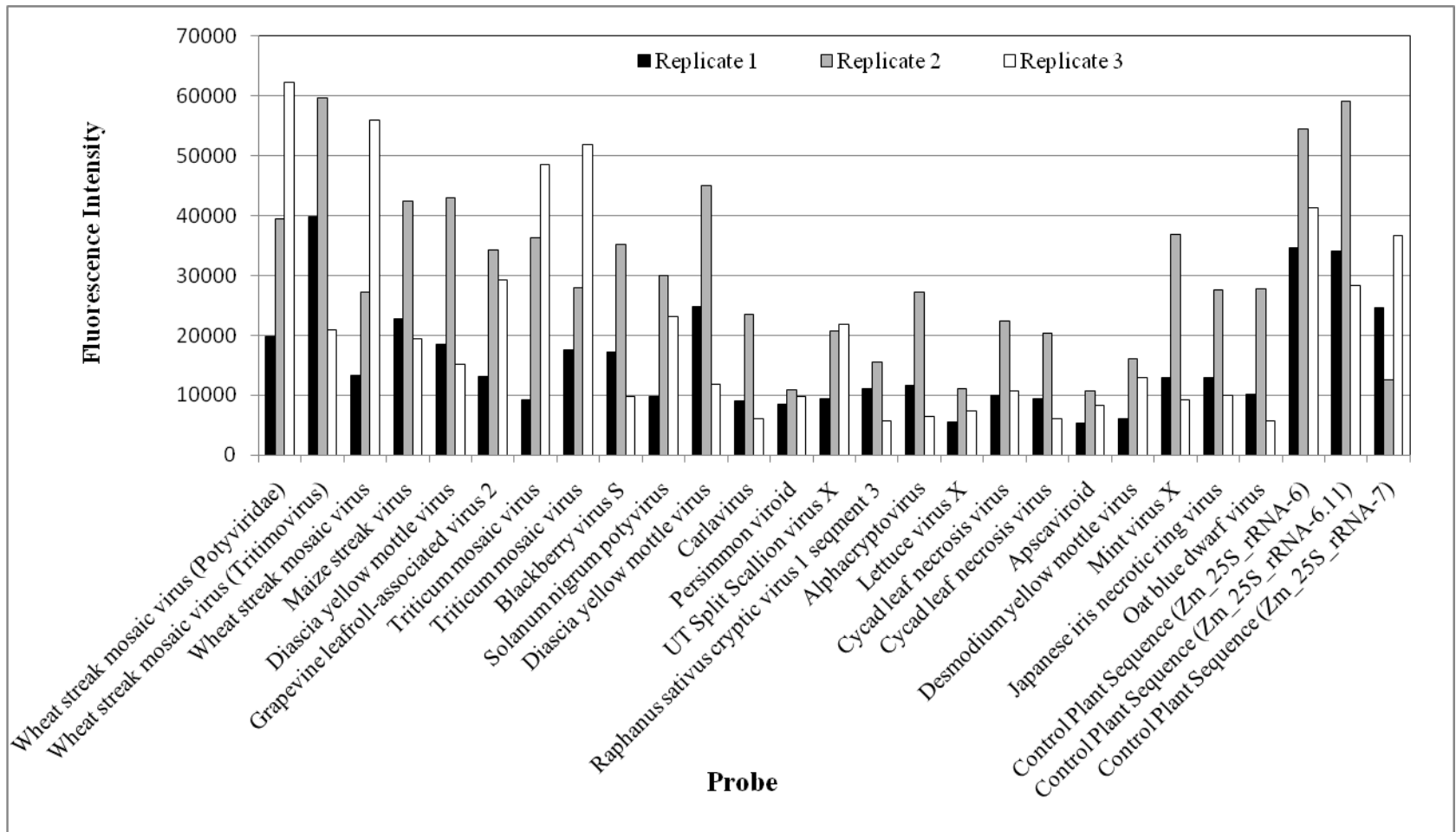


Fig. 3.

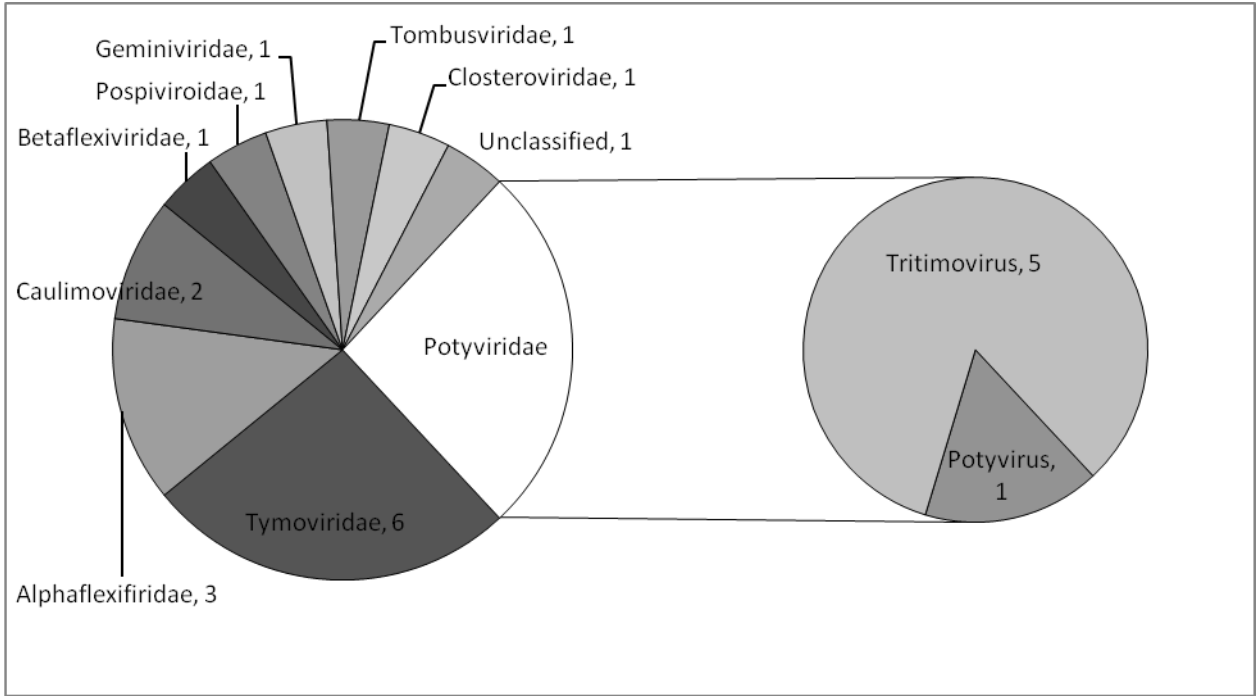


Fig. 4.

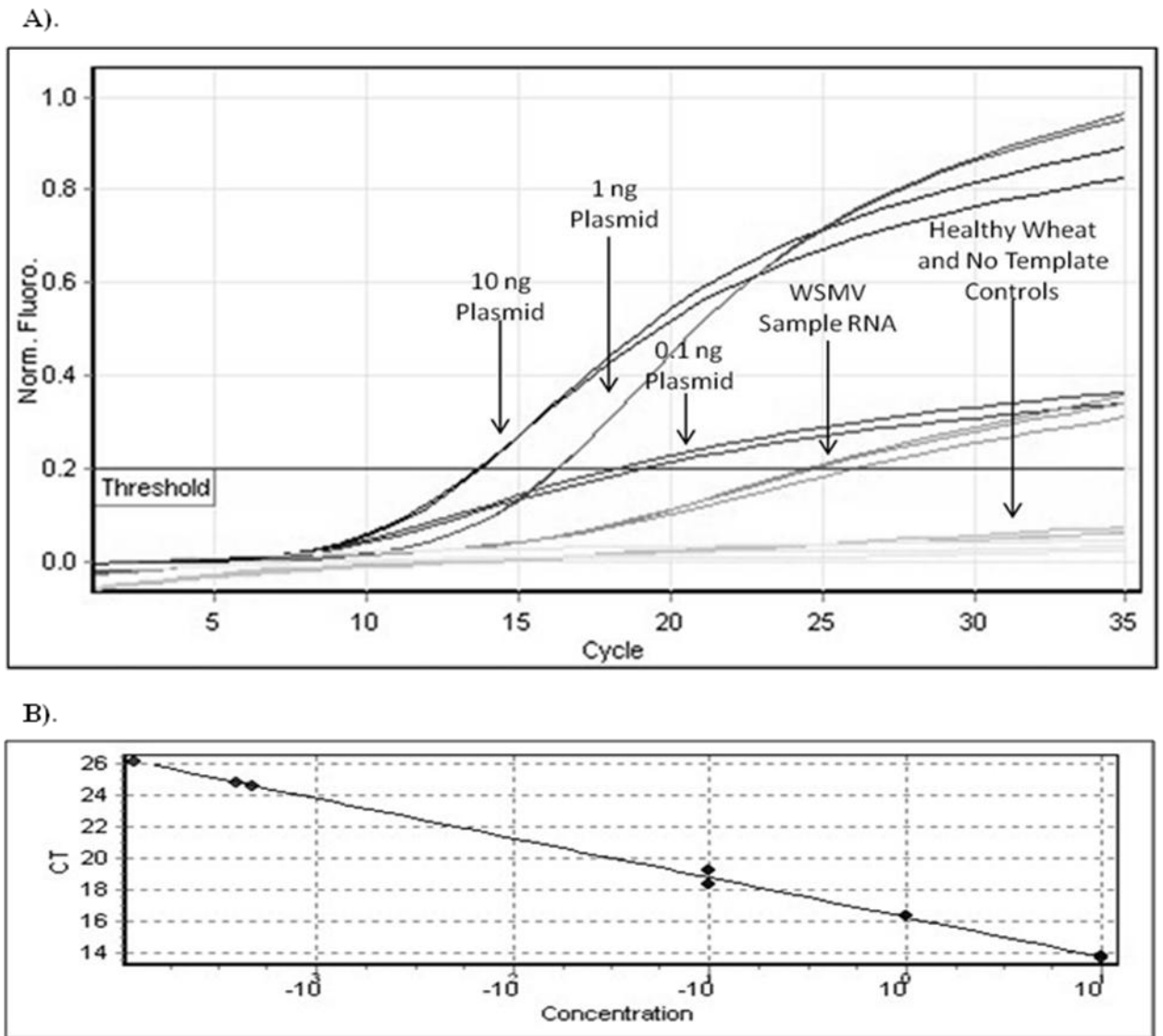
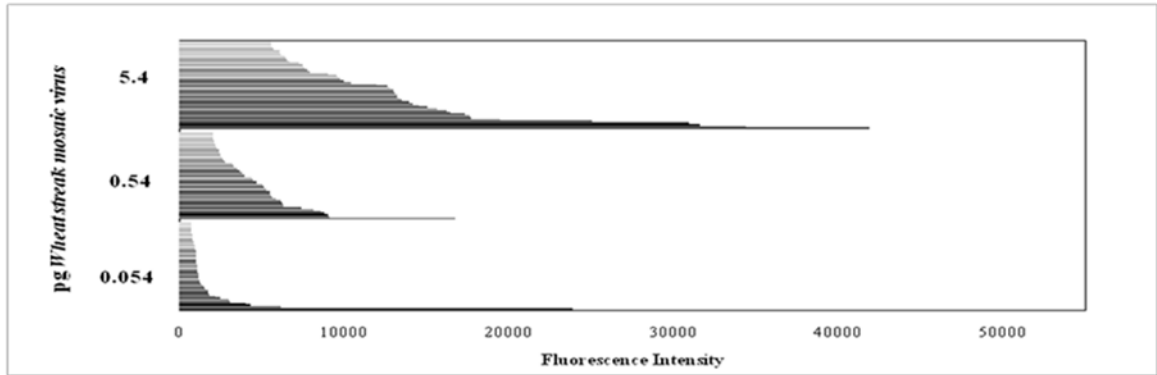


Fig. 5.

A).



B).

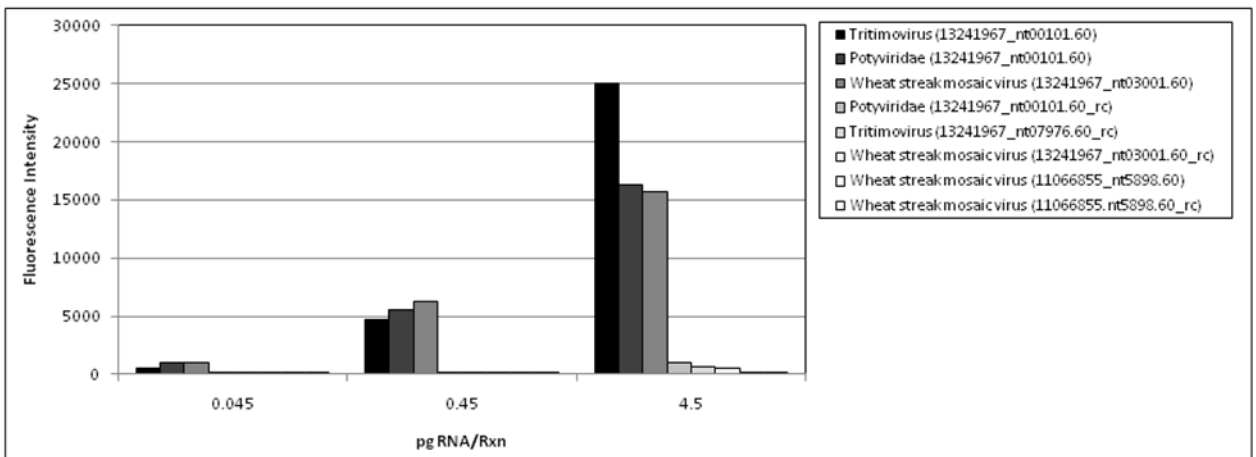


Fig. 6.

Chapter VI Preface

Use of Fatty Acid Methyl Ester Profiles for Discrimination of *Bacillus Cereus* T-Strain Spores Grown on Different Media

This research project was executed at the Federal Bureau of Investigation Counterterrorism and Forensic Science Research Unit (FBI-CFSRU) in Quantico, VA. The manuscript was published in 2010 in Applied and Environmental Microbiology (AEM) and appears in this thesis with the journal's permission (License Number 2657790840032). My contribution to the project was made during a three-month internship at the FBI-CFSRU during the summer of 2008. My specific contributions are as follows:

- 1). Grew *Bacillus cereus* T-strain cells, induced sporulation, performed FAME analyses and documented results.
- 2). Produced the vegetative cell and spore images displayed in the manuscript by capturing images over an 8-hour interval.
- 3). Assisted with qualitative data interpretation.
- 4). Reviewed the draft manuscript and provided grammatical and scientific feedback.

CHAPTER VI

USE OF FATTY ACID METHYL ESTER PROFILES FOR DISCRIMINATION OF *BACILLUS CEREUS* T-STRAIN SPORES GROWN ON DIFFERENT MEDIA

C. J. Ehrhardt, V. Chu, T. Brown, T. Simmons, B. K. Swan, J. Bannan, J. M. Robertson

Abstract

The goal of this study was to determine if cellular fatty acid methyl ester (FAME) profiling could be used to distinguish among spore samples from a single species (*Bacillus cereus* T strain) that were prepared on 10 different medium formulations. To analyze profile differences and identify FAME biomarkers diagnostic for the chemical constituents in each sporulation medium, a variety of statistical techniques were used, including nonmetric multidimensional scaling (nMDS), analysis of similarities (ANOSIM), and discriminant function analysis (DFA). The results showed that one FAME biomarker, oleic acid (18:1 ω 9c), was exclusively associated with spores grown on Columbia agar supplemented with sheep blood and was indicative of blood supplements that were present in the sporulation medium. For spores grown in other formulations, multivariate comparisons across several FAME biomarkers were required to discern profile differences. Clustering patterns in nMDS plots and *R* values from ANOSIM revealed that dissimilarities among FAME profiles were most pronounced

when spores grown with disparate sources of complex additives or protein supplements were compared ($R > 0.8$), although other factors also contributed to FAME differences. DFA indicated that differentiation could be maximized with a targeted subset of FAME variables, and the relative contributions of branched FAME biomarkers to group dissimilarities changed when different media were compared. When taken together, these analyses indicate that *B. cereus* spore samples grown in different media can be resolved with FAME profiling and that this may be a useful technique for providing intelligence about the production methods of *Bacillus* organisms in a forensic investigation.

Introduction

In September 2001, letters containing spores of *Bacillus anthracis*, the causative agent for anthrax, were mailed to television and print media outlets, as well as two U.S. Congressional offices, in an act of bioterrorism. Genetic tests identified a single strain of *B. anthracis*, Ames, in all evidence samples. Since then, considerable effort has gone into developing techniques that can be used to analyze microbiological evidence recovered from a crime scene. Because the Ames strain used in the 2001 attacks was difficult to distinguish genetically from several commonly used Ames strains [1], many recently developed techniques have concentrated on nongenetic signatures associated with the cell that are unique to the methods that were used to culture an organism. Examples include assays that detect the presence of residual agar on spores [2], C/N isotope ranges for different medium components [3-5], and detection of heme in spores grown on blood-containing media [6]. Phenotypic signatures such as these that indicate specific metabolic substrates, characteristic compounds, or defining features of an organism's production

process could help in the attribution of a biocrime by providing leads or excluding suspects during an investigation [2, 7].

One phenotypic system that has not been fully tested in a forensic context is fatty acid methyl ester (FAME) analysis. FAMEs have long been recognized as useful biochemical markers for bacterial classification and characterization [8-10]. The types and relative abundances of fatty acids produced within a cell are largely determined by an organism's genotype and can be used for identification of different species [11] and strains [12, 13] and for discriminating between free spores and vegetative cells [14, 15]. Commercial systems that streamline fatty acid extraction and detection procedures [16] have facilitated the widespread use of fatty acid profiling to identify bacteria in clinical, agricultural, and biodefense settings [9, 17, 18].

Besides aiding in the identification of bacterial species, FAME profiling can potentially provide information on the methods used to grow microorganisms of forensic interest. Within the *Bacillus* group, the amino acid content or the type of complex additives used in the cultivation media can significantly affect the fatty acid composition of bacterial cultures. The relative proportions of branched fatty acids (iso-odd, iso-even, and anteiso), which are prevalent in *Bacillus* spp. [19, 20], are heavily dependent on the ratio of amino acid precursors (leucine, valine, and isoleucine) and the corresponding α -keto acids present in the growth media [21-24]. Accordingly, the complex additives and protein sources that supply these amino acid precursors in growth media also affect the fatty acid compositions of *Bacillus* cultures. For example, it has been reported that inclusion of components such as yeast extract, beef extract, or casein hydrolysate in growth medium formulations can change the relative ratios of iso and anteiso fatty acids

in *Bacillus cereus* cultures [25, 26]. Brain heart infusion (BHI) has been observed to have a similar effect on the relative proportions of branched fatty acids in *Bacillus caldolyticus* [27]. Despite the clear relationship between the fatty acid compositions of vegetative cells and the formulations of growth media, no study has tested whether this could be exploited for investigative purposes by determining whether diagnostic FAME signatures for growth media exist within spores of a forensically relevant organism.

To test whether FAME signatures can be used to infer the compositional characteristics of the sporulation medium, we examined fatty acid profiles among *Bacillus cereus* T-strain (*BcT*) spores grown on 10 different media, spanning nutrient formulations that varied primarily in the source of protein, either in the form of complex additives (yeast extract, beef extracts, brain-heart solids, etc.) or direct protein supplements (peptone, tryptone, or gelatin digest). Formulation pairs that differed in other medium attributes, such as the presence of supplemental sugars, the physical state (agar versus broth), or blood supplements, were included to compare the resulting FAME profile differences with the variation that derives from complex additive/protein components.

The effects that protein components in the sporulation medium have on FAME profiles were specifically targeted in the experimental design because of the direct biosynthetic relationship between amino acids and the three structure classes of branched fatty acids in *Bacillus* [22, 25]. In addition, there are a limited number of common or commercially available complex additive and protein sources used for microbiological media. Identifying forensic signatures based on a reduced number of defining

components rather than the myriad of possible medium formulations makes comprehensive surveys feasible, an added advantage for any potential forensic marker.

To further frame our study in the context of forensic or investigative applications, we chose *B. cereus* as a target organism because of its genetic, structural, and biochemical similarities to *B. anthracis* [28-30]. Also, since evidence from the 2001 anthrax mailings was predominantly composed of *Bacillus* spores [31], we used spore preparations of *B. cereus* for all FAME analyses.

Lastly, analysis of forensic signatures from fatty acid profiles is complicated by the large number of variables (typically >15) and complex interactions among different fatty acid structures during cellular biosynthesis [23]. Therefore, FAME profiles were analyzed with orthogonal multivariate statistical techniques that first considered all variables simultaneously and analyzed the overall dissimilarities among spore FAME profiles and, second, maximized differentiation among groups using a subset of variables and extracted patterns in fatty acid differences that are diagnostic for specific medium formulations.

Materials and Methods

Spore preparation

Cultures of *BcT* were maintained at 30°C on tryptose beef extract agar (TBA) (10 g Bacto tryptose [Becton Dickinson DF0124-17], 5g NaCl, 3 g Bacto beef extract [Becton Dickinson DF0115-17], 16 g agar [Becton Dickinson DF0140-01]). Starter cultures were grown by inoculating single colonies of *BcT* into 125 ml of tryptic soy broth (30 g Trypticase soy broth [Becton Dickinson 211771], 15 g agar [Becton Dickinson 211849])

and incubated for 16 to 18 h at 30°C and 300 rpm. The following sporulation media were used: a modified chemically defined sporulation medium (CDSM) identical to a previous formulation [32] but omitting 50 µg/ml isoleucine-tryptophan, G medium [33, 34], modified G medium (MG) (identical to G medium but excluding glucose [35], CAD medium (8.5 g tryptone, 5.0 g yeast extract, 5.0 g dibasic potassium phosphate, 0.75 g dextrose, 0.025 g thiamine-HCl, 25 g CaCl₂·2H₂O, 5 g FeSO₄·7H₂O, 25 g MgSO₄·7H₂O, 15 g MnSO₄·H₂O), brain heart infusion agar (BHI; Becton Dickinson 211059), Schaeffer's sporulation medium in both liquid and agar form (Sch and SchAg) [36], LL agar (23 g Lab Lemco agar [Fisher Scientific OXCM00115B]), NSM agar [37], Columbia agar (CA) (Becton Dickinson 279240), and Columbia agar with blood (CAB) (22 g Columbia agar base, 2.5% defibrinated sheep blood [Fisher Scientific 212390]). All weights are per liter of nanopure water (18 MΩ). A summary of the sporulation media used in this study is given in Table 1.

For spores grown in broth-based media, 20 ml of starter culture was inoculated into 250 ml of sporulation medium and incubated at 30°C and 300 rpm in an orbital shaker (VWR, PA). The cultures were monitored throughout the incubation period and harvested when the proportion of spores in the cell population reached ~90%. Incubation times varied between medium formulations but ranged between 24 and 48 h. For spores grown on solid media, a loopful of *BcT* colonies from TBA plates was mixed in 10 ml of nanopure H₂O. From this solution, 100 µl was spread on 16 agar plates for each medium type and incubated for 72 to 120 h, depending upon the medium. The sporulation yield was monitored as before and harvested when the yield reached ~90%.

Spores grown in liquid media were harvested by centrifugation at 6,000 x g for 15 min. For spores grown on agar plates, colonies were scraped off with an inoculating loop and placed in 50 ml of cold (4°C) H₂O. Once harvested, all spores were washed three times in 100 ml of sterile H₂O, including a final wash at 4°C on a rocking platform overnight. Following the overnight wash, spore preparations were usually free of any cellular debris or vegetative cells (>98% phase-bright spores). The spores were then washed two more times in 50 ml of H₂O and stored at -20°C until analysis was performed. Additional purification steps, including lysozyme treatment and density centrifugation, did not noticeably increase the spore yield. At least three replicate batches were grown for all media.

Gas chromatography and fatty acid profiling

Prior to chromatographic analysis, 6 to 10 mg of wet spore preparation was thawed and immediately distributed into separate 2-ml glass screw cap tubes (Fisher Scientific NC9174732). Samples were then vacuum dried at room temperature. The dried spore weight in each tube was typically 1 to 3 mg.

Fatty acid extraction and methyl ester generation were performed with the Instant FAME Method kit (part no. 7000 and 7020; MIDI, Inc., Newark, DE) according to the manufacturer's instructions. This protocol was chosen because it requires minimal starting material (~1 mg) and nominal processing time per sample (<15 min) and has been previously validated for other applications [38], all of which are advantageous features of a potential laboratory- or field-based forensic technique [7, 39]. Briefly, the spore materials were first mixed with 250 µl of a KOH-MeOH solution. This step lysed the cells (data not shown) and replaced any polar head group on fatty acids with a methyl

group, creating a FAME. The FAMEs were then partitioned into an organic phase with 200 µl of hexane. Approximately 150 µl of the hexane phase was transferred to a new screw cap vial and loaded onto an HP6890 gas chromatographic (GC) analyzer for FAME profiling. Two replicate GC samples from each spore preparation were processed and included in subsequent analyses to incorporate profile differences arising from variability between different fatty acid extractions and GC runs.

GC FAME profiling was performed with the MIDI Microbial Identification Sherlock software according to the manufacturer's instructions with one exception. In order to increase the sensitivity for the detection of fatty acids in low concentrations, the split ratio on the GC instrument was changed to 40:1 with a split flow of 50.1 ml/min. MIDI calibration standards were used (part no. 1300-AA) for identification and quantification of fatty acid peaks.

Statistical analysis

All fatty acid structures recognized by the MIDI Sherlock software that were common to every profile within a particular medium type were defined as the fatty acid variables in statistical analyses and are listed in Table 2. Values for all variables represent the relative percent contribution of individual fatty acids to the GC profile and are shown in Table 3. Mean and standard deviation values were calculated from the six replicate profiles generated for each spore group (Table 3). For certain comparisons among spore samples, fatty acid variables were grouped into four categories (columns in Table 2 labeled "branched-odd," "branched-even," "anteiso," and "normal") that correspond to the biochemical structure classes and associated biosynthetic pathways in *Bacillus* spp. previously described [25]. These categorical variables were used because the relative

abundance of each of these structure classes is directly affected by the concentrations of amino acid precursors [22, 23] that are likely to be specific for each growth medium.

nMDS and ANOSIM

Multivariate differences in FAME profiles for all spore samples were analyzed by nonmetric multidimensional scaling (nMDS) and analysis of similarities (ANOSIM). Both techniques were performed in Primer v6.0 (Primer-E Ltd., Plymouth, United Kingdom [2004]). The Bray-Curtis similarity measure was used to generate the distance matrix for nMDS. Kruskal fit scheme—option 1, along with 50 restarts, was used for the two-dimensional (2D) plot. Spore media groups (defined in Table 1) spanning both agar and broth forms were compared in the nMDS and ANOSIM analyses. Spores grown on the CDSM formulation were excluded, since this modified recipe is not commonly used for growing/sporulating *Bacillus* organisms.

Discriminant function analysis

Discriminant function analysis (DFA) was used to assess the relative contributions of individual fatty acid biomarkers to the multivariate patterns observed among medium groups and to isolate a subset of variables that would maximize sample group discrimination. DFA was performed in SPSS v.16.0 (SPSS Inc., Chicago, IL [2008]). In contrast to nonparametric techniques, such as nMDS or ANOSIM, the effectiveness of DFA can be influenced by the characteristics of the input variables. Ideally, variables in DFA are normally distributed, have homogeneous variances, and are uncorrelated with each other [40]. Therefore, a subset of FAME markers that minimized these data structure effects was chosen. Variables were first excluded based on the magnitude of correlation with every other variable. Correlations were determined with Pearson's correlation

coefficient (r), which was calculated for all FAME variable pairs. Those variables with high correlations ($|r| > \sim 0.650$) were grouped into categories. Only one FAME variable from each category was included in subsequent analyses so that all remaining variables had approximately equivalent standard deviations. Lastly, uncorrelated variables were subjected to a preliminary round of DFA to identify the FAME biomarkers with the greatest discriminatory power for medium formulations with disparate protein sources. This was judged by calculating an "F-to-remove" statistic for each FAME variable [40, 41]. The eight FAME variables with the largest F-to-remove values and therefore the highest contributions to group separations were included in the final analysis. (See Fig. 4 for the variable subset used for DFA.)

Results

Sporulation yields in different medium formulations

All *BcT* cultures showed sporulation yields greater than 90%. Cultures grown in broth generally showed higher spore yields before water washing/purification (>95%) than cultures grown on agar plates (generally between 90 and 95%). Despite this discrepancy, the washing and purification steps increased the proportion of phase-bright spores in all preparations to above 95% (Fig. 1). Spores grown on Columbia agar and Columbia agar supplemented with blood showed the highest proportion (5 to 10%) of vegetative cells prior to and immediately following the first three water washes. The proportion of vegetative cells was minimized after subsequent water washes (<3%).

Effect of the sporulation medium on the fatty acid composition of *BcT* spores

To investigate the effects of amino acids in the sporulation media on the fatty acid composition of spores, *BcT* cultures were grown in a modified CDSM broth that lacked an exogenous amino acid source compared to other sporulation media that contained either defined or complex protein sources (Table 1). The relative proportions of the four fatty acid structure classes in spores grown in each medium are shown in Fig. 2. While variation existed across each medium preparation for every fatty acid, the largest differences were observed with spores grown in CDSM. The relative proportions of branched-odd fatty acids for spores grown in this medium were generally half of the percentages observed in other sporulation media (Fig. 2a). Conversely, the overall proportion of branched-even fatty acids typically was doubled in CDSM samples compared to spores grown in other media (Fig. 2b). CDSM spore profiles also showed a slight increase in the ratio of normal fatty acids (Fig. 2d). Although the relative percentages of anteiso fatty acids varied considerably between different batches of CDSM spores (between 7 and 16%), the range showed significant overlap with the proportions obtained for other media, including G, LL, and Sch (Fig. 2c).

Multivariate and individual biomarkers in *BcT* FAME profiles

Comparison among cultures grown in media with different types and combinations of complex additives showed more subtle variations in the relative proportions of fatty acids. For example, medium groups containing higher concentrations of tryptone or gelatin digests, such as BHI, CAD, and CA, generally had elevated levels of branched-odd fatty acids (~58 to 63%) compared to certain medium formulations with meat peptone as the sole nitrogen source (i.e., Sch and LL, with ~44% branched-odd) (Fig. 2a and Table 3). In addition, spores grown in G, MG, LL, and Sch media had

slightly elevated proportions of anteiso fatty acids, ranging between 10 and 11% compared to other medium groups (BHI, CAD, and CAB), with anteiso fatty acids ranging between ~5 and 7% (Fig. 2c and Table 3). However, the trends in anteiso FAMES did not directly correlate with the presence of any particular constituents in the sporulation medium.

In addition to differences in the relative proportions of FAME markers, one FAME marker was associated uniquely with one medium. Oleic acid (18:1 ω 9c) was present in appreciable quantities (>1%) only when spores were grown on Columbia agar supplemented with sheep blood (CAB). The concentration varied between 5 and 20% of the total fatty acid composition. Oleic acid was not evident in spores grown on Columbia agar (Table 3). The results from GC-mass spectrometry analysis were consistent with the presence of oleic acid in CAB spore extracts (data not shown).

Multivariate differentiation of *BcT* spore FAME profiles

FAME profiles are complex due to the high number of variables and the fact that the total dissimilarity among profiles may arise from unequal contributions from all the variables. Typically, data from every detected FAME marker are used to generate multivariate distance matrices [15] or functions representing linear combinations of the original variables [42, 43], which are used to analyze dissimilarities among all profiles.

For this study, FAME data were first analyzed in a similar manner by incorporating the relative proportion of every FAME marker identified by GC into the multivariate distances that were calculated for spore profiles. The distances were then analyzed with nMDS and ANOSIM to examine the relationships among spores grown on different medium types and to test for significance between spore-medium groups. These

two nonparametric techniques made no assumptions about data structure or sample groupings prior to analysis. In this way, they provided an unbiased snapshot of the total dissimilarity in the whole FAME composition of each spore sample.

More specifically, nMDS offers a mathematically robust yet conceptually simple method for visualizing multivariate relationships between individual samples and among groups of samples [44]. In a two-dimensional nMDS plot, distances between points reflect the relative rank order dissimilarities among all samples. The spatial arrangement of sample points and the resulting distances among sample groups, therefore, can be interpreted as a measure of the relative dissimilarity in FAME profiles among spores grown in different media. In addition, the goodness of fit between the underlying distance matrix and the observed spatial arrangement of sample points in each nMDS plot is measured with a "stress" value [45].

Figure 3 shows the nMDS plot for spores grown on 10 sporulation media (Table 1, excluding CDSM). Spores grown on media with disparate chemical compositions exhibited well-defined clusters with high intergroup distances with respect to each other, suggesting significant multivariate dissimilarity. These media included LL, G, Sch, CAD, NSM, and BHI. Compositionally similar media, such as G and MG (Table 1), showed closely associated (but separated) clusters with each other. Noncompositional features also appeared to affect the FAME profiles of *BcT* spores. For example, SchAg samples (agar based) did not show a close association with Sch samples (broth based), exhibited high intragroup variation, and overlapped with other medium groups in the center of the nMDS plot (Fig. 3). Similarly, spores grown on CA and CAB showed higher intragroup variation and, in the case of CA, completely overlapped with BHI samples. The stress

value reported in Fig. 3 (0.11) indicates that the nMDS plot is an accurate representation of intersample relationships [45].

To complement the descriptive assessments of multivariate dissimilarities provided by nMDS, ANOSIM was used to test for the statistical significance of grouping spore samples according to the medium formulation. The ANOSIM test statistic, R , is a measure of the magnitude of dissimilarity within and between sample groups and ranges between 0 and 1 [45, 46]. R values near 0 indicate that dissimilarities between sample points within one group are equivalent to the dissimilarities found between different groups. R values closer to 1 indicate strong differences between two groups relative to intragroup variation among FAME profiles. Statistical significances in the form of P values were calculated for each pairwise R value.

The results from the ANOSIM analysis for individual FAME markers were consistent with the spatial trends observed in the nMDS plot (Table 4 versus Figure 3). The R values for most pairwise medium group comparisons showed P values of less than 0.01, indicating that all medium groups had statistically significant dissimilarities in their composite FAME profiles (BHI-CA was the one exception). However, the magnitudes of the R values did show variation across the medium groups. SchAg samples exhibited the lowest R values (<0.8) with all other medium groups, excluding SchAg-LL and SchAg-CAB ($R \geq 0.8$ for both). The R values for G-MG and BHI-CA were also relatively low: 0.39 and 0.30, respectively.

Differentiation of FAME profiles using DFA

While nMDS and ANOSIM are useful for detecting differences among spore profiles, these techniques do not provide criteria for discriminating between different

spore samples. DFA can complement these analyses by identifying individual FAME biomarkers that differentiate spores grown in different medium formulations. In DFA, linear functions (i.e., "discriminant functions," or "canonical variate [CV] functions") that maximize the variation among user-defined sample groups and minimize the variation within each sample group are derived from the original variables. The relative contribution of each variable to the observed group differentiation is proportional to the absolute values of the discriminant function coefficients [40].

The main advantage of DFA for forensic studies is that it incorporates prior data structure into the analysis and allows the user to define sample groupings so that different levels of variation in the data set can be examined. In this way, differences among FAME profiles that may be due to experimental error among replicates or small variations in medium chemistry can be minimized and changes in FAME variables that are driven by distinguishing characteristics of the sporulation medium (e.g., complex additives, supplemental protein, or unusual compounds) can be identified. For example, Sch and SchAg have identical chemical compositions but differ in the physical state (broth and agar, respectively), which results in non-compositionally based variation in FAME profiles (Fig. 3 and Table 4). If all samples grown in either Sch or SchAg are defined in one group, then DFA will generate linear functions that minimize variation between the Sch and SchAg samples and maximize variation between these and other sporulation medium groups. The variables with the largest coefficients in these functions will then represent FAME biomarkers reflecting compositional differences, rather than the total dissimilarity between Sch/SchAg and other media.

To identify FAME variables that are diagnostic for differences in protein content among sporulation media, discriminant functions were built to analyze eight composite medium groups reflecting different complex protein sources. Six of these groups were identical to the sample groups used for nMDS. However, spore samples for MG were grouped with G samples, since both represent media with yeast as the sole protein source (Table 1). Similarly, Sch and SchAg samples were combined into one group. The equations for the first two discriminant functions (or CVs), along with a 2D plot of the distribution of each sample in CV space, are shown in Fig. 4a.

Samples belonging to the LL, G, and CAB groups are clearly differentiated from each other and from the remaining groups in the center of the plot (CA, CAD, NSM, BHI, and Sch). Separation between CAB and G occurs primarily along CV1, whereas the distances between the G and LL groups are found along CV2. The CAB and LL groups are separated along both CV1 and CV2. The coefficients for each CV function (on right of the plots) indicate that different sets of variables drive separation along each axis. CV1 is dominated by the relative proportions of 15:0 anteiso, 17:1 iso ω 10c, and 17:1 anteisoA, and CV2 is dominated by the relative proportions of 17:1 iso ω 10c and 17:1 iso ω 5c.

Because little differentiation was observed among BHI, CA, CAD, NSM, and Sch samples (Fig. 4a), new discriminant functions were constructed using only these groups (Fig. 4b). Similar to the previous plot, two of the sample groups (CAD and Sch) can be easily distinguished from each other along CV1. However, the variables with the largest contribution to the differentiation among these groups are 14:0 iso, 16:0 iso, 17:1 iso ω 10c, and 17:1 iso ω 5c. The remaining groups (CA, BHI, and NSM) cannot be clearly

discriminated from each other along CV1 or CV2 but are separated from CAD and Sch by both CV functions.

When discriminant functions are built for the last three groups (BHI, CA, and NSM) (Fig. 4c), CV1 accounts for nearly all the variation among groups (~90%) and clear separation in the CV plot is observed for every group. Unlike the previous functions, only branched-odd variables, 13:0 iso, 17:1 iso ω 10c, and 17:1 iso ω 5c, account for the variation along CV1.

Discussion

The influence of exogenous protein sources in the sporulation medium on FAME profiles

The relationship between the ratio of amino acid precursors in the culture medium and the proportion of each fatty acid structure in vegetative *Bacillus* cells is well established [22, 23, 25]. The fact that spores grown in CDSM showed obvious changes in the proportions of branched-odd, branched-even, and, to a lesser extent, normal fatty acids compared to spores grown in other media surveyed in this study indicates that amino acid precursors (specifically, leucine and valine) are utilized directly from the sporulation medium. Consequently, the absence of a direct protein source or any complex additive in the formulation can have a significant effect on the types and abundances of branched FAME markers associated with *Bacillus* spores.

Less drastic variation in branched fatty acid composition was also evident among non-CDSM groups, indicating that FAME profiles are also affected by different types of protein sources and complex additives used in the sporulation medium. However, with the

exception of spores grown on meat peptone-based formulations showing decreased branched-odd proportions, FAME abundances did not show clearly defined ranges that corresponded to specific chemical constituents of any growth medium (Table 1). Also, transitions in the relative abundances of FAME biomarkers were small and incremental among the different medium preparations (Fig. 2 and Table 3). Such patterns have been observed previously with bacteria grown on different media [42], suggesting that forensic signatures based on FAME profiles depend on more complex multivariate relationships between variables and samples of interest.

Oleic acid as a biomarker for media with blood supplements.

The only fatty acid biomarker exclusive to any of the surveyed media was oleic acid (18:1 ω 9c) from spores grown on Columbia agar supplemented with sheep blood. Oleic acid is predominantly associated with eukaryotic organisms [47] but can be introduced exogenously through blood supplements [48] or surfactants, such as Tween 80, during the preparation process [49]. Also, oleic acid is a common feature in commercial FAME libraries that contain organisms grown on blood-supplemented substrates (G. Jackoway, MIDI Inc., personal communication). This, combined with the observation that oleic acid did not appear in significant quantities in *BcT* spores grown on the Columbia agar base (CA) (Table 3) but was present when spores were grown on other media containing blood products (tryptic soy agar with blood) (data not shown) suggests that this fatty acid is likely derived from the eukaryotic supplements in the CAB medium. Regardless of the origin, our results indicate that oleic acid may be a promising biomarker for *Bacillus* cultures grown on media that are supplemented with sheep blood products.

Differentiation among medium groups using whole FAME profiles.

Fatty acid analyses based on all FAME biomarkers showed that many of the 10 medium cultures in this study could be differentiated by their FAME profiles on the 2D nMDS plot (Fig. 3) and with pairwise *R* values (Table 4). The largest dissimilarities in FAME profiles were found among spores grown on media with distinctly disparate protein and nitrogen sources (yeast, meat peptone, yeast/casein peptone, brain heart infusion/gelatin digest, and beef extract/meat peptone for G, LL, CAD, BHI, and Sch, respectively). The dissimilarities in FAME profiles likely reflect the distinct differences in fatty acid precursors (amino acids and α -keto acids) inherent in each of the above-mentioned medium formulations.

Other media (BHI, CA, CAB, and SchAg) exhibited less distinct differences, as evidenced by cluster overlap (BHI-CA) or large intersample distances (CA, CAB, and SchAg). Overlapping sample groups could, in part, be indicative of the protein/amino acid composition of the sporulation medium. Both BHI and CA contain a variety of meat digest and beef infusion components (Table 1). Since different beef-derived supplements can have comparable ratios of leucine and isoleucine [50], the similarity in BHI and CA samples may reflect overlapping concentrations of these fatty acid precursors in each medium.

Variation in FAME profiles was also observed between spores grown on media with identical substrate compositions but different physical states (Sch and SchAg). FAME differences between *Bacillus* cultures grown in agar and broth-based media with identical compositions have been observed previously [30, 51] and could be related to microenvironments that are created during sporulation on agar media. In *Bacillus*

organisms, the synthesis of unsaturated fatty acids is mediated by desaturase enzymes that are oxygen dependent [30, 52].

The relative proportions of unsaturated fatty acid markers, such as 16:1 ω 7c, which is abundant in *BcT* FAME profiles (Table 3), is primarily affected by the concentration of the saturated fatty acid precursor (16:0) and oxygen availability [30]. During growth on agar plates, the oxygen concentration may show spatial heterogeneities within *BcT* colonies that can affect the proportions of saturated and unsaturated fatty acid spore profiles compared to organisms grown in liquid media [30, 53]. Similarly, metabolic substrates can show heterogeneities within agar colonies [54]. This hypothesis would be consistent with shifts in the observed proportions of 16:1 ω 7c (SumFeat2), 16:0 iso, and 15:0 iso for Sch and SchAg samples (Table 3). Consequently, statistical analyses (nMDS) based on all individual fatty acids would be prone to these variations, since each marker contributes independently to the calculated dissimilarity among samples. The observed discrepancies in FAME profiles between Sch and SchAg samples indicate that forensic discrimination of spores should incorporate different physical states of the medium.

Similarities and differences in medium composition or physical state would not explain why the SchAg, CA, and CAB groups had larger dissimilarities among replicate samples. This FAME profile heterogeneity displayed within CA and CAB could reflect slight variation in the proportion of vegetative cells in these spore preparations (see Results). Alternatively, the period of incubation prior to spore harvesting varied more for CA, CAB, and SchAg than for broth-based media, which would affect the number of metabolic conversions and potentially the relative proportions of FAMEs. However, this

idea does not hold for all agar groups, since BHI and LL did not show comparable levels of intra-group variation.

Other medium pairs, mainly G-MG, showed distinct but closely related clusters on the nMDS plot and smaller *R* values for ANOSIM. Some dissimilarity between these two groups was expected, since the presence of different sugars in growth media has been found to affect the fatty acid profiles of *Bacillus* organisms [25, 27, 55, 56]. Less dissimilarity between G and MG spore profiles suggests that small variations in supplemental sugar concentrations in the sporulation medium do not affect FAME profiles as significantly as other changes in the growth medium formulation.

Ultimately, these results suggest that using raw calculations of dissimilarity generated from every variable constituting a FAME profile is insufficient to differentiate all spore groups. While the analysis does indicate that FAME differences are most pronounced when spores are prepared on media with different types of complex additives and nonoverlapping protein sources, it also shows that total profile dissimilarities can be affected by variations in other, nondefining attributes of the medium or by the intrinsic variability of certain spore-medium groups. To compensate for these effects, profiles need to be analyzed with a statistical technique that can minimize variation that is nonspecific to the protein/amino acid content of each medium and can extract signatures that are unique to its defining compositional characteristics.

Differentiation of FAME profiles using DFA.

The FAME biomarkers that responded most significantly to variation in the protein source were isolated by combining all spore samples that were grown in media with similar protein components and analyzing the resulting groups with DFA. A

comparison of the clustering patterns in Fig. 3 and 4 indicates that non-compositionally based variation is minimized with DFA. For example, all Sch samples (composites of Sch and SchAg) showed a well-defined cluster with low intersample variation (Fig. 4a and b). Other groups exhibiting heterogeneities among replicates in nMDS (CA and CAB) displayed smaller intragroup variation in CV plots that were comparable to that of other spore groups (Fig. 4a and c). With non-compositionally based variation reduced, canonical variate functions and the corresponding high-magnitude FAME variables can be used to identify promising biomarkers for spore discrimination.

Figure 4a shows that CV1, which was heavily influenced by the proportions of 15:0 anteiso and 17:1 anteisoA, clearly separated G and CAB samples from other media. However, CV1, and therefore these two FAME biomarkers, could not clearly distinguish G samples from LL, or any samples grown in BHI, NSM, Sch, CAD, or CA from each other. Different FAME biomarkers were responsible for differentiating each of these groups. Branched-even FAMES 14:0 iso and 16:0 iso and branched-odd FAMES 17:1 ω 5c and 17:1 ω 10c drove the separation for CAD and Sch groups (Fig. 4b). However, only branched-odd FAMES were significant contributors to the CV functions differentiating NSM, BHI, and CA groups (Fig. 4c).

These results suggest that FAME variation among all spore groups cannot be captured with a single set of equations but requires multiple, successive CV functions to discriminate among all spore groups. Also, the varying contributions of individual FAMES to CV-based differentiation indicate that the same set of FAME variables cannot differentiate all spore groups simultaneously. For example, 15:0 anteiso and 17:1 anteisoA distinguished G-medium samples from other groups but poorly discriminated

Sch and CAD from other samples. Conceptually, this is reasonable, since medium formulations may show variation across different sets of amino acid precursors, and spores grown on these media should vary across different sets of FAME variables.

Due to these observations, DFA-based discrimination of spores grown on different media may necessitate a hierarchically structured analysis similar to that portrayed in Fig. 4. In such a system, FAME profiles would be subjected to a cascading system of discriminant functions that model variation across separate subsets of reference groups [57]. Taking Fig. 4 as an example, the FAME profile of an unknown microbial sample would first be run through an initial set of discriminant functions (Fig. 4a) and classified as belonging to either G, CAB, LL, or the unresolved group composed of BHI, CA, Sch, NSM, and CAD. If the sample was most closely related to the unresolved group, the FAME profile would be analyzed with the second set of discriminant functions (Fig. 4b), classifying it as either Sch, BHI, or the unresolved group composed of BHI, CA, and NSM. If the sample again was most closely related to the unresolved group, subsequent discriminant functions (Fig. 4c) would be used until a singular identification was acquired. Tiered classification systems with DFA have been used successfully in other forensic systems [57, 58] and could be applied to microbial samples provided that the discriminant functions are built from a comprehensive reference database (in this case, a library of FAME profiles from spores grown in various media).

Conclusions and future work.

Overall, the results indicate that variations in *BcT* FAME profiles can be used to discriminate among spores grown on different media. While individual biomarkers, such as oleic acid, may be diagnostic for certain components of the growth medium, the

majority of variation across FAME profiles is driven by differences in the protein or amino acid sources in each sporulation medium. Other characteristics of the medium, such as the physical state or the concentration of glucose, do affect FAME profiles to an extent. Nevertheless, these profile variations can be reduced with DFA so that the FAME signatures that are specific for formulations with unique combinations of complex additives or supplemental protein sources can be detected.

Although this is an important first step for determining whether FAME profiling can be a forensic tool, much work remains before the observed differences among cultures grown on different media can be translated into forensically relevant biosignatures. First, the hypothesis that FAME variation is diagnostic for the combination of complex additives and protein sources that are present in each medium formulation should be explicitly tested by expanding a FAME data set to include formulations that vary in the concentrations of identical protein sources. Second, the effects that environmental factors, such as growth temperature, pH, or dissolved oxygen, have on FAME profiles should be compared to the compositionally based variation reported here. Third, different subsets of FAME variables should be identified with alternative statistical techniques, such as stepwise DFA [41] or Bayesian variable selection [59], and compared to the discriminatory power of the FAME subset used in this study.

Most significantly, multivariate strategies that allow sample profiles to be matched statistically to FAME databases are needed. The discriminant function analyses shown here are descriptive in nature [40] and are not intended to be robust classification schemes. A separate statistical procedure termed "predictive DFA" can be used for this purpose and is a promising strategy for classification of unknown samples [41]. However,

this technique still needs to be tested on a microbial data set. Another possibility is "Bayesian network analysis" [57], which would allow FAME databases to be combined with other orthogonal data sets (isotope, SIMS, etc.) to characterize a microbial sample of unknown origin. Future strategies may need to incorporate a combination of the above-mentioned approaches in order to successfully reduce the intrinsic complexity of FAME profiles and to help this technique become a viable tool in forensic microbiology.

Acknowledgements

We gratefully acknowledge the laboratory contributions of Matt Ducote and Mark Reimers. We also thank Keith Monson, Robert Koons, Bruce Budowle, Robert Bull, Ulrich Melcher, and Lilliana Moreno for their discussions and informal reviews of the manuscript.

This research was supported in part by the Visiting Scientist Program, an educational opportunity administered by the Oak Ridge Institute for Science and Education.

The names of commercial manufacturers are provided for information only and do not imply endorsement by the Federal Bureau of Investigation (FBI) or the U.S. Government. The conclusions are those of the authors and should not be taken as necessarily representing the views, either expressed or implied, of the U.S. Government. This is publication no. 09-08 of the Laboratory Division of the FBI.

LITERATURE CITED

1. Hoffmaster, A., E. Fitzgerald, and L. Mayer, *Molecular subtyping of Bacillus anthracis and the 2001 bioterrorism-associated anthrax outbreak, United States*. Emerging and Infectious Diseases, 2002. 8: p. 1111-1116.
2. Wunschel, D.S., et al., *Detection of agar, by analysis of sugar markers, associated with Bacillus anthracis spores, after culture*. Journal of Microbiological Methods, 2008. 74: p. 57-63.
3. Jarman, K.H., et al., *Bayesian-integrated microbial forensics*. Applied and Environmental Microbiology, 2008. 74: p. 3573-3582.
4. Kreuzer-Martin, H.W., et al., *Stable isotope ratios as a tool in microbial forensics, 2. Isotopic variations among different growth media as a tool for sourcing origins of bacterial cells or spores*. Journal of Forensic Sciences, 2004. 49: p. 961-967.
5. Kreuzer-Martin, H.W., et al., *Microbe forensics: oxygen and hydrogen stable isotope ratios in Bacillus subtilis cells and spores*. Proceedings of the National Academy of Sciences, 2003. 100: p. 815-819.
6. Whiteaker, J. and C. Fenselau, *Quantitative determination of heme for forensic characterization of Bacillus spores using matrix assisted laser desorption/ionization time-of-flight mass spectrometry*. Analytical Chemistry, 2004. 76: p. 2836-2841.
7. Robertson, J., C.J. Ehrhardt, and J. Bannan, *Lipids in microbial forensics*, in *Microbial forensics*, J. Cliff, Editor. In press, Humana Press: Totowa, NJ.

8. Ratledge, C. and S.G. Wilkinson, *Microbial lipids, vol. 1*. 1988, New York, NY: Academic Press, Inc.
9. Welch, D.F., *Applications of cellular fatty acid analysis*. *Clinical Microbiology Reviews*, 1991. 4: p. 422-438.
10. Weyent, R.S., et al., *Bacterial identification by cellular fatty acid analysis*, in *Identification of unusual Gram-negative aerobic and facultatively anaerobic bacteria, 2nd ed.* . 1996, Williams & Wilkins: Baltimore, MD. p. 565-721.
11. Vandamme, P., et al., *Polyphasic taxonomy, a consensus approach to bacterial systematics*. *Microbiological Reviews*, 1996. 60: p. 407.
12. Leonard, R.B., et al., *Comparison of MIDI Sherlock system and pulsed-field gel electrophoresis in characterizing strains of methicillin-resistant Staphylococcus aureus from a recent hospital outbreak* *Journal of Clinical Microbiology*, 1995. 33: p. 2723-2727.
13. Wintzingerode, F.V., et al., *Identification of environmental strains of Bacillus mycooides by fatty acid analysis and species-specific rDNA oligonucleotide probe*. *FEMS Microbiology Ecology*, 1997. 24: p. 201-209.
14. Scandella, C.J. and A. Kornberg, *Biochemical studies of bacterial sporulation and germination XV. Fatty acids in growth, sporulation, and germination of Bacillus megaterium*. *Journal of Bacteriology*, 1969. 98: p. 82-86.
15. Song, Y., et al., *Distinctness of spore and vegetative cellular fatty acid profiles of some aerobic endospore-forming bacilli*. *Journal of Microbiological Methods*, 2000. 39: p. 225-241.

16. Sasser, M., *Identification of bacteria by gas chromatography of cellular fatty acids. MIDI technical note*. 1990: Newark, DE.
17. Kim, W.Y., et al., *Analysis of cellular fatty acid methyl esters (FAMEs) for the identification of Bacillus anthracis*. Journal of the Korean Society for Microbiology, 2001. 35: p. 31-40.
18. White, D.C., et al., *Flash detection/identification of pathogens, bacterial spores and bioterrorism agent biomarkers from clinical and environmental matrices*. Journal of Microbiological Methods, 2002. 48: p. 139-147.
19. Lechevalier, M.P., *Lipids in bacterial taxonomy*. Critical Reviews in Microbiology, 1977. 5: p. 109-210.
20. O'Leary, W.M. and S.G. Wilkinson, *Gram-positive bacteria*, in *Microbial lipids, vol 1*, C. Ratledge and S.G. Wilkinson, Editors. 1988, Academic Press: New York, NY. p. 155-159.
21. Daron, H.H., *Nutritional alteration of the fatty acid composition of a thermophilic Bacillus species*. Journal of Bacteriology, 1973. 116: p. 1096-1099.
22. Kaneda, T., *Fatty acids of the genus Bacillus: an example of branched-chain preference*. Bacteriology Reviews, 1977. 41: p. 391-418.
23. Kaneda, T., *Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance*. Microbiological Reviews, 1991. 55: p. 288-302.
24. Lawrence, D.S., S. Heitefuss, and H.S. Seifert, *Differentiation of Bacillus anthracis from Bacillus cereus by gas chromatographic whole-cell fatty acid analysis*. Journal of Clinical Microbiology, 1991. 29: p. 1508-1512.

25. Kaneda, T., *Factors affecting the relative ratio of fatty acids in Bacillus cereus*. Canadian Journal of Microbiology, 1971. 17: p. 269-275.
26. Kaneda, T., *Fatty acids in the genus Bacillus II. Similarity in the fatty acid compositions of Bacillus thuringiensis, Bacillus anthracis, and Bacillus cereus*. Journal of Bacteriology, 1967. 95: p. 2210-2216.
27. Weerkamp, A. and W. Heinen, *The effect of nutrients and precursors on the fatty acid composition of two thermophilic bacteria*. Archives of Microbiology, 1972. 81: p. 350-360.
28. Carrera, M., et al., *Differences between the spore sized of Bacillus anthracis and other Bacillus species*. Journal of Applied Microbiology, 2007. 102: p. 303-312.
29. Helgason, E., et al., *Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis-one species on the basis of genetic evidence*. Applied and Environmental Microbiology, 2000. 66: p. 2627-2630.
30. Kaneda, T., *Fatty acids in the genus Bacillus II. Similarity in the fatty acid compositions of Bacillus thuringiensis, Bacillus anthracis, and Bacillus cereus*. Journal of Bacteriology, 1968. 95: p. 2210-2216.
31. Beecher, D.J., *Forensic application of microbiological culture analysis to identify mail intentionally contaminated with Bacillus anthracis spores*. Applied and Environmental Microbiology, 2006. 72: p. 5304-5310.
32. Hageman, J.H., et al., *Single, chemically defined sporulation medium for Bacillus subtilis: growth, sporulation, and extracellular protease production*. Journal of Bacteriology, 1984. 160: p. 438-441.

33. Church, B.D., H. Halvorson, and H.O. Halvorson, *Studies on spore germination: its independence from alanine racemase activity*. *Journal of Bacteriology*, 1954. 68: p. 393-399.
34. Hashimoto, T., S.H. Black, and P. Gerhardt, *Development of fine structure, thermostability, and dipicolinate during sporogenesis in bacillus*. *Canadian Journal of Microbiology*, 1960. 6: p. 203.
35. Aronson, A.I., N. Angelo, and S.C. Holt, *Regulation of extracellular protease production in Bacillus cereus T: Characterization of mutants producing altered amounts of protease*. *Journal of Bacteriology*, 1971. 106(3): p. 1016-1025.
36. Schaeffer, P., J. Millet, and J.P. Aubert, *Catabolic repression of bacterial sporulation*. *Proceedings of the National Academy of Sciences*, 1965. 54: p. 704-711.
37. Cliff, J.B., et al., *Differentiation of spores of Bacillus subtilis grown in different media by elemental characterization using time-of-flight secondary ion mass spectrometry*. *Applied and Environmental Microbiology*, 2005. 71: p. 6524-6530.
38. Teska, J.D., et al., *Identification of Bacillus anthracis using gas chromatographic analysis of cellular fatty acids and a commercially available database*. 2003, Agilent Technologies, Inc.: Santa Clara, CA. p. 1-5.
39. Peak, K.K., et al., *Bacillus acidicer sp. nov., isolated from a forensic specimen, containing Bacillus anthracis pX02 genes*. *International Journal of Systematic and Evolutionary Microbiology*, 2007. 57: p. 2031-2036.
40. Huberty, C.J., *Applied discriminant analysis*. 1994, New York, NY: John Wiley and Sons, Inc.

41. Tabachnick, B.G. and L.S. Fidell, *Using multivariate statistics*. 2001, Needham Heights, MA: Allyn & Bacon.
42. Haack, S.K., et al., *Accuracy, reproducibility, and interpretation of fatty acid methyl ester profiles of model bacterial communities*. *Applied and Environmental Microbiology*, 1994. 60: p. 2483-2493.
43. Whittaker, P., et al., *Use of fatty acid profiles to identify food-borne bacterial pathogens and aerobic endospore-forming bacilli*. *Journal of Agricultural and Food Chemistry*, 2005. 53: p. 3735-3742.
44. Manly, B.F.J., *Multivariate statistical methods, 3rd ed.* 2005, Boca Raton, FL: Chapman & Hall.
45. Clarke, K.R., *Non-parametric multivariate analyses of change in community structure*. *Australian Journal of Ecology*, 1993. 18(1): p. 117-143.
46. Clarke, K.R. and R.M. Warwick, *Change in marine communities: an approach to statistical analysis and interpretation*. 1994, Natural Environment Research Council: Plymouth, United Kingdom.
47. Dye, E.S. and F.A. Kapral, *Characterization of a bactericidal lipid developing within staphylococcal abscesses*. *Infection and Immunity*, 1981. 32: p. 98-104.
48. Scherer, C., et al., *Influence of culture conditions on the fatty acid profiles of laboratory-adapted and freshly isolated strains of Helicobacter pylori*. *Journal of Clinical Microbiology*, 2003. 41: p. 1114-1117.
49. Guerzoni, M.E., R. Lanciotti, and P.S. Cocconelli, *Alteration in cellular fatty acid composition as a response to salt, acid, oxidative and thermal stresses in*

- Lactobacillus helveticus*. Microbiology and Molecular Biology Reviews, 2001. 147: p. 2255-2264.
50. Loginova, L.I., V.P. Manuilova, and V.P. Tolstikov, *Content of free amino acids in peptone and the dynamics of their consumption in the microbiological synthesis of dextran*. Pharmaceutical Chemistry Journal, 1974. 8: p. 49-51.
 51. Rose, R., et al., *Comparison of the properties of Bacillus subtilis spores made in liquid or on agar plates*. Journal of Applied Microbiology, 2007. 103: p. 691-699.
 52. Fulco, A.J., *Fatty acid metabolism in bacteria*. Progress in Lipid Research, 1983. 22(133-160).
 53. Wimpenny, J.W.T. and J.P. Coombs, *Penetration of oxygen into bacterial colonies*. Journal of General Microbiology, 1983. 129: p. 1239-1242.
 54. Rieck, V.T., S.A. Palumbo, and L.D. Witter, *Glucose availability and the growth rate of colonies of Pseudomonas fluorescens*. Journal of General Microbiology, 1973. 74: p. 1-8.
 55. Bezbaruah, R.L., et al., *Effect of growth temperature and media composition on the fatty acid composition of Bacillus stearothermophilus AN 002*. Antonie Van Leeuwenhoek, 1988. 54: p. 37-45.
 56. Daron, H.H., *Fatty acid composition of lipid extracts of a thermophilic Bacillus species*. Journal of Bacteriology, 1970. 101: p. 145-151.
 57. Johnson, D.R., et al., *Determination of race and sex of the human skull by discriminant functions analysis of linear and angular dimensions*. Forensic Science International, 1989. 41: p. 41-53.

58. Snow, C.C., et al., *Sex and race determination of crania by calipers and computer: a test of the Giles and Elliot discriminant functions in 52 forensic science cases*. Journal of Forensic Sciences, 1979. 24: p. 448-460.
59. Banowetz, G.M., et al., *Fatty acid methyl ester analysis to identify sources of soil in surface water*. Journal of Environmental Quality, 2006. 35: p. 133-140.

Figure captions

Fig. 1. Vegetative and spore forms of *Bacillus cereus* T strain. (a) Vegetative starter cultures were used to inoculate 12 different sporulation media. (b) After 24 to 48 h of incubation in sporulation media, the cell cultures were >90% phase-bright spores. Scale bar = 5 μm .

Fig. 2. Comparison of spore cultures grown on different medium formulations. (a and b) The spores grown on CDSM showed the most significant differences in the relative proportions of branched-odd and branched-even fatty acids. (a to d) Variation in each of the four fatty acid structure classes was also observed among all the other medium formulations. The error bars indicate 1 standard deviation.

Fig. 3. Two-dimensional nMDS plot of spore FAME profiles grouped by growth medium. Multivariate dissimilarities were generated using every fatty acid variable in the FAME profile. Many of the medium groups showed distinct clusters that were clearly differentiated from other groups (G, MG, Sch, and LL). Some spore groups showed overlapping or poorly defined clusters (BHI, CA, SchAg, and CAB).

Fig. 4. Discriminant function analysis of spore FAME profiles. In each panel, spore samples are plotted against two CV functions. The equation for each CV function is shown on the right of the plot. The cumulative proportion of the total dispersion among groups is included for each CV. (a) Distinct clusters were observed for the G, LL, and

CAB groups across the first two CV functions. (b and c) Successive CV functions separated the remaining medium groups.

Table 1. Sporulation media key.

Sporulation medium	Abbreviation	Reference	Component(s) ^a
Chemically defined sporulation medium ^b	CDSM	16	Ammonium sulfate
G broth	G	6, 18	Yeast extract, ammonium sulfate
Modified G broth	MG	2	Yeast extract, ammonium sulfate (no supplemental sugar source)
Casein acid digest broth	CAD		Yeast extract, tryptone
Brain heart infusion agar	BHI		Meat peptic digest, brain heart solids, gelatin digest
Schaeffer's broth	Sch	34	Beef extract, meat peptone
Schaeffer's agar	SchAg	34	Beef extract, meat peptone
Lab Lemco agar	LL	9	Meat peptone, Lab Lemco powder
New sporulation medium agar	NSM	9	Yeast extract, tryptone, meat peptone, Lab Lemco powder
Columbia agar	CA		Yeast extract, beef heart digest, corn starch, tryptone, meat peptone
Columbia agar supplemented with sheep blood	CAB		Yeast extract, beef heart digest, corn starch, tryptone, meat peptone, defibrinated sheep blood

^a Only complex carbon and nitrogen sources are listed for each medium. Trace metal and salt components are listed in Materials and Methods.

^b Used in broth form.

Table 2. Fatty acid variables.

Branched-odd ^a	Branched-even ^a	Anteiso	Normal
13:0 iso	14:0 iso	13:0 anteiso	14:0
15:0 iso	16:0 iso	15:0 anteiso	15:1 ω5c
17:0 iso		17:1 anteiso A	16:0
17:0 iso ω5c		17:0 anteiso	16:1 ω7c alcohol
17:1 iso ω10c			16:1 ω7c/16:1 ω6c
			18:1 ω9c

^a "Odd" and "even" refer to the number of carbons in each fatty acid (from Kaneda [24]).

Table 3. Relative abundances of fatty acid markers.

Medium	Relative abundance ^a																																		
	13:0 iso		13:0 ante		14:0 iso		14:0		15:0 iso		15:0 ante		15:1 ω5c		16:1 ω7c OH		16:0 iso		SumFeat2 ^b		16:0		17:1 iso ω10c		17:1 iso ω5c		17:1 ante A		17:0 iso		17:0 ante		18:1 ω9c		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean
CDSM	1.9	0.60	1.09	0.41	13.24	1.62	6.94	2.16	11.47	2.87	6.13	2.16	0.71	0.21	2.04	0.42	16.96	2.96	18.89	3.57	7.31	1.16	1.18	0.25	2.31	0.57	1.03	0.42	3.31	0.60	2.20	0.87	0.45	0.50	
G	6.89	0.53	1.50	0.27	6.62	0.17	4.31	0.34	30.68	1.44	6.77	0.66	0.89	0.13	1.01	0.10	6.80	0.16	13.76	0.92	1.97	0.32	3.03	0.29	5.60	0.36	1.10	0.13	6.21	0.52	1.57	0.34	0.02	0.07	
MG	6.15	0.52	1.27	0.13	6.09	0.37	3.86	0.20	32.24	0.43	6.22	0.65	0.58	0.25	0.76	0.08	7.27	0.50	12.22	0.79	2.46	0.51	2.64	0.15	6.19	0.37	1.08	0.09	8.26	0.90	1.69	0.23	0.00	0.00	
CAD	8.56	0.26	0.88	0.05	4.81	0.16	3.46	0.37	34.13	2.09	3.92	0.31	0.84	0.03	0.41	0.08	5.28	0.54	11.54	0.96	2.07	0.32	1.54	0.20	7.67	1.17	1.14	0.11	11.27	0.52	0.83	0.07	0.04	0.08	
BHI	5.82	0.30	0.54	0.06	4.40	0.38	4.46	0.17	34.85	0.88	3.36	0.34	0.04	0.10	0.46	0.04	6.67	0.45	11.76	0.67	5.78	0.87	1.78	0.33	4.81	0.68	0.64	0.20	11.04	0.49	1.68	0.29	0.59	0.32	
Sch	5.65	0.42	1.01	0.05	7.56	0.71	5.59	0.44	25.03	1.75	5.63	1.18	0.50	0.33	0.95	0.13	7.16	0.61	18.85	2.33	2.96	0.17	2.42	0.21	7.16	0.35	1.83	0.87	5.42	0.48	1.54	0.28	0.00	0.00	
SchAg	6.10	1.23	0.98	0.19	5.53	0.37	5.02	1.16	32.40	2.70	5.71	0.60	0.80	0.21	0.52	0.23	4.82	0.45	16.85	3.32	3.17	1.05	1.50	0.22	5.74	0.62	0.99	0.12	6.83	2.77	1.44	0.45	0.23	0.21	
LL	4.45	0.53	1.05	0.18	7.73	0.36	3.72	0.35	23.68	1.65	5.54	0.23	0.23	0.27	0.20	0.22	11.29	0.83	12.15	1.04	6.11	1.08	1.11	0.07	5.74	0.24	0.91	0.11	10.62	1.32	2.90	0.37	0.07	0.16	
NSM	5.69	0.88	0.81	0.06	5.35	0.89	4.76	0.30	27.84	1.80	4.29	0.69	0.64	0.07	0.19	0.21	6.74	0.55	12.74	0.36	8.17	2.09	1.15	0.08	5.30	0.39	0.92	0.06	10.57	0.55	1.89	0.22	0.05	0.13	
CA	5.54	0.63	0.71	0.16	5.46	0.94	3.97	0.41	35.52	3.23	4.25	0.85	0.32	0.30	0.44	0.16	7.25	0.63	8.52	2.64	5.53	0.70	1.61	0.60	5.07	0.84	1.09	0.07	10.59	2.39	1.78	0.41	0.82	0.46	
CAB	5.97	0.76	0.23	0.28	4.31	0.75	4.56	0.43	33.34	3.69	3.07	1.75	0.13	0.22	0.34	0.21	6.02	0.44	7.43	2.04	5.97	0.83	1.46	0.17	3.37	0.88	0.61	0.19	9.48	0.33	1.72	0.96	7.90	5.09	

^a Mean values are calculated from the relative abundance in each profile determined from GC.

^b SumFeat2 represents fatty acids 16:1 ω7c/16:1 ω7c.

Table 4. ANOSIM^a on spore-medium group comparisons.

Medium	Individual fatty acid variable (global $R = 0.84$) ^b									
	G	BHI	CA	CAD	NSM	Sch	SchAg	MG	LL	CAB
G	1	0.94	1	1	0.83	0.61	0.39	1	1	
BHI		1	0.3*	1	0.97	1	0.77	0.99	1	0.75
CA			1	0.83	0.61	0.96	0.48	0.75	0.93	0.87
CAD				1	0.98	1	0.54	1	1	0.96
NSM					1	0.57	1	0.90	0.91	
Sch						1	0.62	0.75	1	1
SchAg							1	0.48	0.80	0.96
MG								1	1	
LL									1	1

^a The matrix lists the R value for each pairwise medium group comparison. The asterisks indicate that the R value was not significant at the $P = 0.01$ level.

^b Global R was calculated from simultaneous comparisons of the rank dissimilarities for all samples and groups.

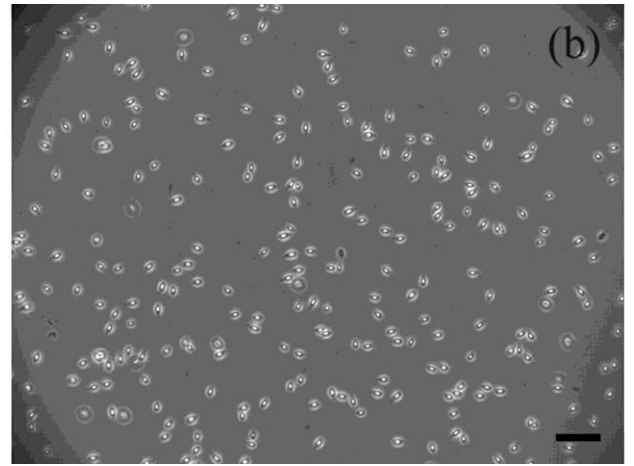
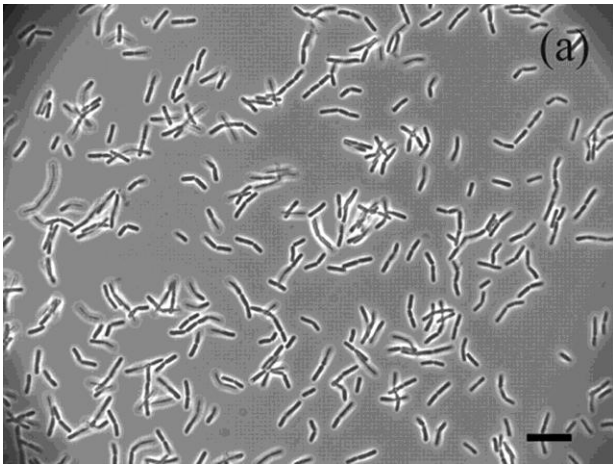


Fig. 1.

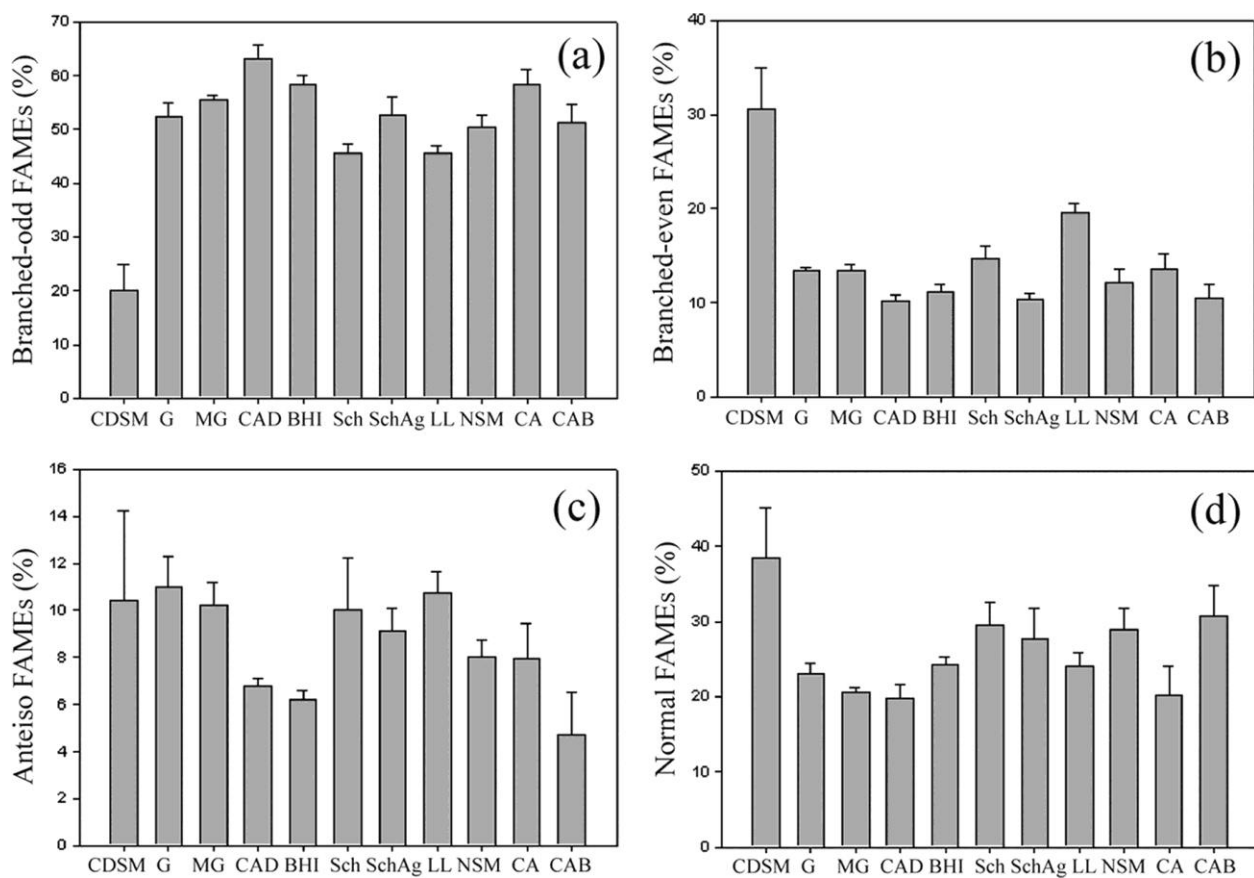


Fig. 2.

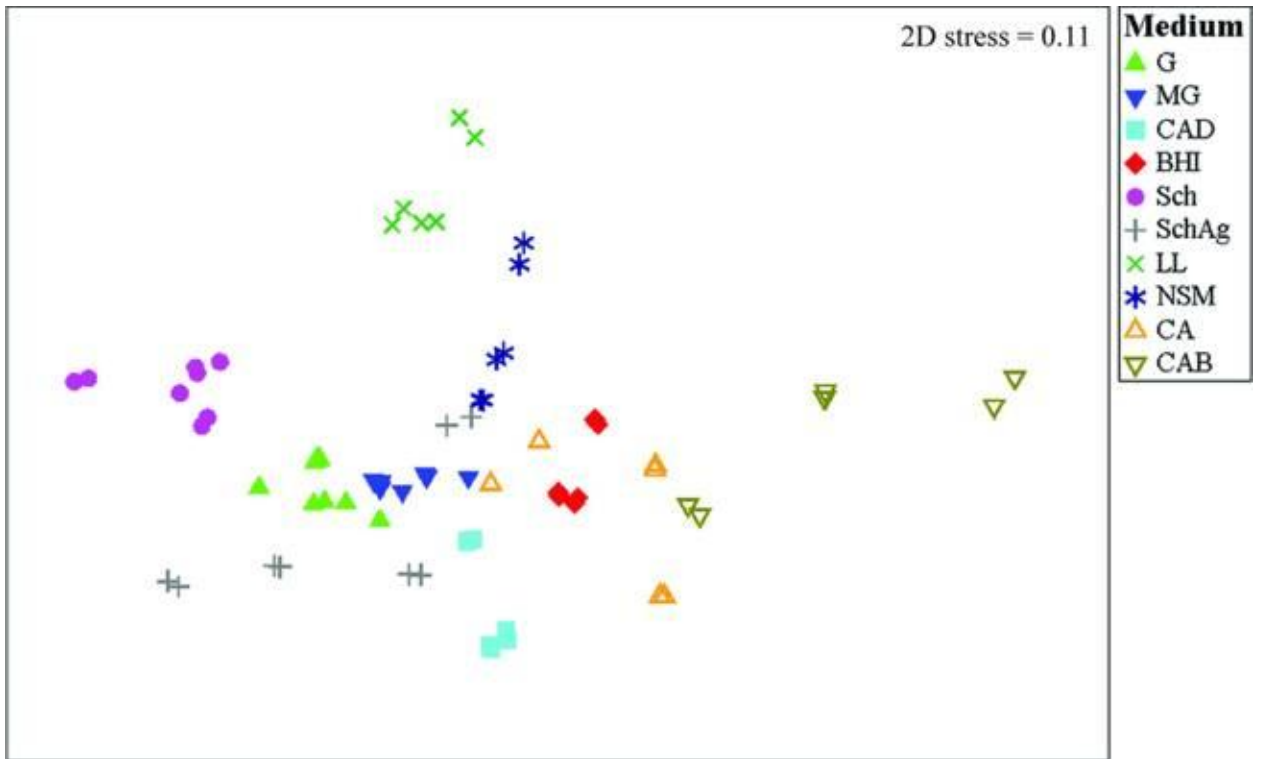


Fig. 3.

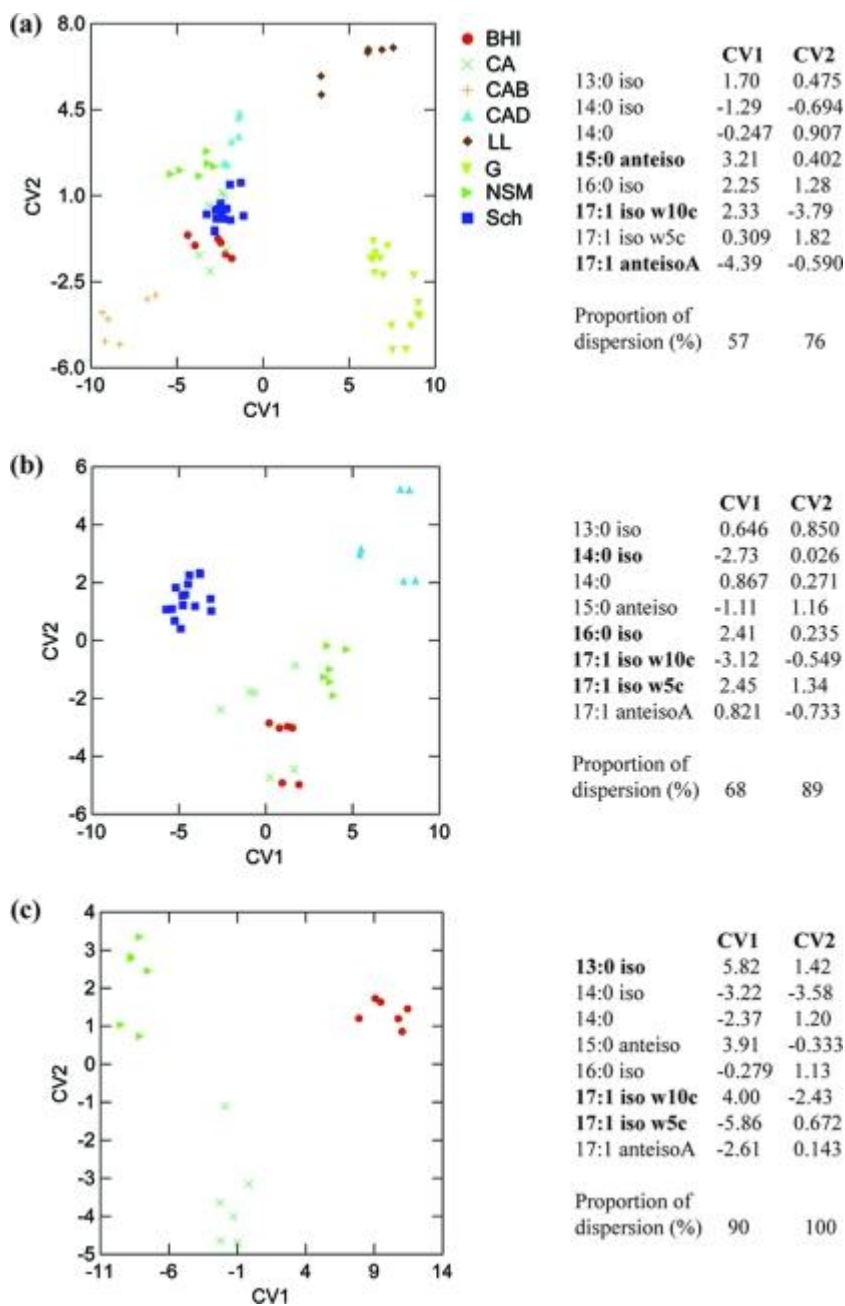





Fig. 4.

APPENDIX

BLAST Results for WSMV-Specific Probes on the Virochip


Probe ID: 13241967_nt00101.60_rc

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
EU914918.1	Wheat streak mosaic virus isolate Saadat-Shahr, complete genome	111	111	100%	2e-22	100%	
EU914917.1	Wheat streak mosaic virus isolate Naghadeh, complete genome	111	111	100%	2e-22	100%	
FJ348359.1	Wheat streak mosaic virus isolate Arg2 polyprotein gene, complete c	111	111	100%	2e-22	100%	
FJ348358.1	Wheat streak mosaic virus isolate WA94 polyprotein gene, complete	111	111	100%	2e-22	100%	
AF511643.2	Wheat streak mosaic virus isolate WA99, complete genome	111	111	100%	2e-22	100%	
AF511630.2	Wheat streak mosaic virus isolate MON96, complete genome	111	111	100%	2e-22	100%	
AF511619.2	Wheat streak mosaic virus isolate ID99, complete genome	111	111	100%	2e-22	100%	
AF511618.2	Wheat streak mosaic virus isolate ID96, complete genome	111	111	100%	2e-22	100%	
AF511615.2	Wheat streak mosaic virus isolate H98, complete genome	111	111	100%	2e-22	100%	
AF511614.2	Wheat streak mosaic virus isolate H95S, complete genome	111	111	100%	2e-22	100%	
AF454455.1	Wheat streak mosaic virus isolate Turkey 1, complete genome	111	111	100%	2e-22	100%	
AF329446.1	Expression vector pWSMV-S1RN, complete sequence	111	111	100%	2e-22	100%	
AF285169.1	Wheat streak mosaic virus type strain complete genome	111	111	100%	2e-22	100%	
AF057533.1	Wheat streak mosaic virus strain Sidney 81, complete genome	111	111	100%	2e-22	100%	
AF454454.1	Wheat streak mosaic virus isolate Czech, complete genome	106	106	100%	1e-20	98%	
AF285170.1	Wheat streak mosaic virus strain El Batan 3 complete genome	99.0	99.0	98%	2e-18	96%	
AY377938.1	Oat necrotic mottle virus isolate Type-NE, complete genome	63.9	63.9	56%	6e-08	100%	
GQ421689.2	Canna Yellow Streak Virus from United Kingdom, complete genome	52.8	52.8	46%	1e-04	100%	

Probe ID: 13241967_nt00101.60_rc

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
EU914918.1	Wheat streak mosaic virus isolate Saadat-Shahr, complete genome	<u>111</u>	111	100%	<u>2e-22</u>	100%	
EU914917.1	Wheat streak mosaic virus isolate Naghadeh, complete genome	<u>111</u>	111	100%	2e-22	100%	
FJ348359.1	Wheat streak mosaic virus isolate Arg2 polyprotein gene, complete c	<u>111</u>	111	100%	2e-22	100%	
FJ348358.1	Wheat streak mosaic virus isolate WA94 polyprotein gene, complete	<u>111</u>	111	100%	2e-22	100%	
AF511643.2	Wheat streak mosaic virus isolate WA99, complete genome	<u>111</u>	111	100%	2e-22	100%	
AF511630.2	Wheat streak mosaic virus isolate MON96, complete genome	<u>111</u>	111	100%	2e-22	100%	
AF511619.2	Wheat streak mosaic virus isolate ID99, complete genome	<u>111</u>	111	100%	2e-22	100%	
AF511618.2	Wheat streak mosaic virus isolate ID96, complete genome	<u>111</u>	111	100%	2e-22	100%	
AF511615.2	Wheat streak mosaic virus isolate H98, complete genome	<u>111</u>	111	100%	2e-22	100%	
AF511614.2	Wheat streak mosaic virus isolate H95S, complete genome	<u>111</u>	111	100%	2e-22	100%	
AF454455.1	Wheat streak mosaic virus isolate Turkey 1, complete genome	<u>111</u>	111	100%	2e-22	100%	
AF329446.1	Expression vector pWSMV-S1RN, complete sequence	<u>111</u>	111	100%	2e-22	100%	
AF285169.1	Wheat streak mosaic virus type strain complete genome	<u>111</u>	111	100%	2e-22	100%	
AF057533.1	Wheat streak mosaic virus strain Sidney 81, complete genome	<u>111</u>	111	100%	2e-22	100%	G
AF454454.1	Wheat streak mosaic virus isolate Czech, complete genome	<u>106</u>	106	100%	1e-20	98%	
AF285170.1	Wheat streak mosaic virus strain El Batan 3 complete genome	<u>99.0</u>	99.0	98%	2e-18	96%	
AY377938.1	Oat necrotic mottle virus isolate Type-NE, complete genome	<u>63.9</u>	63.9	56%	6e-08	100%	G
GQ421689.2	Canna Yellow Streak Virus from United Kingdom, complete genome	<u>52.8</u>	52.8	46%	1e-04	100%	G


Probe ID: 13241967_nt07976.60

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
EU914917.1	Wheat streak mosaic virus isolate Naghadeh, complete genome	111	111	100%	2e-22	100%	
AJ889242.1	Wheat streak mosaic virus partial gene for polyprotein, genomic RNA	111	111	100%	2e-22 	100%	
AF454455.1	Wheat streak mosaic virus isolate Turkey 1, complete genome	111	111	100%	2e-22	100%	
AF329446.1	Expression vector pWSMV-S1RN, complete sequence	111	111	100%	2e-22	100%	
AF285169.1	Wheat streak mosaic virus type strain complete genome	111	111	100%	2e-22	100%	
AF057533.1	Wheat streak mosaic virus strain Sidney 81, complete genome	111	111	100%	2e-22	100%	G
FJ348359.1	Wheat streak mosaic virus isolate Arq2 polyprotein gene, complete c	106	106	100%	1e-20	98%	
FJ348358.1	Wheat streak mosaic virus isolate WA94 polyprotein gene, complete	106	106	100%	1e-20	98%	
AF511643.2	Wheat streak mosaic virus isolate WA99, complete genome	106	106	100%	1e-20	98%	
AF511630.2	Wheat streak mosaic virus isolate MON96, complete genome	106	106	100%	1e-20	98%	
AF511619.2	Wheat streak mosaic virus isolate ID99, complete genome	106	106	100%	1e-20	98%	
AF511618.2	Wheat streak mosaic virus isolate ID96, complete genome	106	106	100%	1e-20	98%	
AF511615.2	Wheat streak mosaic virus isolate H98, complete genome	106	106	100%	1e-20	98%	
U67937.1	Wheat streak mosaic virus polyprotein gene, partial cds	106	106	95%	1e-20	100%	
AF454457.1	Wheat streak mosaic virus isolate Turkey 2 polyprotein gene, partial	106	106	100%	1e-20	98%	
U59478.1	Wheat streak mosaic virus polyprotein (N1b, capsid protein) mRNA, p	106	106	100%	1e-20	98%	
AF511614.2	Wheat streak mosaic virus isolate H95S, complete genome	100	100	100%	5e-19	96%	
AY327870.1	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327867.1	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327866.2	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327868.3	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327869.2	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327865.2	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327872.2	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327871.2	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	

Probe ID: 13241967_nt07976.60_rc

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
EU914917.1	Wheat streak mosaic virus isolate Naghadeh, complete genome	111	111	100%	2e-22	100%	G
AJ889242.1	Wheat streak mosaic virus partial gene for polyprotein, genomic RNA	111	111	100%	2e-22	100%	
AF454455.1	Wheat streak mosaic virus isolate Turkey 1, complete genome	111	111	100%	2e-22	100%	
AF329446.1	Expression vector pWSMV-S1RN, complete sequence	111	111	100%	2e-22	100%	
AF285169.1	Wheat streak mosaic virus type strain complete genome	111	111	100%	2e-22	100%	
AF057533.1	Wheat streak mosaic virus strain Sidney 81, complete genome	111	111	100%	2e-22	100%	
FJ348359.1	Wheat streak mosaic virus isolate Arq2 polyprotein gene, complete c	106	106	100%	1e-20	98%	
FJ348358.1	Wheat streak mosaic virus isolate WA94 polyprotein gene, complete	106	106	100%	1e-20	98%	
AF511643.2	Wheat streak mosaic virus isolate WA99, complete genome	106	106	100%	1e-20	98%	
AF511630.2	Wheat streak mosaic virus isolate MON96, complete genome	106	106	100%	1e-20	98%	
AF511619.2	Wheat streak mosaic virus isolate ID99, complete genome	106	106	100%	1e-20	98%	
AF511618.2	Wheat streak mosaic virus isolate ID96, complete genome	106	106	100%	1e-20	98%	
AF511615.2	Wheat streak mosaic virus isolate H98, complete genome	106	106	100%	1e-20	98%	
U67937.1	Wheat streak mosaic virus polyprotein gene, partial cds	106	106	95%	1e-20	100%	
AF454457.1	Wheat streak mosaic virus isolate Turkey 2 polyprotein gene, partial	106	106	100%	1e-20	98%	
U59478.1	Wheat streak mosaic virus polyprotein (N1b, capsid protein) mRNA, p	106	106	100%	1e-20	98%	
AF511614.2	Wheat streak mosaic virus isolate H95S, complete genome	100	100	100%	5e-19	96%	
AY327870.1	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327867.1	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327866.2	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327868.3	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327869.2	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327865.2	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327872.2	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	
AY327871.2	Wheat streak mosaic virus polyprotein mRNA, partial cds	100	100	95%	5e-19	98%	

Probe ID: 13241967_nt03001.60

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
FJ348359.1	Wheat streak mosaic virus isolate Arq2 polyprotein gene, complete c	111	111	100%	2e-22	100%	
AF511630.2	Wheat streak mosaic virus isolate MON96, complete genome	111	111	100%	2e-22	100%	
AF511619.2	Wheat streak mosaic virus isolate ID99, complete genome	111	111	100%	2e-22	100%	
AF511618.2	Wheat streak mosaic virus isolate ID96, complete genome	111	111	100%	2e-22	100%	
AF511614.2	Wheat streak mosaic virus isolate H95S, complete genome	111	111	100%	2e-22	100%	
AF454455.1	Wheat streak mosaic virus isolate Turkey 1, complete genome	111	111	100%	2e-22	100%	
AF329446.1	Expression vector pWSMV-S1RN, complete sequence	111	111	100%	2e-22	100%	
AF285169.1	Wheat streak mosaic virus type strain complete genome	111	111	100%	2e-22	100%	
AF057533.1	Wheat streak mosaic virus strain Sidney 81, complete genome	111	111	100%	2e-22	100%	
EU914918.1	Wheat streak mosaic virus isolate Saadat-Shahr, complete genome	106	106	100%	1e-20	98%	
EU914917.1	Wheat streak mosaic virus isolate Naghadeh, complete genome	106	106	100%	1e-20	98%	
AF511643.2	Wheat streak mosaic virus isolate WA99, complete genome	106	106	100%	1e-20	98%	
AF454454.1	Wheat streak mosaic virus isolate Czech, complete genome	106	106	100%	1e-20	98%	
FJ348358.1	Wheat streak mosaic virus isolate WA94 polyprotein gene, complete	100	100	100%	5e-19	96%	
AF511615.2	Wheat streak mosaic virus isolate H98, complete genome	100	100	100%	5e-19	96%	

Probe ID: 13241967_nt03001.60_rc

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
FJ348359.1	Wheat streak mosaic virus isolate Arq2 polyprotein gene, complete c	111	111	100%	2e-22	100%	G
AF511630.2	Wheat streak mosaic virus isolate MON96, complete genome	111	111	100%	2e-22	100%	
AF511619.2	Wheat streak mosaic virus isolate ID99, complete genome	111	111	100%	2e-22	100%	
AF511618.2	Wheat streak mosaic virus isolate ID96, complete genome	111	111	100%	2e-22	100%	
AF511614.2	Wheat streak mosaic virus isolate H95S, complete genome	111	111	100%	2e-22	100%	
AF454455.1	Wheat streak mosaic virus isolate Turkey 1, complete genome	111	111	100%	2e-22	100%	
AF329446.1	Expression vector pWSMV-S1RN, complete sequence	111	111	100%	2e-22	100%	
AF285169.1	Wheat streak mosaic virus type strain complete genome	111	111	100%	2e-22	100%	
AF057533.1	Wheat streak mosaic virus strain Sidney 81, complete genome	111	111	100%	2e-22	100%	
EU914918.1	Wheat streak mosaic virus isolate Saadat-Shahr, complete genome	106	106	100%	1e-20	98%	
EU914917.1	Wheat streak mosaic virus isolate Naghadeh, complete genome	106	106	100%	1e-20	98%	
AF511643.2	Wheat streak mosaic virus isolate WA99, complete genome	106	106	100%	1e-20	98%	
AF454454.1	Wheat streak mosaic virus isolate Czech, complete genome	106	106	100%	1e-20	98%	
FJ348358.1	Wheat streak mosaic virus isolate WA94 polyprotein gene, complete	100	100	100%	5e-19	96%	
AF511615.2	Wheat streak mosaic virus isolate H98, complete genome	100	100	100%	5e-19	96%	

Probe ID: 11066855_nt5898.60

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
AF285170.1	Wheat streak mosaic virus strain El Batan 3 complete genome	111	111	100%	2e-22	100%	

Probe ID: 11066855_nt5898.60_rc

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
AF285170.1	Wheat streak mosaic virus strain El Batan 3 complete genome	111	111	100%	2e-22	100%	

VITA

TeeCie Paige Brown

Candidate for the Degree of

Doctor of Philosophy

Thesis: THE DEVELOPMENT AND CHARACTERIZATION OF MOLECULAR
TOOLS FOR MICROBIAL FORENSICS

Major Field: Biochemistry and Molecular Biology

Biographical:

Education:

Received Bachelor of Science in Biology at Tarleton State University, Stephenville, Texas, in May, 2006. Completed the requirements for the Doctor of Philosophy in Biochemistry and Molecular Biology at Oklahoma State University, Stillwater, Oklahoma in May, 2011.

Experience: Research Assistant at Oklahoma State University from August 2006 to May 2011; Visiting Scientist at the FBI-CFSRU from May 2008 to August 2008; Undergraduate Researcher at Tarleton State University from 2004 to 2005.

Professional Memberships:

American Phytopathological Society
American Association for the Advancement of Science

Publications:

Brown, T., Melcher, U. and Fletcher, J. (2010). A tag-array minisequencing-based system for detecting and genetic fingerprinting *Wheat streak mosaic virus*: Implications for plant pathogen forensics. *Phytopathology*. 100:S18.

Ehrhardt, C. J., Chu, V., Brown, T., Simmons, T. L., Swan, B. K., Bannan, J. and J.M. Robertson. Discrimination of *Bacillus cereus* T-strain Spores Grown on Different Media using Fatty Acid Methyl Ester (FAME) Profiles. *Appl. Environ. Microbiol.*, AEM.02443-02409.

Name: TeeCie Paige Brown

Date of Degree: July, 2011

Institution: Oklahoma State University

Location: Stillwater, Oklahoma

Title of Study: THE DEVELOPMENT AND CHARACTERIZATION OF
MOLECULAR TOOLS FOR MICROBIAL FORENSICS

Pages in Study: 199

Candidate for the Degree of Doctor of Philosophy

Major Field: Biochemistry and Molecular Biology

Scope and Method of Study:

The anthrax attacks of 2001 prompted the rapid establishment and growth of the fields of microbial and plant pathogen forensics. A complete forensic capability includes the ability to discriminate between a natural and an intentional disease outbreak, collection of forensic evidence, generation of genetic profiles for use during attribution and storage of samples. This document describes (i) the molecular characterization of plant virus populations derived from plants that were naturally and mechanically-inoculated with a model plant virus, (ii) the application of microarray-based technologies to genetically fingerprint plant viruses, (iii) the characterization of a previously-designed microarray platform for the identification and diagnosis of known and novel plant viruses, and (iv) the use of FAME profiles to aid in the discrimination of media components used to prepare *Bacillus cereus* T-strain spores. Though all of these projects are not directly related, they all fall within the scope of microbial forensics.

Findings and Conclusions:

The molecular characterization of plant virus populations derived from a natural inoculation event and those from mechanically-inoculated plants displayed minor differences in haplotype and pair-wise nucleotide diversities. Additionally, the number of recombination events was found to be lower in the mechanically-inoculated plants than those collected from the natural disease outbreak. These results indicate that differences between the two types of inoculation events exist and may be a direct function of the infection time, source(s) of inoculum(a) or environmental effects.

The solution-based minisequencing and capture array technique demonstrated reproducibility at the same concentration of targets, but was less accurate using variable amounts of synthetic targets. The use of solution-based minisequencing followed by tag-array capture appears to be a promising approach to genotyping plant viruses.

A significant amount of cross-hybridization was observed using the universal plant virus microarray (Virochip). The microarray platform failed to strongly hybridize to most of the known plant viruses that were applied to the array. Hybridization with *Wheat streak mosaic virus*-infected material indicated that the system strongly hybridized with the negative-sense strand, but not the corresponding positive-sense strands.

Discrimination of individual media components was achieved by analyzing fatty acids derived from *Bacillus cereus* T-strain spores prepared in different media. One FAME biomarker, oleic acid, was found to be exclusively associated with media supplemented with blood.

ADVISER'S APPROVAL: Ulrich K. Melcher
