LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Detection of adventitious viruses from biologicals using a broad spectrum Microbial Detection Array

C. Jaing, S. Gardner, K. McLoughlin, J. Thissen, T. Slezak

June 15, 2011

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Detection of adventitious viruses from biologicals using a broad spectrum Microbial Detection Array

Crystal Jaing[1*], Shea Gardner[2], Kevin McLoughlin[2], James B. Thissen[1] and Tom Slezak[2]

[1]Physical & Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550
[2]Computations Directorate, Lawrence Livermore National Laboratory, Livermore, CA 94550

*Corresponding author: Crystal Jaing
Mail address: Lawrence Livermore National Laboratory
Mail Code L-452
7000 East Ave.
Livermore, CA 94550
Phone: (925) 424-6574
Fax: (925) 422-2282
Email: jaing2@llnl.gov

ABSTRACT:

We designed the Lawrence Livermore Microbial Detection Array (LLMDA), which contains 388,000 DNA probes. This array can detect any sequenced viruses or bacteria within 24 hours. In addition, the oligonucleotide probes were selected to enable detection of novel, divergent species with homology to sequenced organisms. We recently used this array to identify an adventitious virus from a vaccine product. We have also used this array to detect viral and bacterial infections from various human clinical samples. Broad spectrum microbial detection microarrays are efficient and cost-effective tools to rapidly screen cell bank samples, raw materials, vaccine samples and clinical samples to ensure drug, food and health safety in the United States and worldwide.

KEYWORD:

Microarray, adventitious, virus, safety, vaccine, clinical

Introduction

Rapid detection and characterization of bacterial and viral pathogens is important for biodefense and public health. The currently available detection techniques require multiple assays to screen for different markers. A more comprehensive and higher throughput method will greatly improve the efficiency and speed of sample analysis. At Lawrence Livermore National Laboratory (LLNL), we have developed several types of high density microarrays to characterize a broad range of microbial sample constituents, with typical processing times under 24 hours. The Lawrence Livermore Microbial Detection Array (LLMDA) contains 388,000 DNA probes designed to detect any of the currently sequenced viruses or bacteria (1). The LLMDA probes are selected to enable detection of novel, divergent species or strains with sufficient DNA homology to sequenced organisms. We recently used this array to identify contaminating porcine circovirus DNA in a vaccine used worldwide in infants to prevent rotavirus infection (2). We have also used this array to detect viral infections from various human clinical samples (3). We developed a maximum likelihood analysis method that enables us to identify multiple viral and bacterial pathogens from the same sample (1). The array is highly sensitive and able to achieve low-copy bacterial and viral genome detection when coupled with whole genome amplification. Our microarrays are efficient and cost-effective tools to rapidly characterize emerging, unknown and engineered pathogens. The arrays can potentially be used for public health, product and food safety and global disease surveillance and can be further customized for specialized need of these missions.

Current pathogen detection techniques

Current detection systems, such as polymerase chain reaction (PCR) technologies, focus on small, prioritized sets of high-risk biological pathogens. PCR analysis is relatively inexpensive and fast. It also has a high sensitivity for known organisms, but it can process no more than about 50 DNA signatures at one time. The likelihood of discovering unknown agents or new species is low with PCR analysis. At the other end of the cost spectrum, high-throughput sequencing provides the most comprehensive information about biological pathogens, both known and unknown. The process, however, is costly and can take several days to produce results.

Microarrays, however, can identify a broad range of organisms, including pathogens on a priority screening list, sequenced bacteria or viruses that might not be anticipated, or even emerging pathogens containing DNA sequences previously identified in other pathogens. Processing samples on an array such as the LLMDA is less expensive and faster than sequencing, and more inclusive than PCR analysis. Figure 1 is a comparison of PCR, microarray and sequencing in speed, cost and resolution.

Several broad spectrum pathogen microarrays have been developed to date. Previously developed microarrays either use small subunit RNA probes for low-resolution identification of bacteria (4), have probes for viruses only (5), or use a combination of these approaches (6). The LLMDA, by contrast, has several dozen probes for each of the thousands of bacteria and viruses sequenced to date, so it can simultaneously examine multiple genomic regions from each organism. Thus, the LLMDA is the only array to date that can provide high-resolution simultaneous characterization of both bacteria and viruses present in a complex sample.

LLMDA probe design for broad spectrum detection of viruses and bacteria

The original LLMDA design covers all bacteria and viruses with sequenced genomes available in spring of 2007, including 38,000 virus sequences representing 2200 species, and 3500 bacterial sequences representing 900 species (1). Updated versions of the LLMDA have been developed to include more recently sequenced genomes. In designing probes for our array, we sought to balance the goals of conservation and uniqueness, prioritizing oligo sequences that were conserved, to the extent possible, within the family of the targeted organism, and unique relative to other families and kingdoms. We designed arrays with larger numbers of probes per sequence (50 or more for viruses, 15 or more for bacteria) than previous arrays having only 2-10 probes per target (5,6). The large number of probes per target was expected to improve sensitivity, an important consideration given possible amplification bias in the random PCR sample preparation protocol, which could result in non-amplification of genome regions targeted by some probes. Probes were selected to avoid sequences with high levels of similarity to human, bacterial and viral sequences not in the target family. Favoring more conserved probes within a family enabled us to minimize the total number of probes needed to cover all existing genomes with a high probe density per target, enhancing the capability to identify the species of known organisms and providing the possibility to detect unsequenced or emerging organisms. Strain or subtype identification was not a goal of probe design for this array. Nevertheless, our ability to combine information from multiple probes in our analysis made it possible to discriminate between strains of many organisms for which multiple genomes are available.

LLMDA processing and analysis

LLMDA processing begins by purifying DNA or RNA from a clinical, environmental or cell culture sample. The purified DNA or RNA is labeled with a fluorescent dye and then hybridized overnight to the microarray, in an incubator heated to 42°C. After we wash off unbound DNA, we scan the array with a laser to excite the labeled DNA targets that are hybridized to matching DNA probes on the array, and measure the fluorescence signal.

To find a combination of organisms whose presence in the sample best explains the observed data, we developed the composite likelihood maximization algorithm (CLiMax), as described in (1). The likelihood maximization algorithm is an iterative process, in which we scan repeatedly through a database of all sequenced microbial genomes. In the first iteration, we look for the target genome that explains the largest portion of the observed detected probe signals, while minimizing the number of probes expected to bind that target that did not have detectable signals. In each subsequent iteration, we choose the target that explains the largest part of the signal not already explained by the first target, while again minimizing the number of expected negative probes. The process continues until a maximal portion of the observed probe signals are explained, or for a specified maximum number of iterations.

The CLiMax algorithm is implemented as part of an automated software pipeline, which incorporates web-based data entry, sample tracking and analysis. The microarray results are shown in a graphical format showing a list of predicted targets organized by viral or bacterial family, together with the log-likelihood scores for each target sequence (Figure 2).

Identification of adventitious agents from vaccines using the LLMDA

In early 2010, our team used the LLMDA to evaluate seven live attenuated virus vaccines including Varivax, MMR-II, Attenuvax, Meruvax, Rotateq, Rotarix, and YF-Vax (2). We combined massively parallel 454 sequencing with LLMDA microarray analysis to evaluate any potential adventitious agents in these vaccines; additionally, an oral polio vaccine (OPV) was tested by 454 sequencing only. The expected attenuated viruses were detected in all vaccines tested. The non-vaccine viruses identified from 454 sequencing and LLMDA analysis are listed in Table 1, together with the cell lines used to produce each vaccine. Both 454 sequencing and LLMDA analysis identified the presence of porcine circovirus (PCV) DNA in the Rotarix rotavirus vaccine. The 454 analysis indicated further that a large fraction of the sequence reads from the Rotarix vaccine (41.6%) consisted of PCV1 sequences. Because serum-free media are used in the manufacture of Rotarix, a possible source for PCV1 DNA is porcine pancreas-derived trypsin used for passaging Vero cells. After introduction of PCV1, a chronic infection may have become established in Vero cells (7). Alternatively, since the porcine-derived trypsin is typically irradiated before use to inactivate adventitious viruses, the PCV1 DNA may simply reflect carryover of noninfectious, lethally irradiated PCV1 (2).

In vaccines grown in chicken embryo fibroblast (CEF) cells, signals to avian leukosis viruses (ALV) and avian endogenous retrovirus (AEV) were detected by LLMDA. The detection of ALV and AEV in live attenuated vaccines derived from CEF has been reported previously and extensive investigations indicated these viruses were non infectious to humans (8,9).

The detection of SRV DNA in the Rotateq vaccine by 454 sequencing reflects the presence of endogenous SRV DNA in the germ line of African green monkeys, from which the Vero cell line used to produce the vaccine is derived. SRV is thought to be a highly prevalent

infection of old-world monkeys and to have endogenized at multiple copies in the genomes of many nonhuman primate species (2,10,11). The removal of SRV nucleic acid from the Rotateq vaccine by DNase indicated that it was present as naked DNA released from Vero cells (2). The detection of a retroviral inactivating mutation in the Rotateq SRV DNA *pol* gene also indicated that this endogenous retrovirus is likely defective (2). LLMDA detected baboon endogenous retrovirus (BEV) sequences from Rotateq. The origin of the BEV signal for Rotateq is probably related to the African green monkey-derived Vero cell used in its manufacture and cross-hybridization of its endogenous retroviruses to the baboon endogenous retrovirus probes. The detection of human endogenous retrovirus K (HEV) in Varivax, MMR-II, and Meruvax was the expected consequence of their manufacture using human cell lines (Table 1).

Clinical diagnostic applications of LLMDA

LLMDA can be used for any applications where the rapid determination of major bacterial and viral components is desired. Recently, we have established a diagnostic platform for unbiased random amplification and microarray identification of viral pathogens in clinical samples (3). We show that Phi29 polymerase-amplification of a diverse set of clinical samples generates enough viral material for successful identification by the LLMDA, demonstrating the potential of the microarray technique for broad-spectrum pathogen detection in human samples. Our studies showed that LLMDA detects both DNA and RNA viruses as well as bacteria and plasmids, present in the same sample, and in some cases can differentiate different subtypes.

Potential of microarray technologies for QC of biological material

High-throughput sequencing and microarrays provide effective means to interrogate current and future vaccines for genetic variants of the targeted viruses and the presence of adventitious viruses. The wide range of sequences detected by these methods, including endogenous retroviruses, bacterial DNA, adventitious viruses such as PCV1, and other nucleic acids of undetermined taxonomic origin, is an expected result given closer scrutiny to the nucleic acids present in vaccines, and not necessarily a reflection of unsafe products.

Advantages and limitations of microarrays

Compared to PCR assays and high-throughput sequencing, microarrays occupy an intermediate range of cost, processing time, and sensitivity for detection of microbial DNA in complex samples. While microarrays are not as sensitive or inexpensive as PCR, they can query hundreds of thousands to several million regions of DNA in parallel (depending on platform and format), compared to at most a few tens of regions in the largest multiplexed PCR assays. Given that approximately 1000 species of human viral and bacterial pathogens have been sequenced at present, PCR assays are too limited for broad-spectrum microbial analysis. However, PCR will continue to be, for the foreseeable future, the most sensitive discriminator of whether a known pathogen is present. Similarly, sequencing will remain the best way to determine the exact nucleotide composition for a novel or mutated pathogen. Nevertheless, we expect that microarrays will evolve during this decade to become the most cost-effective means for quality control of biological products and identifying known pathogens present in clinical samples.

As with any technology based on nucleic acid detection, the capabilities of the LLMDA are limited by the genome sequence information available at the time of design. Many bacterial

and viral species are known but have not yet been sequenced. The LLMDA cannot identify

bacteria from families that have absolutely no genomes sequenced, nor can our probe design

strategy avoid selecting probes that may cross-hybridize to unsequenced organisms. We are

currently seeking funding to update the LLMDA design to take advantage of recent genomic

sequence data. Random amplification used with our array may yield lower sensitivity than arrays

targeting a limited number of specific regions, for which specific PCR can be used. On the other

hand, while specific PCR increases sensitivity for a limited number of targets, it also limits the

breadth of targets that can be detected.

We note that United States government agencies have yet to define formal validation

procedures for high information content assays such as microarrays and genomic sequencing.

This regulatory barrier may delay the widespread use of high-throughput techniques to identify

pathogens in human clinical samples, food, vaccines and other products. However, food and

pharmaceutical companies and diagnostic laboratories in the US and Europe have shown strong

interest in applying these technologies, and we are confident that, with their backing, the

attendant regulatory hurdles can be overcome.

Summary

Microarrays provide a cost-effective way to determine the presence of pathogens with

known sequence at strain resolution in complex samples, without the larger cost and time

required for full genomic sequencing. The LLMDA combines innovative large-scale

bioinformatics with modern oligonucleotide array technology to provide the first high-resolution

pan-microbial microarray that accurately determines both the bacterial and viral content of

unknown clinical or environmental samples. We have demonstrated how this technology can

improve medical product safety (2) and identify viral infections in human samples (3). Future uses of the LLMDA, once permitted by regulatory agencies, will create further cost savings and improvements in research, medical diagnosis, food and product safety applications.

Acknowledgement

Conflict of interest declaration

A patent application has been filed for the LLMDA design and analysis approach.

References

1.   Gardner, S., Jaing, C., McLoughlin, K. and Slezak, T. (2010) A Microbial Detection Array (MDA) for viral and bacterial detection. *BMC Genomics*, **11**, 668doi:610.1186/1471-2164-1111-1668.

2.   Victoria, J.G., Wang, C., Jones, M.S., Jaing, C., McLoughlin, K., Gardner, S. and Delwart, E.L. (2010) Viral Nucleic Acids in Live-Attenuated Vaccines: Detection of Minority Variants and an Adventitious Virus. *Journal of Virology*, **84**, 6033-6040.

3.   Erlandsson, L., Rosenstierne, M.W., McLoughlin, K., Jaing, C. and Fomsgaard, A. (2011) The Microbial Detection Array combined with random Phi29-amplification used as a diagnostic tool for unbiased virus detection in clinical samples. *Submitted to PLoS ONE*.

4.   DeSantis, T., Stone, C., Murray, S., Moberg, J. and Andersen, G. (2005) Rapid
        quantification and taxonomic classification of environmental DNA from both prokaryotic
        and eukaryotic origins using a microarray. *FEMS Microbiol Lett* **245**, 271-278.

5.   Wang, D., Coscoy, L., Zylberberg, M., Avila, P.C., Boushey, H.A., Ganem, D. and
        DeRisi, J.L. (2002) Microarray-based detection and genotyping of viral pathogens. *Proc
        Natl Acad Sci U S A*, **99**, 15687–15692.

6.   Palacios, G., Quan, P., Jabado, O.J., Conlan, S., Hirschberg, D.L., Liu, Y., Zhai, J.,
        Renwick, N., Hui, J., Hegyi, H., Grolla, A. *et al.* (2007) Panmicrobial oligonucleotide
        array for diagnosis of infectious diseases. *Emerg. Infect. Dis.*, **13**, 73-81.

7.   Allan, G.M., Mackie, D.P., McNair, J., Adair, B.M. and McNulty, M.S. (1994)
        Production, preliminary characterisation and applications of monoclonal antibodies to
        porcine circovirus. *Vet. Immunol. Immunopathol.*, **43**, 357–371

8.   Hussain, A.I., Johnson, J.A., Freire, M.D.S. and Heneine, W. (2003) Identification and
        characterization of avian retroviruses in chicken embryoderived yellow fever vaccines:
        investigation of transmission to vaccine recipients. *J. Virol.*, **77**, 1105–1111.

9.   Hussain, A.I., Shanmugam, V., Switzer, W.M., Tsang, S.X., Fadly, A., D.Thea, Helfand,
        R., Bellini, W.J., Folks, T.M. and Heneine, W. (2001) Lack of evidence of endogenous
        avian leukosis virus and endogenous avian retrovirus transmission to measles, mumps,
        and rubella vaccine recipients. *Emerg. Infect. Dis.*, **7**, 66–72.

10.   van der Kuyl, A.C., Mang, R., Dekker, J.T. and Goudsmit, J. (1997) Complete nucleotide
        sequence of simian endogenous type D retrovirus with intact genome organization:
        evidence for ancestry to simian retrovirus and baboon endogenous virus. *J. Virol.*, **71**,
        3666–3676.

11.   Sommerfelt, M.A., Harkestad, N. and Hunter, E. (2003) The endogenous langur type D retrovirus PO-1-Lu and its exogenous counterparts in macaque and langur monkeys. *Virology*, **315**, 275–282.

Table 1. Nonvaccine viruses identified by 454 sequencing and LLMDA

Vaccine	Cell line	Nonvaccine viruses	
		454 sequencing	**Microarray**
Meruvax	WI-38		HEV
Attenuvax	CEF	ALV	ALV, AEV
YF-Vax	CEF		AEV
MMR-II	WI-38 and CEF		HEV, AEV
Varivax	MRC-5		HEV
Rotateq	Vero	SRV	BRV
Rotatrix	Vero E6	PCV	PCV

ALV: avian leukosis virus

AEV: avian endogenous retrovirus

HEV: human endogenous retrovirus

SRV: simian endogenous retrovirus

BRV: baboon endogenous retrovirus

PCV: porcine circovirus

Figure legends

Figure 1. Comparison of PCR, microarray and high-throughput sequencing technologies. PCR, microarray and sequencing are three of the most used DNA detection technologies. PCR is fast and cheap but limited to a few known agents. DNA sequencing has the highest resolution, but the cost is high and it takes a minimum of three days to get results. Microarrays fill some gaps between PCR and sequencing. It has medium cost and results can be obtained within 24 hours. Microarrays can analyze known pathogens and discovery emerging and unknown pathogens.

Figure 2. An example of the LLMDA data from a human clinical sample. The LLMDA was analyzed using the CliMax algorithm developed at LLNL. The results are shown in a list of predicted targets organized by viral or bacterial family, together with the log-likelihood scores for each target sequence.