Title

A Survey on Plant Viruses in Natural Brassicaceae Communities Using RNA-Seq

The names of authors

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

The affiliations and addresses of the authors

[1] Center for Ecological Research, Kyoto University, Hirano 2-509-3, Otsu 520-2113, Japan

[2] Faculty of Agriculture, Ryukoku University, Yokotani 1-5, Seta Oe-cho, Otsu 520-2914, Japan

The e-mail address and telephone numbers of the corresponding authors

Mari Kamitani

kamitani@ecology.kyoto-u.ac.jp

Hiroshi Kudoh

kudoh@ecology.kyoto-u.ac.jp

TEL: +81-77-549-8200 FAX: +81-77-549-8201

ORCID

Hiroshi Kudoh: 0000-0001-9777-4886

Mari Kamitani: 0000-0002-5803-9747

Acknowledgements

We would like to thank Dr. Y. Sato for his helpful comments on statistical analysis.

Funding information

25 This work was supported by KAKENHI (JP26221106, 16H06171, JP16H01473), CREST (JPMJCR15O1,
26 JPMJCR15O2), and Grant-in-Aid for JSPS Fellows (15J00628).
27
28

Abstract

Studies on plant viruses are biased towards crop diseases and little is known about viruses in natural vegetation. We conducted extensive surveys of plant viruses in wild Brassicaceae plants occurring in three local plant communities in central Japan. We applied RNA-Seq with selective depletion of rRNA, which allowed us to detect infections of all genome-reported viruses simultaneously. Infections of *Turnip mosaic virus* (TuMV), *Cucumber mosaic virus* (CMV), *Brassica yellows virus*, *Pelargonium zonate spot virus* and *Arabidopsis halleri partitivirus 1* were detected from the two perennial species, *Arabidopsis halleri* subsp. *gemmifera* and *Rorippa indica*. *De novo assembly* further detected partial sequences of a putative novel virus in *Arabis fragellosa*. Virus species composition and infection rate differed depending on site and plant species. Viruses were most frequently detected from the perennial clonal plant, *A. halleri*, in which a high clonal transmission rate of viruses across multiple years was confirmed. Phylogenetic analysis of TuMV and CMV showed that virus strains from wild Brassicaceae were included as a major clade of these viruses with other reported strains from crop plants, suggesting that viruses were shared among wild plants and crops. Our studies indicated that distribution of viruses in natural plant populations are determined by the combinations of life histories of viruses and hosts. Revealing viral distribution in the natural plant communities improves our knowledge on the ecology of plant viruses.

Keywords **4 to 6 keywords**

Asymptomatic infection, Brassicaceae, plant communities, Plant viruses, RNA-Seq, Viral ecology

Introduction

Plants and viruses have developed their interactions over the evolutionary time scale in natural environments. Therefore, the distribution patterns of plant viruses in natural vegetation is likely to be different from those found in agricultural fields, although studies on plant viruses have traditionally targeted to viruses that cause crop diseases. Some earlier studies investigated viruses in wild plants based on the idea that natural vegetation may have served as a reservoir for diseases [1, 2]. More recently, increasing number of studies have become investigate plant viruses in natural vegetation aiming to understand their ecology [3-5].

In the natural plant-virus interactions, infections producing severe symptoms are considered to be adaptive for neither plants nor viruses [6]. Indeed, asymptomatic infections are often observed in natural plant communities [7]. Most of them are caused by low-titer viruses, which are rarely transmitted horizontally by vectors [8]. Although many previous studies have been conducted on annual plants, viral infections in perennial plants should also be examined. Plant species with longer lifespan might have a greater influence on virus epidemiology in natural plant communities [9]. In perennial plants, high infection rates by non-seed-borne viruses are expected, because transmission through clonal propagation of host plants allows long-term persistence of viruses through clonal linage of the hosts. Although non-seed-borne viruses usually show more severe symptoms than seed-borne viruses in annual plants [8, 10], this may not be the case in perennial plants when the virus-host interaction lasts for multiple years.

The complexity of natural plant communities (diverse plant species inhabiting the same location) might be an important determinant of distribution and abundance of viruses. Mixed cultures, including host and non-host plants, have been reported to reduce the risk of infection in crops or cultivated plants [11, 12] as the presence of non-host plants lowers the probability of vectors carrying viruses among host plants [13]. Similarly, plant communities with high species complexity have been expected to have low rates of viral infection, because the complexity is expected to lower the frequency of host-virus encounters thereby

reducing virus transmission rates [13].

Multiple infections by more than one virus species to the same host plants are also expected to occur in natural vegetation when interactions between viruses and their hosts last long. Such co-existence might cause competitive or facilitative interactions among viruses, which are mediated by competition for host resources or suppression of host defense mechanisms [14]. Facilitation is expected to instigate co-existence of multiple viruses while competition will prevent it [15, 16]. Spatial aggregation of virus distributions is another common pattern that can be expected in natural vegetation. Localized transmission of a pathogen among hosts or transmission by the same vector species leads to the spatial aggregation of pathogens [17, 18]. Analysis of spatial patterns of infected plants and host-virus combinations in their natural habitats will contribute to estimation of the risk of viral transmission to the susceptible hosts inhabiting the surrounding area.

To examine viral diversity in natural plant communities, it is necessary to use a methodology that enables the comprehensive detection of existing virus species. RNA sequencing (RNA-Seq), by incorporating *de novo* genome/transcriptome assembly, became a powerful tool for identifying both known and unknown viral RNA and DNA from plant samples [19]. This technique allows us to survey existing viruses from natural plant communities without any prior information on their distribution. In our previous study, we found a high rate of viral infection in wild Brassicaceae, *Arabidopsis halleri* subsp. *gemmifera* (hereinafter referred to as *A. halleri*), in its natural habitat in central Japan [20]. Based on this, we hypothesized that Brassicaceae species coexisting with *A. halleri* might be potential hosts for the same viruses or they may host distinct sets of viruses.

In the present study, to reveal co-infections, differences in viral infection rates among sites, and viruses shared among different plant species, we conducted a comprehensive survey of viruses in wild Brassicaceae species occurring in three local natural plant communities in central Japan. We employed RNA-Seq method modified for comprehensive viral detection [20] and reverse transcription quantitative polymerase chain

reaction (RT-qPCR), to detect infections. We identified virus species, their natural host range, and the spatial distribution of infected plants within each locality. Co-infections, differences in viral infections among sites, and viruses shared among Brassicaceae species within each community were also analyzed. For the representative viruses that were detected from multiple localities, we applied phylogenetic analysis to relate the properties of their genomes with geographic pattern and host rages [21, 22]. Moreover, possibility of sharing of same viruses between crops and wild plants was discussed.

Materials and Methods

Plants and sampling

Plant samples were collected from three local plant communities occurring along small valleys in hills of Osaka, Shiga, and Hyogo Prefectures in central Honshu, Japan (Table 1). The first, second and third sites were upstream of Mino-gawa River, Mino, Osaka Prefecture (34° 50' N, 135° 28' E, alt. ca. 200 m); in Gongen-dani Valley, Taga, Shiga Prefecture (35° 15' N, 136° 21' E, alt. ca. 320 m); and upstream of Inado-gawa River, Tamba, Hyogo Prefecture (35° 16' N, 134° 57' E, alt. ca. 310 m). Hereinafter, these sites will be referred to as Mino-gawa, Gongen-dani, and Inado-gawa, respectively.

All Brassicaceae species in the three sites were sampled at 5–20 m intervals along the valley to provide a broad picture of the viral presence within each area. Samples were collected within the flowering season of Brassicaceae when all study plant species, both annuals and perennials, were in bloom at least in part. At Mino-gawa, 73 plants representing four Brassicaceae species were sampled on June 18, 2014 and June 11, 2016. At Gongen-dani, 125 plants representing six species were sampled on June 28, 2014 and May 13, 2016. At Inado-gawa, 91 plants representing seven species were collected on May 30, June 3, and June 16, 2014. Samples collected in 2014 were used to identify virus species infecting plants, and samples collected in 2016 were used to survey a wider area within each site, using both RNA-Seq and RT-qPCR. No samples were collected in 2016 in Inado-gawa because no viruses were detected in the initial survey.

121 Sample size differed depending on abundance of each plant species in each study site. A single fully
122 developed young leaf blade was collected from each plant. Cauline leaves were collected in most cases,
123 because most individuals were flowering; rosette leaves were collected in the remaining cases. Immediately
124 after collection, each leaf was immersed in 1.0 mL RNAlater (Life Technologies, CA, USA) to avoid RNA
125 degradation and samples were maintained at 0 °C during the transfer, and then stored at −20 °C in the
126 laboratory until RNA extraction.
127
128 *Total RNA extraction, RNA-Seq library preparation, and sequencing*
129 Leaf samples were completely ground by cylinder-shaped metal beads, using the multi-beads shocker
130 (Yasui Kikai, Japan). Total RNA was extracted from each sample using the Maxwell® 16 Total RNA
131 Purification Kit (Promega Corporation, Madison, WI, USA) according to the manufacturer's instructions,
132 which included DNase I treatment. To identify the virus species in each plant community, RNA-Seq was
133 performed for 30, 35, and 30 samples from Mino-gawa, Gongen-dani, and Inado-gawa, respectively. These
134 samples included perennial plants and plants showing symptom-like phenotypes, because such plants are
135 expected to display high infection rates.
136 The RNA-Seq library was prepared using selective enzymatic depletion of rRNA by thermostable
137 RNaseH (Takara Bio, Japan) to detect different types of viruses simultaneously [23]. The conventional
138 methods with oligo-dT-beads could not be applied for the viruses without poly-A tails on the 3′ end of their
139 genomes. Detailed method of RNA-Seq library preparation is described in our previous study [24]. Single-
140 end 50 bases and index sequencing were performed in HiSeq 2000 (Illumina, CA, USA) platform.
141 Viruses were identified to be present when high number of reads covered majority of viral genome
142 regions [20]. This procedure avoids misdetections caused by partial sequence identity. The number of reads
143 from viruses turned out to be highly distinctive between infected and non-infected plants. Therefore, we
144 calculated the number of virus reads (rpm) using the total reads derived from host genes (excluding rRNA)

as the denominator, instead of using the total reads including virus reads [20]. Details on the methods used for mapping and calculating coverage and/or read numbers are provided in a previous publication [20].

De novo assembly

To identify infection by a novel virus, *de novo* assembly of unmapped reads was conducted. Reads that were not mapped to the reference sequences (transcriptome sequences of *A. halleri* as representatives of Brassicaceae host sequences and genome sequences of known viruses) were treated as unmapped reads. The unmapped reads of all RNA-Seq samples were pooled within each site (28,489,176 reads for Mino-gawa, 81,378,310 reads for Gongen-dani, and 14,339,016 reads for Inado-gawa) and assembled by Trinity v2.0 [25]. Contigs with at least 200 nucleotides (nt) were annotated using a Blastn homology search (NCBI) to identify virus-like sequences. Deduced amino acid (AA) sequences of these contigs were obtained using EMBOSS software provided by European Molecular Biology Laboratory [26]. A homology search of amino acid sequences was conducted using an NCBI Blastp homology search. Sequences that were annotated to any reported virus sequences with moderate similarity (50%–90% similarity) were treated as candidate sequences of a novel virus. If a putative novel virus is present, multiple reads derived from its genome should be found among unmapped reads. To reconstruct the putative novel virus, remapping of the RNA-Seq data on the references with the candidate contigs were conducted using RSEM v1.2.15.

Virus detection by RT-qPCR

To validate viral infections detected by RNA-Seq and to determine the extent of infection (Table 1), 200 ng of RNA from each sample was reverse transcribed (High-Capacity cDNA Reverse Transcription Kit, Life Technologies, CA, USA) with random primers. The RT reaction volume was 20 µL and the reaction profile was 25°C for 10 min, 37°C for 120 min, and 85°C for 5 min. The qPCR that followed was conducted using 1 µL cDNA (derived from 10 ng RNA) and Fast SYBR® Green Master Mix (Life

169 Technologies), according to the manufacturer's instructions. Specific primers for each virus and an internal
170 control gene were designed here or constructed based on the previous studies (see Table S1) and used in
171 RT-qPCR reactions (10 µM each). Purified fragments of *Turnip mosaic virus* (TuMV), *Cucumber mosaic*
172 *virus* (CMV), and *Brassica yellows virus* (BrYV) with known molarity were used to estimate molarity and
173 set the cycle threshold (CT) value of each infection.
174
175 *Analysis of viral infection patterns*
176 All statistical analyses were conducted in R v3.2.2. Differences in virus species composition among
177 sites and plant species were examined using multivariate analysis of variance with 9,999 permutations
178 (perMANOVA). The *vegan* package in R and its *adonis* function using Bray-Curtis similarity were used
179 [27] by setting virus species composition (the presence/absence of five viruses) within plants as the
180 response variable and site, host species, and the interaction between site and host species as the explanatory
181 variables.
182 Fisher's tests were conducted to examine if multiple infection by a particular set of viruses was more
183 frequent than expected by chance. Tests were conducted for each site using the *fisher.test* function
184 implemented in R, with a two-tailed test.
185
186 *Phylogenetic analysis of viruses*
187 Phylogenetic analysis of TuMV was conducted using first protein (P1) and coat protein (CP)
188 nucleotide sequences, because TuMV strains can be divided into four major lineages based on the sequences
189 [22]. Well-sequenced (>80% of full genome determined) strains, including six TuMV from Mino-gawa,
190 six TuMV strains from our previous study [20], and reported TuMV strains (46 sequences) [22], were used.
191 *Japanese yam mosaic virus* was added as an out-group (NCBI accession no. AB016500 and NC_000947).
192 Phylogenetic analysis of CMV was conducted using sequence fragments consisting of a CP open reading

193 frame (ORF), in addition to, 92 and 123 bp of 5′ and 3′ untranslated region, respectively [28, 29]. Four
194 strains of CMV from Mino-gawa, four strains from our previous study [20], and reported CMV strains (46
195 sequences) were used, setting *Peanut stunt virus* as out-group. Representative sequences of TuMV and
196 CMV obtained in the present study were deposited in GenBank (Accession Nos. LC368037-LC368038 and
197 LC368039, respectively). Alignments were performed in MAFFT v7.273 with default settings
198 (http://mafft.cbrc.jp/alignment/software/). Phylogenetic analysis was conducted in MEGA v7 [30] using
199 the maximum-likelihood method based on the Tamura-Nei model for TuMV and the General Time
200 Reversible model for CMV, both chosen based on the Akaike information criterion. There were 616
201 positions for TuMV and 799 positions for CMV in the final datasets. Branch support was estimated by
202 1,000 bootstrap replications.
203
204 *Analysis of viral transmission through clonal propagation of* A. halleri
205 Because *A. halleri* displayed high infection rates in our previous study [20], it was necessary to
206 determine whether viruses are transmitted through clonal propagation to interpret the results of the present
207 study and of the previous one. *Arabidopsis halleri* propagates clonally by producing aerial daughter rosettes
208 from the apical and lateral meristems of flowering stems (Fig. S1), which later become established in the
209 ground after flowering stems fall down. To examine the transmission rate of each virus through the clonal
210 propagation of *A. halleri*, 56 pairs of leaves from the mother and its aerial daughter rosettes were collected
211 on June 2 in 2015, at the Omoide-gawa site, Hyogo Prefecture, Japan (Table 2). To obtain samples in which
212 infecting virus species were already determined, we collected leaves for this experiment from a natural
213 population of *A. halleri*, in which TuMV, CMV, BrYV, and AhPV1 were previously detected [20]. Daughter
214 rosettes derived from the shoot apical meristems were used because they were most distant along the
215 flowering stems, from the mother rosettes. RNA was extracted from each sample and viral infection was
216 detected by the RT-PCR method described above.

Results

Detection of viruses

To determine which viruses were present in the three plant communities, the maximum read number and maximum coverage for each virus across samples within the site, were examined for each virus (Fig. 1). To screen for viruses in Brassicaceae plant communities, data from all plants within each site was pooled. Five, two, and zero viruses were identified from the samples collected at Mino-gawa, Gongen-dani, and Inado-gawa, respectively (Fig. 1).

At Mino-gawa, TuMV (genus, *Potyvirus*; family, Potyviridae), BrYV (genus, *Polerovirus*; family Luteoviridae), and *Pelargonium zonate spot virus* (PZSV; genus, *Anulavirus*; family, Bromoviridae) had more than 80 % coverage and high number of reads (Fig. 1a). AhPV1 (genus, *Alphapartitivirus*; family, Partitiviridae) and CMV (genus, *Cucumovirus*; family, Bromoviridae) showed 74% and 46% coverage, respectively, and their reads were widely distributed within the viral genome. Viruses that showed 10–20% coverage (TuYV, *Gayfeather mild mottle virus*, *Peanut stunt viru*s, and *Tomato aspermy virus*) were considered mis-mappings due to partial sequence homology with BrYV (TuYV) and CMV (the other three viruses). Therefore, five virus species were considered to infect the Brassicaceae community at Mino-gawa.

At Gongen-dani, PZSV and AhPV1 had more than 80% coverage and a high number of reads (Fig. 1b). No other viruses were detected with >10% coverage. Therefore, these two species were considered to infect the Brassicaceae community at Gongen-dani. At Inado-gawa, no viral infections were detected by RNA-Seq (Fig. 1c). Infected plants were spatially clustered in both Mino-gawa and Gongen-dani (Fig. 2), and these clusters partly represent spatial aggregation of *A. halleri* and *Rorippa indica* distribution, within the sites (Fig. 2).

Survey of novel viral infection by de novo *assembly*

241 Unmapped reads were assembled and 19, 27, and 1 contig(s) were annotated to viruses of Mino-gawa,
242 Gongen-dani, and Inado-gawa, respectively (Table S2). Among these, reads annotated to TuMV, CMV,
243 BrYV, and PZSV or other closely related viruses, in the reference sequence, were not included in further
244 analyses, because these sequences were already determined to be derived from five viruses in the first
245 analyses (previous subsection). One contig from Mino-gawa and seven contigs from Gongen-dani were
246 annotated to viruses with moderate similarities and treated as candidates for putative viruses.
247 The seven contigs from Gongen-dani were annotated to *Burdock mottle virus* (BdMoV, Table S3).
248 Among these, five and two contigs correspond to *RNA1* and *RNA2* of BdMoV, covering 33% and 27% of
249 the two genomes, respectively (Table S3). When the seven contigs were mapped on BdMoV genome
250 sequences, their positions did not overlap; they were widely distributed within the genome. Deduced amino
251 acid (AA) sequences of the contigs had putative coding sequences that generated putative polypeptides
252 similar to those produced by *RNA1* and *RNA2* of BdMoV (70% and 62% AA sequence similarity,
253 respectively). To examine the co-occurrence of these contigs in each sample, the longest two contigs each
254 from the *RNA1* and *RNA2* of the putative novel virus (four contigs in total) were remapped using RNA-Seq
255 data from Gongen-dani. The BdMoV-like contigs were detected in only one sample, *Arabis flagellosa*;
256 therefore, the putative virus was named *Arabis flagellosa Virus 1* (AflV1). The remaining 1 contig from
257 Gongen-dani was annotated to *Aspergillus foetidus slow virus 2*, but the aligned sequence was short, 191
258 nt of a 3,634-nt reference sequence (5.3%), and therefore this virus was not considered to be present (Table
259 S3).
260
261 *Distribution of viruses across Brassicaceae species*
262 At Mino-gawa, viruses were detected in *A. halleri* and *R. indica* with infection rates of 80% and 8%,
263 respectively (Table 1). Five and two viruses were detected in *A. halleri* and *R. indica*, respectively (Table
264 1). No viral infection was detected in *Cardamine scutata* or *C. occulta*. At Gongen-dani, AhPV1 was

265 detected in *A. halleri* with an infection rate of 8%, and PZSV was detected in *R. indica* with an infection
266 rate of 37.5%. No viral infection was detected in *C. leucantha*, *C. impatiens*, or *C. hirsuta*. *A. flagellosa*
267 plants were not infected by the five virus species (Table 1), but putative infection by AflV1 was detected in
268 a single plant individual. None of the seven plant species sampled at Inado-gawa (*C. leucantha*, *C.*
269 *impatiens*, *C. scutata*, *A. flagellosa*, *R. indica*, *C. hirsuta*, and *C. appendiculata*) were infected by viruses
270 (Table 1). Virus species composition was significantly different among the three sites and among host plant
271 species (perMANOVA, R^2 = 0.019, 0.098, and 0.11 and *p*-value = 0.016, 0.0006, and 0.0001 for site, plant,
272 and site-plant interaction, respectively).
273 Multiple infections were only observed in Mino-gawa, and therefore analyses on viral co-infections
274 were performed on the data at the Mino-gawa site. Multiple infections by TuMV, BrYV, and CMV were
275 significantly more frequent than that expected by chance (Fisher's test, $p < 0.01$), especially when the plant
276 was infected by TuMV (Fisher's test, $p < 0.001$, Table S4). No significant association was observed between
277 infections by PZSV and other viruses (Table S5).
278
279 *Phylogenetic analysis of viruses*
280 All six strains of TuMV from Mino-gawa formed a single clade with 95% bootstrap probability (Fig.
281 S2). The clade was included in the world-B group, one of the four phylogenetic lineage of TuMV defined
282 in the previous study [22], along with other strains from Omoide-gawa site (Ahg Plot) and those reported
283 from crops (Fig. S2).
284 All strains of CMV obtained from *A. halleri* in Mino-gawa belonged to subgroup IA (sub-IA), a
285 phylogenetic lineage of CMV (Fig. S3) [21, 28]. The sub-A group formed a clade with 87% bootstrap
286 probability. CMV strains from Mino-gawa, like other strains in sub-IA, encoded 2b proteins that were 10
287 AA longer than that encoded by strains in subgroup II (sub-II). Within the sub-IA group, Mino-gawa strains
288 formed a single clade by themselves with 96% bootstrap support.

289

290 *Clonal transmission of virus through clonal reproduction of* A. halleri

291 Among the 56 mother plants, 50, 37, 6, and 36 were infected by TuMV, CMV, BrYV, and AhPV1, and
292 the transmission rates of viruses from these infected plants to the aerial daughter rosettes were 92%, 92%,
293 83%, and 100%, respectively (Table 2). In eight cases when mother-daughter transmission failed (4, 3, and
294 1 for TuMV, CMV, and BrYV, respectively), virus accumulation levels in mother plants were lower than
295 those found in cases of successful transmissions. Quadruple, triple, double, and single infections were
296 observed in 4, 20, 21, and 11 plants, respectively. Among these plants, 3, 18, 16, and 11 transmitted whole
297 virus sets (Table S6). Reduction in the number of transmitted species was observed; TuMV, CMV, and
298 BrYV failed to be transmitted occasionally (Table S6).

299

300 **Discussion**

301 Infection rates differed notably among the combinations of virus, plant species and sampling sites.
302 Viruses turned out to infect only two perennial plant species, *A. halleri* and *R. indica*. Infection rates
303 detected in the present study were similar to those reported in previous studies conducted in semi-natural
304 plant populations. Six virus species have been reported to infect 0%–3.8% plants of the wild *Brassica rapa*
305 populations growing close to crop fields in southern England [31]. CMV was reported to infect 18% to 25%
306 plants of the biennial and perennial populations of weedy Brassicaceae species in North America [32]. The
307 most frequently detected virus in Mino-gawa and Gongen-dani was PZSV; it was reported for the first time
308 in Japan, in these sites [33]. PZSV was known to be transmitted vertically via pollen and seeds in tomato,
309 in addition to horizontal transmission [34]. Therefore, high infection rate of PZSV is likely to be maintained
310 through vertical transmission in the *A. halleri* populations.

311 Viral infection rate was exceptionally high in *A. halleri* plants, and we have previously reported similar
312 levels of infection rates [20]. Compared to other Brassicaceae members, *A. halleri* seemed to accumulate

313 infected plants in its population even for non-seed-borne viruses. We expect that the high transmission rate
314 of viruses through clonal propagation of the host plant contributes primarily to the long-term maintenance
315 of infection in the clonal linages. Then, the long-term persistence of virus in the clonal linage is likely to
316 allow successful horizontal transmission to take place. We found that the infected plants distribute across
317 the distance of over several hundred meters, and the pattern should require horizontal transmission. Overall,
318 long-term persistence of viruses through clonal propagation of host plants and occasional horizontal
319 transmission are likely to have formed the observed spatial distribution of infected plants in the host plant
320 populations.

321 Phylogenetic analysis of TuMV and CMV revealed that all virus strains obtained from *A. halleri* in
322 natural plant communities were closely related to the strains reported from crops. The pattern in the
323 phylogeny suggested the presence of virus exchanges between natural plant communities and agricultural
324 fields. Further experimental analyses are required to determine whether the strains in natural plant
325 communities are infectious to crop plants. Previous studies also reported some evidence of the virus
326 exchange between crops and wild plants, which implies future-emergent viruses in crops can be
327 investigated from viruses in natural habitats [2, 42].

328 The presence of *A. halleri* in a plant community might have enhanced infection risk for the other related
329 plant species. This, however, was not the case in our study. Viral infection rate in *R. indica* was low although
330 it seemed to be a host of three of the five detected viruses in Mino-gawa (TuMV, BrYV, and PZSV). This
331 suggests that viruses were horizontally transmitted to a certain degree within *A. halleri*, but this was not
332 high enough for spread of viral infections to other Brassicaceae species. TuMV and BrYV were reported to
333 be transmitted by aphids [35, 36], and PZSV by a thrips, which carries virus-included pollen grains on its
334 body [34]. No viral detection from *Cardamin*e species may represent a lack of effective means of
335 transmission and maintenance that matches the life cycle of both host plants and vectors. It is likely that
336 vector density is not high enough in natural vegetation, which results in low infection rate of vector-

dependent viruses. Further studies are required on the density and distribution of vector insects in the natural plant communities.

In some cases, multiple viruses were co-existed in their shared host species, which may be the results of a certain mechanism that enhance multiple-infections. In this study, multiple infections were significantly frequent in TuMV infected leaves in Mino-gawa, which suggested that TuMV might facilitate infection of other virus species. In previous studies, promotion of replication of other viruses by *potyviruses* has been reported to be achieved by suppression of host RNA-silencing machinery [20, 37, 38].

From Gongen-dani, one *A. flagellosa* plant yielded a putative novel virus, tentatively named AfV1, which was similar to BdMoV. This member of genus *Benyvirus* was first reported in an edible burdock plant (*Arctium lappa*) in Japan and has a bisegmented RNA1 (7038 nt) and RNA2 (4315 nt) genome [39]. While RNA1 has a single ORF encoding a 249-kDa polypeptide [39], RNA2 has six ORFs encoding a coat protein, a coat-protein readthrough, three movement proteins, and a cysteine-rich protein [39]. In the present study, putative AA sequences were identified from seven contigs corresponding to a 249-kDa polypeptide and to first and third movement proteins in the triple-gene block of BdMoV. We could not obtain the whole genome sequence of AflV1, due to the low number of AflV1 reads (2.4×10^3 rpm), as this is a cryptic virus [40-42] and therefore, difficult to detect.

In this study, we determined the distribution of viruses in natural Brassicaceae plant communities and analyzed the difference among plant species and localities. We also detected a novel virus candidate in this study. Further update of the virus database is clearly required, especially by accumulating sequence information of novel viruses from natural plant vegetation. Revealing ecology of viruses, host plants, and vectors will be a promising way to understand mechanisms that determine virus dynamics in natural plant communities.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Tomlinson JA, Carter AL, Dale WT, Simpson CJ (1970) Weed plants as sources of cucumber mosaic virus. Annals of Applied Biology 66: 11-16. https://doi.org/10.1111/j.1744-7348.1970.tb04597.x
2. Mueller EE, Groves RL, Gratton C (2011) Crop and Non-Crop Plants as Potential Reservoir Hosts of Alfalfa mosaic virus and Cucumber mosaic virus for Spread to Commercial Snap Bean. Plant Disease 96: 506-514. https://doi.org/10.1094/PDIS-02-11-0089
3. Roossinck MJ (2012) Plant virus metagenomics: biodiversity and ecology. Annu Rev Genet 46: 359-369. https://doi.org/10.1146/annurev-genet-110711-155600
4. Aranda MA, Freitas-Astúa J (2017) Ecology and diversity of plant viruses, and epidemiology of plant virus-induced diseases. Annals of Applied Biology 171: 1-4. https://doi.org/10.1111/aab.12361
5. Islam W, Zhang J, Adnan M, Noman A, Zainab M, Jian W (2017) Plant virus ecology: a glimpse of recent accomplishments. Appl Econ Environ Res 15: 691-705. http://dx.doi.org/10.15666/aeer/1501_691705
6. May RM, Anderson RM (1983) Epidemiology and genetics in the coevolution of parasites and hosts. Proc R Soc Lond B Biol Sci 219: 281-313. http://dx.doi.org/10.1098/rspb.1983.0075
7. Prendeville HR, Ye X, Morris TJ, Pilson D (2012) Virus infections in wild plant populations are both frequent and often unapparent. Am J Bot 99: 1033-1042. http://dx.doi.org/10.3732/ajb.1100509
8. Li L, Liu J, Zhang Q, Fu R, Zhu X, Li C, Chen J (2016) Seed-borne viral dsRNA elements in three cultivated Raphanus and Brassica plants suggest three cryptoviruses. Can J Microbiol 62: 287-295. http://dx.doi.org/10.1139/cjm-2015-0788
9. Alexander HM, Bruns E, Schebor H, Malmstrom CM (2017) Crop -associated virus infection in a native perennial grass: reduction in plant fitness and dynamic patterns of virus detection. J Ecol https://doi.org/10.1111/1365-2745.12723
10. Pagán I, Montes N, Milgroom MG, García-Arenal F (2014) Vertical transmission selects for reduced virulence in a plant virus and for increased resistance in the host. PLoS Pathogens 10: e1004293. http://dx.doi.org/10.1371/journal.ppat.1004293
11. Bottenberg H, Irwin ME (1992) Using mixed cropping to limit seed mottling induced by soybean mosaic virus. Plant Dis 76: 304-306. http://dx.doi.org/10.1094/PD-76-0304
12. Jones RAC, Ferris DG (2000) Suppressing spread of alfalfa mosaic virus in grazed legume pasture swards using insecticides and admixture with grass, and effects of

400 insecticides on numbers of aphids and three other pasture pests. Ann Appl Biol 137:
401 259-271. http://dx.doi.org/10.1111/j.1744-7348.2000.tb00067.x
402 13. Hooks CRR, Fereres A (2006) Protecting crops from non-persistently aphid-
403 transmitted viruses: A review on the use of barrier plants as a management tool.
404 Virus Research 120: 1-16. http://dx.doi.org/10.1016/j.virusres.2006.02.006
405 14. Pedersen AB, Fenton A (2007) Emphasizing the ecology in parasite community
406 ecology. Trends Ecol Evol 22: 133-139. http://dx.doi.org/10.1016/j.tree.2006.11.005
407 15. Callaway RM (1995) Positive interactions among plants. Bot Rev 61: 306-349.
408 https://doi.org/10.1007/BF02912621
409 16. Seabloom EW, Borer ET, Gross K, Kendig AE, Lacroix C, Mitchell CE, Mordecai EA,
410 Power AG (2015) The community ecology of pathogens: coinfection, coexistence and
411 community composition. Ecol Lett 18: 401-415. https://doi.org/10.1111/ele.12418
412 17. Filipe J, Maule M (2004) Effects of dispersal mechanisms on spatio-temporal
413 development of epidemics. J Theoret Biol 226: 125-141.
414 https://doi.org/10.1016/S0022-5193(03)00278-9
415 18. Gibson G (1997) Investigating mechanisms of spatiotemporal epidemic spread using
416 stochastic models. Phytopathology 87: 139-146.
417 http://dx.doi.org/10.1094/PHYTO.1997.87.2.139
418 19. Barba M, Czosnek H, Hadidi A (2014) Historical Perspective, Development and
419 Applications of Next-Generation Sequencing in Plant Virology. Viruses 6: 106-136.
420 http://dx.doi.org/10.3390/v6010106
421 20. Kamitani M, Nagano AJ, Honjo MN, Kudoh H (2016) RNA-Seq reveals virus-virus
422 and virus-plant interactions in nature. FEMS Microbiol Ecol 92.
423 http://dx.doi.org/10.1093/femsec/fiw176
424 21. Palukaitis P, Roossinck MJ, Dietzgen RG, Francki RI (1992) Cucumber mosaic
425 virus. Adv Virus Res 41: 281-348.
426 22. Ohshima K, Yamaguchi Y, Hirota R, Hamamoto T, Tomimura K, Tan Z, Sano T,
427 Azuhata F, Walsh JA, Fletcher J, Chen J, Gera A, Gibbs A (2002) Molecular
428 evolution of Turnip mosaic virus: evidence of host adaptation, genetic recombination
429 and geographical spread. J General Virol 83: 1511-1521. doi:
430 http://dx.doi.org/10.1099/0022-1317-83-6-1511
431 23. Morlan JD, Qu K, Sinicropi DV (2012) Selective depletion of rRNA enables whole
432 transcriptome profiling of archival fixed tissue. PLoS ONE 7: e42882.
433 http://dx.doi.org/10.1371/journal.pone.0042882
434 24. Kamitani M, Nagano AJ, Honjo MN, Kudoh H (2016) RNA-Seq reveals virus–virus
435 and virus–plant interactions in nature. FEMS Microbiol Ecol 92.

http://dx.doi.org/10.1093/femsec/fiw176

25. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnol 29: 644-652. http://dx.doi.org/10.1038/nbt.1883

26. Rice P, Longden I, Bleasby A (2000) EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 16: 276-277. http://dx.doi.org/10.1016/s0168-9525(00)02024-2

27. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara R, Simpson GL, Solymos P, Stevens MHH, Wagner H (2015) Vegan: Community ecology package. R package version 2.3-5. 2016.

28. Roossinck MJ, Zhang L, Hellwald KH (1999) Rearrangements in the 5' nontranslated region and phylogenetic analyses of cucumber mosaic virus RNA 3 indicate radial evolution of three subgroups. J Virol 73: 6752-6758.

29. Bashir NS, Kalhor MR, Zarghani SN (2006) Detection, differentiation and phylogenetic analysis of cucumber mosaic virus isolates from cucurbits in the northwest region of Iran. Virus Genes 32: 277-288. http://dx.doi.org/10.1007/s11262-005-6912-2

30. Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33: 1870-1874. http://dx.doi.org/10.1093/molbev/msw054

31. Pallett D, Thurston M, Cortina - Borja M, Edwards ML, Alexander M, Mitchell E, Raybould A, Cooper J (2002) The incidence of viruses in wild Brassica rapa ssp. sylvestris in southern England. Ann Appl Biol 141: 163-170. https://doi.org/10.1111/j.1744-7348.2002.tb00209.x

32. Rist DL, Lorbeer JW (1989) Occurrence and overwintering of cucumber mosaic virus and broad bean wilt virus in weeds growing near commercial lettuce fields in New York. Phytopathology 79: 65-69. http://dx.doi.org/10.1094/Phyto-79-65

33. Kamitani M, Nagano AJ, Honjo MN, Kudoh H (2017) First report of Pelargonium zonate spot virus from wild Brassicaceae plants in Japan. J General Plant Pathol http://dx.doi.org/10.1007/s10327-017-0727-6

34. Lapidot M, Guenoune-Gelbart D, Leibman D, Holdengreber V, Davidovitz M, Machbash Z, Klieman-Shoval S, Cohen S, Gal-On A (2010) Pelargonium zonate spot virus is transmitted vertically via seed and pollen in tomato. Phytopathology 100: 798-804. http://dx.doi.org/10.1094/phyto-100-8-0798

35. Schliephake E, Graichen K, Rabenstein FF (1999) Investigations on the vector transmission of the Beet mild yellowing virus (BMYV) and the Turnip yellows virus (TuYV)
36. Whitfield AE, Falk BW, Rotenberg D (2015) Insect vector-mediated transmission of plant viruses. Virology 479-480: 278-289. http://dx.doi.org/10.1016/j.virol.2015.03.026
37. Choi SK, Yoon JY, Ryu KH, Choi JK, Palukaitis P, Park WM (2002) Systemic movement of a movement-deficient strain of Cucumber mosaic virus in zucchini squash is facilitated by a cucurbit-infecting potyvirus. J General Virol 83: 3173-3178. http://dx.doi.org/10.1099/0022-1317-83-12-3173
38. Pruss G, Ge X, Shi XM, Carrington JC, Bowman Vance V (1997) Plant viral synergism: the potyviral genome encodes a broad-range pathogenicity enhancer that transactivates replication of heterologous viruses. Plant Cell 9: 859-868.
39. Kondo H, Hirano S, Chiba S, Andika IB, Hirai M, Maeda T, Tamada T (2013) Characterization of burdock mottle virus, a novel member of the genus Benyvirus, and the identification of benyvirus-related sequences in the plant and insect genomes. Virus Res 177: 75-86. http://dx.doi.org/10.1016/j.virusres.2013.07.015
40. Nibert ML, Ghabrial SA, Maiss E, Lesker T, Vainio EJ, Jiang D, Suzuki N (2014) Taxonomic reorganization of family Partitiviridae and other recent progress in partitivirus research. Virus Res 188: 128-141. https://doi.org/10.1016/j.virusres.2014.04.007
41. Boccardo G, Lisa V, Luisoni E, Milne RG (1987) Cryptic plant viruses. Adv Virus Res 32: 171-214. https://doi.org/10.1016/S0065-3527(08)60477-7
42. Roossinck MJ (2010) Lifestyles of plant viruses. Philos Trans R Soc Lond B Biol Sci 365: 1899-1905. http://dx.doi.org/10.1098/rstb.2010.0057

Legends to Figures

Fig. 1 Detection of viruses from Brassicaceae communities at Mino-gawa (a), Gongen-dani, (b) and Inado-gawa (c). *Turnip mosaic virus* (TuMV), *Cucumber mosaic virus* (CMV), *Brassica yellows virus* (BrYV), *Peralgonium zonate spot virus* (PZSV) and *Arabidopsis halleri partitivirus 1* (AhPV1) were detected from Mino-gawa. PZSV and AhPV1 were detected from Gongen-dani. No virus infection was detected in Inado-gawa. Maximum values of $\log_{10}$ of the read number (Y-axis) and the genome coverage (X-axis) for the 3,981 reported virus sequence (NCBI database) are shown. Each point indicates a single virus. Gray-filled points represent infecting viruses.

Fig. 2 Maps showing positions of sampled and infected plants at Mino-gawa (a), Gongen-dani, (b) and Inado-gawa (c). Closed circles, open circles, and open triangles indicate positions of infected perennial, non-infected perennial, and non-infected annual plants, respectively. There were no infected annuals in the three localities. Alphabets represent the plant species: A, *Arabidopsis halleri*; F, *Arabis flagellosa*; D, *Cardamine appendiculata*; H, *C. hirsuta*; I, *C. impatiens*; L, *C. leucantha*; O, *C. occulta*; S, *C. scutata*; and R, *Rorripa indica*. Infected viruses were also listed by red letters. Solid and gray shaded lines represented streams and sampling trails, respectively. Scale bars in each map represented 50 m in distance. Directions are indicated by bearing marks.

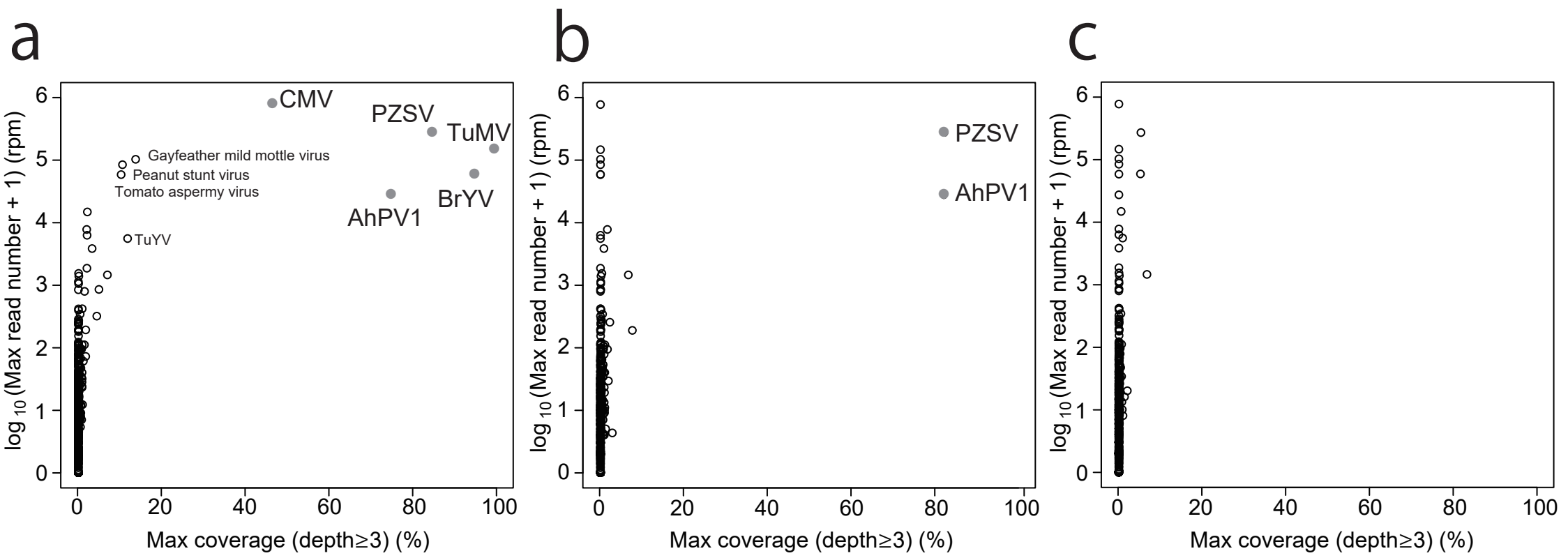
a
b
c
CMV
PZSV
TuMV
BrYV
AhPV1
Gayfeather mild mottle virus
Peanut stunt virus
Tomato aspermy virus
TuYV
log 10 (Max read number + 1) (rpm)
Max coverage (depth≥3) (%)

Figure 1

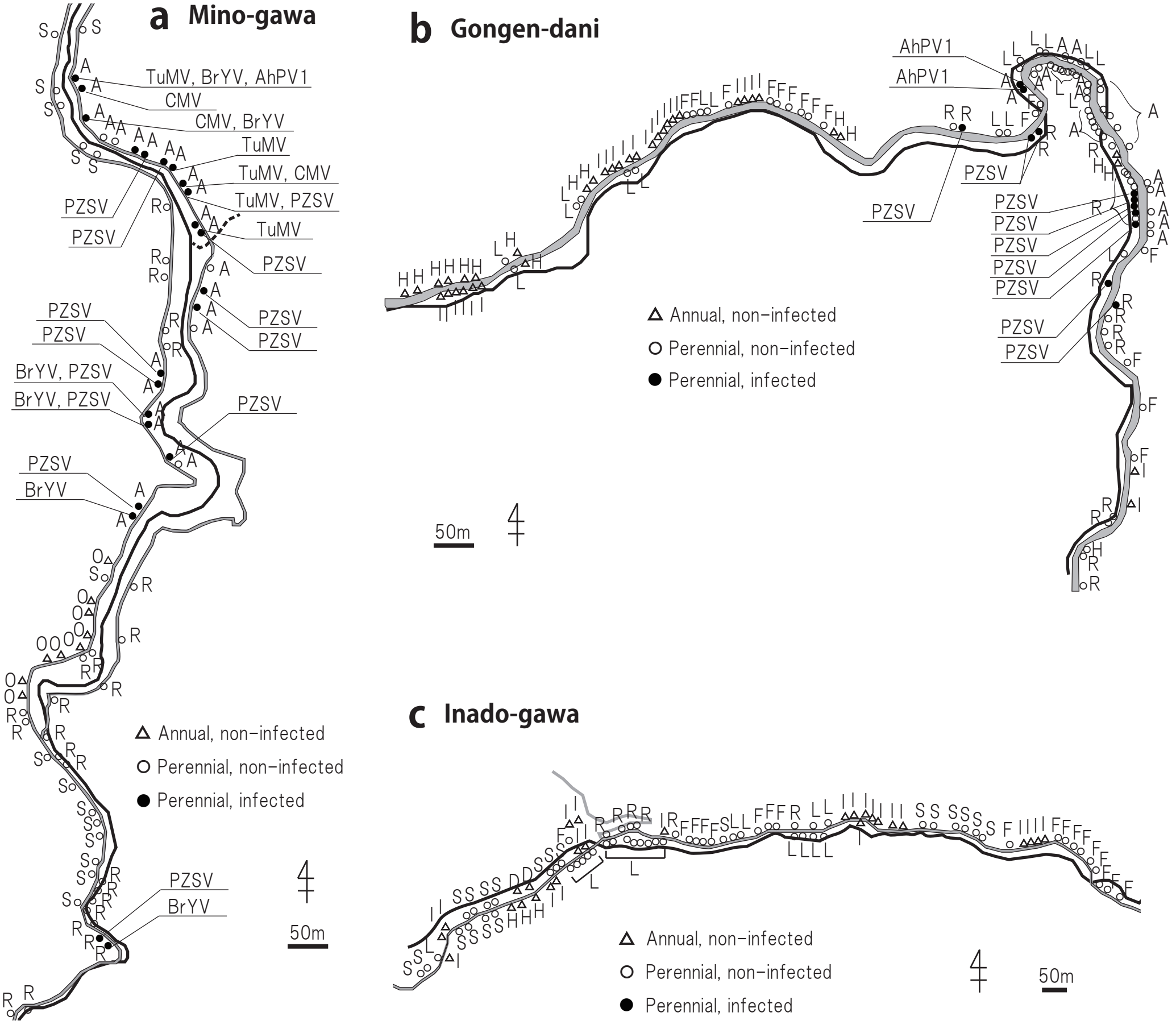
a Mino-gawa
TuMV, BrYV, AhPV1
CMV
CMV, BrYV
TuMV
TuMV, CMV
TuMV, PZSV
TuMV
PZSV
PZSV
PZSV
PZSV
PZSV
PZSV
PZSV
BrYV, PZSV
BrYV, PZSV
PZSV
PZSV
BrYV
PZSV
BrYV
Annual, non-infected
Perennial, non-infected
Perennial, infected
50m
b Gongen-dani
AhPV1
AhPV1
PZSV
PZSV
PZSV
PZSV
PZSV
PZSV
PZSV
PZSV
PZSV
Annual, non-infected
Perennial, non-infected
Perennial, infected
50m
c Inado-gawa
Annual, non-infected
Perennial, non-infected
Perennial, infected
50m

Figure 2

Table 1 Results of virus screening of Brassicaceae communities from the three sites in this study and one site from the previous study

Site	Plant species	Perennial/ Annual	Examined plants[(1)]	Number of plants infected by viruses							Infection rate (pooled)
				TuMV	CMV	BrYV	PZSV	AhPV1	AflV1[(2)]	Pooled[(3)]	
Mino-gawa	*Arabidopsis halleri* (L.) O'Kane et Al-Shehbaz subsp. *gemmifera* (Matsum.) O'Kane et Al-Shehbaz	Perennial	25 [12]	**6**	**4**	**5**	**13**	**1**	**0**	**20**	80%
Mino-gawa	*Rorripa indica* (L.) Hiern	Perennial	25 [12]	**1**	0	**1**	**1**	0	0	2	8%
Mino-gawa	*Cardamine scutata* Thunb. subsp. *regeliana* (Miq.) H.Hara	Perennial	15 [6]	0	0	0	0	0	0	0	0%
Mino-gawa	*Cardamine occulta* Hornem.	Annual	8	0	0	0	0	0	-	0	0%
Gongen-dani	*A. halleri*	Perennial	25 [12]	0	0	0	0	**2**	**0**	**2**	8%
Gongen-dani	*R. indica*	Perennial	24 [11]	0	0	0	**9**	0	**0**	**9**	37.5%
Gongen-dani	*Cardamine leucantha* (Tausch) O.E.Schulz	Perennial	25 [6]	0	0	0	0	0	0	0	0%
Gongen-dani	*Cardamine impatiens* L.	Annual	22	0	0	0	0	0	-	0	0%
Gongen-dani	*Cardamine hirsuta* L.	Annual	15	0	0	0	0	0	-	0	0%
Gongen-dani	*Arabis flagellosa* Miq.	Perennial	15 [6]	0	0	0	0	0	1	0	16.7%
Inado-gawa	*C. leucantha*	Perennial	23 [6]	0	0	0	0	0	0	0	0%
Inado-gawa	*C. impatiens*	Annual	19 [6]	0	0	0	0	0	0	0	0%
Inado-gawa	*C. scutata*	Perennial	20 [6]	0	0	0	0	0	0	0	0%
Inado-gawa	*Arabis flagellosa*	Perennial	18 [6]	0	0	0	0	0	0	0	0%
Inado-gawa	*R. indica*	Perennial	6 [6]	0	0	0	0	0	0	0	0%
Inado-gawa	*C. hirsuta*	Annual	3	0	0	0	0	0	-	0	0%
Inado-gawa	*Cardamine appendiculata* Franch. et Sav.	Perennial	2	0	0	0	0	0	-	0	0%
Omoide-gawa[(4)]	*A. halleri*	Perennial	68	39	18	21	0	56	-	62	91%

[(1)] All plants were analyzed by real-time PCR and numbers in brackets indicate the samples analyzed also by RNA-Seq

[(2)] The presence of a novel putative virus, AflV1, was determined only for RNA-Seq-analyzed samples

[(3)] Number of plants infected by any of the six detected viruses

[(4)] Data from Kamitani *et al.* 2016, and there were no other Brassicaceae species in Omoide-gawa site [25]

Table 2 Clonal transmission of four viruses from mother to aerial daughter rosettes of *A. halleri*

	TuMV	CMV	BrYV	AhPV1
Examined pairs of mother and daughter rosettes	56	56	56	56
Infected mother rosettes	50	37	6	36
Infected daughter rosettes	46	34	5	36
Uninfected pairs	6	19	50	20
Infection rate of mother rosettes	89%	66%	11%	64%
Transmission rate from mother to daughter rosettes	92%	92%	83%	100%

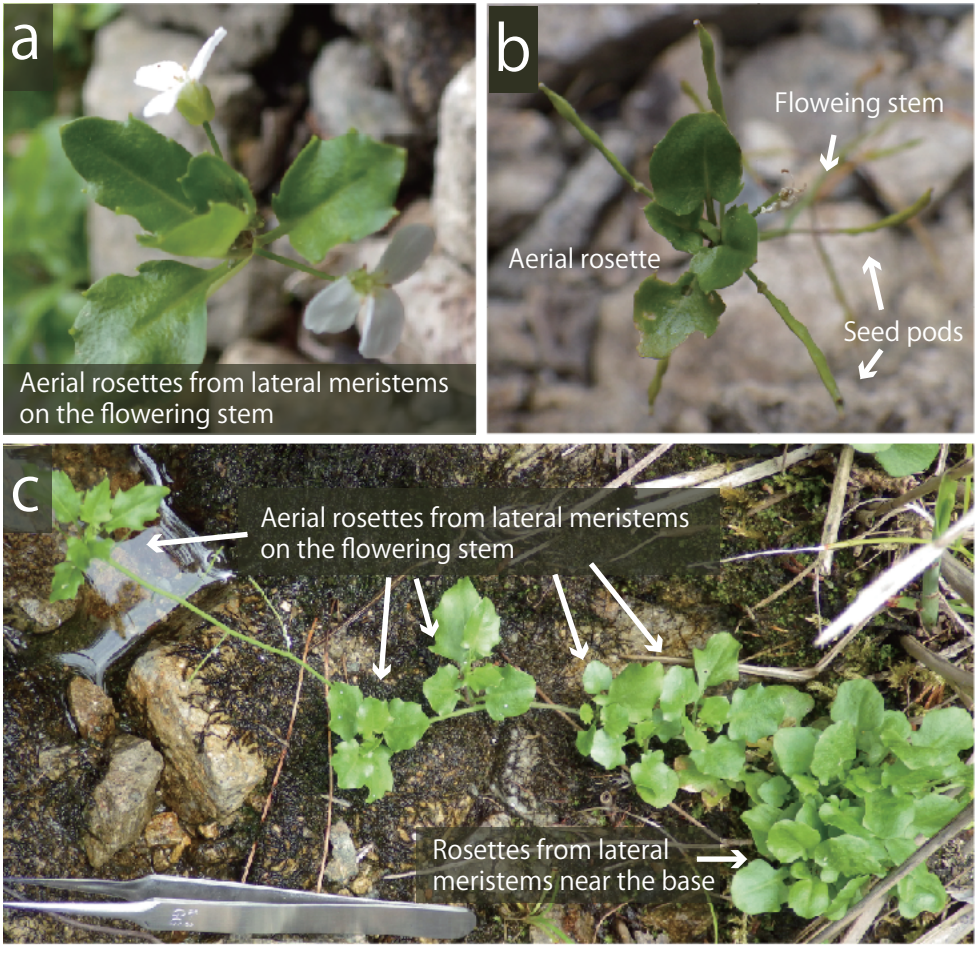

Fig. S1 Clonal propagation of *Arabidopsis halleri*. An aerial daughter rosette formed at the shootapical meristem (i.e. top of inflorescence) during flowering (a) and after flowering (b). Clonal daughter rosettes from shoot apical meristems and lateral meristems on a flowering stem, and basal lateral meristems (c)

Kamitani M*, Nagano AJ, Honjo M, Kudoh H* "A Survey on Plant Viruses in Natural Brassicaceae Communities Using RNA-Seq" Microbial Ecology. Hiroshi Kudoh: kudoh@ecology.kyoto-u.ac.jp

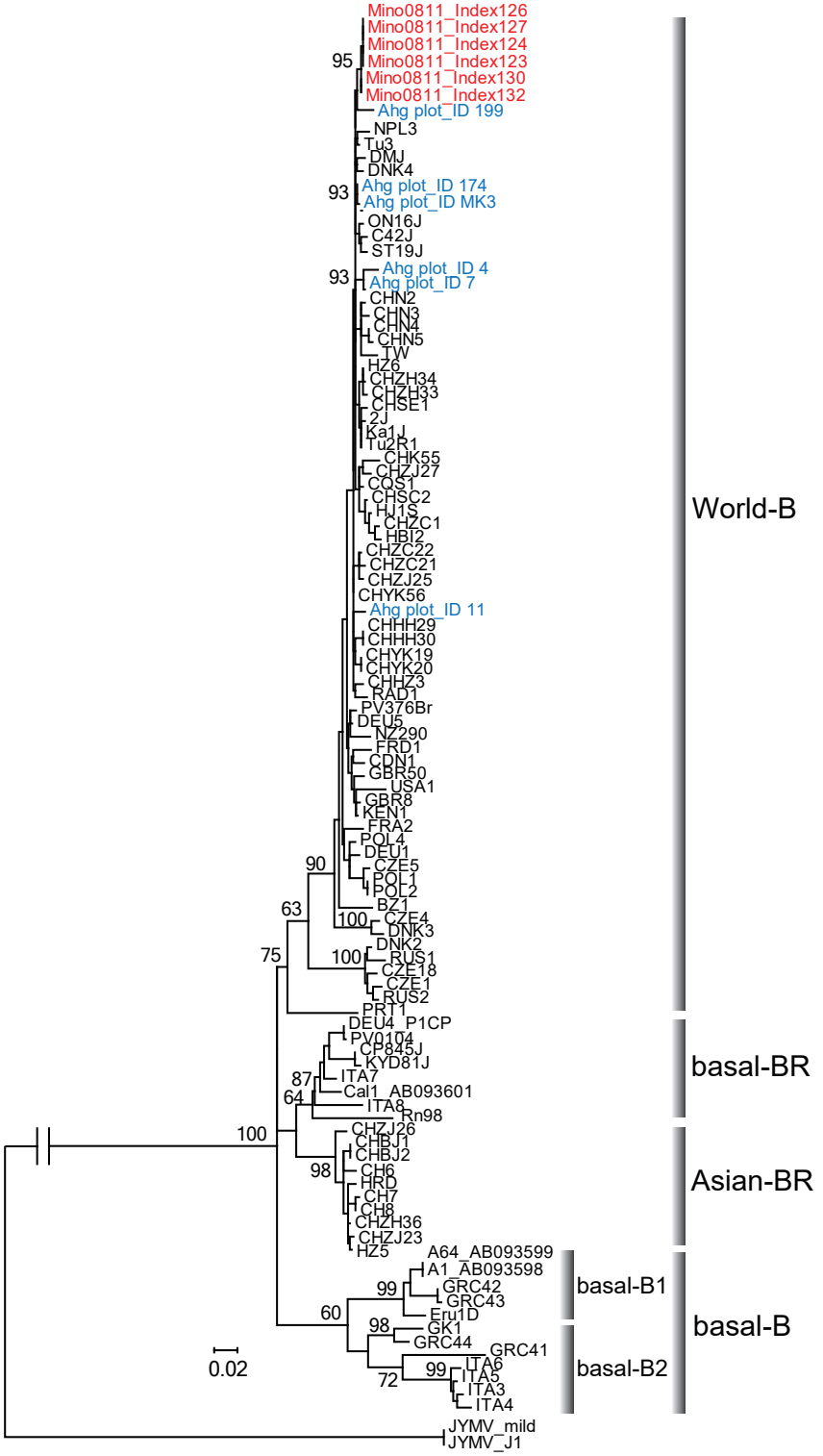

Fig. S2 Phylogenetic analysis on six strains of TuMV from *A. halleri* in Mino-gawa (red) based on P1 and CP sequences. Strains obtained from another natural *A. halleri* population (Omoide-gawa, Hyogo Pref., Japan) in our previous study are shown in blue characters (Kamitani et al. 2016). Sequences of other reported virus strains were obtained from Ohshima et al. (2002). Mild strain and J1 strain of *Japanese yam mosaic virus* (JYMV) were used as out-groups (Fuji and Nakamae, 2000, Fuji and Nakamae, 1999). Numbers represent bootstrap values (percentages) and values more than 60% are listed.

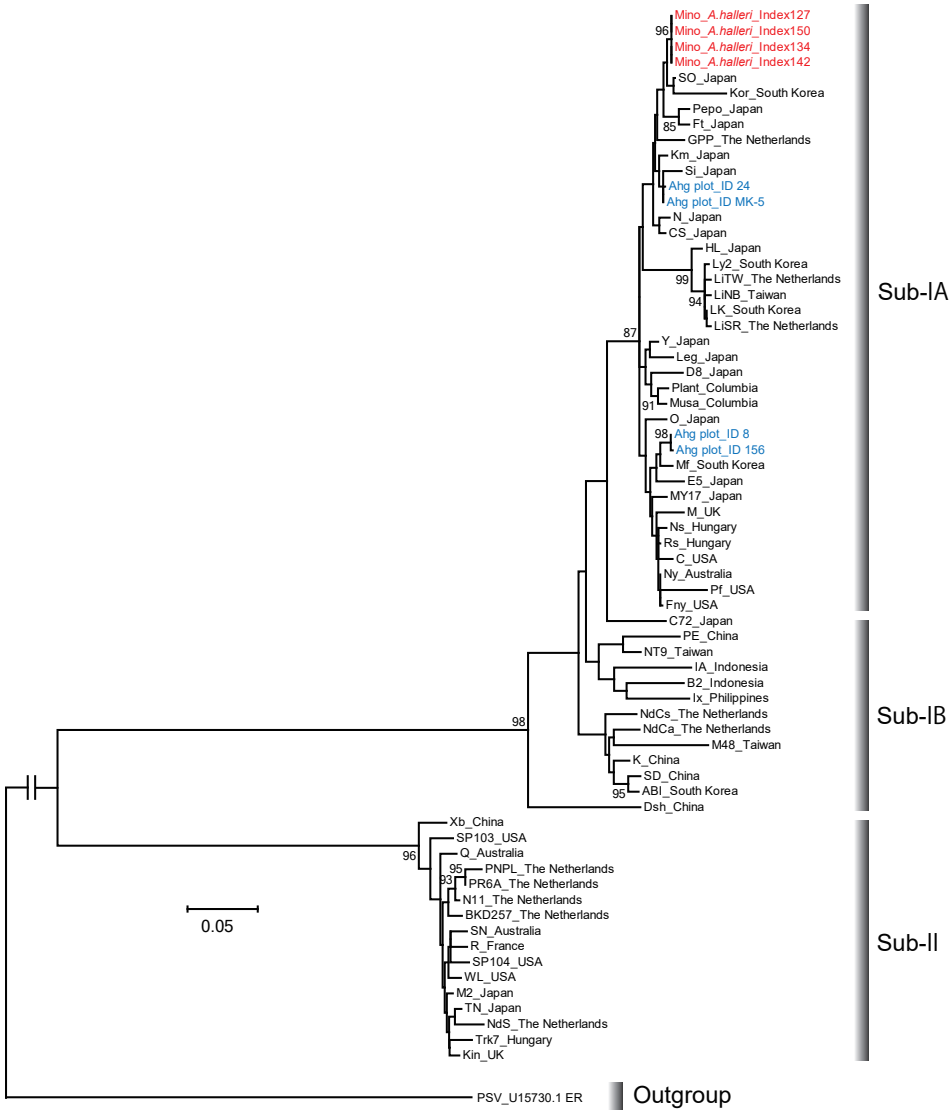

Fig. S3 Phylogenetic locations of four strains of CMV from *A. halleri* of Mino-gawa (red) based on sequences of CP open reading frame (ORF) plus 92 and 123 bp of the 5′ and 3′ untranslated region, respectively. Strains with blue color indicate CMV detected from the Omoide-gawa *A. halleri* population (Kamitani et al. 2016). Previously reported CMV strains were included (Bashir et al. 2006). *Peanut stunt virus* (PSV, GenBank Accession No. U15730.1) was used as an out-group. Numbers represent bootstrap values (percentages) and values more than 80% are listed.

Kamitani M*, Nagano AJ, Honjo M, Kudoh H* "A Survey on Plant Viruses in Natural Brassicaceae Communities Using RNA-Seq" Microbial Ecology. Hiroshi Kudoh: kudoh@ecology.kyoto-u.ac.jp

Kamitani M*, Nagano AJ, Honjo M, Kudoh H* "A Survey on Plant Viruses in Natural Brassicaceae Communities Using RNA-Seq" Microbial Ecology.
*Corresponding author Hiroshi Kudoh: kudoh@ecology.kyoto-u.ac.jp

Table S1 Primers used for virus detection and internal control in RT-qPCR

	forward (5′–3′)	**reverse (5′–3′)**
***Turnip mosaic virus* (TuMV)** [(1)]	TGGCTGATTACGAACTGACG	CTGCCTAAATGTGGGTTTGG
***Cucumber mosaic virus* (CMV)**	CAACTGGCTCGTATGGTGGA	CCTTCTCGCTGGGACTTTTG
***Brassica yellows virus* (BrYV)**	TGAAAGCAGACAACTCATGGAAG	GACCGACCAGAAAGAACGATG
***Pelargonium zonate spot virus* (PZSV)**	AGATTTTTCCGGGCTCTCTA	GTTCAACTGTTTTACCAGGATAG
***Arabidopsis halleri partitivirus 1* (AhPV1)**	ATGAAGAACACCGTCGTTCTC	GACTTCAGTTTCCCGTCATAC
***ACT2* (host gene, positive control)** [(2)]	TCCCTCAGCACATTCCAGCAGAT	AACGATTCCTGGACCTGCCTCATC

[(1)] Wei T, Zhang C, Hou X, Sanfaçon H, Wang A (2013) The SNARE Protein Syp71 Is Essential for Turnip Mosaic Virus Infection by Mediating Fusion of Virus-Induced Vesicles with Chloroplasts. PLOS Pathogens 9(5): e1003378. https://doi.org/10.1371/journal.ppat.1003378

[(2)] Skirycz A, Jozefczuk S, Stobiecki M, Muth D, Zanor MI, Witt I, Mueller‐Roeber B (2007) New Phytol. 175:425-438, https://doi.org/10.1111/j.1469-8137.2007.02129.x

*Corresponding author Hiroshi Kudoh: kudoh@ecology.kyoto-u.ac.jp

Table S2 Break down of the contigs obtained by *de novo* assembly

		Mino-gawa	Gongen-dani	Inado-gawa
Number of unmapped-read		28,489,176	81,378,310	14,339,016
Number of contigs (Total)		827	3,831	415
Viruses (high similarity with reported ones) (1)		19	19	1
Viruses (moderate similarity with reported ones) (2)		0	8(3)	0
Number of contigs annotated to other organisms	**Plants**	580	1457	346
	Animals	2	1758	1
	Fungi	20	182	11
	Bacteria	10	17	9
Not annotated		165	314	36

(1) Number of contigs annotated to viruses in reference list and considered to be derived from TuMV, CMV, BrYV, and PZSV

(2) Number of contigs annotated to other viruses with moderate similarities and treated as candidates for putative novel viruses

(3) Seven contigs (971, 404, 372, 353, and 210 nt) were annotated to *Burdock mottle virus* and regarded as a putative novel virus, *Arabis flagellosa Virus 1* (AflV1). The remained one contig was annotated to *Aspergillus foetidus slow virus 2*, but the sequence length was too short to represent the presence of the virus

*Corresponding author Hiroshi Kudoh: kudoh@ecology.kyoto-u.ac.jp

Table S3. The length and annotation of the eight candidate contigs of novel virus

No.	Length of the contigs	Annotation	Identity(1)	Coverage(2)
1	971	*Burdock mottle virus* (BdMoV) RNA2	70%	33%
2	404	BdMoV RNA2	77%	
3	372	BdMoV RNA2	71%	
4	353	BdMoV RNA2	67%	
5	210	BdMoV RNA2	80%	27%
6	929	BdMoV RNA1	67%	
7	237	BdMoV RNA1	73%	
8	191	*Aspergillus foetidus slow virus 2*	68%	5.3%

(1) Sequence identity of contigs with the annotated reference sequence

(2) The percentage of length of reference viral genome covered by the contigs

*Corresponding author Hiroshi Kudoh: kudoh@ecology.kyoto-u.ac.jp

Table S4 Contingency tables for infections by pairs of viruses (TuMV, CMV, and BrYV) at Mino-gawa. Significance levels in Fisher's exact test are also listed

Contingency table			**Significance**
Infection	TuMV+	TuMV-	
CMV+	2	2	*
CMV-	5	64	
	TuMV+	TuMV-	
BrYV+	5	3	**
BrYV-	11	181	
	CMV+	CMV-	
BrYV+	2	4	*
BrYV-	2	65	

* $p < 0.05$

** $p < 0.01$

Table S5 Contingency tables for infections by PZSV and other three virus species at Mino-gawa. Significance levels in Fisher's exact test are also listed

Contingency table			**Significance**
	PZSV+	PZSV-	
TuMV+	1	6	n.s.
TuMV-	13	53	
	PZSV+	PZSV-	
CMV+	0	4	n.s.
CMV-	14	55	
	PZSV+	PZSV-	
BrYV+	2	4	n.s.
BrYV-	2	65	

n.s., no significance at $p < 0.05$

*Corresponding author Hiroshi Kudoh: kudoh@ecology.kyoto-u.ac.jp

Table S6 Combinations of clonal transmission of viruses observed from mother to daughter rosettes of *Arabidopsis halleri*

Viruses detected in mother rosettes	**Viruses detected in daughter rosettes**	**Number of cases**
TuMV, CMV, BrYV, AhPV1	TuMV, CMV, BrYV, AhPV1	3
TuMV, CMV, BrYV, AhPV1	TuMV, CMV, AhPV1	1
TuMV, CMV, AhPV1	TuMV, CMV, AhPV1	16
TuMV, CMV, AhPV1	TuMV, AhPV1	1
TuMV, CMV, AhPV1	CMV, AhPV1	1
TuMV, CMV, BrYV	TuMV, CMV, BrYV	1
TuMV, BrYV, AhPV1	TuMV, BrYV, AhPV1	1
TuMV, CMV	TuMV, CMV	8
TuMV, CMV	TuMV	2
TuMV, CMV	CMV	1
TuMV, AhPV1	TuMV, AhPV1	7
TuMV, AhPV1	AhPV1	2
CMV, AhPV1	CMV, AhPV1	1
TuMV	TuMV	6
CMV	CMV	2
AhPV1	AhPV1	3