

Animal Viruses Probe Dataset (AVPDS) for Microarray-Based Diagnosis and Identification of Viruses

Brijesh S. Yadav · Mayank Pokhriyal ·
Dinesh P. Vasishtha · Bhaskar Sharma

Received: 15 August 2013 / Accepted: 2 September 2013 / Published online: 16 October 2013
© Springer Science+Business Media New York 2013

Abstract AVPDS (Animal Viruses Probe dataset) is a dataset of virus-specific and conserve oligonucleotides for identification and diagnosis of viruses infecting animals. The current dataset contain 20,619 virus specific probes for 833 viruses and their subtypes and 3,988 conserved probes for 146 viral genera. Dataset of virus specific probe has been divided into two fields namely virus name and probe sequence. Similarly conserved probes for virus genera table have genus, and subgroup within genus name and probe sequence. The subgroup within genus is artificially divided subgroups with no taxonomic significance and contains probes which identifies viruses in that specific subgroup of the genus. Using this dataset we have successfully diagnosed the first case of *Newcastle disease virus* in sheep and reported a mixed infection of *Bovine viral diarrhea* and *Bovine herpesvirus* in cattle. These dataset also contains probes which cross reacts across species experimentally though computationally they meet specifications. These probes have been marked. We hope that this dataset will be useful in microarray-based detection of viruses. The dataset can be accessed through the link <https://dl.dropboxusercontent.com/u/94060831/avpds/HOME.html>.

Keyword AVPDS (Animal viruses probe dataset) · Unique probes · Conserve probes · Microarray

B. S. Yadav · M. Pokhriyal · B. Sharma (✉)
Division of Biochemistry, Indian Veterinary Research Institute,
Izatnagar, India
e-mail: bhaskar@ivri.res.in

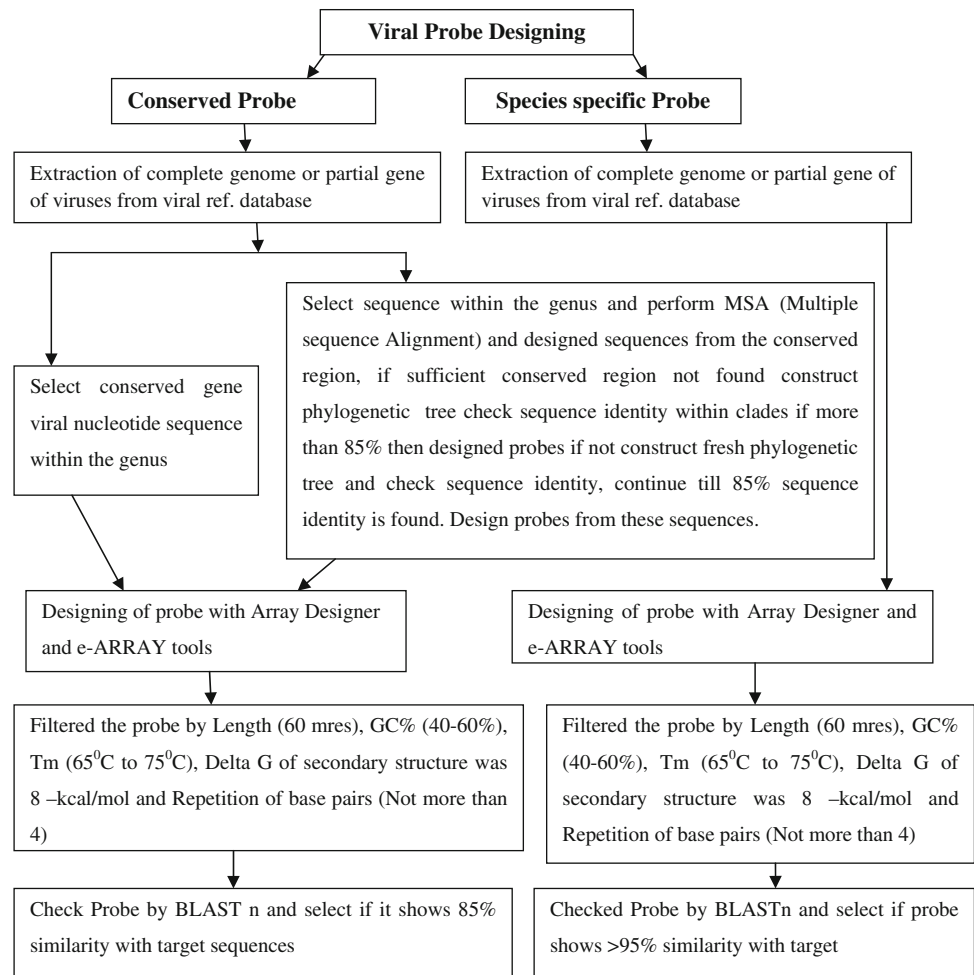
B. S. Yadav
e-mail: brijeshbioinfo@gmail.com

B. S. Yadav · D. P. Vasishtha
Department of Microbiology, H.N.B. Garhwal University,
Srinagar, Garhwal, India

Introduction

The microarray chip used for diagnostic purpose contain oligonucleotides specific to pathogens and may contain thousands of oligonucleotide probes. The probes designed for diagnostic assays are unique to a specific pathogen with respect to all other pathogen genomes and also to host and other non-specific genome sequence present in the clinical samples. As a result, the designing of pathogen-specific probes require computationally expensive comparison of target genomes with all known non-target sequences. Many methods have been developed for designing probes for pathogen diagnostic assays; some methods are intended for PCR-based assays, whereas others are intended for microarray-based assays [7]. Several groups have designed microarrays containing probes for microbial detection, discovery or a combination of both [2, 3, 5, 9]. The viro-chip discovery array was one of the first to target a broad range of pathogens; it is best known for its role in characterizing SARS as a coronavirus [10]. Chou et al. [1] designed conserved genus probes and species-specific probes covering 53 viral families and 214 genera. Palacios et al. [3] built the Greenechip, an array targeting vertebrate viruses and rRNA sequences of fungi, bacteria and protozoa, containing approximately 30,000 probes. Viral probes were designed to target a minimum of three genomic regions for each family or genus, including at least one highly conserved region coding for polymerase or structural proteins, and two or more variable regions. Pan-Microbial Detection Array (MDA) [2] is the most comprehensive chip designed for virus and other pathogen identification. Chou et al. have computationally designed virus-specific and conserved probes for microarray-based diagnosis of viruses using a specifically designed algorithm. We had used conserved probes designed by them

Flow diagram 1 Designing strategy of oligonucleotides probes



and found that some of them do not work experimentally and felt that there is need for a new dataset [1].

Materials and Methods

Preparation of the List of Animal Viruses

In this study, we report a new dataset of microarray probes for diagnosis of both viruses and virus genera. The viruses included in the database were compiled after exhaustive search and personal discussion with the virologist working in India in different institutes. Complete sequence of listed viruses was extracted from NCBI (National Center for Biotechnology Information) reference sequence viral database using a Perl script. In case of viruses with partial genomes known structural gene sequences were downloaded from NCBI nucleotide database.

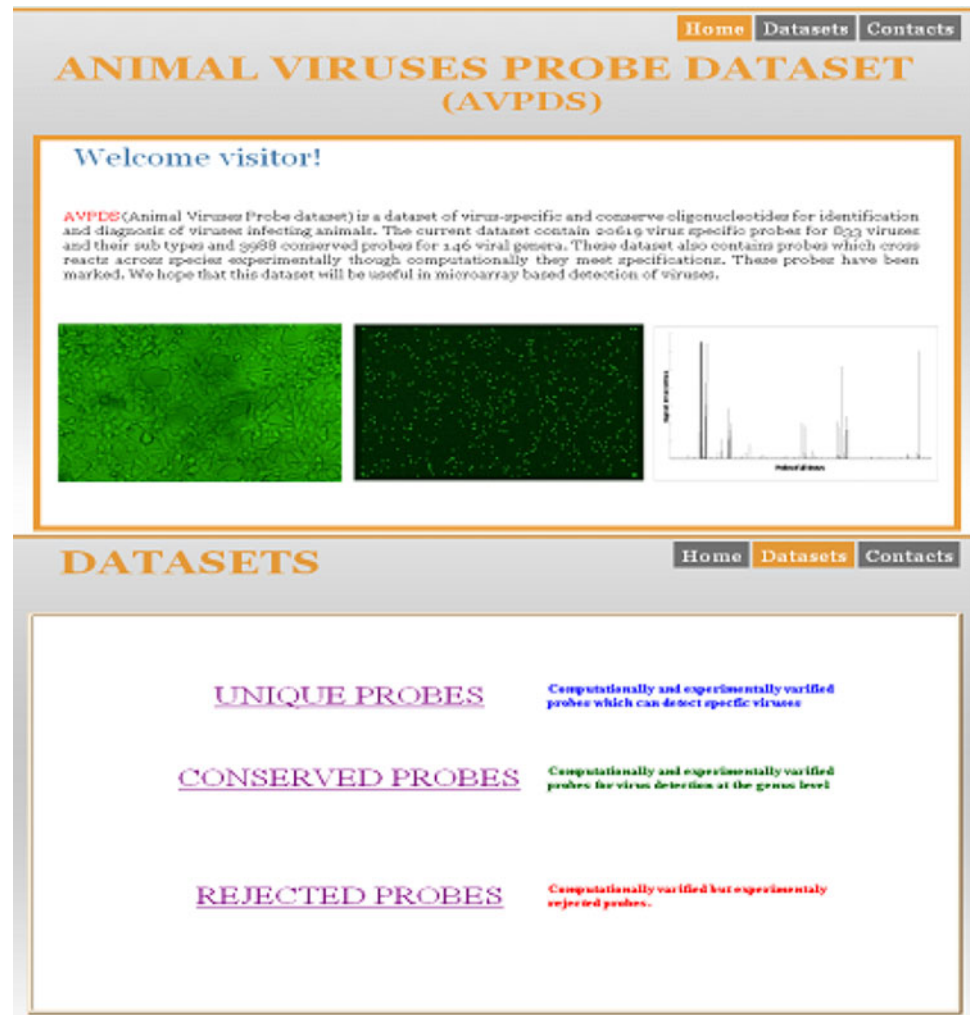
Microarray Probe Design

Our aim to for designing microarray chip was to identify a virus at species level and to detect an unknown virus at genus level. The probe designing strategy is given flow diagram 1.

Development of Dataset

In the first phase of dataset development, the probe sequences were collected, arranged alphabetically and structured into three tables unique probes conserved probes (both computationally as well as experimentally verified probes) and rejected probes (passed in computational verification but failed experimentally); so that it could be compiled into a common database. The unique probes table is divided into two fields namely virus name and probe sequence. Similarly, conserved probes table has genus,

Fig. 1 The outlook of home page and dataset page



Sub-group name and probe sequence. All this Information was compiled in MS EXCEL. In the current form, the web database has total of 20,619 unique and 3,988 conserved probes. In the phase two, structured relational database was designed for an easy access by the users of database. For designing the core relational database MS access 2007 was used. In the phase, three databases in MS access were integrated into HTML web pages so that the database could be accessed through out the internet. The dataset in the current form has the look shown in Fig. 1.

Result

The database contains probes for viruses, virus genera and a rejected probe list. The rejected probe list contains all those probes which have been experimentally shown to cross react across species but are computationally correct. These cross reactive or sticky probes have been reported by others also [4]. We have provided the list of cross reactive probes so that others can avoid using them in their probe

dataset. This dataset was used to identify an unexpected case of *Newcastle disease virus* in sheep [8] and for identifying a mixed infection of *Bovine viral diarrhea* and *Bovine herpesvirus* in cattle [6]. The dataset can be accessed by user through the link <https://dl.dropboxusercontent.com/u/94060831/avpds/HOME.html>.

Acknowledgments The authors thank Indian council of Agricultural Research (ICAR) for providing funding to all the authors for carrying out this work.

References

1. Chou CC, Lee T, Chen CH, Hsiao HY et al (2006) Design of microarray probes for virus identification and detection of emerging viruses at the genus level. *BMC Bioinform* 7:232
2. Gardner SN, Jaing CJ, McLoughlin KS, Slezak TR (2011) A microbial detection array (MDA) for viral and bacterial detection. *BMC Genom* 11:668
3. Palacios G, Quan PL, Jabado O, Conlan S, Hirschberg D, Liu Y (2007) Panmicrobial oligonucleotide array for diagnosis of infectious diseases emerging infectious disease. *Emerg Infect Dis* 13:73–81

4. Pozhitkov AE, Tautz D, Noble PA (2007) Oligonucleotide microarrays: widely applied–poorly understood. *Brief Funct Genom Prot* 6:141–148
5. Quan PL, Palacios G, Jabado OJ et al (2007) Detection of respiratory viruses and subtype identification of influenza A viruses by green chip resp oligonucleotide microarray. *J Clin Microbiol* 45:2359
6. Ratta B, Yadav BS, Pokhriyal M, Saxena M, Sharma B (2013) Microarray chip based identification of mix infection of bovine herpesvirus 1 and bovine viral diarrhoea 2 from Indian cattle. *Curr Microbiol*. doi:[10.1007/s00284-013-0448-9](https://doi.org/10.1007/s00284-013-0448-9)
7. Ravi et al (2008) In silico microarray probe design for diagnosis of multiple pathogens. *BMC Genom* 9:496
8. Sharma B, Pokhriyal M, Rai GK, Saxena M et al (2012) Isolation of newcastle disease virus from a non-avian host (sheep) and its implications. *Arch Virol* 157:1565
9. Wang D, Coscoy L, DeRisi JL et al (2002) Microarray-based detection and genotyping of viral pathogens. *PNAS* 99:15687–15692
10. Wang D, Urisman A, Liu Y-T, Springer M, Ksiazek TG et al (2003) Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* 1:e2