RESEARCH ARTICLE

RNA-Seq reveals virus–virus and virus–plant interactions in nature

Mari Kamitani[1,*], Atsushi J. Nagano[1,2,3], Mie N. Honjo[1] and Hiroshi Kudoh[1]

[1]Center for Ecological Research, Kyoto University, Hirano 2-509-3, Otsu, Shiga 520-2113, Japan, [2]Faculty of Agriculture, Ryukoku University, Yokotani 1-5, Seta Oe-cho, Otsu, Shiga 520-2914, Japan and [3]JST PRESTO, Japan Science and Technology Agency, Honcho, Kawaguchi, Saitama 332-0012, Japan

*Corresponding author: Center for Ecological Research, Kyoto University, Hirano 2-509-3, Otsu 520-2113, Japan. Tel: +81-77-549-8200, Fax: +81-77-549-8201, E-mail: kamitani@ecology.kyoto-u.ac.jp

One sentence summary: RNA-Seq, coupled with *de novo* assembly, enables the comprehensive and solid determination of virus infection, and revealed the virus–virus and virus–host interactions in naturally growing plants.

Editor: Rolf Kümmerli

ABSTRACT

As research on plant viruses has focused mainly on crop diseases, little is known about these viruses in natural environments. To understand the ecology of viruses in natural systems, comprehensive information on virus–virus and virus–host interactions is required. We applied RNA-Seq to plants from a natural population of *Arabidopsis halleri* subsp. *gemmifera* to simultaneously determine the presence/absence of all sequence-reported viruses, identify novel viruses and quantify the host transcriptome. By introducing the criteria of read number and genome coverage, we detected infections by *Turnip mosaic virus* (TuMV), *Cucumber mosaic virus* and *Brassica yellows virus*. Active TuMV replication was observed by ultramicroscopy. *De novo* assembly further identified a novel partitivirus, *Arabidopsis halleri partitivirus 1*. Interestingly, virus reads reached a maximum level that was equivalent to that of the host's total mRNA, although asymptomatic infection was common. *AhgAGO2*, a key gene in host defence systems, was upregulated in TuMV-infected plants. Multiple infection was frequent in TuMV-infected leaves, suggesting that TuMV facilitates multiple infection, probably by suppressing host RNA silencing. Revealing hidden plant–virus interactions in nature can enhance our understanding of biological interactions and may have agricultural applications.

Keywords: Argonaute2; *Arabidopsis halleri*; *Arabidopsis halleri partitivirus 1*; multiple infection; RNA-Seq; *Turnip mosaic virus*

INTRODUCTION

Plant virus research has mainly focused on crop diseases, as viruses often have serious impacts on agriculture and the economy (Rybicki 2015). Although viruses play an important role in ecosystems (Mitchell and Power 2003; Malmstrom *et al.* 2005a), little is known about the ecology of viruses in natural environments (King *et al.* 2011; Roossinck 2011). Early studies on viruses in wild plants often targeted agriculturally important pathogenic viruses in plant communities adjacent to crop fields (Cooper and Jones 2006; Mueller, Groves and Gratton 2011; Pagán

et al. 2012). However, considering recent finding that viruses can be common across diverse plants, virus ecology will provide a novel perspective on the biology of viruses and their interactions with surrounding organisms (Alexander *et al.* 2014; Stobbe and Roossinck 2014). Recently, several studies have been conducted in natural habitats to better understand the multiple roles of viruses in ecosystems (Roossinck 2012, 2015).

Virus ecology research typically follows one of two approaches. One involves revealing the 'virus flora' in natural habitats (Hackett *et al.* 2009; Kreuze *et al.* 2009; Barba, Czosnek

and Hadidi 2014). RNA sequencing technology (RNA-Seq) enables different types of viruses to be screened simultaneously in a single experiment, because the majority of plant viruses are RNA genome viruses, and DNA genome viruses transcribe mRNA during their life cycles. Incorporating *de novo* assembly can make RNA-Seq a powerful tool for identifying both known and novel viruses (Kreuze *et al.* 2009; Barba, Czosnek and Hadidi 2014; Roossinck 2015). The other approach is capturing virus–host and virus–virus interactions (Thurston *et al.* 2001; Remold 2002; Prendeville *et al.* 2012). These interactions can be major determinants of host physiology (Hull 2014) as well as host demography (Malmstrom *et al.* 2005a,b). In natural ecosystems, asymptomatic infection might be common (Boccardo *et al.* 1987; Thurston *et al.* 2001; Prendeville *et al.* 2012), and virus pathogenesis can be altered by surrounding environments (Fraile and Garcia-Arenal 2016). However, infection that produces severe symptoms is generally not adaptive for either plants or viruses (Remold 2002; Pagan *et al.* 2014), and under particular conditions, viruses may have beneficial effects on their hosts (Xu *et al.* 2008).

By applying RNA-Seq to natural host–virus systems, these two approaches in virus ecology can be accomplished together. For example, in the study of host–pathogen interactions, RNA-Seq has provided novel opportunities to simultaneously analyse host and pathogen transcriptomes, which is referred to as 'dual RNA-Seq' (Tierney *et al.* 2012; Westermann, Gorski and Vogel 2012; Westermann *et al.* 2016). The application of dual RNA-Seq to a rice and blast fungus system detected upregulation of infection-related fungus genes and defence-related host genes (Kawahara *et al.* 2012). Although dual RNA-Seq has been detected dual transcriptomes in fungal/bacterial pathogen systems, it can be applied to reveal virus diversity and host transcriptome simultaneously in a host–virus system. It allows the efficient collection of key information, such as the full genome sequences of all infected viruses, virus species combinations in multiple infections and transcriptomic responses of host plants.

In virus–host plant interactions, RNA silencing is a major host defence mechanism against viruses, in which Dicer or an RNA-induced silencing complex (RISC) recognises and degrades the virus genome (Soosaar, Burch-Smith and Dinesh-Kumar 2005; Shiekh 2014). Viruses have developed counter-defences involving suppressor genes against this host defence mechanism (Incarbone and Dunoyer 2013). Such counter-defence facilitates the replication of co-infecting viruses under multiple infections (Pruss *et al.* 1997; Fukuzawa *et al.* 2010; Syller 2012). To study this system in natural habitats, it is necessary to detect viruses comprehensively in symptomatic and asymptomatic hosts, and to analyse the interaction between host defence and virus counter-defence on site.

In this study, we used RNA-Seq to analyse the interactions of viruses with their hosts and other viruses in nature for the first time. We used a natural population of *Arabidopsis halleri* subsp. *gemmifera* (hereafter referred to as *A. halleri*) in central Japan. Because this species is closely related to *A. thaliana*, high-quality reference sequences and gene annotations are available. We used the enzymatic depression of rRNA instead of oligo-dT bead purification in RNA-Seq, which allowed us to analyse viruses both with and without a polyA-tail (Nagano *et al.* 2015). By combining multiple indices of virus genome information, we introduced novel criteria to identify viruses. Such rigorous identification of viruses in multiple individual samples provides novel insights into many aspects of virus ecology, such as spatial patterns of infection, multiple infections, the relationship between infection and host phenotypes and host responses.

MATERIALS AND METHODS

Plants and sampling

Arabidopsis halleri is a perennial plant native to Japan. It is a metallophyte that often inhabits soils contaminated by heavy metals (Kubota and Takenaka 2003). The study site was located at Omoide River in Hyogo prefecture in Japan (35°06′N, 134°56′E, Alt. 200 m, Aikawa *et al.* 2010). The plants grew along a small, shallow creek running through secondary forests and Japanese cedar (*Cryptomeria japonica*) plantations on the south-facing slope of a hill. Vegetation was sparse along the creek, due to heavy metal contamination of the soil from previous mining activities near the site, and no cruciferous species were observed except for *A. halleri*.

We collected and analysed leaf samples from 68 different individuals (Supporting Information, Table S1). The first set of samples (22 leaves) was collected along the creek for ca. 150 m, with 5–10-m intervals between plants. The second set of samples (46 leaves) was collected from a 20 m by 25 m rectangular plot, in which the locations of all sampled individuals were recorded precisely. The study plot contained more than 200 *A. halleri* individuals, and we selected 46 sampled plants covering the whole study plot. The first and second sets were collected on 27 May and 30 June, 2014, respectively, between noon and 3:00 p.m. on both days. New plastic disposable gloves were used to sample the leaves of each individual not to spread viruses artificially during sampling. We collected one leaf blade per plant. The majority of sampled leaves were rosette leaves, but cauline leaves 1.5–2.0 cm in length were collected from individuals that lacked rosette leaves (n = 8).

Our preliminary results indicated that the presence/absence of virus from a single leaf mostly corresponded with infection/non-infection at a whole-plant level as long as we avoided collecting newly expanded leaves. During sampling, the presence of herbivory by insects and chlorosis on sampled leaves were recorded and photographed (Fig. S1, Supporting Information). We also recorded leaf chlorosis as a potential visible symptom of viral infection. The major herbivore at the study site was a leaf beetle, *Phaedon brassicae*, the larvae and adults of which attacked leaves primarily in May–June (Kawagoe and Kudoh 2010). Immediately after sampling, each leaf was placed in a 1.5-ml micro-tube with 1.0 ml RNA later (Ambion, Life Technologies, CA, USA) to avoid RNA degradation. The samples were immersed in the RNA later at 4°C for 1 day and then transferred to –20°C for 1.5 months until RNA extraction was performed.

Total RNA preparation, RNA-Seq library preparation and sequencing

Total RNA was extracted from each leaf using 300 μl TRI-zol reagent (Life Technologies) according to the manufacturer's instructions. Leaves were completely crushed with cylinder-shaped metal beads using the multi-beads shocker (YASUI KIKAI, Japan). During RNA extraction, the DNase treatment was not applied because it was included in the following library preparation. The amount of RNA was measured using Qubit (Life Technologies), and the quality was assessed using a Bioanalyzer (Agilent Technologies, CA, USA). For RNA-Seq library preparation, 200 ng of total RNA per sample was used. To obtain all plant virus RNA that was present in the wild plants, we used selective depression of rRNA by thermostable RNaseH (Morlan, Qu and Sinicropi 2012) instead of mRNA purification by oligo-dT beads, which cannot trap virus genomes that lack polyA tails. A detailed protocol for the library preparation was described in our

previous study (Nagano *et al.* 2015). The quality check of the library was performed by the Bioanalyzer. The 68 tagged libraries were mixed to generate a final concentration of 0.33 nmol l^{-1} for each sample. The single-end 50 bases and index sequencing was conducted using two lanes of the HiSeq 2500 (Illumina) with the TruSeq v3 platform. The sequence data were deposited to the short read database, under accession number DRA003823.

Mapping and quantification of the short read sequences

Analysis of the RNA-Seq data was conducted as follows (Fig. S2, Supporting Information). FASTQ files from RNA-Seq were preprocessed by removing adapter sequences and low-quality fragments using trimmomatic-0.32 (Bolger, Lohse and Usadel 2014) with the following parameters: [ILLUMINA-CLIP:${ADAPTER:2:30:10 LEADING:19 TRAILING:19 SLIDINGWIN-DOW:30:20 AVGQUAL:20 MINLEN:40]. In the brackets, command lines indicate cutting adapter and other Illumina sequences with 2, 30 and 10 for allowed seed mismatches, palindrome clip threshold and simple clip threshold, respectively, to cut bases off the start and end of a read if the quality is <19, to perform a sliding 30-base window trimming with a quality threshold 20, to drop the read if the average quality is <20, and to drop the read if it is below 40 bases in length. Virus genome sequences, *A. halleri* transcript sequences and ERCC (External RNA Controls Consortium) spike in control (Life Technologies) were used as reference sequences. The virus reference sequences were composed of complete genome sequences of 3981 viruses obtained from the NCBI GenBank (Table S2, Supporting Information). The reference transcriptome sequences of *A. halleri* were prepared according to the following strategy. RNA-Seq reads of *A. halleri* from our study site were mapped to the *A. lyrata* transcriptome JGI release v. 1.0 (Hu *et al.* 2011) by bowtie2 (Langmead and Salzberg 2012). We found 370464 single-nucleotide polymorphisms (SNPs) supported by at least two reads. The SNP loci of the *A. lyrata* gene model were replaced by *A. halleri* alleles. The replaced sequences were used as the reference transcriptome sequences of *A. halleri* (submitted to Dryad Digital Repository, doi: 10.5061/dryad.4pf96). The preprocessed sequences were mapped on the reference and quantified using RSEM-1.2.15 (Li and Dewey 2011).

Detection of virus infection

The output of RSEM was analysed using R 3.1.1 software (R Development Core Team 2011). In current multiplex sequencing experiments with the Illumina genome analyser, sample misidentification occurred at a rate of 0.03% (Kircher, Sawyer and Meyer 2012). In the present experiment, the rate was approximately 0.01%. In order to avoid missorting of reads, we employed 0.05% as an empirical threshold. We removed the effect of missorted reads in quantification according to the following procedure. Let R_i denote the total number of reads mapped on gene i for each lane, that is,

$$R_i = \sum_{j=1}^{m} r_{ij},$$

where m is the number of samples and r_{ij} is the raw read count of gene i in sample j. The read number of gene i in sample j (q_{ij}) is estimated as $q_{ij} = r_{ij} - 0.0005R_i$ if $r_{ij} > 0.0005R_i$, otherwise $q_{ij} = 0$. All genes were treated in the same way. To evaluate virus read amounts for the following analysis, the virus read number (rpm) was calculated. In this study, rpm was defined as $\log \frac{v_{kj}}{\sum_{i\in H} q_{ij}}$, where v_{kj} is the number of reads mapped on virus k in sample j. H is the host gene set and the denominator indicates the number of reads mapped on the host transcriptome sequences (total host gene expression). The number of virus reads was so large that the total read number ($\sum_{i\in H} q_{ij} + \sum_{i\in V} v_{kj}$, where V is the virus genome set) was very different between infected and non-infected plants. Therefore, we used the total reads derived from host genes, except for rRNA, as a denominator, instead of the total reads including virus reads. The genome coverage of each virus was calculated. The percentage of the virus genome region covered by at least three mapped reads was defined as the coverage (Fig. 1a). Scripts for analysis and visualisation were written using default packages including 'base,' 'graphics' and 'stats' of R 3.1.1 software.

To examine infection by a novel virus(es), *de novo* assembly of unmapped reads was conducted. The unmapped reads from 68 samples were pooled (5885060 reads in total) and were assembled by Trinity v2.0 (Grabherr *et al.* 2011). Contigs with 200 nt or more were annotated by the Blastn homology search to identify virus-like sequences. The pre-processed RNA-Seq data from each of 68 leaf samples were mapped again on the reference, including sequences of a novel virus using RSEM-1.2.15 (Fig. S2, Supporting Information). RNA from five putative infected and five non-infected plants was reverse transcribed (High-Capacity cDNA Reverse Transcription Kit, Life Technologies) and co-occurrence of the two genome segments of the putative novel virus, RdRP (RNA-dependent RNA polymerase) and CP (capsid protein), was examined. PCRs (KOD-plus, TOYOBO, Japan) were conducted with the specific primer sets that amplify 428-nt RdRP and 450-nt CP fragments (forward ATGAAGAACACCGTCGTTCTC, reverse GACTTCAGTTTCC-CGTCATAC; and forward ATGTCGATGAAAAGAAAGTCAAG, reverse CTGAACCCAGAAAAGAACTCC, respectively).

Deduced amino acid (AA) sequences were determined for RdRP and CP of the novel partitivirus. The RdRP AA sequences of the novel virus, reported partitivirus (58 sequences) and a cryspovirus as an outgroup were aligned using MAFFTv.7.273 (the default settings; http://mafft.cbrc.jp/alignment/software/). The species used for the analysis are listed in Table S3 (Supporting Information). The phylogenetic analysis was conducted using MEGA7 (Kumar, Stecher and Tamura 2016) using the maximum-likelihood method based on the LG model (Le and Gascuel 2008). We chose the model according to the model selection based on the Akaike information criterion (AIC). Initial tree(s) for the heuristic search was obtained by applying neighbour-join and BioNJ algorithms to the matrix and then selecting the topology with the superior log-likelihood value. A discrete Gamma distribution was applied to model differences in evolutionary rate among sites (+G, 5 categories, parameter = 2.3). The rate variation model allowed for some sites to be evolutionarily invariable (+I, 1.7% sites). The branch lengths were determined by the number of substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 356 positions in the final dataset. Confidence of branch points was estimated by 1000 bootstrap replications.

TEM analysis of infected leaves

Transmission electron microscopy (TEM) was conducted to observe viruses in plant tissue. Leaf sections measuring 2 mm × 2 mm were cut from plants 13 and 24, in which viruses had been
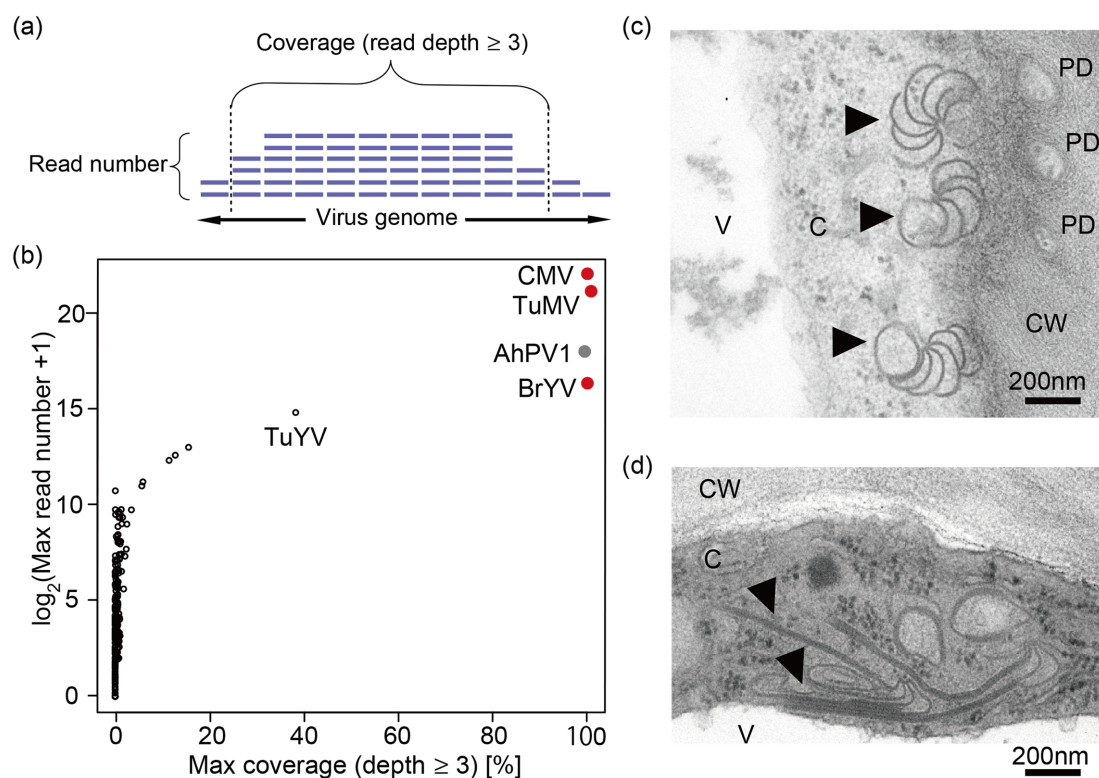
Figure 1. Detection of TuMV, CMV and BrYV in a natural population of *A. halleri*. (**a**) Definitions of the read number and the coverage for a particular virus genome used in this study. Reads obtained from RNA-Seq were mapped onto each virus genome. Read depth refers to the number of reads covering the same sequence. The coverage indicates the percentage of the genome area covered with more than three reads. (**b**) Determination of infection by sequence-reported viruses. Maximum values of the log2 of the read number and the genome coverage for the 3981 sequence reported viruses (NCBI database) are shown. Each point indicates a single virus. Red-filled points represent infecting viruses. A novel virus identified by the *de novo* assembly, AhPV1, is shown by a grey-filled point after re-mapping. (**c** and **d**) TEM images of an infected leaf. Filled arrowheads show the pinwheel (c) and rod-shaped (d) structures of TuMV. V, C and CW indicate the vacuole, cytoplasm and cell wall, respectively. PD indicates the secondary plasmodesma.

detected by RNA-Seq. The sampling and the primary fixation were performed in the field. The primary fixation was done with 2% paraformaldehyde and 2% glutaraldehyde in 0.05 M cacodylate buffer pH 7.4 on ice. Procedures that followed the primary fixation were conducted as described in a previous study (Goto *et al.* 2014).

Analysis of virus genome variation

To characterise the sequence of viruses from each individual plant, a consensus sequence within each leaf sample was constructed from RNA-Seq data. The sequenced reads were mapped on each virus genome in reference virus sequences, using bowtie in RSEM as mentioned above. The aligned reads of the detected viruses were extracted from the bam files by samtools. Next, consensus sequences of viruses were constructed in a single host plant using samtools and bcftools, with the criteria of mapQ score >10 and depth $\geq$3 for each read. For each detected virus species, we obtained consensus sequences from all infected plants (numbers of infected plants are listed in Table 1).

Analysis of virus infection patterns

To examine whether similarity of virus species composition within hosts depended on the distance between plants, we conducted a spatial autocorrelation analysis using the data set from the rectangular plot. Pearson correlation between Euclidean distances (cm) and the Jaccard dissimilarities of virus species com-

position was examined using all pairs of plant individuals. The Jaccard dissimilarities were calculated using the vegan package and its vegdist function in R (Oksanen *et al.* 2016). We conducted the Mantel test for the correlation with 9999 permutations by using the mantel function in the vegan package (Oksanen *et al.* 2016).

Fisher tests were conducted to examine the relationship between infection by each virus and the evidence of herbivory or leaf chlorosis. We also applied the Fisher test to examine whether coinfection by a particular set of viruses was more frequent than expected by chance. In these analyses, we used the fisher.test function (implemented in R) with a two-tailed test.

Analysis of the host gene expression

Analysis of host gene expression was conducted for the reads mapped on the *A. halleri* transcriptome (32648 genes). Samples with fewer than 10^5 reads were excluded from the host gene analysis. Genes on which fewer than 10 reads were mapped on average were excluded from the analysis. The differences in expression of the remaining 10603 genes were examined using edgeR (Robinson, McCarthy and Smyth 2010). Multiple testing corrections were performed by setting the FDR (false discovery rate; Benjamini and Hochberg 1995).

Differentially expressed genes were further analysed using a generalised linear model with a negative binomial family and a log link function. The response variable was the read count of a focal gene. The presence/absence of chlorosis in the host and

Table 1. Infection rate of each virus and relationship between chlorosis and infection.

Virus species		No. of plant	Infection rate(%)	No. of plant with chlorosis	No. of plant without chlorosis	Asymptomatic infection (%)	P value[a]
TuMV	+	39	57	22	17	44	0.013
	−	29		7	22		
CMV	+	18	26	13	5	28	0.0050
	−	50		16	34		
BrYV	+	21	31	12	9	43	0.12
	−	47		17	30		
AhPV1	+	56	82	25	31	55	0.54
	−	12		4	8		
Any	+	62	91	28	34	44	0.23
	−	6		1	5		

[a]P value by Fisher's exact test to examine relationship between chlorosis and infection.

viral infection were candidate explanatory variables. We evaluated the effects of the two explanatory variables by comparing AICs in the following five models: a *null* model and models with the chlorosis term only, the virus term only, chlorosis and virus terms, and the interaction between chlorosis and viral infection.

RESULTS

In order to identify the diversity and features of viruses in wild plants, RNA-Seq was conducted for *A. halleri* individuals in a natural habitat. On average, 2121142 reads per sample were sequenced. From 1% to 82% of the total read (1% to 93% of the total mapped reads) was mapped on the 3981 virus reference genomes.

Detection of viruses by RNA-Seq

Reliable detection of viruses using RNA-Seq depends on the exclusion of false detections resulting from mismapped reads. To solve this problem, we applied a combined analysis of read number and coverage against the virus genome (Fig. 1a). To determine the viruses that were present in the plant population, we focused on the maximum values of the read number and coverage of each virus among all the plant samples (max read-number and max coverage, Figs 1b and S3, Supporting Information). We found that three viruses had more than 95% max coverage (red-filled points in Fig. 1b): *Turnip mosaic virus* (TuMV, genus *Potyvirus*, family *Potyviridae*), *Cucumber mosaic virus* (CMV, genus *Cucumovirus*, family *Bromovirus*) and *Brassica yellows virus* (BrYV, proposed genus *Polerovirus*, family *Luteoviridae*). For most viruses in the reference sequences, the max coverage, if any, was <10% of the coverage accompanied by distinctively low read numbers (Fig. 1b). For four viruses, the max coverage was marginal (10%–38%, Fig. 1b). For example, *Turnip yellows virus* (TuYV) showed 38% max coverage, but we attributed this to the high-sequence similarity with BrYV (81% identity between BrYV NC_016038 and TuYV NC_003743). Because BrYV showed much higher max coverage, we concluded that the apparent 38% max coverage of TuYV was derived from mismapping. The combination of data from the max read-number and max coverage indicated that TuMV, CMV and BrYV were the infecting viruses in *A. halleri* at the Omoide River site. These three viruses are known to infect plants in the Brassicaceae family, but to our knowledge, this is the first report of their infection in *A. halleri*. In addition, as far as we know, this is the first report of BrYV in Japan, although its closely related species TuYV has been reported in Japan (Ohki *et al.* 1977; Stevens, McGrann and Clark 2008).

For each sample, we examined the read number and coverage of these three viruses (Fig. S3). The coverage varied from 39% to 99%, 90% to 100% and 12% to 99% among samples for TuMV, CMV and BrYV, respectively (Fig. S3). Although some samples showed reduced coverage, in these samples the mapped reads tended not to be localised but were widely distributed on the virus genome. Therefore, we considered the samples with low virus coverage to be infected by the viruses as long as they were TuMV, CMV and BrYV. Even when the viruses that had marginal coverage (for example, <30% coverage) were excluded from the analysis, the results led to the same conclusion.

Ultramicroscopic observation of two infected leaves of *A. halleri* (from plants 13 and 24) revealed the pinwheel structures of TuMV (Fig. 1c). This structure, an inclusion body of the virus, corresponds to its replication stage (Sorel, Garcia and German-Retana 2014). A rod-like structure, another typical structure of TuMV, was also observed (Fig. 1d). TuMV structures were found in the cytoplasm of leaves and were frequently observed near the secondary plasmodesmata (Fig. 1c), which are reported to be channels for cell-to-cell virus transport (Kumar *et al.* 2015). It is difficult to identify CMV and BrYV in plant cells using electron microscopy without performing immunogold staining. Therefore, we could not identify CMV and BrYV through electron microscopy, although their sequences were present in the two plant samples.

Detection of a novel virus by *de novo* assembly

We assembled 5885060 unmapped reads and obtained 3898 contigs. The majority of the contigs were annotated to plants and the viruses in the reference sequence (3366 and 43 contigs, respectively). In all, 51, 33 and 12 contigs were annotated to fungi, bacteria and insects, respectively. There were 338 contigs that were not annotated to any organisms. In addition, three contigs were annotated to putative virus sequences. The first of these, a contig with a length of 1959 nucleotides (nt) (putative RdRP, Fig. 2a), was similar to RdRP of *Vicia faba partitivirus 1* (VfPV1) with 67.2% nt identity, and the aligned 1698 nt covered 88.7% of the 1915 nt full length of the corresponding VfPV1 genome fragment. The second contig, with a length of 1763 nt (putative CP, Fig. 2a), was similar to a partitivirus CP-like sequence reported from *Olimarabidopsis pumila* with 67.5% nt identity (1181 nt cover against 1251 nt, 94.4%). The last contig was annotated to *Aristotelia chilensis virus 1* (AcV1), but the aligned sequence was short (148 nt cover against 7122 nt, 2.1%).
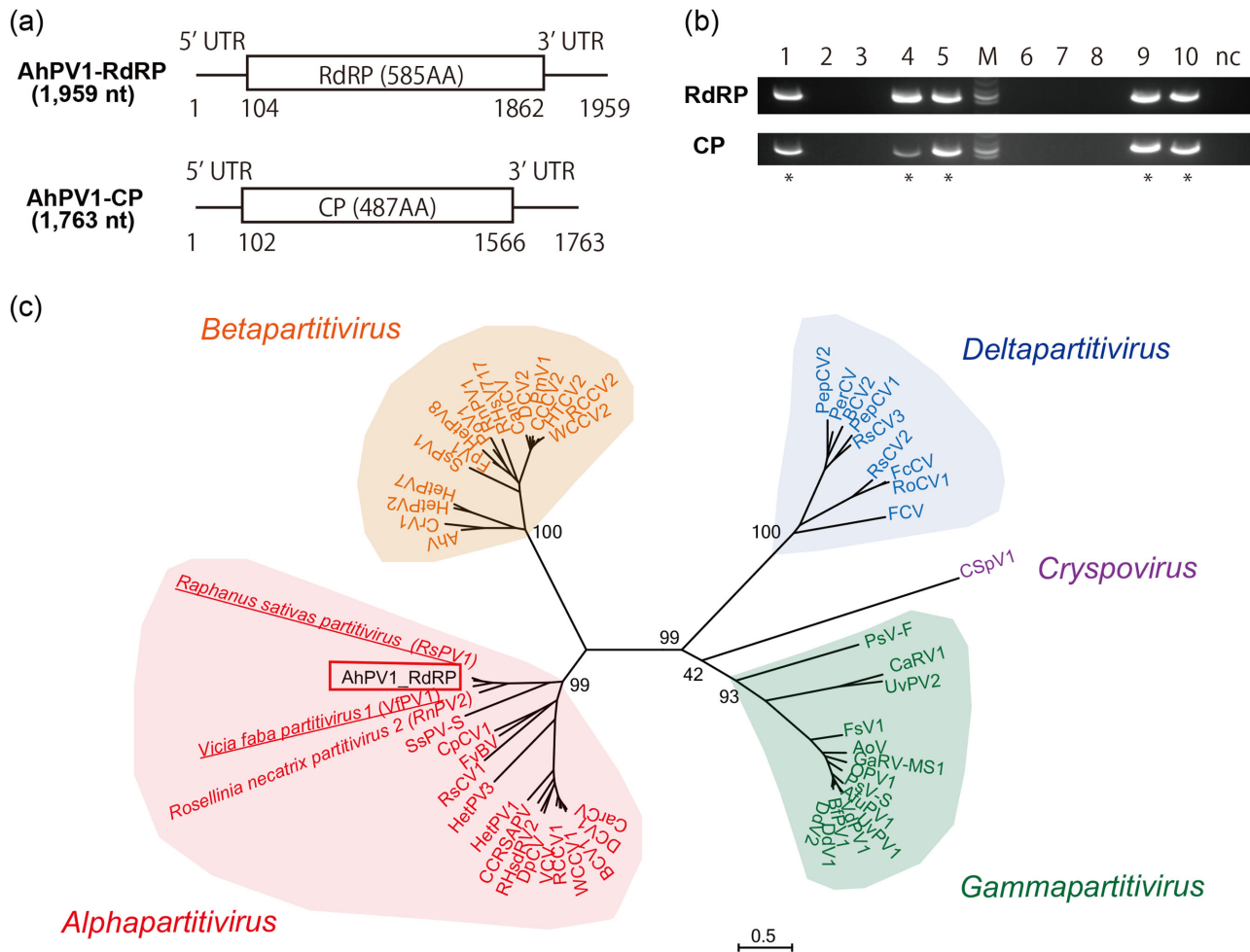
Figure 2. AhPV1, a novel partitivirus identified by the *de novo* assembly. (**a**) Schematic diagrams of AhPV1-RdRP and AhPV1-CP fragments. The total nucleotide lengths of the fragments are shown in parentheses. Coding regions of RdRP and CP are shown by boxes with amino acid lengths in parentheses. Numbers under the diagrams indicate the starting and terminative positions of the fragment and coding region. (**b**) Amplifications of RdRP- and CP-specific fragments for five putative infected (shown by asterisks) and five non-infected leaves. M and nc represent the size marker (400 and 500 bp) and no-template negative control, respectively. (**c**) Phylogenetic locations of AhPV1 (boxed) based on RdRP AA sequences. Corresponding viruses shown by abbreviations are listed in Table S3. Colours correspond to the four genera of partitivirus. *Cryspovirus* was used as an outgroup. Underlining indicates related *Alphapartitivirus* for which the CP sequence is unknown. Numbers represent bootstrap values (percentages).

We confirmed colocalisation of RdRP and CP sequences in infected plants only (Fig. 2b). On the basis of this co-occurrence and the following phylogenetic analyses, we considered that two contigs were derived from a novel partitivirus, named *Arabidopsis halleri partitivirus 1* (AhPV1). The putative full-length AhPV1-RdRP and AhPV1-CP fragments were 1959 nt and 1763 nt with 585 and 487 AA lengths, respectively (Fig. 2a; NCBI GenBank Accession Nos. LC151461, LC151462). The phylogenetic tree of RdRP AA sequences with the highest log likelihood (–26845) suggested that AhPV1 belongs to the genus *Alphapartitivirus* (Fig. 2c). In RdRP sequences, AhPV1 showed the highest homology (80.7% AA identity and 90.6% AA similarity) to *Raphanus sativus partitivirus 1* (RsPV1), the second highest homology (66.9% identity and 78.5% similarity) to VfPV1 and the third highest homology (52.1% identity and 66.4% similarity) to *Rosellinia necatrix partitivirus 2* (RnPV2, Fig. 2c). CP sequences of the RsPV1 and VfPV1 have not been reported. Among the partitiviruses with reported CP sequences, AhPV1 had the highest homology to that of RnPV2 with 31.2% AA identity (50.7% similarity). We found that AhPV1 showed >95% max coverage (a grey-filled point in Fig. 1b).

Analysis of virus genome variation

RNA-Seq enabled us to survey whole viral genome sequences from the plants examined. The data allowed us to evaluate diversity within virus species, especially between host individuals. One consensus sequence was determined for each virus species in each host individual. The length of consensus sequences that passed the criteria (see Methods section) ranged from 83% to 100%, 92% to 97%, 74% to 97% and 75% to 99% for TuMV, CMV, BrYV and AhPV1, respectively. The nucleotide identities of consensus sequences within the same virus species ranged from 96.1% to 99.8%, 92.3% to 100%, 95.3% to 100% and 99.8% to 100% in TuMV, CMV, BrYV and AhPV1, respectively.

Spatial pattern of infected plants, herbivory and chlorosis

In 62 out of 68 examined plants, we found at least one virus species (Table 1). The infection rates varied among four viruses. The most commonly found virus was AhPV1 with an infection
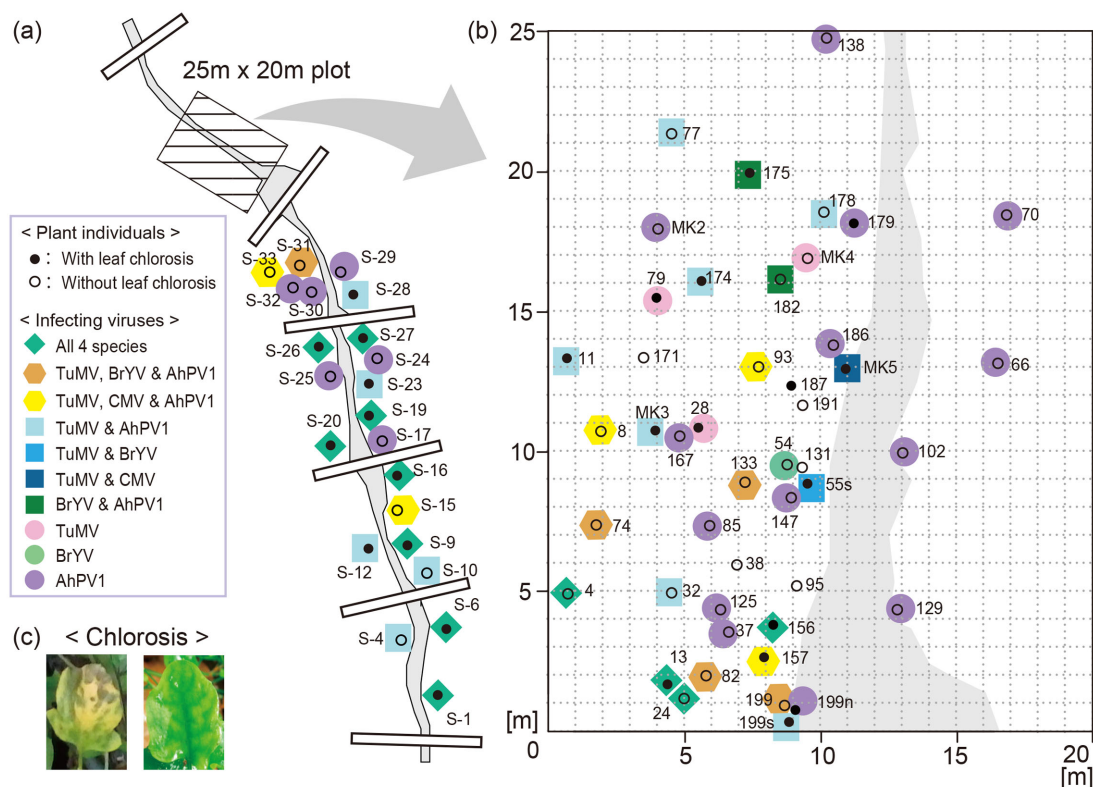
Figure 3. Spatial distributions of *A. halleri* plants infected by different combinations of viruses, and chlorosis phenotypes in the study site. Spatial pattern in samples collected along the creek (**a**) and in the 20 m × 25 m rectangular plot (**b**). Each small circle represents a single plant. Presence/absence of leaf chlorosis is shown by filled/open circles, respectively. Combinations of infecting virus species are shown by coloured shadings. Circle, square, hexagon and diamond shadings represent single, double, triple and quadruple infections, respectively. The grey area indicates a stream that flows from the top to the bottom of diagrams. White bars crossing the stream indicate the positions of erosion-control dams. (**c**) Typical chlorosis of *A. halleri* leaves.

rate of 82% (56/68), comprising 90% (56/62) of the virus infections (Table 1). TuMV, CMV and BrYV infected 57%, 26% and 31% of plants, respectively (Table 1). TuMV, CMV, BrYV and AhPV1 were found in both sets of samples along the creek (Fig. 3a) and in the rectangular plot (Fig. 3b). Viruses were spatially spread widely within the examined area, and no significant correlation was detected between distance among host plants and similarity of virus species composition (Mantel test, $r = 0.16$, $P > 0.05$). No infected plants showed lethal or severe symptoms at the whole-plant level. Representative examples of leaf chlorosis in our samples are shown in Fig. 3c. Chlorosis was observed in 29 out of 68 leaf samples. A positive relationship between the presence of chlorosis and infection by a single virus was detected for TuMV and CMV (Fisher's exact test, $P < 0.05$, Table 1). Viral infection without apparent chlorosis was observed in 44%, 28%, 43% and 55% of plants infected with TuMV, CMV, BrYV and AhPV1, respectively (Table 1).

We observed herbivory by leaf beetles in 38 out of 68 observed individuals (Table S4, Supporting Information). Viral infection and herbivory showed no significant relationship (Fisher's exact test, $P > 0.05$; Table S4). Herbivory and chlorosis showed no significant relationship between plants (Fisher's exact test, $P > 0.05$, data not shown).

Multiple infection

In some cases, multiple virus species were found in a single host plant individual (Fig. 3). Multiple infections by 2, 3 and 4 viruses were observed in 16, 10 and 12 plants, respectively (Fig. 4a).

Infections by a single virus species were observed in 3, 0, 1 and 20 plants in the cases of TuMV, CMV, BrYV and AhPV1, respectively (Fig. 4a). Significantly higher coinfections than would have occurred by chance were detected in the pairs between TuMV, CMV and BrYV (Fig. 4b). These three viruses are known to be transmitted horizontally, whereas partitiviruses are rarely transmitted horizontally. Among the three viruses with horizontal transmission, multiple infections were often observed when plants were infected by TuMV (Fig. 4a).

Using the RNA-Seq data, we compared the expression of host genes in TuMV-infected and non-infected plants. Out of 10603 host genes, 4 were detected as candidates in which expression was altered by viral infection (Fig. 4c). Expression of the *Argonaute 2* orthologue (*AhgAGO2*) was 3.2 times higher, on average, in infected plants than in non-infected plants (Fig. 4d, Table 2). Because TuMV infection was associated with the presence of host chlorosis (Table 1), we examined whether infection by TuMV or the presence of chlorosis explained the *AhgAGO2* expression level by comparing five models with different combinations of terms as explanatory variables: *null*, chlorosis only, virus only, chlorosis and virus, and the interaction between chlorosis and virus. The AICs of these models were 493.2, 492.7, 487.8, 489.0 and 491.0, respectively. The model that included the TuMV infection term alone showed the smallest AIC. The differences in the remaining three genes are ambiguous because they contained many samples without reads for both infected and non-infected plants (Table 2, Fig. S4). In *A. thaliana*, AGO2 is involved in RNA silencing, a predominant antiviral mechanism of the host plant,
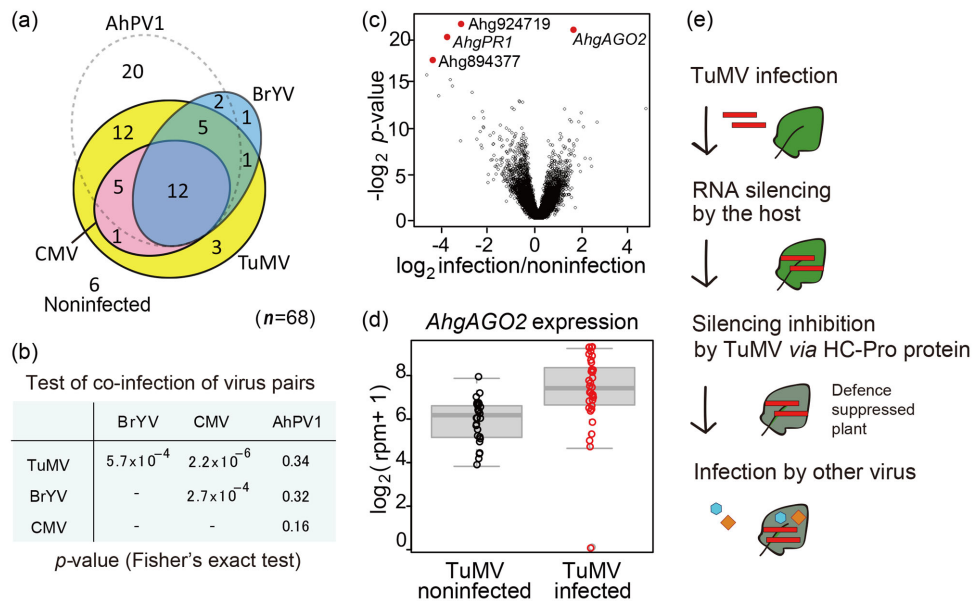
Figure 4. Multiple infections and host response. (**a**) Frequency of multiple infection at the Omoide River site. Numbers of plants infected by the viruses are shown in a Venn diagram ($n = 68$). The overlapping areas represent multiple infections from corresponding viruses. (**b**) Test of coinfection of virus pairs. The results of Fisher's exact test (P-value) between all virus pairs are shown. (**c**) Differentially expressed genes in the comparison between TuMV-infected and non-infected host transcriptomes. In the volcano plot, statistical significance ($-\log_2$ P-value) is plotted against the log ratio of the average expression (rpm, infection/non-infection) for each host gene. Red dots represent the genes with FDR-adjusted P-values of <0.05. (**d**) Difference in *AhgAGO2* gene expression between TuMV-non-infected and TuMV-infected plants (black and red circles, respectively). The vertical axis is the read number of the *AhgAGO2* gene obtained from RNA-Seq. The box and lines denote the median, first quartile, and third quartile values of each sample group. (**e**) The hypothesised mechanism of facilitation in multiple infection by TuMV. TuMV-infected plants recognise the infection and respond with RNA silencing. TuMV inhibits the host RNA silencing with HC-Pro protein. A plant whose defence is suppressed becomes susceptible to infection by other viruses.

Table 2. List of upregulated or downregulated genes by virus infection.

Gene name	Gene alias	AGI code	Fold change of mean value	Fold change of medium value	Adjusted P (FDR)[a]
Ahg473645	*AhgAGO2*	AT1G31280.1	3.2	2.2	6.1×10^{-3}
Ahg894377	–	–	0.045	–[b]	0.028
Ahg924719	–	AT1G62760.1	–[b]	–[b]	6.1×10^{-3}
Ahg931178	*AhgPR1*	AT2G14610.1	6.3×10^{-4}	–[b]	6.7×10^{-3}

[a]Adjusted P-value by Fisher's exact test using edgeR to test relationship between each gene expression and infection.
[b]The value of the denominator was 0.

and is a component of RISC. *AhgAGO2* expression did not differ significantly between infected and non-infected plants when they were tested separately for CMV, BrYV and AhPV1 infections (data not shown). This result suggests that transcription of *AhgAGO2* was induced by TuMV infection.

Sequence analysis of TuMV found that the strains from the Omoide River site contained an AA sequence motif, 'FRNK' (Table S5, Supporting Information) that has been reported to be important in RNA-silencing suppressor activity (Tan *et al.* 2005; Shiboleth *et al.* 2007). The AA sequence of TuMV from our site also included other reported mutations that are known to enhance systematic infection (Table S5). We did not find a reported sequence that increases virulence (Table S5).

DISCUSSION

To understand the ecology of viruses, it is necessary to obtain comprehensive information on virus–virus and virus–host interactions in natural systems. In this study, we showed that RNA-Seq realised such analyses without any prior assumptions

regarding infecting viruses, virus symptoms or responding host genes. In a natural population of *A. halleri*, we developed a comprehensive and solid listing of infecting viruses, including a novel virus, for each individual plant. The lists allowed us to identify a potential facilitator of multiple infection. RNA silencing-mediated host anti-virus responses were suggested to be operating under natural conditions. Interestingly, virus reads reached a maximum level that was equivalent to that of the host's total mRNA, although asymptomatic infection was common.

In studies of plant viruses in the field, ELISA- or PCR-based methods have commonly been used (Prendeville *et al.* 2012). However, these methods can only detect a specific targeted virus and cannot be used for comprehensive detection. In this study, the use of RNA-Seq enabled the comprehensive determination of infection by all sequence-reported viruses. By considering the coverage of virus genomes, we avoided misdetection of viruses caused by similarity between short sequence parts. The high-coverage genome allowed us to identify infecting viruses to the species level, which is often difficult when partial sequences are used.

Furthermore, we successfully identified a novel partitivirus, AhPV1, by *de novo* assembly. Partitiviruses are encapsidated dsRNA viruses with two genome segments of RdRP and CP that infect plants, fungi and protozoa (Nibert *et al.* 2014). Partitiviruses are divided into four genera: *Alphapartitivirus, Betapartitivisue, Gammapartitivirus* and *Deltapartitivirus* (Nibert *et al.* 2014), and AhPV1 was identified as a member of *Alphapartitivirus*. AhPV1 showed a high infection rate in the *A. halleri* population, although we found no obvious symptomatic effects. Previous studies also reported that partitiviruses are cryptic and cause few symptoms in their hosts (Roossinck 2010). In Brassicaceae crops, *de novo* assembly analyses of dsRNA have identified multiple partitiviruses (Li *et al.* 2016). In another example using wild plants, *de novo* assembly identified novel viruses belonging to 12 families, including partitiviruses (Roossinck *et al.* 2010). Partitivirus has been reported to be transmitted vertically through seeds (Li *et al.* 2016). Indeed, in a preliminary experiment, we observed vertical transmission of AhPV1 (9/15 infected seedlings from one infected plant collected in the wild and surface-sterilised seeds).

Our data showed that asymptomatic infection is common for all four virus species detected in this study. TuMV and CMV are well known to widely infect agricultural crops and severely damage their hosts (Tomlinson 1987; Rybicki 2015). TuMV can cause severe symptoms in *Brassica* and *Raphanus* in agricultural fields (Tomlinson 1987; Walsh and Jenner 2002). Of note, we recorded infections by TuMV with active replication accompanied by only slight or no symptoms. TEM analysis of TuMV-infected leaves revealed the presence of inclusion bodies, structures that are specific to the replication stage of the virus. Furthermore, virus reads reached a maximum level that was equivalent to that of the host's total mRNA. The observation of inclusion bodies and the vast amount of virus reads indicate that the viruses were actively replicating in *A. halleri*. Therefore, the current case is different from asymptomatic infection caused by cryptic or low-titre viruses, in which active replication does not occur (Boccardo *et al.* 1987). Infection that produces severe symptoms is generally not adaptive for the viruses, whose replication depends on the host (Pagan *et al.* 2014). Meanwhile, asymptomatic infection can be adaptive for both plants and viruses (May and Anderson 1983; Marquez and Roossinck 2012). To examine whether these viruses affect long-term host fitness, further field studies are required.

The patterns of multiple infections suggested that TuMV was a facilitator of infection by other viruses. In laboratory experiments, TuMV was reported to possess HC-Pro, a potyvirus-specific suppressor of RNA silencing in the host plant (Anandalakshmi *et al.* 1998). RNA silencing of plants is the predominant host defence mechanism against viruses. In RNA silencing, Dicer and RISC bind to virus RNA and degrade it. HC-Pro protein, produced by TuMV, inhibits the binding of Dicer and RISC to virus RNA. It was reported that HC-Pro generated by potyvirus could facilitate the replication of other viruses in some cases (Pruss *et al.* 1997; Fukuzawa *et al.* 2010; Syller 2012). On the basis of these reports and our observations, we hypothesised that TuMV facilitates multiple infection in nature (Fig. 4e). The hypothesised process involves silencing inhibition (suppression of host RNA silencing against viruses) by TuMV viaHC-Pro (Fig. 4e). As a result, infection by other viruses to the defence-suppressed plants is facilitated (Fig. 4e).

This hypothesis is consistent with the results of host gene expression analysis. The *AhgAGO2* gene, a key gene in RNA silencing (Alvarado and Scholthof 2011), was significantly upregulated by TuMV infection. This result was supported by the find-

ing that AGO2 works as a component of RISC, a protein complex that incorporates viral RNA in RNA silencing (Harvey *et al.* 2011). This complex is a target of the HC-Pro protein produced by TuMV, which suppresses the binding of RISC to viral RNA. Upregulation of *AhgAGO2* appears to mitigate the suppression. However, upregulation and accumulation of *AhgAGO2* mRNA was reasonable, because suppression by HC-Pro works at a protein level, not at a transcription level. The presence of RNA-silencing suppressors has been reported in CMV (Brigneti *et al.* 1998; Roth, Pruss and Vance 2004) and viruses in Luteoviridae to which BrYV belongs (Roth, Pruss and Vance 2004; Liu *et al.* 2012). Any hypotheses on the facilitation of multiple infection by TuMV, CMV and BrYV should be tested experimentally.

RNA-Seq data from infected plants represent the densest type of information in the plant–virus system, the virus genome and host transcriptome. Therefore, RNA-Seq will be a promising way to reveal the molecular basis of host–virus interactions. It is important to note that the application of dual RNA-Seq to host–virus systems differs from its application to host–other pathogen systems. In previous studies that applied dual RNA-Seq analyses to fungal and bacterial pathogens, expression analyses were carried out separately for the host and the pathogen (Kawahara *et al.* 2012; Tierney *et al.* 2012; Westermann *et al.* 2016). In host–virus systems, the number of virus-derived reads represents the multiplication process that potentially increases virus sequences exponentially. Indeed, we observed that virus reads reached a maximum level that was equivalent to that of the host's total mRNA, even when the host showed no visible symptoms. Because the total number of reads was highly dependent on virus reads, we calculated the rpm of both host genes and virus genomes against total reads derived from host genes (mRNA), instead of the total reads including virus reads.

Changes in the transcriptome due to infection may not be obvious in natural conditions, although experimental viral infections in the laboratory have caused drastic changes in host transcriptomes (Marathe *et al.* 2004; Yang *et al.* 2007; Postnikova and Nemchinov 2012). A number of factors have been found to affect plant transcriptomes, such as the weather, host genotype and microenvironments (Nagano *et al.* 2012; Richards *et al.* 2012; Plessis *et al.* 2015). Such noise is likely to inhibit the statistical detection of transcriptomic changes by viruses. Because RNA-Seq opens up opportunities to study plant–virus interaction in diverse natural habitats, a sophisticated statistical analysis should be developed to extract the effect of virus infection on hosts from the noisy transcriptome data. Revealing hidden plant–virus interactions in the natural environments is important not only because it adds to our understanding of the ecology of biological interaction, but also because it can improve agricultural management by revealing the natural history of crop pests.

DATA ACCESSIBILITY

AUTHOR CONTRIBUTION

MK, AJN, MNH and HK designed the research. MK and HK conducted the field work. MK and MNH conducted the laboratory experiment. MK and AJN conducted the data analysis. MK, AJN

and HK wrote the manuscript. HK supervised the study. All authors discussed the results and approved the manuscript.

SUPPLEMENTARY DATA

Supplementary data are available at FEMSEC online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

Aikawa S, Kobayashi MJ, Satake A *et al*. Robust control of the seasonal expression of the Arabidopsis FLC gene in a fluctuating environment. *P Natl Acad Sci USA* 2010;**107**:11632–7.

Alexander HM, Mauck KE, Whitfield AE *et al*. Plant-virus interactions and the agro-ecological interface. *Eur J Plant Pathol* 2014;**138**:529–47.

Alvarado VY, Scholthof HB. AGO2: a new argonaute compromising plant virus accumulation. *Front Plant Sci* 2011;**2**:112.

Anandalakshmi R, Pruss GJ, Ge X *et al*. A viral suppressor of gene silencing in plants. *P Natl Acad Sci USA* 1998;**95**:13079–84.

Barba M, Czosnek H, Hadidi A. Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses* 2014;**6**:106–36.

Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 1995;**57**:289–300.

Boccardo G, Lisa V, Luisoni E *et al*. Cryptic plant viruses. *Adv Virus Res* 1987;**32**:171–214.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20.

Brigneti G, Voinnet O, Li WX *et al*. Viral pathogenicity determinants are suppressors of transgene silencing in Nicotiana benthamiana. *EMBO J* 1998;**17**:6739–46.

Cooper I, Jones RAC. Wild plants and viruses: under-investigated ecosystems. *Adv Virus Res* 2006;**67**:1–47.

Fraile A, Garcia-Arenal F. Environment and evolution modulate plant virus pathogenesis. *Curr Opin Virol* 2016;**17**:50–6.

Fukuzawa N, Itchoda N, Ishihara T *et al*. HC-Pro, a potyvirus RNA silencing suppressor, cancels cycling of Cucumber mosaic virus in Nicotiana benthamiana plants. *Virus Genes* 2010;**40**:440–6.

Goto C, Tamura K, Fukao Y *et al*. The novel nuclear envelope protein KAKU4 modulates nuclear morphology in Arabidopsis. *Plant Cell* 2014;**26**:2143–55.

Grabherr MG, Haas BJ, Yassour M *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 2011;**29**:644–52.

Hackett J, Muthukumar V, Wiley GB *et al*. Viruses in Oklahoma Euphorbia marginata. *Proc Okla Acad Sci* 2009;**89**:49–54.

Harvey JJ, Lewsey MG, Patel K *et al*. An antiviral defense role of AGO2 in plants. *PLoS One* 2011;**6**:e14639.

Hu TT, Pattyn P, Bakker EG *et al*. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat Genet* 2011;**43**:476–81.

Hull R. *Plant Virology*. Amsterdam: Elsevier Academic Press, 2014.

Incarbone M, Dunoyer P. RNA silencing and its suppression: novel insights from in planta analyses. *Trends Plant Sci* 2013;**18**:382–92.

Kawagoe T, Kudoh H. Escape from floral herbivory by early flowering in Arabidopsis halleri subsp. gemmifera. *Oecologia* 2010;**164**:713–20.

Kawahara Y, Oono Y, Kanamori H *et al*. Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PLoS One* 2012;**7**:e49423.

King AMQ, Elliot L, Michael JA *et al*. *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Amsterdam: Elsevier, 2011.

Kircher M, Sawyer S, Meyer M. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* 2012;**40**:e3.

Kreuze JF, Perez A, Untiveros M *et al*. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology* 2009;**388**:1–7.

Kubota H, Takenaka C. Arabis gemmifera is a hyperaccumulator of Cd and Zn. *Int J Phytoremediat* 2003;**5**:197–201.

Kumar D, Kumar R, Hyun TK *et al*. Cell-to-cell movement of viruses via plasmodesmata. *J Plant Res* 2015;**128**:37–47.

Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016;**33**:1870–4.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;**9**:357–9.

Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol* 2008;**25**:1307–20.

Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;**12**:323.

Li L, Liu J, Zhang Q *et al*. Seed-borne viral dsRNA elements in three cultivated Raphanus and Brassica plants suggest three cryptoviruses. *Can J Microbiol* 2016;**62**:287–95.

Liu Y, Zhai H, Zhao K *et al*. Two suppressors of RNA silencing encoded by cereal-infecting members of the family Luteoviridae. *J Gen Virol* 2012;**93**:1825–30.

Malmstrom CM, Hughes CC, Newton LA *et al*. Virus infection in remnant native bunchgrasses from invaded California grasslands. *New Phytol* 2005a;**168**:217–30.

Malmstrom CM, McCullough AJ, Johnson HA *et al*. Invasive annual grasses indirectly increase virus incidence in California native perennial bunchgrasses. *Oecologia* 2005b;**145**:153–64.

Marathe R, Guan Z, Anandalakshmi R *et al*. Study of Arabidopsis thaliana resistome in response to cucumber mosaic virus infection using whole genome microarray. *Plant Mol Biol* 2004;**55**:501–20.

Marquez LM, Roossinck MJ. Do persistent RNA viruses fit the trade-off hypothesis of virulence evolution? *Curr Opin Virol* 2012;**2**:556–60.

May RM, Anderson RM. Epidemiology and genetics in the co-evolution of parasites and hosts. *P Roy Soc Lond B Bio* 1983;**219**:281–313.

Mitchell CE, Power AG. Release of invasive plants from fungal and viral pathogens. *Nature* 2003;**421**:625–7.

Morlan JD, Qu K, Sinicropi DV. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. *PLoS One* 2012;**7**:e42882.

Mueller EE, Groves RL, Gratton C. Crop and non-crop plants as potential reservoir hosts of Alfalfa mosaic virus and Cucumber mosaic virus for spread to commercial snap bean. *Plant Dis* 2011;**96**:506–14.

Nagano AJ, Honjo MN, Mihara M *et al.* Detection of plant viruses in natural environments by using RNA-Seq. *Methods Mol Biol* 2015;**1236**:89–98.

Nagano AJ, Sato Y, Mihara M *et al.* Deciphering and prediction of transcriptome dynamics under fluctuating field conditions. *Cell* 2012;**151**:1358–69.

Nibert ML, Ghabrial SA, Maiss E *et al.* Taxonomic reorganization of family Partitiviridae and other recent progress in partitivirus research. *Virus Res* 2014;**188**:128–41.

Ohki ST, Yamashita S, Arai K *et al. Ann Phytopath Soc Japan* 1977;**43**:46–54.

Oksanen J, Blanchet FG, Kindt R *et al. vegan: Community Ecology Package. R Package Version 2.3-5* 2016, http://CRAN.R-project.org/package=vegan.

Pagán I, González-Jara P, Moreno-Letelier A *et al.* Effect of biodiversity changes in disease risk: exploring disease emergence in a plant-virus system. *PLoS Pathog* 2012;**8**:e1002796.

Pagan I, Montes N, Milgroom MG *et al.* Vertical transmission selects for reduced virulence in a plant virus and for increased resistance in the host. *PLoS Pathog* 2014;**10**:e1004293.

Plessis A, Hafemeister C, Wilkins O *et al.* Multiple abiotic stimuli are integrated in the regulation of rice gene expression under field conditions. *eLife* 2015;**4**:e08411.

Postnikova OA, Nemchinov LG. Comparative analysis of microarray data in Arabidopsis transcriptome during compatible interactions with plant viruses. *Virol J* 2012;**9**:101.

Prendeville HR, Ye X, Morris TJ *et al.* Virus infections in wild plant populations are both frequent and often unapparent. *Am J Bot* 2012;**99**:1033–42.

Pruss G, Ge X, Shi XM *et al.* Plant viral synergism: the potyviral genome encodes a broad-range pathogenicity enhancer that transactivates replication of heterologous viruses. *Plant Cell* 1997;**9**:859–68.

R Development Core Team. *R: A Language and Environment for Statistical Computing.* 2011, https://www.R-project.org/.

Remold SK. Unapparent virus infection and host fitness in three weedy grass species. *J Ecol* 2002;**90**:967–77.

Richards CL, Rosas U, Banta J *et al.* Genome-wide patterns of Arabidopsis gene expression in nature. *PLos Genet* 2012;**8**:e1002662.

Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**:139–40.

Roossinck MJ. Lifestyles of plant viruses. *Philos T Roy Soc B* 2010;**365**:1899–905.

Roossinck MJ. The big unknown: plant virus biodiversity. *Curr Opin Virol* 2011;**1**:63–7.

Roossinck MJ. Plant virus metagenomics: biodiversity and ecology. *Annu Rev Genet* 2012;**46**:359–69.

Roossinck MJ. Plants, viruses and the environment: Ecology and mutualism. *Virology* 2015;**479–480**:271–7.

Roossinck MJ, Saha P, Wiley GB *et al.* Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. *Mol Ecol* 2010;**19**:81–8.

Roth BM, Pruss GJ, Vance VB. Plant viral suppressors of RNA silencing. *Virus Res* 2004;**102**:97–108.

Rybicki EP. A Top Ten list for economically important plant viruses. *Arch Virol* 2015;**160**:17–20.

Shiboleth YM, Haronsky E, Leibman D *et al.* The conserved FRNK box in HC-Pro, a plant viral suppressor of gene silencing, is required for small RNA binding and mediates symptom development. *J Virol* 2007;**81**:13135–48.

Shiekh MA. Plant defense and the counter defense by viruses. *Global J Res Rev* 2014;**1**:37–44.

Soosaar JL, Burch-Smith TM, Dinesh-Kumar SP. Mechanisms of plant resistance to viruses. *Nat Rev Microbiol* 2005;**3**:789–98.

Sorel M, Garcia JA, German-Retana S. The Potyviridae cylindrical inclusion helicase: a key multipartner and multifunctional protein. *Mol Plant Microbe In* 2014;**27**:215–26.

Stevens M, McGrann G, Clark B. Turnip yellows virus (syn Beet western yellows virus): an emerging threat to European oilseed rape production? *HGCA Res Rev* 2008;**69**:1–36.

Stobbe AH, Roossinck MJ. Plant virus metagenomics: what we know and why we need to know more. *Front Plant Sci* 2014;**5**:150.

Syller J. Facilitative and antagonistic interactions between plant viruses in mixed infections. *Mol Plant Pathol* 2012;**13**:204–16.

Tan Z, Gibbs AJ, Tomitaka Y *et al.* Mutations in Turnip mosaic virus genomes that have adapted to Raphanus sativus. *J Gen Virol* 2005;**86**:501–10.

Thurston MI, Pallett DW, Cortina-Borja M *et al.* The incidence of viruses in wild Brassica nigra in Dorset (UK). *Ann Appl Biol* 2001;**139**:277–84.

Tierney L, Linde J, Muller S *et al.* An interspecies regulatory network inferred from simultaneous RNA-seq of Candida albicans invading innate immune cells. *Front Microbiol* 2012;**3**:85.

Tomlinson JA. Epidemiology and control of virus diseases of vegetables. *Ann Appl Biol* 1987;**110**:661–81

Walsh JA, Jenner CE. Turnip mosaic virus and the quest for durable resistance. *Mol Plant Pathol* 2002;**3**:289–300.

Westermann AJ, Forstner KU, Amman F *et al.* Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature* 2016;**529**:496–501.

Westermann AJ, Gorski SA, Vogel J. Dual RNA-seq of pathogen and host. *Nat Rev Microbiol* 2012;**10**:618–30.

Xu P, Chen F, Mannas JP *et al.* Virus infection improves drought tolerance. *New Phytol* 2008;**180**:911–21.

Yang C, Guo R, Jie F *et al.* Spatial analysis of arabidopsis thaliana gene expression in response to Turnip mosaic virus infection. *Mol Plant Microbe In* 2007;**20**:358–70.