IOWA STATE UNIVERSITY
Digital Repository

Graduate Theses and Dissertations | Graduate College

2015

Nucleic acid detection techniques for adventitious agent testing

Kaitlin Elizabeth Brien
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd

 Part of the Genetics Commons, and the Virology Commons

Recommended Citation
Brien, Kaitlin Elizabeth, "Nucleic acid detection techniques for adventitious agent testing" (2015). *Graduate Theses and Dissertations*. 14322.
http://lib.dr.iastate.edu/etd/14322

This Thesis is brought to you for free and open access by the Graduate College at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

Nucleic acid detection techniques for adventitious agent testing

by

Kaitlin Elizabeth Brien

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Co-Majors: Microbiology, Genetics

Program of Study Committee:
Joseph Hermann, Co-Major Professor
Cathy Miller, Co-Major Professor
Karin Dorman
Bradley Blitvich

Iowa State University

Ames, Iowa

2015

Copyright © Kaitlin Elizabeth Brien, 2015. All rights reserved.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my committee chairs, Dr. Joseph Hermann and Dr. Cathy Miller, and my committee members, Dr. Karin Dorman and Dr. Bradley Blitvich, for their guidance and support throughout the course of this research. Thanks also to several wonderful individuals who without their scientific expertise and mentorship this project would not have been possible: Dr. Roger Barrette, Tod Stuber, Alethea Fry, Alexa Scupham, and Patrick Camp.

In addition, I would like to thank the USDA for the Veterinary Services Daniel E. Salmon Scholarship, which provided financial support over the course of my Master's program. I would also like to thank my friends and colleagues at the USDA and the faculty and staff in the Iowa State University Interdepartmental Microbiology and Genetics programs for making my time at Iowa State University a wonderful experience. Finally, thanks to my family and friends for their encouragement.

ABSTRACT

The availability of safe and effective animal vaccines is critical for the prevention of animal disease. Adventitious agent testing is done on master seed viruses prior to vaccine licensure to ensure that no biological contaminants were introduced during manufacture. Traditional adventitious agent testing is performed using a variety of cell culture lines and a panel of polymerase chain reaction tests. The purpose of this research was to determine if new technologies like DNA microarray and next-generation sequencing (NGS) could be of any benefit for adventitious agent testing. A literature review describes the state of the field and the challenges that will have to be addressed to use these technologies in a regulatory environment. Both techniques were tested on a panel of mammalian and avian viruses, and each virus was tested individually and in combination with other viruses. NGS was found to be a more reliable method of screening for adventitious agents than microarray.

CHAPTER 1. GENERAL INTRODUCTION

Introduction

The Center for Veterinary Biologics is responsible for ensuring that pure, safe, potent, and effective biologics are available for the diagnosis, treatment, and prevention of animal diseases. Prior to licensure of a new vaccine, manufacturers must submit a sample of their master seed virus, which will be used as the source material to produce the vaccine. Master seed viruses are tested for adventitious agents to ensure that no biological contaminants were inadvertently introduced. Primary methods of adventitious agent testing for viruses include cell culture testing conducted according to the 9 Code of Federal Regulations and polymerase chain reaction (PCR) tests [1]. However, cell culture detection is lengthy, labor intensive, and relies on the availability of a permissive cell line. PCR tests can only be performed for a selected number of agents. If a potential adventitious agent is not cell-culture adapted or is not targeted by a PCR, it may not be detected. Recent technologies based on the broad detection of viral nucleic acids, like microarray and next-generation sequencing (NGS), could be used to increase the range of detectable viruses in adventitious agent testing.

Microarray and NGS have been successfully used to detect unknown pathogens in diagnostic laboratories [2-6] . However, diagnostic samples are often the result of an acute infection with high levels of a single pathogen. Adventitious agent testing requires the detection of both the vaccine virus and any adventitious viruses, with the adventitious viruses potentially present at much lower concentrations. Additional work is needed to determine if these technologies could be suitable screening tools for adventitious agents. Many recent studies focus on recovery of a single virus at a time and determining a limit

of detection [7-12]. While it is critical to have an estimate of sensitivity for new methods, determining whether multiple viruses can be detected in combination is equally critical. To determine whether microarray and NGS can reliably detect multiple viruses in a sample, a novel DNA microarray and Illumina sequencing were used to detect a panel of mammalian and avian viruses. All viruses were tested both individually and in combinations. In most cases, each method was able to detect both viruses, although both methods had failures. NGS had a higher detection rate than the microarray. Furthermore, NGS provides higher confidence in the results because it is able to provide more information than microarray, like the specific strain of virus detected, percent genome coverage, and average depth of sequencing.

Additional work is needed to determine the reliability and sensitivity of NGS before it can be used to confidently rule out the presence of adventitious agents in master seeds. However, these preliminary results show that this novel microarray and NGS can be used in conjunction with traditional methods for adventitious agent testing.

Thesis Organization

This thesis is organized into four chapters. Chapter one provides a general introduction to the rationale guiding this research. The literature review following the general introduction describes the challenges facing this area of research and details the accomplishments of other laboratories to date. Chapter two details the work I have done to answer the question of whether microarray and NGS would be suitable for use in adventitious agent testing. Chapter three provides an example of using microarray and NGS to detect bovine viral diarrhea virus, a common adventitious agent. Figures in

chapter two and three can be found following the acknowledgments and references at the end of each chapter. Chapter four contains general conclusions of this research.

Literature Review

Introduction

Animal vaccination has greatly improved the health of humans and animals since the 1800s when Edward Jenner's smallpox vaccine was introduced [13]. Vaccination has been the driving force for the global eradication of rinderpest and the elimination of foot-and-mouth disease from Europe [14]. However, the reasons for veterinary vaccination go beyond animal health and welfare. Human health is protected from zoonotic agents like rabies [15]. Veterinary vaccination also promotes sustainable livestock production by preventing the use of mass slaughter in response to a disease outbreak [16].

The many benefits of vaccination, however, are dependent on the availability of safe and effective vaccines. Vaccines, whether killed, modified-live, or recombinant, are all vulnerable to contamination by biological organisms throughout the course of manufacturing [17]. These contaminating biological organisms are referred to as adventitious agents, and the World Health Organization (WHO) describes them as "contaminating microorganisms of cell culture or source materials ... that have been unintentionally introduced into the manufacturing process of a biological product" [18]. Contaminants can be introduced through source material including cell substrates and serum that may contain endogenous viruses. In addition, cross-contamination by infectious materials from the laboratory or equipment can introduce contaminants to the vaccine [17].

Current testing for adventitious viruses

Whatever the source of contamination, both vaccine manufacturers and regulators have a vested interest in ensuring that vaccines are free of adventitious agents. Testing for adventitious agents in master seed viruses is governed by the 9 Code of Federal Regulations [1]. Master seed viruses are the pure viral stocks that will be propagated to produce a vaccine. After a master seed virus is approved, a vaccine will not be tested again for adventitious agents because the finished product is often composed of several viruses and adjuvants that can complicate testing.

Requirements for master seed testing are determined by the species of the vaccine recipient as well as the animal origin of any products that were used during production. For example, if a poultry vaccine is produced in eggs, but fetal bovine serum was used at any point in production, it would have to undergo testing for poultry and bovine adventitious agents. Testing for specific viruses is done using detection techniques including fluorescent antibody staining and polymerase chain reaction (PCR). All mammalian master seed viruses are tested for rabies virus, reovirus, and bovine viral diarrhea virus. Testing for nonspecific adventitious viruses is done by inoculating the neutralized master seed virus into cell culture or eggs and observing the cells or eggs for signs of infection after several passages [19].

Although the scope of adventitious agent testing is fairly broad, it is well recognized that there are instances when it has failed to detect an adventitious virus. Cases from the human literature include the detection of SV40 virus in poliomyelitis virus vaccines and the detection of avian leukosis virus in yellow fever and measles vaccines [20, 21]. More recently, porcine circovirus-1 was detected in a rotavirus vaccine

[22]. Veterinary vaccines have also been found to be contaminated. Reticuloendotheliosis virus and Newcastle disease virus have been detected in live poultry vaccines [23, 24]. Bluetongue virus was found in a canine vaccine, causing abortions and deaths in pregnant bitches [25]. A bovine viral diarrhea virus (BVDV) vaccine strain transmission study was disrupted by a vaccine recall caused by the presence of an additional strain of BVDV in the vaccine [26]. As others have pointed out, the detection of adventitious agents in biologicals is rare due to good manufacturing practices, including testing of raw materials and substrates [27, 28]. However, the potential consequences of an adventitious virus being present in a vaccine warrant a high degree of caution.

In addition to these documented cases of adventitious agents slipping through the regulatory process, there are several general problems with the current testing methodology. Cell culture testing for adventitious viruses takes a minimum of three weeks to complete, and it is common to have to repeat testing due to difficulty neutralizing the master seed virus. If there were to be a disease outbreak and a vaccine needed to be licensed quickly, it would not be ideal for the master seed virus to have to go through several weeks of testing. In addition, it is possible that a replication-competent adventitious virus would not be able to grow in the cell cultures that are used for testing. Nucleic acid detection techniques like PCR are more able to detect virus in this situation, but due to practical constraints, PCR tests can only be conducted for a limited number of viruses. Therefore, there is a chance that traditional adventitious agent testing would miss a novel or unexpected virus. A virus detected by nonspecific testing techniques may take weeks to identify. These cases of contamination and general

disadvantages highlight the shortcomings of current methods of adventitious agent testing.

Alternative broad-spectrum strategies to detect adventitious viruses: microarray and next-generation sequencing

Microarray and next-generation sequencing (NGS) may be able to address some of the limitations of traditional adventitious agent testing. Results can be obtained in a few days, as opposed to weeks for cell culture testing. Both of these techniques, like PCR, are based on detection of viral nucleic acids to signify the presence of viral particles. The difference between these techniques and PCR, however, is that these techniques are broad-spectrum, meaning they can simultaneously detect a multitude of viruses. No prior knowledge about the sample is required for testing because the viral nucleic acid is amplified using random primers. This random primer based approach to amplification could almost certainly expand the range of viruses detected, as well as decreasing the time to results.

One specific case highlights the benefits of using random priming to screen for all viral nucleic acids. In 2004, a novel pestivirus, 'D32/00_HoBi', was isolated from fetal bovine serum (FBS) originating from Brazil [29]. The virus is extremely different from other pestiviruses in terms of both genome sequence and antigenic properties. These differences prompted Schirrmeier *et al*. to propose that the virus belongs to a new pestivirus species, bovine viral diarrhea virus 3 (BVDV-3) [29]. Another BVDV-3 isolate was discovered in Switzerland during routine screening of FBS [30]. The FBS was negative for BVD by virus isolation and ELISA, but the BVD PCR detected virus. In neither instance were the conventional 324/326 primers used, and further investigation determined these primers are incapable of HoBi virus amplification [30, 31]. Had the

laboratories had been using the conventional primers, the variant virus might not have been detected. Random primer amplification followed by microarray or NGS could have detected the variant viruses because sequence differences between the virus and the primer pairs would not impede amplification.

Although broad-spectrum nucleic acid detection could improve the range of detectable viruses by using random priming amplification, it may also complicate follow-up analysis. If viral nucleic acid is detected by microarray or NGS, the sample would have to undergo further testing to determine if replication-competent viral particles are present in the sample, or if not, whether any risk is incurred by the presence of viral nucleic acid. Many commonly used cell substrates for vaccine production contain endogenous viruses (usually retroviruses) or latent DNA viruses, which would be a source for viral nucleic acids in master seed stocks. For example, the feline endogenous retrovirus RD-114 was detected in canine vaccines [32]. All seed stocks contained RNA from the virus, but only 7/18 had infectious particles. Species-specific pathogenicity of any viruses found must also be considered. RD-114 is nonpathogenic in cats, but it can replicate in canine cell culture and therefore could potentially cause disease in dogs. Risk analyses will have to be carefully conducted for any viruses identified by microarray or NGS [33].

The benefits and concerns of using microarray and NGS are shared between both technologies, but they also have significant differences. One difference is in the timing of bioinformatics analysis. The bioinformatics analysis is done up-front to design a microarray. Microarray features are designed based on conserved regions of the genomes of interest, which are obtained from databases [34]. In contrast, all analysis on NGS data

is done after sequencing has been completed. Time to complete the test is another difference. It takes less time to complete a microarray test than NGS. Sensitivity and off-target tolerance is another difference between microarray and NGS [35]. PCR is a good example of a technology with very high sensitivity and low off-target tolerance. The sensitivity is high because minute quantities of DNA can be detected using PCR, but the off-target tolerance is low because small changes to the genome sequence in the primer-binding region may inhibit detection. On the other hand, NGS has low sensitivity and high off-target tolerance. The sensitivity is low because most of the reads (without any viral enrichment) will be eukaryotic or bacterial, which can adversely affect virus detection. The off-target tolerance is high because there is no targeting of detection with the use of primers. Microarrays fall in the middle. They have a higher off-target tolerance than PCR because they have features designed to target multiple points of many viruses. The sensitivity is lower than PCR, though, because of higher background fluorescence in the assay [35].

In addition to these differences, the type of data produced and method of analysis is different between the two technologies. Microarray produces a spreadsheet in which each feature's fluorescence is quantified. Microarray data is analyzed by clustering the positive features into groups that correspond to viruses. The end result is a list of possible viruses with the percentage of positive features for each virus. It is also possible to do statistical analysis by predicting virus-specific hybridization profiles to determine whether those percentages are higher than chance [36]. The specificity of identification is unique to each microarray, but typically it is less specific than NGS. NGS produces a very large text file with millions of 250 base pair DNA sequences. NGS data analysis

relies on matching the 250 base pair reads against a database of virus genome sequences. The results are the number of reads that align to a specific virus genome, the percentage of the genome that is covered by the reads, and the average depth of sequencing.

Review of previous research

Several previous studies have investigated the possibility of using microarray and NGS for adventitious agent testing. The most notable effort in the microarray arena has been the group from Lawrence Livermore National Laboratory with the development of their microbial detection array (MDA) [36, 37]. They surveyed seven publicly available vaccines for adventitious agents using the MDA and discovered that one of the rotavirus vaccines they tested was contaminated with porcine circovirus-1 [22]. The results were confirmed with NGS. The MDA was also used in a study to determine sensitivity of detection compared to quantitative PCR (qPCR) for five viruses [8]. It was concluded that the MDA has sensitivity rivaling qPCR if the background nucleic acids are reduced. In addition, it was determined that hybridization time can be reduced to as little as one hour, although with a small decrease in sensitivity [38].

Testing cell substrates for adventitious agents is another important application of NGS. The Food and Drug Administration (FDA) Center for Biologics Evaluation and Research has done research into using NGS to detect viruses in cell substrates [39]. They tested six cell lines known to express viral genes or particles using both 454 and Illumina sequencing. Prior to sequencing, the samples were amplified and sequencing adapters were added with degenerate oligonucleotide-primed PCR. This amplification strategy combined with viral enrichment and Illumina sequencing was found to be the most sensitive method. Additionally, a novel rhabdovirus in the Sf9 insect cell line was

discovered using the same amplification protocol and NGS [40]. Chinese hamster ovary cells have been involved in multiple bioreactor contaminations, and one manufacturer, Genzyme, was able to use NGS without amplification of the nucleic acid to determine the identity of the contaminant [10, 41]. In addition, they performed spike recovery experiments with six bacteriophage or viral vectors. The sensitivity with their method was quite high, less than one copy per cell for each sample, although their threshold for detection was set fairly low at twenty reads.

In addition to using NGS to screen cell substrates for viruses, a variety of experiments have been performed by several other researchers. Cabannes *et al.* have proposed a method to use different filtration and nucleic acid extraction protocols prior to NGS to determine if viral particles, viral genomes, or both, are present in a sample [9]. David Onions and John Kohlman developed a degenerate oligonucleotide-primed PCR amplification protocol for 454 sequencing with which they discovered a novel bovine parvovirus in bovine serum [42]. Gagnieur *et al*. sequenced several batches of bovine serum and trypsin using NGS in order to catalog the diversity found there [43]. Ng *et al*. used human cytomegalovirus as a model virus to compare detection by conventional cell culture methods, qPCR, and NGS [44]. These studies are only a small representation of the ways NGS is being studied for virus detection in biologics and in clinical samples.

The experiments detailed above are certainly valuable, but for technologies like microarray and NGS to gain traction in the global biologics regulatory arena, a coordinated effort by the regulatory community and industry must occur. This was the topic of the workshop “Advanced Technologies for Virus Detection in the Evaluation of Biologicals- Applications and Challenges” held in 2013 by the Parenteral Drug

Association (PDA) and the FDA [45]. One of the studies presented there by Modrof *et al*. attempts to address the problem of how to evaluate and standardize methods used by different researchers [7]. Four blinded samples, spiked at different concentrations into different media, were sent to four laboratories, which either used microarray, PCR-electrospray ionization mass spectrometry, or NGS to identify the samples. The majority of the samples were identified correctly regardless of technology (83% of the time), but false-positives were frequently reported (31% of the time). More of these types of studies will have to be coordinated as the field moves towards developing best practices for these methods. Currently only the human vaccine industry and regulators are involved in this working group, but it would be advantageous to include veterinary regulators in the interest of the One Health Initiative [46].

Computational considerations for implementing microarray or NGS for adventitious agent testing

The analysis of microarray and NGS data requires a much larger reliance on computational resources than typical laboratory techniques. A basic microarray analysis involves calculating the mean and standard deviation of the thousands of microarray features. The positive features are clustered into groups based on which viral genus or species each feature was designed to detect. More complicated algorithms exist to assign probability to the results [36, 47, 48]. They require determining predicted hybridization profiles for each virus that is represented on the array. Then, the actual hybridization profile is compared to the predicted one to determine a probability that the virus is present in the sample. These types of analyses are specific to each microarray design. For microarray analysis, it is helpful to have a bioinformatician involved, but the analysis can be done on a standard desktop computer.

NGS data analysis is considerably more complex. Compared to microarray, an NGS experiment generates several times more data. It is also a more open-ended analysis than microarray analysis. There are two overarching themes of NGS analysis: assembly and alignment [49]. Assembly algorithms are used to reconstruct full genome sequences represented by the short sequencing reads NGS produces. Assembly is primarily used when the virus being sequenced is from an unknown sample or when no reference genome is available. Alignment uses a reference genome and matches the short sequencing reads to their proper position along the genome. This generates additional information compared to assembly, including the position of single nucleotide polymorphisms (SNPs), depth of sequencing coverage, and percentage of genome coverage. Depending on the size of the data files containing the sequencing reads and complexity of the desired analysis, it could be done on a desktop computer with commercially available software, or it may require a multi-core computer and a custom bioinformatics analysis.

A challenge of bioinformatics analysis of viral NGS datasets is that typically there is a high degree of host-derived DNA that is sequenced along with the virus, so assembly will mostly produce contigs matching the host genome. In addition, in the case of adventitious agent testing, the identity of the adventitious agent may be unknown at the time of sequencing, so alignments are not possible. Therefore, databases such as NCBI GenBank are critical to the analysis of NGS data because they can be used to suggest an identity for each short read. Negative selection can be done using a database of the host genome to remove any reads that are not viral. Positive selection can likewise be done to select viral reads for analysis. The resulting reduced dataset is computationally more

manageable. Bekkari *et al*. discuss the development and use of such databases in their analyses [50]. However, this reliance on databases means that the quality of the analysis is dependent on the quality of the database being used. Tom Slezak discusses some of the challenges with developing and maintaining a high quality public database of sequence data, including the capture of associated metadata and dealing with changing nomenclature [34]. This topic will be of continuing relevance as the rate of deposition of data to public databases increases [51].

Finding ways to query the databases discussed above very quickly is also an active area of research. It is not feasible to use an algorithm like the Basic Local Alignment Search Tool (BLAST) to find a match for each of millions of sequencing reads in a database the size of GenBank because the analysis would run for days [52]. Therefore, alternative sequence classification strategies have been developed, and they have been reviewed previously [53]. Most recently taxonomic classifiers based on classifying smaller k-mers within the sequencing reads have been developed [54, 55]. These classifiers run very quickly because they use a modified database to search for exact matches between k-mers in the individual sequencing reads and k-mers in the database. They then use a taxonomy tree to identify the lowest common ancestor of all k-mers contained in each read. These sequencing read classifiers are useful to quickly identify which viruses may be contained in the sample, and identified reads can be isolated for further analysis.

Although the mechanics of analyzing microarray and NGS data are important, developing the bioinformatics expertise necessary to do these analyses is critical for success. Commercial analysis tools are able to help, but analyzing many samples that

way is not feasible. In addition, the tools available to perform NGS data analysis are changing rapidly, and integration of the new tools into commercial software packages can take years. Being able to use a basic scripting language opens up many more possibilities for analyzing data, as most software is released with only a command line interface. However, spending too much time writing a program to analyze data can take time away from data generation. Bioinformaticians at the FDA have developed a cloud-based platform to allow scientists without bioinformatics programming knowledge to analyze NGS data while still utilizing the latest tools. [56]. Computational power requirements and storage requirements are additional considerations. File transfers between collaborators are not a trivial issue, and currently the USDA is upgrading its network bandwidth to compensate for the increased need to transfer large files. Policies also need to be developed for the retention of sequencing data if it is to ever be reanalyzed with tools available in the future.

Moving forward: the role of microarray and NGS in federal adventitious agent testing

Much has been accomplished in the last few years to bring microarray and NGS closer to the goal of routine use for adventitious agent testing, but there are still many unanswered questions. The first major question is related to laboratory preparation of the virus sample prior to sequencing or hybridization to a microarray. Regulatory environments require the use of standardized protocols for all samples. Many researchers are using different approaches to enrich for viral nucleic acid, amplify viral nucleic acid, and fluorescently label nucleic acid or prepare sequencing libraries. Collaboration will be the key to establishing a method that is most sensitive for detecting any type of viral nucleic acid in master seed samples. Collaboration will allow for more studies similar to

Modrof *et al*. where different protocols are tested on the same samples, as well as testing the same protocol in multiple laboratories to ensure repeatability across different samples [7]. This issue of protocol development is made more difficult by the fact that it is a moving target, as laboratories in academia and industry are actively working to improve amplification and sequencing methodologies.

It is important to realize that complete validation will not be possible with either microarray or NGS, but reliability will have to increase for either of these to serve as an effective screening tool. Experiments should be carefully designed and realistically mimic conditions in which an adventitious agent might be detected. Unfortunately, although cost has decreased significantly for both technologies, it is still a consideration in designing experiments. Regulatory officials will have to commit to support the implementation of microarray and NGS, including the time and expense of thoroughly evaluating the methods prior to their use.

Regulatory officials also must support the hiring and training of personnel to analyze the data generated by microarray and NGS. The amount of data that these technologies produce is unparalleled by any other laboratory technique but is meaningless without significant intellectual investment in the analysis. IT infrastructure development is also critical, including purchasing high performance computers and storage to handle the terabytes of data produced.

The success of federal initiatives to use microarray or NGS as screening tools to expand the scope of adventitious agent testing is dependent on a high degree of collaboration within and between agencies and the biologics industry. It will involve laboratory staff, bioinformaticians, administrators, and IT professionals. The research

detailed here shows that microarray and NGS would be a beneficial addition to adventitious agent testing, but it is still very early in the developmental process. With a commitment to open communication and collaboration, it should be no problem to develop these technologies for adventitious agent testing in order to improve the health and safety of animal biologics.

References

[1] Viruses, Serums, Toxins, and Analogous Products; Organisms and Vectors. 9 Code of Federal Regulations2015.

[2] Erlandsson L, Rosenstierne MW, McLoughlin K, Jaing C, Fomsgaard A. The Microbial Detection Array Combined with Random Phi29-Amplification Used as a Diagnostic Tool for Virus Detection in Clinical Samples. PloS one. 2011;6:e22631.

[3] Hwang BH, Shin HH, Seo JH, Cha HJ. Specific Multiplex Analysis of Pathogens Using a Direct 165 rRNA Hybridization in Microarray System. Anal Chem. 2012;84:4873-9.

[4] Jiang YH, Guo Y, Wang P, Dong QF, Opriessnig T, Cheng JH, et al. A novel diagnostic platform based on multiplex ligase detection-PCR and microarray for simultaneous detection of swine viruses. Journal of virological methods. 2011;178:171-8.

[5] Palacios G, Quan PL, Jabado OJ, Conlan S, Hirschberg DL, Liu Y, et al. Panmicrobial oligonucleotide array for diagnosis of infectious diseases. Emerging Infect Dis. 2007;13:73-81.

[6] Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Moller N, et al. Rapid Whole-Genome Sequencing for Detection and Characterization of Microorganisms Directly from Clinical Samples. J Clin Microbiol. 2014;52:139-46.

[7] Modrof J, Berting A, Kreil TR. Parallel evaluation of broad virus detection methods. PDA journal of pharmaceutical science and technology / PDA. 2014;68:572-8.
[8] Feng SF, Jaing C, Thissen JB, Wang C, Duncan P. Modeling an approach to define sensitivity of viral detection in sample matrices-examples with microarray readout. PDA journal of pharmaceutical science and technology / PDA. 2014;68:589-94.

[9] Cabannes E, Hebert C, Eloit M. Whole genome: next-generation sequencing as a virus safety test for biotechnological products. PDA journal of pharmaceutical science and technology / PDA. 2014;68:631-8.

[10] Richards B, Cao S, Plavsic M, Pomponio R, Davies C, Mattaliano R, et al. Detection of adventitious agents using next-generation sequencing. PDA journal of pharmaceutical science and technology / PDA. 2014;68:651-60.

[11] Wang CJ, Feng SF, Duncan P. Defining a sample preparation workflow for advanced virus detection and understanding sensitivity by next-generation sequencing. PDA journal of pharmaceutical science and technology / PDA. 2014;68:579-88.

[12] Cheval J, Sauvage V, Frangeul L, Dacheux L, Guigon G, Dumey N, et al. Evaluation of High-Throughput Sequencing for Identifying Known and Unknown Viruses in Biological Samples. J Clin Microbiol. 2011;49:3268-75.

[13] Andre FE. Vaccinology: past achievements, present roadblocks and future promises. Vaccine. 2003;21:593-5.

[14] Lombard M, Pastoret PP, Moulin AM. A brief history of vaccines and vaccination. Revue Scientifique Et Technique-Office International Des Epizooties. 2007;26:29-48.

[15] Human Rabies. http://www.cdc.gov/rabies/location/usa/surveillance/human_rabies.html: Centers for Disease Control. January 26, 2015.

[16] Pastoret PP, Jones P. Veterinary vaccines for animal and public health. In: Schudel A, Lombard M, editors. Control of Infectious Animal Diseases by Vaccination2004. p. 15-29.

[17] Khan AS. Current testing methods and challenges for detection of adventitious viruses. PDA journal of pharmaceutical science and technology / PDA. 2011;65:627-33.

[18] Recommendations for the evaluation of animal cell cultures as cell substrates for the manufacture of biological medicinal products and for the characterization of cell banks. WHO Tech Rep Ser: World Health Organization (WHO); 2013.

[19] Marcus-Sekura C, Richardson JC, Harston RK, Sane N, Sheets RL. Evaluation of the human host range of bovine and porcine viruses that may contaminate bovine serum and porcine trypsin used in the manufacture of biological products. Biologicals. 2011;39:359-69.

[20] Hilleman MR. Discovery of Simian Virus 40 (SV40) and its relationship to poliomyelitis virus vaccines. In: Brown F, Lewis AM, editors. Simian Virus 401998. p. 183-90.

[21] Harris RJC, Doughert.Rm, Biggs PM, Payne LN, Goffe AP, Churchil.Ae, et al. Contaminant Viruses in Two Live Virus Vaccines Produced in Chick Cells. Journal of Hygiene-Cambridge. 1966;64:1-7.

[22] Victoria JG, Chunlin W, Jones MS, Jaing C, McLoughlin K, Gardner S, et al. Viral Nucleic Acids in Live-Attenuated Vaccines: Detection of Minority Variants and an Adventitious Virus. Journal of Virology. 2010;84:6033-40.

[23] Fadly AM, Witter RL, Smith EJ, Silva RF, Reed WM, Hoerr FJ, et al. An outbreak of lymphomas in commercial broiler breeder chickens vaccinated with a fowlpox vaccine contaminated with reticuloendotheliosis virus. Avian Pathol. 1996;25:35-47.

[24] Jorgensen PH, Handberg KJ, Ahrens P, Manvell RJ, Frost KM, Alexander DJ. Similarity of avian paramyxovirus serotype 1 isolates of low virulence for chickens obtained from contaminated poultry vaccines and from poultry flocks. Vet Rec. 2000;146:665-8.

[25] Wilbur LA, Evermann JF, Levings RL, Stoll IR, Starling DE, Spillers CA, et al. Abortion and Death in Pregnant Bitches Associated with a Canine Vaccine Contaminated with Bluetongue Virus. J Am Vet Med Assoc. 1994;204:1762-5.

[26] Palomares RA, Marley SM, Givens MD, Gallardo RA, Brock KV. Bovine viral diarrhea virus fetal persistent infection after immunization with a contaminated modified-live virus vaccine. Theriogenology. 2013;79:1184-95.

[27] Kerr A, Nims R. Adventitious Viruses Detected in Biopharmaceutical Bulk Harvest Samples over a 10 Year Period. PDA journal of pharmaceutical science and technology / PDA. 2010;64:481-5.

[28] Nims RW. Detection of adventitious viruses in biologicals - A rare occurrence. In: Sheets R, Petricciani J, editors. Vaccine Cell Substrates 20042006. p. 153-64.

[29] Schirrmeier H, Strebelow G, Depner K, Hoffmann B, Beer M. Genetic and antigenic characterization of an atypical pestivirus isolate, a putative member of a novel pestivirus species. J Gen Virol. 2004;85:3647-52.

[30] Peletto S, Zuccon F, Pitti M, Gobbi E, De Marco L, Caramelli M, et al. Detection and phylogenetic analysis of an atypical pestivirus, strain IZSPLV_To. Res Vet Sci. 2012;92:147-50.

[31] Vilcek S, Herring AJ, Herring JA, Nettleton PF, Lowings JP, Paton DJ. Pestiviruses Isolated From Pigs, Cattle and Sheep Can Be Allocated Into At Least 3 Genogroups Using Polymerase Chain-Reaction and Restriction-Endonuclease Analysis. Arch Virol. 1994;136:309-23.

[32] Yoshikawa R, Sato E, Miyazawa T. Presence of Infectious RD-114 Virus in a Proportion of Canine Parvovirus Isolates. J Vet Med Sci. 2012;74:347-50.

[33] Regulatory risk evaluation on finding an adventitious agent in a marketed vaccine. World Health Organization (WHO); 2014.

[34] Slezak T. Bacterial genome reference databases: progress and challenges. PDA journal of pharmaceutical science and technology / PDA. 2014;68:619-25.

[35] Kolman JL. Massively parallel sequencing for the detection of adventitious viruses. PDA journal of pharmaceutical science and technology / PDA. 2011;65:663-7.

[36] Gardner SN, Jaing CJ, McLoughlin KS, Slezak TR. A microbial detection array (MDA) for viral and bacterial detection. BMC genomics. 2010;11:668.

[37] Jaing C, Gardner S, McLoughlin K, Thissen JB, Slezak T. Detection of adventitious viruses from biologicals using a broad-spectrum microbial detection array. PDA journal of pharmaceutical science and technology / PDA. 2011;65:668-74.

[38] Thissen JB, McLoughlin K, Gardner S, Gu P, Mabery S, Slezak T, et al. Analysis of sensitivity and rapid hybridization of a multiplexed Microbial Detection Microarray. J Virol Methods. 2014;201:73-8.

[39] McClenahan SD, Uhlenhaut C, Krause PR. Optimization of virus detection in cells using massively parallel sequencing. Biologicals. 2014;42:34-41.

[40] Ma H, Galvin TA, Glasner DR, Shaheduzzaman S, Khan AS. Identification of a Novel Rhabdovirus in Spodoptera frugiperda Cell Lines. Journal of Virology. 2014;88:6576-85.

[41] Berting A, Farcet MR, Kreil TR. Virus Susceptibility of Chinese Hamster Ovary (CHO) Cells and Detection of Viral Contaminations by Adventitious Agent Testing. Biotechnol Bioeng. 2010;106:598-607.

[42] Onions D, Kolman J. Massively parallel sequencing, a new method for detecting adventitious agents. Biologicals. 2010;38:377-80.

[43] Gagnieur L, Cheyal J, Gratigny M, Hebert C, Muth E, Dumarest M, et al. Unbiased analysis by high throughput sequencing of the viral diversity in fetal bovine serum and trypsin used in cell culture. Biologicals. 2014;42:145-52.

[44] Ng SH, Azizi A, Edamura K, Malott RJ, Charlebois RL, Logvinoff C, et al. Preliminary Evaluation of Next-Generation Sequencing Performance Relative to qPCR and In Vitro Cell Culture Tests for Human Cytomegalovirus. PDA journal of pharmaceutical science and technology / PDA. 2014;68:563-71.

[45] Khan AS, Vacante DA. Introduction and workshop summary: advanced technologies for virus detection in the evaluation of biologicals-applications and challenges. PDA journal of pharmaceutical science and technology / PDA. 2014;68:546-7.

[46] One Health Initiative. http://www.onehealthinitiative.com/index.php. January 26, 2015.

[47] Allred AF, Wu GA, Wulan T, Fischer KF, Holbrook MR, Tesh RB, et al. VIPR: A probabilistic algorithm for analysis of microbial detection microarrays. BMC bioinformatics. 2010;11.

[48] Urisman A, Fischer KF, Chiu CY, Kistler AL, Beck S, Wang D, et al. E-Predict: a computational strategy for species identification based on observed DNA microarray hybridization patterns. Genome biology. 2005;6:R78.

[49] Flicek P, Birney E. Sense from sequence reads: methods for alignment and assembly. Nat Methods. 2009;6:S6-S12.

[50] Bekkari K, Shpungin J, Thompson JR. A practical approach to a viral detection pipeline using existing viral and non-viral sequence resources. PDA journal of pharmaceutical science and technology / PDA. 2014;68:595-601.

[51] Lathe W, Williams, J., Mangan, M., Karolchik, D. Genomic Data Resources: Challenges and Promises. Nature Education; 2008.

[52] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol. 1990;215:403-10.

[53] Bazinet AL, Cummings MP. A comparative evaluation of sequence classification programs. BMC bioinformatics. 2012;13:92.

[54] Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome biology. 2014;15:R46.

[55] Ames SK, Hysom DA, Gardner SN, Lloyd GS, Gokhale MB, Allen JE. Scalable metagenomic taxonomy classification using a reference genome database. Bioinformatics. 2013;29:2253-60.

[56] Wilson CA, Simonyan V. FDA's Activities Supporting Regulatory Application of "Next Gen" Sequencing Technologies. PDA journal of pharmaceutical science and technology / PDA. 2014;68:626-30.

CHAPTER 2. EVALUATION OF A VIRAL MICROARRAY AND NEXT-GENERATION SEQUENCING FOR DETECTION OF ADVENTITIOUS AGENTS IN VETERINARY VACCINES

A paper to be submitted to *Biologicals*

Kaitlin Brien[a,b]*, Roger Barrette[c], Tod Stuber[d], Joseph Hermann[a, e]

Abstract

Vaccination is an effective strategy for preventing and reducing the severity of disease in animals, but vaccines contaminated with adventitious agents are a health and liability hazard. All manufacturers seeking licensure of vaccines for sale in the U.S. must submit their master seed viruses for adventitious agent testing by the Center for Veterinary Biologics to ensure that no biological contaminants were inadvertently introduced. Primary methods to conduct adventitious agent testing for viruses include cell culture conducted according to the 9 Code of Federal Regulations and polymerase chain reaction tests. New technologies including microarray and next-generation sequencing (NGS) that are capable of detecting any virus promise to expand the scope of adventitious agent testing. A novel viral microarray called the extraneous veterinary viral microarray and next-generation sequencing (NGS) were tested for their ability to detect a panel of licensed vaccines and a panel of mock adventitious agents individually and in combination. NGS was found to be a reliable approach for detection of

[a] Center for Veterinary Biologics, U.S. Dept. of Agriculture, Ames, IA 50010, USA
[b] Primary author
[c] Foreign Animal Disease Diagnostic Laboratory, U.S. Dept. of Agriculture, Plum Island, NY 11957, USA
[d] National Veterinary Services Laboratories, U.S. Dept. of Agriculture, Ames, IA 50010, USA
[e] Present address: Boehringer Ingelheim Vetmedica, Inc., Ames, IA 50010, USA
* Corresponding author: (515)337-7124, Kaitlin.e.brien@aphis.usda.gov, 1920 Dayton Ave, Ames, IA 50010

this panel of viruses, although further work is needed to increase the confidence in results before unknown samples are tested.

1. Introduction

Several incidents, both historical and recent, of contaminants being detected in animal and human vaccines highlight the need for expansion of the scope of adventitious agent testing [1-5]. Traditional adventitious agent testing is performed according to the 9 Code of Federal Regulations and involves neutralizing the master seed virus and inoculating it onto a variety of permissive cell lines [6]. Examining cells for cytopathic effect or using various detection techniques determines the presence of virus. In addition, a panel of polymerase chain reaction (PCR) tests exists for commonly encountered contaminant viruses. However, cell culture detection is lengthy, labor intensive, and relies on the availability of a permissive cell line. PCR tests can only be performed for a selected number of agents. If a potential adventitious agent is not cell-culture adapted or is not targeted by a PCR, it may not be detected.

In order to address the limitations of traditional adventitious agent testing methods, nucleic acid detection strategies like microarray and next-generation sequencing (NGS) have been explored. Potential benefits and limitations of using these new technologies for adventitious agent testing are well documented [7-9]. These techniques are broad-spectrum, meaning they can simultaneously detect a multitude of viruses. No prior knowledge about the sample is required for testing because the viral nucleic acid is amplified using random primers. This random primer based approach to amplification could almost certainly expand the range of viruses detected, as well as decreasing the time to results.

The objective of this research was to investigate whether the extraneous veterinary viral microarray or NGS could be used as screening tools to supplement adventitious agent testing at the Center for Veterinary Biologics (CVB). To accomplish this, a panel of licensed vaccines and a panel of positive control viruses representing adventitious agents were selected for testing. Positive control viruses and vaccines were tested individually and in combinations using both microarray and NGS.

2. Materials and Methods

2.1 Viruses

Positive control viruses were used as mock adventitious viruses (obtained from CVB and the National Veterinary Services Laboratories). The viruses were selected to represent diverse virus families and common contaminants of biologicals. The mock adventitious viruses included bovine viral diarrhea virus-1 (BVDV), rabies virus (RABV), reovirus (REO), avian leukosis virus (ALV), chicken anemia virus (CAV), and Marek's disease virus (MDV) (Table 1). Vaccines were selected to represent diverse virus families. The vaccines selected were porcine respiratory and reproductive syndrome virus (PRRSV), porcine circovirus-2 (PCV), pseudorabies virus (PRV), infectious bronchitis virus (IBV), infectious bursal disease (IBD) virus, hemorrhagic enteritis virus (HEV), and fowlpox virus (FPV) (Table 2). Vaccines selected comprised live, modified live, and killed vaccines. Vaccines were hydrated using included diluent or 25 mL of water if no diluent was included.

2.2 Virus enrichment and nucleic acid extraction

Virus enrichment was performed to reduce the amount of host DNA in the viral samples and was done according to the protocol presented in Erlandsson, et al [10]. Briefly, viruses were mixed at a 1:1 ratio by volume (if applicable). A volume of 800-1000 µl of virus

was centrifuged for ten minutes at 14,000 RCF. Samples were then filtered through a 0.45 μm or 0.22 μm spin filter (Corning Costar Spin-X sterile cellulose acetate filter) for two minutes at 1,400 RCF. Six μl each of DNase I reaction buffer and DNase I (Invitrogen) were added to 400 μl of filtered sample and shaken at 300 rpm at room temperature for ninety minutes. Six μl of 25 mM ethylenediaminetetraacetic acid (EDTA) was added, and enzyme was inactivated at 65° C for ten minutes. Subsequently, total nucleic acid was extracted from 400 μl of treated sample using the Purelink Viral RNA/DNA Mini Kit (Invitrogen) according to manufacturer's instructions.

2.3 Multiple displacement amplification

Viral RNA/DNA was amplified from total nucleic acid using the Qiagen REPLI-g Cell WGA and WTA Kit. Briefly, samples were divided, and the RNA fraction underwent a DNA digestion followed by reverse transcription, ligation of cDNA fragments, and amplification using oligo(dT) primer, random primers, and Phi29 polymerase. The DNA fraction is modified for high efficiency ligation, followed by the ligation and amplification reactions used in the RNA fraction (Figure 1). After amplification, DNA was purified using the Qiagen QIAamp DNA Mini Kit with a modified protocol for purification of REPLI-g amplified DNA.

2.4 Novel microarray design- the extraneous veterinary viral microarray

Microarray design for specific viral agents was performed using an oligonucleotide selection algorithm based on a previous approach by Barrette et al [11]. The microarray feature selection algorithm was scripted in Python (v2.7) with database management through MySQL (v5.6). Variations of the originally described method included targeting a limited subset of the NCBI nucleotide database to include common viral adventitious agents of

interest. Briefly, full genome sequences were obtained from the NCBI nucleotide database for all target viruses of interest. Viral genome sequences were 'fragmented' into overlapping 60 nucleotide sequences to provide the initial candidate library. Candidate oligonucleotide sequences were filtered to remove redundancy, and compared using the tblastx algorithm of the Basic Local Alignment Search Tool (BLAST) to other related viral genomes within the NCBI nucleotide database to target coding regions for taxonomically conserved viral peptides [12]. Sequences identified by tblastx as unique, but similar (>90% identity at the nucleotide level) were included as unique microarray features in the final array design. This pipeline identified 109,795 oligonucleotides, which were submitted to Agilent using E-array software (Agilent Technologies), for synthesis on a custom 4 x 180K oligonucleotide array.

2.5 Microarray

Amplified DNA was fluorescently labeled using the Invitrogen BioPrime Array CGH Genomic Labeling System with Alexa-Fluor 555-aha-dUTP or Alexa-Fluor 647-aha-dUTP (Molecular Probes) according to manufacturer's instructions. Unbound fluors were removed with the included purification module. The Agilent Oligo aCGH/ChIP-on-Chip Hybridization Kit was used to prepare the hybridization solution. Fifty-five µl of blocking agent, 275 µl of hybridization buffer, 25 µl of salmon sperm DNA (Invitrogen), and 39 µl of labeled DNA were mixed and hybridized on the slide at 67° C for 4-16 hours. The slide was washed with the Agilent Oligo aCGH/ChIP-on-chip Wash Buffer Kit, and data was collected using Molecular Devices GenePix 4400 scanner with Genepix Pro 7 software.

Analysis was performed using a Microsoft Access database using a modified Genepix Array List (.gal) file which included fluorescence intensity and taxonomic information, including species and genus taxonomy IDs, for each microarray locus. Microarray loci with

mean fluorescence intensity greater than 3.5 standard deviations over the mean across all loci were considered ‘positive’ for this analysis. Finally, the percentage of positive features with a given species level NCBI taxonomy ID number were reported.

2.6 Next-generation sequencing

Amplified DNA was prepared for sequencing using the Nextera XT library preparation kit (Illumina) according to manufacturer’s instructions. Sequencing was performed using the Illumina MiSeq instrument with v2 reagents at the National Veterinary Services Laboratory (Ames, IA).

Raw fastq files from the sequencer were analyzed using Kraken software (v0.10.5-beta) using the standard database (all RefSeq bacterial and viral sequences) to sort reads into phylogenetic groups (Figure 2) [13]. Reads that were classified to an NCBI taxonomy identification that contained the word “virus” were selected for further analysis. Selected taxonomies that included the words “unclassified”, “DNA”, “RNA”, “transcribing”, “pea”, “invertebrate”, “molluscum”, “moorhen”, “T4”, “T5”, “T7”, “like”, “cryptic”, “globosa”, “occidentalis”, “viridae”, “mosaic”, “mottle”, “human”, “baculo” and “ichno” were excluded from further analysis. These groups were excluded either because they signified the broadest levels of the phylogenetic tree and were not useful for selecting a specific genome to analyze further, or they signified bacteriophages, insect viruses, plant viruses, or human viruses that were not relevant to this analysis. Clusters reflecting phylogenetic relationships of selected taxonomy ID numbers were generated (e.g. a genus level cluster included reads assigned by Kraken to that genus taxonomy ID and reads identified more specifically to its species and subspecies taxonomies). Reads in each cluster were assembled using ABySS (v1.5.2), and the resulting contigs were compared individually to previously sequenced viruses in the

NCBI nucleotide database using the blastn algorithm of the Basic Local Alignment Search Tool (BLAST) with default parameters [12, 14]. The genome that was identified as the top result by BLAST for the most contigs that also contained the word “genome” (most often used in the phrase “complete genome”) was then used as a reference genome to do an alignment using the full fastq file and the Burrows-Wheeler Aligner (v0.7.10, bwa-mem algorithm with a mismatch penalty of one) [15]. In the case of segmented genomes, all genomic segments were concatenated and treated as one reference genome. The number of aligned reads, percent of total reads that aligned, percent genome coverage, and average depth of sequencing were reported from this reference alignment. Accession numbers are reported in the results.

3. Results

3.1 Mock contaminants and vaccines tested individually

Positive control viruses were selected to act as mock contaminants (Table 1), and several licensed vaccines were selected (Table 2) for the experiment. The viruses were chosen with consideration given to representing diverse virus families. Nucleic acid from each mock contaminant or vaccine was extracted individually, amplified with multiple displacement amplification, and tested in tandem using the microarray and Illumina sequencing (Figure 1). Mock contaminants had between 4-100% positive microarray loci for the expected virus species (Table 3). Between 8-69% percent of total sequencing reads were identified as belonging to the respective mock contaminants (Table 3). Genome coverage for the mock contaminants ranged between 92-98%. Vaccines had between 30-99% positive microarray loci for the expected virus species (Table 4). Between 0.5-85% of total sequencing reads were identified as belonging to the respective vaccine virus. Genome

coverage for the vaccine ranged between 73-98%, although all vaccines but the Pseudorabies virus vaccine had at least 94% genome coverage.

3.2 Mock contaminated vaccine trials

Mock contaminants and vaccines were divided into mammalian and avian categories. A mock contaminant and a vaccine from the same category were mixed together 1:1 by volume prior to virus enrichment and nucleic acid extraction. The amplification procedure was followed as for individual viruses above (Figure 1). The three reovirus combination samples were not tested with the microarray because of the very low hybridization percentage of the individual Reovirus microarray.

It was determined based on the results of the individual virus trials that 30% positive microarray loci for the expected virus species would be required to consider a microarray trial successful. Using this cutoff, half of the avian trials and 1/6 of the mammalian trials using the microarray were able to detect both viruses (Table 5 and 6).

Genome coverage is an informative metric because it allows some indication of the quality of sequencing results and whether the entire virus genome is present or only a portion of it. Near-full genome coverage was required to determine successful detection of viruses by NGS because the experimental goal was to recover the full genome of the known viruses spiked into the sample. Based on the individual virus sequencing results, where the lowest percent genome coverage was 73%, 80% genome coverage for each virus was considered successful. Greater than 80% genome coverage was obtained for both viruses in 6/10 avian trials and 6/9 mammalian trials (Table 5 and 6). In the avian trials, 2/3 IBD contaminants had less than 15% genome coverage (Table 5). The two avian double-stranded DNA virus vaccines spiked with MDV did not meet the 80% genome coverage threshold for both viruses

(Table 5). The three mammalian trials containing PRV failed because they had low coverage of the PRV genome. This is not surprising because greater than 80% genome coverage was never obtained for PRV (Table 4). At least 92% genome coverage was generated for the mock contaminant in all mammalian trials.

4. Discussion

4.1 General conclusions

Using the amplification method outlined in Figure 1, both the extraneous veterinary viral microarray and NGS were able to detect all individual positive control and vaccine viruses, with the exception that the microarray failed to detect Porcine Reovirus and NGS failed to detect Pseudorabies virus. When the viruses were combined, the detection rate was considerably lower. Both viruses were detected in half to two-thirds of samples tested by sequencing and in a sixth to half of the samples tested by microarray. However, these conclusions are limited as it is difficult to determine an appropriate threshold for determining presence or absence of a given virus without a larger set of samples, including biological and technical replicates.

Generally speaking, it is more possible to draw conclusions based on the NGS data. The microarray is less able than NGS to make a distinction between true virus presence and non-specific binding of background nucleic acid. This microarray was designed with a select number of viruses represented in the hope of maximizing the sensitivity of virus detection. However, the large number of loci per virus resulted in a less stringent selection of oligonucleotide sequences, which may have increased the amount of non-specific binding. It is possible that an improved analysis algorithm to predict cross-species binding for the microarray would give more confidence in differentiating non-specific binding from true

virus presence. In contrast, the specificity of NGS results makes even very low percent genome coverage indicative of viral presence. NGS was able to make correct strain level assignments in several cases (data not shown). Background nucleic acid is less of a problem for NGS because it is either specifically identified or discarded in the NGS analysis. If further analysis is needed to examine the specific sequence of the virus present for some reason, existing data may be revisited. For these reasons, further efforts will be focused on developing a reliable method for virus detection by NGS.

4.2 Strengths and limitations of this study

Several recent studies have been published on the topic of using microarrays and/or NGS for the purpose of detecting adventitious agents [8, 16-20]. These studies have been very informative but have typically tested a very limited number of samples in favor of doing a titration to determine sensitivity. This information is critical, but it was determined that before those types of studies could be performed at CVB, we needed to determine the feasibility of using these technologies with a broader set of viruses. Rather than titrating single viruses, we tested the viruses primarily in combinations, mimicking the scenario in which adventitious agents are detected. Overall, the method used seems to have been successful. A few real adventitious agents were detected, including porcine circovirus 1 in a CVB positive control, Marek's disease virus in a CVB positive control, and endogenous retroviruses in several samples (data not shown).

Although our study used a wide variety of samples, there are several disadvantages of this approach. Using well-characterized samples with genome copy information determined by qPCR is the strategy implemented by most recent publications. We determined that it was not feasible to use qPCR for this number of viruses because CVB had no preexisting qPCRs

developed for these agents, nor reference standards with which to compare experimental samples. Infectivity concentrations were not compared because this information was not available for all samples, and the methods used to determine infectivity are not comparable across all samples (e.g. ELISA relative potency vs $TCID_{50}$). Our choice of samples reflected the most realistic mock contaminants and allowed us to test a wide variety of samples without regard to the availability of control data. However, not having control data makes it more difficult to draw conclusions about failed samples.

It was initially hypothesized that difference in genome type (DNA vs. RNA, double-stranded vs. single stranded, etc.) would influence amplification and therefore likelihood of detection. We encountered several combinations of viruses with different genome types in which both viruses were not detected at the 80% genome coverage threshold. Without rigorous genome concentration data and an even larger pool of virus samples, we are unable to determine if the failure of detection was caused by a virus specific problem, too large of a difference in concentration between viruses, or the difficulty in amplifying different genome types with one method.

The amplification method includes a ligation step prior to amplification with phi29 polymerase. During method development, the only goal of the study was to determine presence or absence of viruses. However, in some situations, the difference between the vaccine virus and an adventitious agent is as small as a few hundred base pair insertion of host DNA, as in the case of cytopathic and noncytopathic BVD viruses [21]. In this case, it is difficult for the data generated by this ligation-mediated amplification to be used for further analysis. Any host DNA in the sample could conceivably be ligated to the virus DNA in the same position as the expected insertion. In addition, contigs generated by ligation-amplified

DNA could be misleading. The DNA strand that is amplified is likely composed of DNA from multiple sources that had been ligated together. When it is broken in order to sequence, and contigs are assembled after sequencing (by matching overlapping ends of the sequencing reads), the resulting contigs may or may not be an accurate reflection of a single virus genome. A good reference genome would be needed to ensure that contigs are able to align to the genome appropriately. When in-depth analysis of the sequence data is required, it may be necessary to perform a virus specific amplification to obtain higher quality data. However, for assessing viral presence/absence, it was determined based on a limited number of samples that this method was the most robust (data not shown).

4.3 Remaining questions and future directions

Further investigation is certainly needed in a variety of areas. The first area is in method development. Using this random priming amplification method results in background nucleic acid from host being amplified along with the vaccine virus, potentially decreasing the likelihood of detecting adventitious agents. In general, 20-99% of NGS reads are non-viral. The low percent genome coverage of some virus combinations in our study indicates that more work is needed to develop a robust amplification method. In addition, viral enrichment procedures will need to be studied further. Some viruses, like fowlpox virus, are very large (0.3 μM), and detection was inhibited when virus was filtered through a 0.22 μM filter prior to nucleic acid extraction (data not shown). However, it's possible that using a .45 μM filter or no filtration would result in a higher degree of background nucleic acid that could inhibit virus detection. The best method for performing nuclease treatment prior to nucleic acid extraction will also need to be examined. Preliminary data suggests it is helpful to obtain a higher number of viral sequencing reads, but it was not studied extensively.

Another area needing further investigation is conducting comparisons of NGS for virus detection with traditional, validated methods. Performing titrations of several representative viruses in combination with other viruses will help to determine the relative sensitivities of NGS and traditional methods. One of the challenges of broad-spectrum detection methods is that there is no way to validate the method. It is not feasible to test all viruses, never mind all combinations of viruses, due to cost and time. However, additional studies should provide information about the risks and benefits of including NGS as a screening tool for adventitious agent testing.

The third area of concern is related to analysis of the sequencing data. One of the challenges of this project was to determine criteria to classify a successful test and an unsuccessful test. Whether the number of aligned reads (and percent of total reads) is an important metric to consider, or whether percent genome coverage is the more important indicator of presence of virus should be investigated. The percent genome coverage cutoff of 80% was chosen based on the sequencing results for single viruses. More work is needed to determine an appropriate percent coverage cutoff when an adventitious agent contaminates a vaccine at a very low concentration.

Detection of nucleic acid does not necessarily mean that infectious viral particles are present. Therefore, much time could be wasted on investigating nonreplicative adventitious viruses that would not have been detected using traditional methods. Regulatory officials will need to establish parameters for additional testing and determine where NGS fits into the adventitious agent testing landscape.

This preliminary study on the feasibility of using microarray and NGS at CVB for the purpose of detecting adventitious agents in vaccine master seeds presents evidence that NGS

is especially worthy of additional investigation. All mock contaminant viruses were detected at some level in combination with a vaccine, although several did not meet the 80% genome coverage threshold. In addition, NGS is capable of providing more information than the microarray about any viruses found. Once a robust amplification method is determined, all routine samples that are tested in CVB should be tested with the NGS protocol in order to start building a larger pool of data to make regulatory decisions about the future of NGS for adventitious agent testing. It may be difficult to reach the level of confidence required to rule a sample negative for virus by NGS alone, but NGS can supplement traditional testing. NGS platforms continue to improve library preparation protocols and increase their depth of sequencing, which should aid in the detection of low concentration viruses. Research is ongoing on how NGS can be used by the Center for Veterinary Biologics to improve the safety of veterinary biologics.

Acknowledgments

I wish to thank Dr. Roger Barrette for his work on designing the microarray and providing technical assistance, Tod Stuber for his help with analyzing NGS data, Patrick Camp and David Farrell for their technical assistance with NGS, Dr. Joseph Hermann and Alethea Fry for their advice in experimental design, and Alexa Scupham for critical reading of this manuscript.

The views expressed in this material are those of the authors and do not reflect the official policy or position of the U.S. government or the Department of Agriculture.

References

[1] Victoria JG, Chunlin W, Jones MS, Jaing C, McLoughlin K, Gardner S, et al. Viral Nucleic Acids in Live-Attenuated Vaccines: Detection of Minority Variants and an Adventitious Virus. Journal of Virology. 2010;84:6033-40.

[2] Wilbur LA, Evermann JF, Levings RL, Stoll IR, Starling DE, Spillers CA, et al. Abortion and Death in Pregnant Bitches Associated with a Canine Vaccine Contaminated with Bluetongue Virus. J Am Vet Med Assoc. 1994;204:1762-5.

[3] Yoshikawa R, Sato E, Miyazawa T. Presence of Infectious RD-114 Virus in a Proportion of Canine Parvovirus Isolates. J Vet Med Sci. 2012;74:347-50.

[4] Farsang A, Kulcsar G. Extraneous agent detection in vaccines - A review of technical aspects. Biologicals. 2012;40:225-30.

[5] Marcus-Sekura C, Richardson JC, Harston RK, Sane N, Sheets RL. Evaluation of the human host range of bovine and porcine viruses that may contaminate bovine serum and porcine trypsin used in the manufacture of biological products. Biologicals. 2011;39:359-69.

[6] Viruses, Serums, Toxins, and Analogous Products; Organisms and Vectors. 9 Code of Federal Regulations2015.

[7] Onions D, Kolman J. Massively parallel sequencing, a new method for detecting adventitious agents. Biologicals. 2010;38:377-80.

[8] Cheval J, Sauvage V, Frangeul L, Dacheux L, Guigon G, Dumey N, et al. Evaluation of High-Throughput Sequencing for Identifying Known and Unknown Viruses in Biological Samples. J Clin Microbiol. 2011;49:3268-75.

[9] McClenahan SD, Krause PR. The potential role of advanced technologies for virus detection in development and regulation of vaccines. PDA journal of pharmaceutical science and technology / PDA. 2014;68:552-5.

[10] Erlandsson L, Rosenstierne MW, McLoughlin K, Jaing C, Fomsgaard A. The Microbial Detection Array Combined with Random Phi29-Amplification Used as a Diagnostic Tool for Virus Detection in Clinical Samples. PloS one. 2011;6:e22631.

[11] Barrette RW, Metwally SA, Rowland JM, Xu L, Zaki SR, Nichol ST, et al. Discovery of swine as a host for the Reston ebolavirus. Science. 2009;325:204-6.

[12] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol. 1990;215:403-10.

[13] Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome biology. 2014;15:R46.

[14] Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: A parallel assembler for short read sequence data. Genome Res. 2009;19:1117-23.

[15] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754-60.

[16] Modrof J, Berting A, Kreil TR. Parallel evaluation of broad virus detection methods. PDA journal of pharmaceutical science and technology / PDA. 2014;68:572-8.

[17] Feng SF, Jaing C, Thissen JB, Wang C, Duncan P. Modeling an approach to define sensitivity of viral detection in sample matrices-examples with microarray readout. PDA journal of pharmaceutical science and technology / PDA. 2014;68:589-94.

[18] Cabannes E, Hebert C, Eloit M. Whole genome: next-generation sequencing as a virus safety test for biotechnological products. PDA journal of pharmaceutical science and technology / PDA. 2014;68:631-8.

[19] Richards B, Cao S, Plavsic M, Pomponio R, Davies C, Mattaliano R, et al. Detection of adventitious agents using next-generation sequencing. PDA journal of pharmaceutical science and technology / PDA. 2014;68:651-60.

[20] Wang CJ, Feng SF, Duncan P. Defining a sample preparation workflow for advanced virus detection and understanding sensitivity by next-generation sequencing. PDA journal of pharmaceutical science and technology / PDA. 2014;68:579-88.

[21] Kummerer BM, Tautz N, Becher P, Thiel HJ, Meyers G. The genetic basis for cytopathogenicity of pestiviruses. Vet Microbiol. 2000;77:117-28.

Table 1. Mock contaminants selected represent diverse viral families and genome types.

Virus family	Virus	Abbreviation	Genome Type	Genome Size
Retroviridae	Avian leukosis virus	ALV	ssRNA (rt)	7.5 kb
Circoviridae	Chicken anemia virus	CAV	ssDNA	2.3 kb
Herpesviridae	Marek's disease virus	MDV	dsDNA	166 kb
Flaviviridae	Bovine viral diarrhea virus	BVDV	ssRNA (+)	12.6 kb
Rhabdoviridae	Rabies virus	RABV	ssRNA (-)	11.9 kb
Reoviridae	Porcine reovirus	REO	dsRNA	20 kb (segmented)

Table 2. Vaccines selected represent diverse viral families and genome types.

Virus family	Virus	Abbreviation	Genome Type	Genome Size
Coronaviridae	Infectious bronchitis virus	IBV	ssRNA (+)	27.5 kb
Birnaviridae	Infectious bursal disease virus	IBD	dsRNA	5.9 kb (segmented)
Adenoviridae	Hemorrhagic enteritis virus	HEV	dsDNA	26.3 kb
Poxviridae	Fowlpox virus	FPV	dsDNA	288.5 kb
Circoviridae	Porcine circovirus	PCV	ssDNA	1.8 kb
Herpesviridae	Pseudorabies	PRV	dsDNA	143.4 kb
Arteriviridae	Porcine reproductive and respiratory syndrome virus	PRRSV	ssRNA (+)	15.4 kb

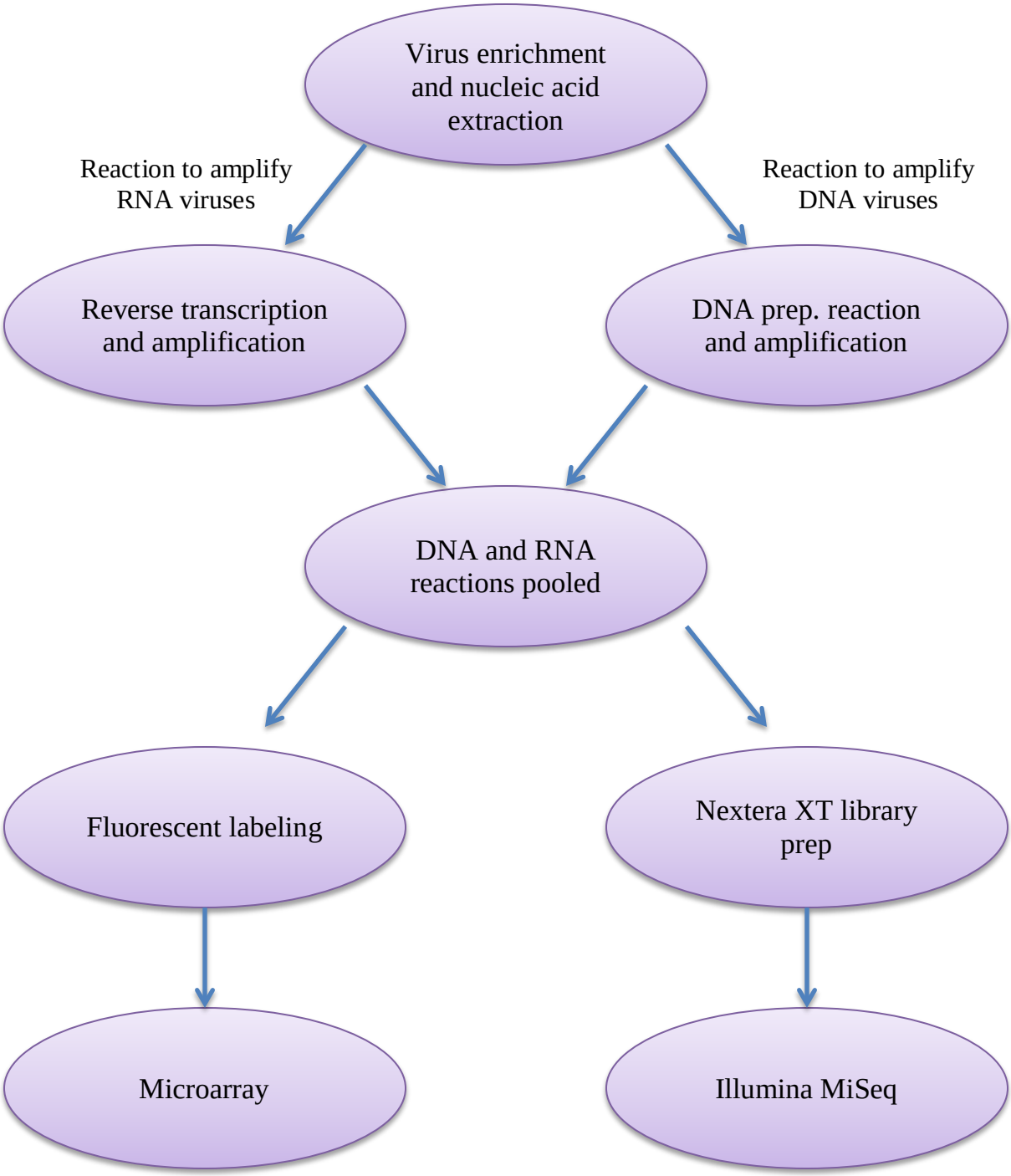

Figure 1. Sample preparation workflow followed to amplify viral RNA/DNA prior to microarray or sequencing.

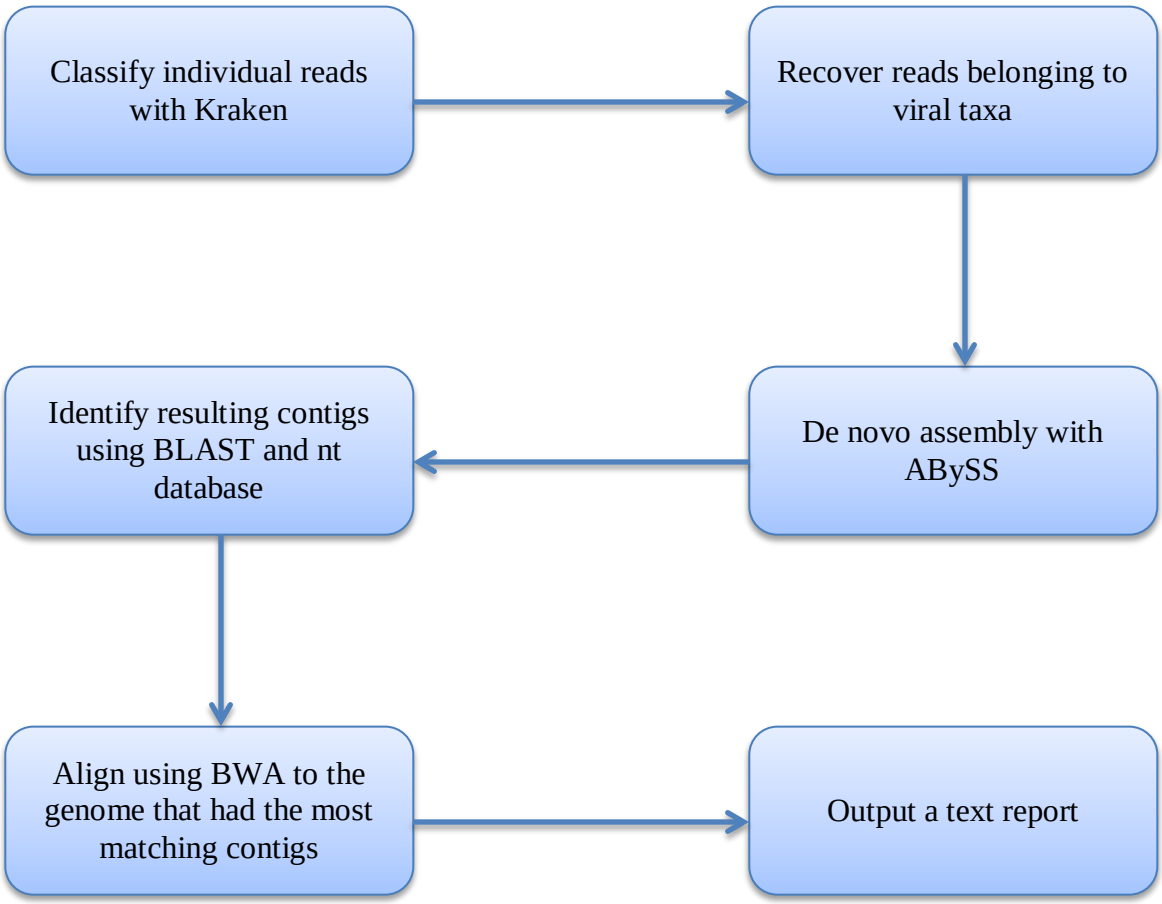

Figure 2. NGS data analysis workflow followed starting with raw FASTQ files from the sequencing instrument.

Table 3. Mock contaminants tested independently by microarray and sequencing.

Sample	Species Percentage of Positive Microarray Loci	Number of NGS Reads (% of total)	Percent Genome Coverage	NCBI Accession number
ALV	98	504,284 (23% total)	97.7	KF866225
CAV	100	161,626 (8% total)	98.5	DQ991394
MDV	35	209,826 (11% total)	95.2	HQ840738
BVDV	75	957,622 (43% total)	98.6	M31182
RABV	72	527,502 (21% total)	98.5	JQ944709
REO	4	787,590 (69% total)	92.3	JX415465-JX415474

Abbreviations: ALV (avian leukosis virus), CAV (chicken anemia virus), MDV (Marek's disease virus), BVDV (bovine viral diarrhea virus), RABV (rabies virus), REO (porcine reovirus)

Table 4. Licensed vaccines tested independently by microarray and sequencing.

Sample	Species Percentage of Positive Microarray Loci	Number of NGS Reads (% of total)	Percent Genome Coverage	NCBI Accession number
IBV	63	1,514,952 (62% total)	98.6	KJ435286
IBD	53	13,298 (0.5% total)	98.4	NC_004178-NC_004179
HEV	99	1,245,860 (85% total)	98.6	AF074946
FPV	30	520,740 (25% total)	94.9	AF198100
PCV	39	75,698 (6% total)	96.8	KJ128274
PRV	31	25,354 (2% total)	73.2	KJ717942
PRRSV	82	337,246 (19% total)	98.6	KF771273

Abbreviations: IBV (infectious bronchitis virus), IBD (infectious bursal disease virus), HEV (hemorrhagic enteritis virus), FPV (fowlpox virus), PCV (porcine circovirus), PRV (pseudorabies virus), PRRSV (porcine reproductive and respiratory syndrome virus)

Table 5. Avian mock contaminant and vaccine combinations tested by microarray and sequencing.

Mock Contaminant (MC)	Vaccine (V)	Species Percentage of Positive Microarray Loci	Number of NGS Reads (% of total)	Percent Genome Coverage
ALV	IBV	MC: 78 V: 47	MC: 64,706 (4%) V: 567,518 (33%)	MC: 97.4 V: 98.6
	IBD	MC: 0 V: 55	MC: 250 (0.01%) V: 5,176 (0.3%)	MC: 89.6 V: 98.4
	HEV	MC: 30 V: 83	MC: 3,738 (0.7%) V: 80,078 (14%)	MC: 94.3 V: 98.4
CAV	IBV	MC: 97 V: 30	MC: 115,386 (7%) V: 463,568 (28%)	MC: 98.5 V: 98.6
	IBD	MC: 0 V: 3	MC: 2 (0.00%) V: 586 (0.04%)	MC: 14.4 V: 95.6
	HEV	MC: 100 V: 97	MC: 59,438 (2%) V: 184,362 (5%)	MC: 98.5 V: 98.6
MDV	IBV	MC: 52 V: 78	MC: 125,008 (5%) V: 1,001,312 (40%)	MC: 88.9 V: 98.6
	IBD	MC: 0.03 V: 25	MC: 380 (0.04%) V: 3,288 (0.3%)	MC: 10.7 V: 98.4
	HEV	MC: 1 V: 34	MC: 3,684 (0.5%) V: 33,588 (4%)	MC: 63.0 V: 98.3
	FPV	MC: 35 V: 0.5	MC: 104,072 (9%) V: 14,304 (1%)	MC: 94.2 V: 53.1

Abbreviations: ALV (avian leukosis virus), CAV (chicken anemia virus), MDV (Marek's disease virus), IBV (infectious bronchitis virus), IBD (infectious bursal disease virus), HEV (hemorrhagic enteritis virus), FPV (fowlpox virus)

Table 6. Mammalian mock-contaminant and vaccine combinations tested by microarray and sequencing.

Mock Contaminant (MC)	Vaccine (V)	Species Percentage of Positive Microarray Loci	Number of NGS Reads (% of total)	Percent Genome Coverage
BVDV	PCV	MC: 66 V: 21	MC: 543,866 (31%) V: 12,366 (0.7%)	MC: 93.7 V: 98.5
	PRV	MC: 39 V: 6	MC: 51,174 (3%) V: 11,534 (0.7%)	MC: 93.7 V: 52.2
	PRRSV	MC: 57 V: 68	MC: 172,478 (17%) V: 46,078 (5%)	MC: 98.6 V: 98.6
REO	PCV	N/A	MC: 474,270 (31%) V: 49,540 (3%)	MC: 92.4 V: 95.4
	PRV	N/A	MC: 660,866 (44%) V: 101,750 (7%)	MC: 92.4 V: 70.0
	PRRSV	N/A	MC: 1,118,320 (54%) V: 386,108 (19%)	MC: 92.3 V: 98.6
RABV	PCV	MC: 53 V: 0	MC: 50,390 (3%) V: 1,102 (0.07%)	MC: 98.0 V: 81.0
	PRV	MC: 74 V: 4	MC: 142,032 (12%) V: 6,944 (0.6%)	MC: 98.5 V: 34.4
	PRRSV	MC: 60 V: 16	MC: 90,086 (6%) V: 11,050 (0.8%)	MC: 98.5 V: 98.6

Abbreviations: BVDV (bovine viral diarrhea virus), REO (porcine reovirus), RABV (rabies virus), PCV (porcine circovirus), PRV (pseudorabies virus), PRRSV (porcine reproductive and respiratory syndrome virus)

CHAPTER 3. EVALUATION OF A VIRAL MICROARRAY AND NEXT-GENERATION SEQUENCING FOR DETECTION OF BOVINE VIRAL DIARRHEA VIRUSES

Abstract

Bovine viral diarrhea viruses are common contaminants in vaccines, as most viruses are grown in cell culture with the use of fetal bovine serum. Two major genotypes are present in North America, type 1 and type 2. A novel viral microarray, the extraneous veterinary viral microarray, and next-generation sequencing (NGS) were previously tested for their ability to detect combinations of unrelated viruses. This study investigated the ability of the microarray and NGS to detect and distinguish between related viruses. Both the microarray and NGS were able to distinguish the two genotypes when presented individually, and NGS was able to give strain level identification. Both the microarray and NGS were also able to detect both genotypes when they were tested in combination. Determining the presence of multiple strains of the same virus genotype requires further investigation.

1. Introduction

Bovine viral diarrhea (BVD) viruses have a single stranded RNA genome and belong to the Flaviviridae family in the *Pestivirus* genus [1]. BVD viruses are a genetically and antigenically diverse group, and they infect ruminants primarily [1]. Two biotypes of the virus exist, cytopathic virus, which causes cell lysis when grown in cell culture, and noncytopathic virus, which does not [2]. Persistently infected cows are born when they are exposed to noncytopathic BVDV in the first 120 days of gestation, and a recombination event generating a cytopathic strain of the virus can trigger fatal mucosal

disease [3, 4]. The prevalence of persistently infected animals is estimated to be between 0.5-2%, but 60-85% of cattle have antibody for BVDV, showing past exposure [5].

The high prevalence of BVDV in cattle populations means that it is a very common contaminant found in fetal bovine serum (FBS). FBS is used as a nutrient source to support growth of cell culture and virus during vaccine production. If contaminated FBS is used during vaccine production, the final product may also be contaminated with BVDV. All mammalian master seed viruses are tested for BVDV for this reason. One study tested one thousand lots of FBS for BVDV and found 203 lots were contaminated [6]. More recently, atypical (non-type 1 or type 2) BVD viruses have been isolated from FBS [7-9]. The conventional PCR tests used to detect BVD viruses may be unsuited to detect these atypical viruses because mismatches in primer binding regions can inhibit amplification [7].

Microarray and next-generation sequencing (NGS) can be used in conjunction with random primer amplification to supplement other methods for BVDV detection. This random primer amplification method is well suited for detecting emerging BVD viruses because virus can be amplified without prior knowledge of the sequence. In addition, NGS can provide a specific identification if the genome sequence is available in a public database. The purpose of this research was to determine if a novel viral microarray and NGS could be used to detect both BVDV-1 and BVDV-2. The viruses were examined individually to determine the ability of the microarray and NGS to distinguish between genotypes. In addition, the viruses were combined to test the ability of the microarray and NGS to determine the presence of multiple viruses.

2. Materials and Methods

2.1 Viruses

BVDV-1 strain NADL, BVDV-1 strain Singer, and BVDV-2 strain 125c were obtained from CVB stocks. Complete genome sequences for type 1 BVD viruses were available publically from NCBI GenBank (NADL: accession M31182.1, Singer: accession DQ088995.2). A complete genome sequence for BVDV-2 strain 125c was made available by Dr. John Neill, Agricultural Research Service.

2.2 Virus amplification, microarray, and NGS

Materials and methods were the same as presented in chapter 2.

2.3 Data analysis

Analysis of microarray data was done according to the method presented in chapter 2. The NGS data analysis workflow presented in chapter 2 was done, with additional alignments generated to include the other BVDV strains detailed in section 2.1.

3. Results

3.1 BVDV-1 and BVDV-2 tested individually

Both BVDV-1 NADL and BVDV-2 were tested individually to determine relative cross-detection using the microarray and NGS (Table 7). The microarray had a higher percentage of true positive loci for the expected viruses than false positive loci for the absent viruses, with a difference of at least 58%. Both trials had false positive loci percentages less than the 30% cutoff discussed in chapter 2 for the absent viruses. NGS data showed that there were at least 75% fewer reads that aligned to the genome sequence of the absent virus, and at least thirteen percentage points lower genome coverage. However, the BVDV-2 virus sample had greater than 80% genome coverage for BVDV-

1, suggesting presence of BVDV-1 according to the criteria established in chapter 2. The relatively high number of BVDV-1 reads in the BVDV-2 sample may be attributed to barcode cross-talk from the sequencer, as BVDV-1 isolates were run on the sequencer at the same time [10]. This is supported by the presence of 62,036 reads that aligned to the BVDV-2 virus genome (covering 89.4% of the genome) in the sample with both strains of BVDV-1 virus. The BVDV-1 NADL only sample presented here was sequenced separately with unrelated viruses, and the percent genome coverage for BVDV-2 is below the percent identity shared between BVDV-1 and BVDV-2.

3.2 BVDV-1 and BVDV-2 tested in combination

BVDV-1 NADL and BVDV-2 were combined 1:1 by volume prior to extraction and amplified according to the procedure in Figure 1. The results show that both BVDV-1 and BVDV-2 could be detected (Table 8). The microarray had a high percentage of positive loci for both viruses: 72% of BVDV-1 loci were positive and 86% for BVDV-2 loci were positive. This is in contrast to the results when only one virus was present, where the absent virus had a percentage of positive loci below the 30% threshold. The NGS results for this sample showed that there were approximately 400,000 reads that aligned to each virus genome, and greater than 97% genome coverage for each virus. This is suggestive of the presence of both viruses, as an alignment between the genome sequences of BVDV-2 125c and BVDV-1 NADL using the megablast algorithm of the Basic Local Alignment Search Tool showed a 73% identity between the two viruses [11].

3.3 Two strains of BVDV-1 tested in combination

Two historically distinct strains of BVDV-1, BVDV-1 strain NADL and BVDV-1 strain Singer, were tested as above (Table 9). The results of this test are more difficult to

interpret, as the microarray was not designed to give strain level distinction. Seventy-one percent of BVDV-1 species-specific loci were positive, but 100% of BVDV-1 NADL strain-specific loci were positive. BVDV-1 Singer was not used as a reference genome to design the array, so strain specific data was not available. The NGS data showed that each strain had at least 473,000 aligned reads and greater than 98.4% genome coverage. However, the standard analysis detailed in Figure 2 failed to detect BVDV-1 NADL, as more contigs were assembled matching BVDV-1 Singer and only the top result is chosen for alignment. An alignment of the two sequences using megablast showed the two strains share a 96% identity, so it is difficult to determine if the high level of genome coverage observed is caused by the presence of both viruses or the fact that many reads align to both genomes.

4. Discussion

Both the microarray and NGS were able to distinguish between BVDV-1 and BVDV-2 viruses when they were tested individually, as evidenced by the higher percentage of positive microarray loci and higher percentage of genome coverage for the present virus than the absent virus. Both technologies were also able to detect BVDV-1 and BVDV-2 in combination, as both viruses had high percentages of positive microarray loci and high percentages of genome coverage. However, in the case of two strains of BVDV-1 tested in combination, the microarray was not able to detect both strains, as one of the strains was not used in the design process for the array. NGS data showed a high percentage of genome coverage for both strains, which in this case indicates the presence of both viruses. However, it was known in advance which virus strains were present, so a targeted approach could be used to generate genome alignments.

Alignment may not be the best tool to determine absence of a related virus, as many reads from the present virus will also align to the related genome. This is especially relevant in the case highlighted here of attempting to detect multiple strains of BVDV-1 virus in one sample. They share a very high percent identity (96%), so the same NGS reads will likely align to both genomes. In the case of an unknown sample, it would be very difficult to determine whether multiple viruses were present without targeting specific conserved loci that distinguish the viruses. This is the approach used by most BVDV PCRs, which target the conserved 5' un-translated region to distinguish BVDV-1 from BVDV-2 [12, 13]

Further work will be needed to determine if other related viruses tested individually can be distinguished by microarray and NGS. Determining the presence of multiple viruses in a sample will require further investigation into specific phylogenies of the viruses of interest to determine conserved regions that are diagnostic of viral presence. This task is only suited to NGS data because the microarray may not contain oligonucleotides matching those conserved regions. NGS is also the platform most capable of detecting atypical BVD viruses because new BVDV sequences can be integrated into analyses as they are added to public databases. The NGS data analysis protocol currently in use seems to be most useful for identifying the presence of unrelated viruses. Detecting BVD viruses in master seed viruses is of high importance, and research is ongoing to determine how to best utilize NGS for this purpose.

References

[1] Ridpath JF. BVDV genotypes and biotypes: practical implications for diagnosis and control. Biologicals. 2003;31:127-31.

[2] Kummerer BM, Tautz N, Becher P, Thiel HJ, Meyers G. The genetic basis for cytopathogenicity of pestiviruses. Vet Microbiol. 2000;77:117-28.

[3] Meyers G, Thiel HJ. Molecular characterization of pestiviruses. Advances in Virus Research, Vol 47. 1996;47:53-118.

[4] Darweesh MF, Rajput MKS, Braun LJ, Ridpath JF, Neill JD, Chase CCL. Characterization of the cytopathic BVDV strains isolated from 13 mucosal disease cases arising in a cattle herd. Virus Res. 2015;195:141-7.

[5] Houe H. Epidemiological features and economical importance of bovine virus diarrhoea virus (BVDV) infections. Vet Microbiol. 1999;64:89-107.

[6] Bolin SR, Ridpath JF. Prevalence of bovine viral diarrhea virus genotypes and antibody against those viral genotypes in fetal bovine serum. Journal of Veterinary Diagnostic Investigation. 1998;10:135-9.

[7] Schirrmeier H, Strebelow G, Depner K, Hoffmann B, Beer M. Genetic and antigenic characterization of an atypical pestivirus isolate, a putative member of a novel pestivirus species. J Gen Virol. 2004;85:3647-52.

[8] Stahl K, Beer M, Schirrmeier H, Hoffmann B, Belak S, Alenius S. Atypical 'HoBi'-like pestiviruses-Recent findings and implications thereof. Vet Microbiol. 2010;142:90-3.

[9] Peletto S, Zuccon F, Pitti M, Gobbi E, De Marco L, Caramelli M, et al. Detection and phylogenetic analysis of an atypical pestivirus, strain IZSPLV_To. Res Vet Sci. 2012;92:147-50.

[10] Mir K, Neuhaus K, Bossert M, Schober S. Short Barcodes for Next Generation Sequencing. PloS one. 2013;8.

[11] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol. 1990;215:403-10.

[12] Vilcek S, Herring AJ, Herring JA, Nettleton PF, Lowings JP, Paton DJ. Pestiviruses Isolated From Pigs, Cattle and Sheep Can Be Allocated Into At Least 3 Genogroups Using Polymerase Chain-Reaction and Restriction-Endonuclease Analysis. Arch Virol. 1994;136:309-23.

[13] Letellier C, Kerkhofs P, Wellemans G, Vanopdenbosch E. Detection and genotyping of bovine diarrhea virus by reverse transcription-polymerase chain amplification of the 5 ' untranslated region. Vet Microbiol. 1999;64:155-67.

Table 7. Cross-detection of bovine viral diarrhea virus 1 (BVDV-1) and bovine viral diarrhea virus 2 (BVDV-2) in samples in which only one virus is present.

Sample	Species Percent Positive Microarray Loci	Number of NGS Reads Identified by Kraken	Number of Aligned NGS Reads	Percent Genome Coverage
BVDV-1	BVDV-1: 75 BVDV-2: 17	BVDV-1: 953,634 BVDV-2: 2	BVDV-1: 957,622 BVDV-2: 130,046	BVDV-1: 98.6 BVDV-2: 63.3
BVDV-2	BVDV-2: 89 BVDV-1: 21	BVDV-2: 843,360 BVDV-1: 50,972	BVDV-2: 1,053,050 BVDV-1: 263,494	BVDV-2: 97.5 BVDV-1: 84.8

Table 8. Detection of BVDV-1 and BVDV-2 viruses in a sample in which both viruses are present.

Virus 1	Virus 2	Species Percentage of Positive Microarray Loci	Number of Aligned NGS Reads	Percent Genome Coverage
BVDV-1 strain NADL	BVDV-2 strain 125	BVDV-1: 72 BVDV-2: 86	BVDV-1: 441,850 BVDV-2: 399,394	BVDV-1: 98.6 BVDV-2: 97.5

Table 9. Detection of two strains of BVDV-1 in a sample in which both strains are present.

Virus 1	Virus 2	Species Percentage of Positive Microarray Loci	Number of Aligned NGS Reads	Percent Genome Coverage
BVDV-1 strain NADL	BVDV-1a strain Singer	BVDV-1 (all): 71 NADL: 100	NADL: 473,410 Singer: 473,896	NADL: 98.6 Singer: 98.5

CHAPTER 4. GENERAL CONCLUSIONS

The research presented in this thesis shows that the extraneous veterinary viral microarray presented here and NGS are tools worthy of further investigation for adventitious agent testing. NGS is especially useful because it is possible to capitalize on the increasing volume of DNA sequences being added to public databases. For this reason, it is more feasible to detect novel and emerging viruses using NGS. Incorporating novel and emerging viruses would likely require a redesign of the microarray. NGS also has the ability to identify contaminants with strain level specificity.

Nucleic acid from mock contaminant and vaccine mixtures was detected in all samples, although several did not meet the high percentage of positive microarray loci or high percentage of genome coverage criteria. Additional investigation is needed to determine an appropriate threshold for detection. BVDV-1 and BVDV-2 could also be distinguished by both microarray and NGS.

Further development is needed in sample preparation and data analysis. The Center for Veterinary Biologics is committed to embracing new technologies in order to increase the safety of veterinary biologics, so additional work on implementing NGS as a screening tool will be conducted.