Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Virology

Insect-specific viruses: from discovery to potential translational applications

Shahideh Nouri[1], Emilyn E Matsumura[2], Yen-Wen Kuo[2] and Bryce W Falk[2]

Over the past decade the scientific community has experienced a new age of virus discovery in arthropods in general, and in insects in particular. Next generation sequencing and advanced bioinformatics tools have provided new insights about insect viromes and viral evolution. In this review, we discuss some high-throughput sequencing technologies used to discover viruses in insects and the challenges raised in data interpretations. Additionally, the discovery of these novel viruses that are considered as insect-specific viruses (ISVs) has gained increasing attention in their potential use as biological agents. As example, we show how the ISV Nhumirim virus was used to reduce West Nile virus transmission when co-infecting the mosquito vector. We also discuss new translational opportunities of using ISVs to limit insect vector competence by using them to interfere with pathogen acquisition, to directly target the insect vector or to confer pathogen resistance by the insect vector.

Addresses
[1] Department of Plant Pathology, Kansas State University, Manhattan, Kansas 66506, USA
[2] Department of Plant Pathology, University of California-Davis, CA 95616, USA

Corresponding author: Falk, Bryce W (bwfalk@ucdavis.edu)

Current Opinion in Virology 2018, **33**:33–41

This review comes from a themed issue on **Virus-vector interactions**

Edited by **Anna E. Whitfield** and **Ralf Dietzgen**

https://doi.org/10.1016/j.coviro.2018.07.006

Introduction

Viruses are the most abundant microbes on our planet [1] and are found associated with all types of organisms, even other viruses. Insects have long been recognized as vectors of many important viruses that affect humans, animals and plants, but insects also are hosts to many different viruses. In general, viruses belonging to several taxa, including the *Baculoviridae*, *Parvoviridae*, *Flaviviridae*, *Ascoviridae*, *Togaviridae*, *Bunyavirales* (formerly *Bunyaviridae*), and *Rhabdoviridae*, have most commonly been considered as being associated with insects [2]. Most viruses have been discovered because they are pathogens and cause disease in their hosts, but recently many viruses have been discovered from non-diseased plants and animals (including insects) suggesting that viruses are much more abundant than previously recognized. Recent estimates suggest that currently less than 1% of the virosphere — the total universe of viruses — has been discovered and classified [3], and traditional virus detection and discovery methods, which require prior knowledge of the viral genome sequences, are considered as one of the causes of this restricted view of the virosphere. These limitations indicate a need for new, improved and unbiased virus discovery approaches. With the introduction of high-throughput next-generation sequencing (NGS) in 2005 [4], a new age of virus discovery has commenced and novel viruses/viral sequences have been identified in many organisms including insects [5–25].

Many of the newly discovered insect viruses are referred to as insect-specific viruses (ISVs). ISVs replicate only in their insect hosts, give persistent infections and are likely transmitted vertically (transovarially) within the population [24,26]. Phylogenetic studies indicate that many of the newly discovered ISVs belong to several different virus families, as well as many which are as yet unclassified [8,24]. These newly discovered viruses/viral sequences provide novel opportunities for gaining a greater understanding of virus evolution and virus-host interactions, but also many may have potential for translational applications for manipulating traits in insects. Whether ISVs have potential to be used for novel translational applications as wild-type or even recombinant viruses is a new chapter of research in virology, and offer new opportunities for efforts to target insects and the pathogens they transmit.

In this review, we focus on ISVs discovered among different insects. We discuss the different methods used to collect metagenomics data, and the challenges in interpretation of viral metagenomes. We also discuss/propose possible translational applications of ISVs; although, we are just at the beginning of this interesting but challenging research area.

NGS technologies for insect virus discovery

NGS technologies are high throughput and highly sensitive non-Sanger-based methodologies that generate

millions to billions of nucleic acid sequences in a single run [27,28]. These massively parallel sequencing platforms are capable of sequencing complex mixtures of genetic material [28] and have been extremely efficient in detecting DNA and RNA viruses, even when their titers are low in the infected host/tissue [29]. Furthermore, NGS is comparatively inexpensive, allows for accurate sequencing and is currently considered the fastest available approach for virus discovery [27,28,30].

Given the high diversity of viruses in both animal and plant hosts and their different replication strategies, several NGS sample and library preparation methods and types of bioinformatics analyses have been used to identify viral sequences in diverse types of host tissues. Some have used different types of templates in order to enrich the libraries for viral sequences [31], and the most commonly used are: total RNA or DNA; virus-derived small interfering RNAs (siRNAs); double-stranded RNAs (dsRNAs); polyadenylated RNAs (poly(A) RNAs); ribosomal RNA depleted total RNA, and RNA from purified or partially purified viral particles [31,32]. All of these have shown advantages and disadvantages and are reviewed in Roossinck *et al.* [32]. Different sample and library preparation methods and currently available NGS platforms have also been reviewed in detail [27,28,32–37]. In general, transcriptome and small RNA (sRNA) libraries appear to be the most suitable for virus discovery [7,8,27], and in this section, we compare these two types of NGS technologies for detecting/discovering insect viruses.

Transcriptome sequencing

The transcriptome is the complete set of transcripts in cells [38,39]. In a virus-infected sample, this will include host and viral RNAs (genomic or mRNAs). Transcriptome sequencing or RNA sequencing (RNA-Seq) by using NGS technology has allowed characterization and quantification of the total transcriptome from different organisms and has shown clear advantages over other approaches in terms of detection of viral nucleic acids in host cells, providing also insights into host responses to virus infection (virus-host interactions) [31]. In general, transcriptome libraries are constructed from total or fractionated RNAs, which are converted to a library of cDNA fragments that have adaptors linked to one or both ends [40]. cDNA fragments, PCR-amplified or not, are then sequenced by using NGS techniques, which can generate short sequences (30–400 nucleotides) from one end (single-end sequencing) or both ends (paired-end sequencing) [40]. Several bioinformatics tools are available for NGS data analysis/processing (see below). Basically, the adaptor sequences and low-quality reads are removed from the libraries and the high-quality reads are generally assembled by two different assembly methods: reference-based (transcriptome is assembled by mapping reads to previously known sequences) and *de novo* assembly (transcriptome is reconstructed without the aid of a reference genome) [39,40]. These methods, combined with Basic Local Alignment Search Tool (BLAST) searches against non-redundant databases or viral databases, have successfully led to virus detection and the discovery of new viruses, respectively [27].

In insects, the first application of NGS technology for transcriptome sequencing was in 2007, when a metagenomics survey of the honey bee (*Apis mellifera* L.) was conducted in a search for a causal agent of colony collapse disorder [21]. Several viruses were identified, which indicated the potential of using this approach [27]. Since then, several groups have sequenced the transcriptomes of different insects and detected not only the known viruses and their variants, but also, in some cases, discovered novel viruses [8,17,30,41–46]. However, an important shortcoming of sequencing transcriptome libraries for detecting viral sequences is that the majority of the generated reads are derived from the host. The percentage of viral RNA genomes within the host RNA background can be very low (less than 1%) and viruses present at low titers may not be detected by this approach [31,32]. In order to reduce the host RNA background and enrich the transcriptome libraries with viral sequences, methods for removing the abundant host ribosomal RNA from the total RNA, or for removing RNA derived from a control (or healthy) host by subtractive hybridization, have been explored [31,32,34]. Other strategies, such as using dsRNA and/or RNA from partially purified viral particles for transcriptome sequencing, have also been used [31,32].

After sequences are obtained, an important challenge is how to manage the large amount of data generated by transcriptome sequencing. There is no standardized pipeline to analyze the data, and viral sequences assemble differently depending on the program and parameter settings used [31]. The challenges of bioinformatics have demanded the development of efficient methods mainly to overcome concerns including storage, quality control, mapping, error correction, single nucleotide variant calling, haplotype reconstruction and data integration [47].

Small RNA sequencing

Deep sequencing of sRNAs has also been used for discovering DNA and RNA viruses in both animal and plant hosts [8,9,48–50]. This approach takes advantage of the natural host antiviral defense system found in most eukaryotes as a component of a more general sequence-specific mechanism for gene expression regulation, which is known as RNA interference (RNAi) [32,50,51]. In arthropods, RNAi consists of at least three different pathways: the small interfering (si), the micro- (mi) and the piwi-interacting (pi) RNA pathways [53]. These three pathways generate sRNAs that have distinct molecular characteristics, including their biogenesis, length, targets and modes of action, and also reflect the virus from which they originated [51].