De novo assembly of honey bee RNA viral genomes by tapping into the innate insect antiviral response pathway

Elisabeth Fung[a,*], Kelly Hill[a,b], Katja Hogendoorn[a], Richard V. Glatz[a,c,d], Kathryn R. Napier[e], Matthew I. Bellgard[e], Roberto A. Barrero[e,*]

[a] School of Agriculture, Food and Wine, The University of Adelaide, Waite Campus, Urrbrae, SA 5064, Australia
[b] South Australian Research and Development Institute (SARDI), Waite Campus, Urrbrae, SA 5064, Australia
[c] D'Estrees Entomology & Science Services, Kingscote, SA 5223, Australia
[d] South Australian Museum, Terrestrial Invertebrates, Adelaide, SA 5000, Australia
[e] Centre for Comparative Genomics, Murdoch University, WA 6150, Australia

ARTICLE INFO

ABSTRACT

Bee pollination is critical for improving productivity of one third of all plants or plant products consumed by humans. The health of honey bees is in decline in many countries worldwide, and RNA viruses together with other biological, environmental and anthropogenic factors have been identified as the main causes. The rapid genetic variation of viruses represents a challenge for diagnosis. Thus, application of deep sequencing methods for detection and analysis of viruses has increased over the last years. In this study, we leverage from the innate Dicer-2 mediated antiviral response against viruses to reconstruct complete viral genomes using virus-derived small interfering RNAs (vsiRNAs). Symptomatic A. mellifera larvae collected from hives free of Colony Collapse Disorder (CCD) and the parasitic Varroa mite (Varroa destructor) were used to generate more than 107 million small RNA reads. We show that de novo assembly of insect viral sequences is less fragmented using only 22 nt long vsiRNAs rather than a combination of 21–22 nt small RNAs. Our results show that A. mellifera larvae activate the RNAi immune response in the presence of Sacbrood virus (SBV). We assembled three SBV genomes from three individual larvae from different hives in a single apiary, with 1–2% nucleotide sequence variability among them. We found 3–4% variability between SBV genomes generated in this study and earlier published Australian variants suggesting the presence of different SBV quasispecies within the country.

1. Introduction

The European honey bee (Apis mellifera) is the best known and most important crop pollinator worldwide (Delaplane and Mayer, 2000). Because of its value as honey producer and pollinator, A. mellifera has been deliberately introduced into Australia and the Americas which has resulted in an almost global distribution with exception of Antarctica (Goulson and Hughes, 2015). Till date 31 RNA viruses have been reported infecting A. mellifera worldwide, primarily positive-strand RNA viruses from the families Dicistroviridae and Iflaviridae (de Miranda et al., 2013; McMenamin and Genersch, 2015; Mordecai et al., 2016a; Remnant et al., 2017) although not all are recognised and classified by the International Committee on Taxonomy of Viruses (King et al., 2012). These viruses affect the health and fitness of the bees and have been linked to widespread disease and loss of managed A. mellifera colonies in USA, Europe and Asia (e.g. Carreck, 2016; de Miranda et al., 2013; Goulson et al., 2015; Neumann and Carreck, 2010; Potts et al., 2010a, 2010b; vanEngelsdorp and Meixner, 2010).

It has been recently reported that A. mellifera individuals can reduce the impact of infection of RNA viruses through their antiviral immune response. Among all the antiviral responses, RNA interference (RNAi) is the most important in plants and insects (Brutscher et al., 2015; Brutscher and Flenniken, 2015; Vijayendran et al., 2013). RNAi is a post-transcriptional gene silencing mechanism that involves three distinct pathways, namely small interfering RNA (siRNA), microRNA (miRNA), and piwi-interacting RNA (piRNA). They have distinct biological functions and characteristics such as biogenesis, length of mature functional small RNA sequences, developmental timing, tissue

specificity, and regulation of different downstream targets. The major antiviral response of *A. mellifera* is siRNA and it is well characterised in comparison with the other pathways (Brutscher and Flenniken, 2015; Ding, 2010; Hammond et al., 2001; Vijayendran et al., 2013).

Small interfering RNA is triggered by occurrence of double-stranded RNA (dsRNA), which can be either the viral genome itself or an intermediate dsRNA product generated during virus replication. The host recognizes the dsRNA and uses ribonuclease III (*Dicer-like*) to cleave the viral genome into 21–22 nucleotide (nt) long pieces called virus-derived small interfering RNAs (vsiRNAs; Brutscher et al., 2015; Niu et al., 2014). siRNA antiviral response was first demonstrated when mortality level and *Israel acute paralysis virus* (IAPV) titres reduced in individual *A. mellifera* fed with IAPV-dsRNA prior to infection under controlled conditions (Maori et al., 2009). Later, this response was demonstrated in *A. mellifera* colonies under normal beekeeping conditions (Hunter et al., 2010).

Deep sequencing of vsiRNAs reported siRNA pathway in naturally infected *A. mellifera*. Adult *A. mellifera* from colonies affected by Colony Collapse Disorder (CCD) had abundant vsiRNAs of 21–22 nt matching *Deformed wing virus* (DWV), IAPV, and *Kashmir bee virus* (KBV). vsiRNAs matching *Acute bee paralysis virus* (ABPV), *Sacbrood virus* (SBV), and *Varroa destructor virus-1* (VDV; unassigned) were observed in the same study but in low incidence (Chejanovsky et al., 2014).

Deep sequencing of siRNAs and subsequent assembly of viral genomes was previously demonstrated for plants, mosquitoes, fruit flies and nematodes. This process was used successfully to reassemble entire or partial genomes of known viruses and discovery of novel viruses (Barrero et al., 2017; Kreuze et al., 2009; Wu et al., 2010). The use of deep sequencing of vsiRNAs for diagnosis and genome assembly of bee viruses is still limited although there is an increasing interest in *A. mellifera* antiviral defence which will likely see this technique used more widely for this purpose.

SBV was the first virus identified in *A. mellifera* and it has become distributed in all continents where *A. mellifera* is present (Chen and Siede, 2007). Three serotypes of this single-strand RNA (ssRNA) virus have been characterised worldwide, namely European, Asian and New Guinea serotypes (Allen and Ball, 1996; Bailey et al., 1982, 1964; Roberts and Anderson, 2014). The New Guinea serotype was briefly mentioned in late 1970's (Bailey et al., 1979), and it was recently characterised from an *A. mellifera* isolate. It was hypothesised that it evolved from the European serotype via mutations (Bailey et al., 1979; Roberts and Anderson, 2014). The European serotype has been detected in *A. mellifera* and is normally reported as SBV (Bailey et al., 1964; de Miranda et al., 2013). The Asian serotype has been detected in the Asian honey bee *A. cerana*, and it includes several variants, namely Thai, Chinese and Korean SBV (TSBV, CSBV, KSBV; Allen and Ball, 1996; Bailey et al., 1982; de Miranda et al., 2013). Although this serotype has been mainly reported in *A. cerana,* recent studies have reported it in *Apis* bees other than *A. mellifera* (Gong et al., 2016; Roberts and Anderson, 2014). While the European and Asian serotypes cause the same disease symptoms, there are physiochemical, pathogenic and genome variation between them (Allen and Ball, 1996; de Miranda et al., 2013). However, the variation between genomes of both serotypes has not exceeded 10% (e.g. Choe et al., 2012; Grabensteiner et al., 2001; Ma et al., 2011; Reddy et al., 2016).

The genomic variability of around 10% between the Asian and European serotypes and the phenotypic differences of symptoms caused in infected hosts, separate them as two distinctive viral *strains* (Hull, 2014). Clouds of *mutant genomes* (*mutant spectra*) that form viral strains are called *quasispecies* (Carter and Gensch, 2007; Eigen, 1996; Hull, 2014). Viral quasispecies are a collection of non-identical but closely related viral genomes that can be generated by mutation or recombination, and are continuously subjected to genetic variation, competition and selection. Quasispecies surround a central master species that is the most frequently occurring and fitted variant (Andino and Domingo, 2015; Chen and Siede, 2007; Domingo and Holland,

1997; Domingo et al., 2006). Although the term quasispecies was introduced to describe variants of RNA virus of a phylogenetically related population present in one infected organism, it has been used to describe genome heterogeneity of RNA virus populations (Carter and Gensch, 2007). The term quasispecies was used to describe genome diversity of DWV variants detected in *A. mellifera* from the same apiary (Mordecai et al., 2016b) and hence, in this study, we will also use 'quasispecies' to reflect variants of RNA viruses in several species and individuals. In Australia, very little information is available on quasispecies of RNA viruses circulating in the country including SBV, which was one of the most common and consistently detected viruses in the most recent national survey (Roberts et al., 2015, 2017). It is important to know the sequence variability of the viruses present in Australia to better evaluate the impact of distinct variants.

This study investigates the occurrence of the antiviral RNAi pathway in symptomatic SBV-infected *A. mellifera* larvae, and generate full bee RNA virus genomes using a *de novo* assembly approach that exclusively uses a subset of small RNAs derived from the innate insect immune response. Three SBV-infected larvae from different hives within one South Australian apiary were analysed in order to (i) confirm SBV replication detected using strand-specific reverse-transcription polymerase chain reaction (RT-PCR), (ii) confirm the occurrence of antiviral response by deep sequencing of siRNAs, and (iii) use SBV as model to confirm our methodological approach of generating full RNA virus genomes for comparison with those available in public databases.

2. Materials and methods

2.1. Sample collection

Three symptomatic larvae of *A. mellifera* were collected by the South Australian state apiary inspector from different hives in the same apiary located near Scott Creek Conservation Park, Mount Lofty Ranges, near Adelaide, South Australia. Since CCD and the parasitic Varroa mite (*Varroa destructor*) have not been reported in Australia (Neumann and Carreck, 2010; Roberts et al., 2017; Rosenkranz et al., 2010; Wilfert et al., 2016), we consider these hives were free of both CCD and Varroa mite. Samples were immersed in RNALater and stored in −20 °C until transported to Waite Research Institute in Adelaide, where the larvae were stored at −80 °C until molecular analysis.

2.2. Molecular analyses and library preparation

Prior to RNA extraction, larvae were washed in 1% sodium dodecyl sulphate to guarantee elimination of any brood food and any other possible contaminants. Total RNA and enriched small RNA were extracted from individual larvae simultaneously using mirVana miRNA Isolation kit (Life Technologies) following the manufacturer's instructions. Concentration and purity of the total RNA was measured using Qubit®3.0 Fluorometer and a NanoDrop ND-1000 spectrophotometer (both Thermo Fisher Scientific), respectively. Solubilised RNA was stored at −80 °C until used.

Total RNA was used to test for (+)SBV (positive-strand) and (−) SBV (negative-strand) using the strand-specific RT-PCR method based on the work of Boncristiani et al. (2009) but with several modifications (Fig. 1). Two types of RT reactions were performed using Bioscript Reverse Transcriptase kit (Bioline). First, RT was conducted using either a biotinylated (fwd) or biotinylated (rev) to generate complementary DNA (cDNA) of (−)SBV and (+)SBV, respectively. Second, RT was performed using non-specific random hexamer primers (Bioline) to generate conventional (non-biotinylated) cDNA (Fig. 1). Prior to PCR amplification, biotinylated-cDNA were magnetically separated from any non-target cDNA using Dynabeads® kilobaseBINDER™ kit (Invitrogen) according to the manufacturer's instructions (Fig. 1). Two µl of both conventional- and biotinylated-cDNA were added to a final 25 µl PCR reaction solution: 2.5 µl 10x NH_4 buffer, 0.7 µl of 10 mM