Virus Research

Sequence and phylogenetic analyses of novel totivirus-like double-stranded RNAs from field-collected powdery mildew fungi

Hideki Kondo [a] *, Sakae Hisano [a], Sotaro Chiba [a, 1], Kazuyuki Maruyama [a], Ida Bagus Andika [a], Kasuhiro Toyoda [b], Fumihiro Fujimori [c] and Nobuhiro Suzuki [a]

[a] *Institute of Plant Science and Resources (IPSR), Okayama University, Kurashiki 710-0046, Japan*

[b] *Graduate School of Environmental and Life Science, Okayama University, Okayama 700-8530, Japan*

[c] *Department of Environmental Education, Tokyo Kasei University, 1-18-1 Kaga, Itabashi, Tokyo 173-8062, Japan*

[1] *Current address for S. Chiba: Asian Satellite Campuses Institute, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8601, Japan*

* Corresponding author; **H. Kondo**
E-mail: hkondo@okayama-u.ac.jp
Tel: +81-86-434-1232 / Fax: +81-86-434-1249

Part of special issue: Advances in Fungal Virus Research. The Third International Mycovirus Symposium organized by Angus Dawe, Nobuhiro Suzuki, and Bradley Hillman Edited by Nobuhiro Suzuki

MS information: Word counts: 299 words (Abstract); 5676 words (Main Text) Table: 3; Figure: 5; Supplementary Table: 2; Supplementary Figure: 6

ABSTRACT

The identification of mycoviruses contributes greatly to understanding of the diversity and evolutionary aspects of viruses. Powdery mildew fungi are an important and widely studied obligate phytopathogenic agents, but there has been no report on mycoviruses infecting these fungi. In this study, we used a deep sequencing approach to analyze the double-stranded RNA (dsRNA) segments isolated from field-collected samples of powdery mildew fungus-infected red clover plants in Japan. Database searches identified the presence of at least ten totivirus (genus *Totivirus*)-like sequences, termed red clover powdery mildew-associated totiviruses (RPaTVs). The majority of these sequences shared moderate amino acid sequence identity with each other (<44%) and with other known totiviruses (<59%). Nine of these identified sequences (RPaTV1a, 1b and 2–8) resembled the genome of the prototype totivirus, Saccharomyces cerevisiae virus-L-A (ScV-L-A) in that they contained two overlapping open reading frames (ORFs) encoding a putative coat protein (CP) and an RNA dependent RNA polymerase (RdRp), while one sequence (RPaTV9) showed similarity to another totivirus, Ustilago maydis virus H1 (UmV-H1) that encodes a single polyprotein (CP-RdRp fusion). Similar to yeast totiviruses, each ScV-L-A-like RPaTV contains a –1 ribosomal frameshift site downstream of a predicted pseudoknot structure in the overlapping region of these ORFs, suggesting that the RdRp is translated as a CP-RdRp fusion. Moreover, several ScV-L-A-like sequences were also found by searches of the transcriptome shotgun assembly (TSA) libraries from rust fungi, plants and insects. Phylogenetic analyses show that nine ScV-L-A-like RPaTVs along with ScV-L-A-like sequences derived from TSA libraries are clustered with most established members of the genus *Totivirus*, while one RPaTV forms a new distinct clade with UmV-H1, possibly establishing an additional

genus in the family. Taken together, our results indicate the presence of diverse, novel totiviruses in the powdery mildew fungus populations infecting red clover plants in the field.

***Key words*:** Powdery mildew; Deep sequencing; Double stranded RNA virus; Totivirus; *Saccharomyces cerevisiae virus L-A*; *Ustilago maydis virus H1*

1. Introduction

Mycoviruses or fungal viruses are widespread in all major groups of fungi (Ghabrial et al., 2015; Ghabrial and Suzuki, 2009). Although most mycoviruses have no significant impact on their fungal hosts, some viruses are associated with hypovirulence in plant pathogenic fungi, and thus can be potential biological control agents against fungal diseases (Ghabrial and Suzuki, 2009; Pearson et al., 2009; Kondo et al., 2013b; Xie and Jiang, 2014). The largest number of mycoviruses have double-stranded RNA (dsRNA) genomes, which are encapsidated in rigid virus particles. This dsRNA mycovirus group currently consists of six families (*Totiviridae*, *Partitiviridae*, *Chrysoviridae*, *Reoviridae*, *Megabirnaviridae* and *Quadriviridae*) and a recently proposed family *Botybirnaviridae* (http://talk.ictvonline.org/files/proposals/m/mediagallery/default.aspx). In paticular, members of the families *Totiviridae* (genus *Victorivirus*) and *Partitiviridae* are most frequently found in filamentous fungi.

Powdery mildew fungi (family Erysiphaceae, order Erysiphales) cause disease on a wide range of plant species including crop (e.g., wheat, barley), vegetable (e.g., cucumbers, pea), fruit (e.g., grapes, strawberry) and ornamental (e.g., rose) plants. This fungal family consists of

873 species belonging to 16 genera (Braun and Cook, 2012). These filamentous ascomycetous fungi are obligate biotrophs that depend on living host tissue for their nutrition. *Erysiphe* is the largest genus in the Erysiphaceae and constitutes more than 50% of the members of this family (Braun and Cook, 2012; Takamatsu et al., 2015). *Erysiphe pisi* was once believed to be the only causal agent of powdery mildew disease in pea (*Pisum sativum* L.) that led to heavy yield losses. However, recent studies demonstrated that *Erysiphe trifoliorum* (syn. *E. trifolii*), a closely related species to *E. pisi*, is a second causal agent of pea powdery mildew and is able to overcome *er1* and *Er3* resistance genes against *E. pisi* in pea lines (Attanayake et al., 2010; Fondevilla et al., 2013). *E. trifoliorum* has also been reported to infect other legumes, such as clover (*Trifolium* spp.) and lentil plants (*Lens culinaris*) (Attanayake et al., 2009). The recent draft genome sequencing of some powdery mildew species, including *E. pisi* and *E. necator* (pathogenic on grapes), revealed that their genomes (~120 Mb) are more than four times larger than the normal ascomycetes (Douchkov et al., 2014; Wicker et al., 2013; Spanu et al., 2010). These genomes include a massive expansion of transposable elements, and lost a set of carbohydrate-active enzymes, probably reflecting its evolution towards an obligate biotrophic life style (Spanu et al., 2010). Until now, there have been no reports regarding mycoviruses or virus-like agents infecting powdery mildew fungi.

The family *Totiviridae* is currently divided into five genera, consisting of *Totivirus*, *Victorivirus*, *Giardiavirus*, *Trichomonasvirus* and *Leishmaniavirus*. Viruses that infect fungi have been classified to the first two genera, and those that infect protozoa to the remaining three genera (Goodman et al., 2011; Wickner et al., 2011). Some recently identified viruses infecting arthropods and fish are also proposed as members of a tentative genus *Artivirus* of the family

Totiviridae (Dantas et al., 2016; Haugland et al., 2011; Zhai et al., 2010). Members of family *Totiviridae* are characterized by the non-enveloped isometric virions, 30 to 40 nm in diameter (Goodman et al., 2011; Wickner et al., 2011). Although many characterized dsRNA viruses have segmented genomes, members of family *Totiviridae* have a non-segmented dsRNA genome (4.6–6.7 kb) that commonly contains two overlapping open reading frames (ORFs) encoding a capsid protein (CP or gag) and an RNA-dependent RNA polymerase (RdRp or pol) (Wickner et al., 2011).

Members of genus *Totivirus* infect yeast strains (*Saccharomyces cerevisiae*, *Scheffersomyces segobiensis* and *Xanthophyllomyces dendrorhous*) (Baeza et al., 2012; Bruenn, 1993; Taylor et al., 2013), the corn smut fungus (*Ustilago maydis*) and the subterranean fungus (*Tuber aestivum*, the black summer truffle) (Kang et al., 2001; Stielow and Menzel, 2010). Saccharomyces cerevisiae virus-L-A (ScV-L-A), the so called “yeast LA virus” is the type strain of this genus and is one of the best studied mycoviruses at the molecular level (Wickner et al., 2013). The ScV-L-A replicase is expressed as a Gag-Pol-like fusion protein (analogous to retoroviral Gag-Pol fusion proteins) through a minus one (–1) ribosomal frameshifting (Dinman et al., 1991), while the Ustilago maydis virus H1 (UmV-H1) encodes a large polyprotein and its RdRp is predicted to be released from the fusion by proteolysis (Kang et al., 2001). ScV-L-A and ScV-L-BC (another yeast totivirus closely related to L-A virus) possess a novel cap-snatching mechanism in which the viral Gag (CP) can remove a 7-methyl-guanosine “cap” from cellular mRNA and transfer it to the 5′ end of *de novo* viral transcripts (Fujimura and Esteban, 2011, 2013). Some strains of ScV-L-A and UmV-H1 are known to have additional small satellite dsRNA that code for secreted “killer” protein toxins (Bostian et al., 1984; Koltin

and Day, 1976).

In this study, we report new ten complete or near-complete totivirus-like genome sequences determined by deep sequencing on dsRNA isolated from field-collected powdery mildew fungus that infected red clover plants (the term “totivirus” refers to putative members of the genus, not the family). Interesting insights into the evolution of members of the genus *Totivirus* was provided by phylogenetic analyses of the newly identified genome sequences along with several other totivirus-related sequences detectable in the transcriptome shotgun assembly (TSA) libraries from some species of fungal, plants and insects. This study represents the first report of mycovirus infection of obligate ascomycetes.

2. Materials and method

2.1. Collection of fungal samples and species determinations

Samples of powdery mildew fungus were obtained from red clover (*Triticum aestivum* L. cv. Hokuseki) plants that grew in an experimental field (approximately 3 m × 3 m area) at the Institute of Plant Science and Resources (IPSR), Okayama University, Japan in the early spring of 2013. Conidia (and conidiophores) of powdery mildew fungus were collected from heavily infected leaves by washing the leaves with sterilized water and the samples were stored at –80 ˚C.

For species determination, internal transcribed spacer (ITS) sequences were analyzed. Total genomic DNA was extracted from conidia (and conidiophore)-enriched samples using DNeasy® Blood and Tissue Kit (Qiagen) following the manufacturer’s instructions, to generate a PCR template. ITS sequences of ribosomal DNA were amplified by PCR using the

Erysiphe-specific ITS primer pair, EryF and EryR (Attanayake et al., 2009) (Table S1). Amplified DNA fragments were sequenced using an ABI3100 DNA sequencer (Applied Biosystems, Foster City, CA, USA).

2.2. Nucleic acid extraction and analysis

Total nucleic acid was extracted as described previously (Sun and Suzuki, 2008). After precipitation of total nucleic acids in ethanol, samples were treated with S1 nuclease (Fermentas) and subsequently with RQ1 RNase-free DNase I (Promega) to digest single-stranded RNAs (rRNAs and mRNAs) and DNAs (genomic DNA), respectively. The undigested nucleic acids were further purified by affinity column chromatography using CC41 cellulose powder (Whatman) to obtain dsRNA-enriched fractions as described previously (Eusebio-Cope and Suzuki, 2015).

Reverse transcription (RT)-PCR was performed as described previously (Lin et al., 2012). cDNAs were synthesized with Superscript II RT (Invitrogen) using random hexamers following the manufacturer's instructions. The resultant cDNA was used as templates for PCR amplification. We took advantage of the viral genomic dsRNA to simultaneously determine both 5′- and 3′-terminal of totivirus-related sequences using 3′-RLM-RACE (Potgieter et al., 2009 and Chiba et al., 2009). PCR products were then sequenced using the Sanger sequencing method.

2.3. RNA-seq and reads assembly

The mixture of dsRNA samples (total 3.7 μg) including three independent dsRNA samples derived from other phytopathogenic fungi (sequence data from these fungi will be reported elsewhere) was used for cDNA library construction using the TruSeq RNA Sample Preparation Kit (Illumina). In this method, ds-cDNA is end-repaired and adenylated prior to adaptor ligation, library construction and amplification. The sequenced-ready library was then subjected to paired-end sequencing of 100 nucleotide (nt) reads using Illumina HiSeq 2000 technology (Illumina). The cDNA library construction and deep sequencing analysis were carried out by Hokkaido System Science Co.

After deep sequencing, adapter sequences attached to cDNA sequences were removed from sequence reads (61,387,280 raw reads) using cutadapt (https://code.google.com/p/cutadapt/). *De novo* assembly of contiguous sequences was carried out using the Velvet *de novo* assembly algorithm (http://www.ebi.ac.uk/~zerbino/velvet) with hash lengths from 55, 65, 75, 85 or 95. The assembly yielded 1293 (hash length=95) to 82,217 (hash length=55) contigs.

2.4. Database search and sequence analysis

These contigs (~1.0 kb) were subjected to BLAST searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi) against the non-redundant (nr) DNA and protein databases (see below) of National Center for Biotechnology Information (NCBI). To construct larger contigs, the contigs (1.0–4.9 kb) that show e-values >$1e^{-5}$ with totiviruses were assembled with the overlapped surrounding contigs, including those <1.0 kb long, using GENETYX-MAC (GENETYX Co., Tokyo, Japan). Mapping of sequence reads to each of the totivirus-like contigs

was done using the Read Mapping algorithm of the CLC Genomics Workbench ver. 8 (http://www.clcbio.com/) with settings for the length fraction and similarity of 0.9 and 0.98, respectively.

ORFs were identified using GENETYX-MAC or EnzymeX version 3.3.3 (http://nucleobytes.com/enzymex/index.html). Sequence similarities were calculated using the BLAST program available from NCBI. The prediction of RNA pseudoknot structure was done using the DotKnot program (Sperschneider and Datta, 2010) and drawn with PseudoViewer3 (Byun and Han, 2009).

BLAST searches using totivirus-like sequences as queries were conducted against sequence databases available from NCBI (nucleotide collection, nr/nt; non-human, non-mouse expressed sequence tags, EST; transcriptome shotgun assembly, TSA; whole-genome shotgun contigs, WGS) (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Pairwise amino acid identity was calculated using SDT v1.2 (Muhire et al., 2014) with the MUSCLE (Edgar, 2004) alignment. Conserved domain searches were performed using NCBI's conserved domain database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

2.5. Phylogenetic analysis

Phylogenetic tree construction was based on a maximum-likelihood (ML) method as described previously with minor modifications (Kondo et al., 2015; Kondo et al., 2013a). Multiple alignments were performed using MAFFT version 7 set to the default parameters (Katoh and Toh, 2008). Ambiguously aligned regions were removed using Gblocks 0.91b (Talavera and Castresana, 2007), with the stringency levels lowered for all parameters. ML

phylogenetic trees were estimated using PhyML 3.0 (Guindon et al., 2010), with automatic model selection by SMS (Smart Model Selection) (http://www.atgc-montpellier.fr/phyml-sms/). The phylogenetic tree (mid-point rooted) was visualized and refined with FigTree version 1.3.1 software (http://tree.bio.ed.ac.uk/software/).

3. Results

3.1. The presence of dsRNA in the powdery mildew fungus infecting red clover plants

In a preliminary study, we detected the presence of dsRNA molecules in the total RNA isolated from conidia of powdery mildew fungus (*E. trifoliorum*) infecting red clover plants that grew in our experimental field in IPSR and were collected in the early spring of 2012 (Fig. S1A). In contrast, no dsRNA was detected in the powdery mildew (*E. pisi*) samples harvested from pea plants in the other IPSR field (Fig. S1A and data not shown). Analyses of the isolated dsRNA using standard cDNA library preparation and sequence analysis identified the presence of several totivirus-like fragments (~1 kb), suggesting that undescribed totiviruses might predominantly infect those red clover powdery mildew (H.Kondo and S. Hisano, unpublished results).

For further confirmation and more detailed analyses of the viruses, in the early spring of 2013, the conidia (and conidiophores) of powdery mildew fungus (*E. trifoliorum*) were again collected from the leaves of heavily infected red clover plants that grew in the same experimental field described above (Fig. 1A). Microscopic observation showed that the field-collected conidia were predominantly ellipsoid-cylindrical shape and estimated to be approximately 30–40 μm in length (Fig. 1B and data not shown). Analysis of the ITS sequence by genomic PCR confirmed

that this powdery mildew fungus was of the *E. trifoliorum* species (data not shown). The total nucleic acid extraction from collected conidia showed the presence of two major and some minor dsRNA segments (around 5–6 kbp length) based on their resistance to digestion with both S1 nuclease and DNase I (Fig. 1C and S1B). The dsRNA molecules were further purified from the total nucleic acid fraction and subjected to cDNA library preparation and deep sequencing analysis (Fig. S1C).

3.2. Identification of putative totiviruses infecting red clover powdery mildew fungus

After deep sequencing and assembly of the sequence reads, the resulting contigs were used as queries in BLAST searches (tblastn) against nonredundant NCBI sequence databases. BLAST searches revealed that many contigs bore similarity to the yeast LA virus (ScV-L-A) and its related sequences as well as to other mycoviral sequences. The contigs having significant identities (e-value $<2e^{-20}$) to members of the genus *Totivirus* and with lengths greater 4.0 kb were selected for further analyses. Note that most of the contigs obtained were composed of several overlapping items (data not shown).

Because deep sequencing analysis included other fungal samples (see the Materials and Method section), we performed RT-PCR to verify the presence of those totivirus-related RNA species in the powdery mildew sample using the specific primer sets for each of the totivirus-like sequences (Table S1 and data not shown). Ten sets of primers successfully targeted DNA fragments for amplification from the original dsRNA sample derived from the red clover powdery mildew, but not from the dsRNA derived from other fungal samples (Fig. 1D and data not shown). Sequence analysis confirmed that the amplified fragments were identical to the corresponding

virus-like sequences obtained by deep sequencing.

Hereafter, these totivirus-like sequences (>4.1 kbp in length) derived from the powdery mildew fungus are referred to as red clover powdery mildew-associated totivirus 1 to 9 (RPaTV1 to 9) (Table 1). Mapping of the reads to RPaTVs showed that the highest number of reads (6808 reads) were mapped to RPaTV9, suggesting that RPaTV9 was likely to be the most predominant totivirus-like segment present in the original dsRNA pool (Table 1 and Fig. S2).

3.3. Sequence analysis of putative totiviruses

To determine the terminal sequences of RPaTVs, 3′-RLM-RACE was performed using the dsRNA sample derived from the red clover powdery mildew (Fig. S1D and data not shown). After direct sequencing of RLM-RACE products, the 3′-end sequences corresponding to both strands of five RPaTVs (1a, 3 and 5 to 7) and the positive strand of two RPaTVs (2 and 9) were obtained (Table 1). The complete nucleotide sequences of RPaTVs (1a, 3 and 5 to 7), ranging from 4781 to 5056 nt, and near complete sequences of other RPaTVs, ranging for 4178 to 5997 nt, were submitted to the DNA Data Bank of Japan (DDBJ) (http://www.ddbj. nig.ac.jp/) and the accession numbers are listed in Table 1.

Using the NCBI blastn-suite-2, no significant nucleotide sequence similarity among the RPaTV sequences was observed, except for RPaTV1a and 1b, which showed high similarity (87% identity) (data not shown). The PASC (Pairwise Sequence Comparison) analysis (Bao et al., 2014) showed that RPaTV sequences have moderate levels of nucleotide sequence identity (50–57%) to known totiviruses in the pairwise global alignment, but these degrees of similarity are much higher than those identified using BLAST-based alignments (Table S2).

The 5′-terminal sequences of RPaTVs 1a, 3 and 5 (5′-CGAAAUUUU…-3′, 5′-GGAAUUUUU…-3′ and 5′-UGAGAUUUUU…-3′ respectively) appeared to be partially conserved with that of ScV-L-A (5′-GAAAAAUUUUU…-3′) and some other totiviruses (Figs. 2, S3 and data not shown). RPaTVs 6 and 7 shared the first 13 nt of 5′-terminal sequences (UGAGAAAAACUCA) and the first 5-nt sequence (5′-UGAGA…-3′) is identical to that of RPaTV5 (Figs. 2 and S3). Likewise, RPaTVs 1a, 3, 6 and 7 shared the 3 nt of 3′-terminal sequences (UCA) (Figs. 2 and S3).

Of these ten RPaTVs analyzed (Table 1), seven (RPaTVs 1a, 1b, 3, 5–7 and 9) and one (RPaTV9) encode or probably encode two complete ORFs (ORF1 and ORF2) and a single ORF, respectively (Figs. 2 and S3), which are the typical genome structures of totiviruses infecting yeast (ScV-L-A) and smut fungi (UmV-H1), respectively. Our observations that the N-terminal sequence of RPaTV ORF1s or RPaTV9 ORF coded proteins are well aligned with those of known totivirus CPs support this (Fig. S4). In addition, the sequences surrounding the AUG start codons (Kozack sequence) of these RPaTV ORFs are mostly in the favorable context for translation initiation in the yeasts and plants (Hamilton et al., 1987; Lutcke et al., 1987) (Fig. 3A). The remaining three sequences (RPaTVs 2, 4 and 8) are also likely to code for two proteins, however, they lack the sequences coding for N- or C-terminal regions of ORF1 and ORF2, respectively (Figs. 2 and S3).

The proteins encoded by ORFs 1 and 2 of RPaTVs 1–8 and the ORF of RPaTV9 have conserved domains of CP (pfam09220, but not all RPaTVs) and RdRp (pfam02123), respectively (Figs. 2 and S3). Pair-wise comparisons among proteins encoded by RPaTVs 1–9 revealed that moderate levels (23–44%) of amino acid sequence identity among ORF2 proteins (RdRp, 88–95

kDa), including the RdRp domain of the predicted RPaTV9 polyprotein (ORF, 204 kDa), while little relationship (18–34%) was found among ORF1 proteins (CP, 76–91 kDa), including the putative RPaTV9 CP domain (Fig. S5). An exception, RPaTV1a proteins share a high level of amino acid sequence identity (96% or 97%) with those of RPaTV1b (Fig. S5). The proteins predicted by RPaTVs 1–8 show a moderate level of amino acid sequence identities (21–44% for CP and 34–59% for RdRp) with those of known totiviruses infecting yeasts (*Sa. cerevisiae, Sc. segobiensis* and *X. dendrorhous*), fungi (*Us. maydis* and *T. aestivum*) and an unknown host (probably fungus, yeast or plants, *Rubus* sp.) (Table 2). The multiple alignment analysis suggests that the amino acid residues crucial for the cap-snatching function of totivirus CP, such as the His-154 for ScV-L-A and His-156 for ScV-BC (Fujimura and Esteban, 2011, 2013), are mostly conserved or replaced with related similar amino acids in the majority of RPaTV ORF1-encoded proteins, but not in the RPaTV9 ORF-encoded protein (Fig. S4). In addition, the RPaTV9 ORF- and other RPaTV ORF2-encoded proteins contain eight conserved domains (I to VIII), including the GDD motif, which are widely conserved in the RdRp of dsRNA viruses (Routhier and Bruenn, 1998) (Fig. S4).

Notably, a possible –1 ribosomal frameshift site (canonical slippery sites “X XXY YYZ”, triplets are shown for the preshifted reading frame), which is similar to the –1 ribosomal frameshift site in the ScV-L-A genome (Dinman et al., 1991), was recognized in the overlap region of ORF1 and ORF2 of each of RPaTV1–8 (Figs. 2 and S3). The slippery site “GGG/AUUUU” sequence instead of ScV-L-A’s “GGGUUUA” heptanucleotide (Brierley, 1995), was predicted in RPaTV1–8 sequences (Fig. 3B). Using the DotKnot program, a pseudoknot structure, which could help pause the translating ribosomes and increasing the

frequency of frameshifting (Brierley et al., 1989), is predicted to exist immediately downstream of each putative slippery site (Fig. 3C and data not shown), Thus, if all of them employ the same ribosomal frameshift mechanism, the putative fusion proteins encoded by each RPaTVs range from 1528 (RPaTV1) to 1644 amino acids (RPaTV5) long, with the predicted molecular mass of 172–186 kDa.

3.4. The discovery of further totivirus-like sequences in transcriptome databases

Recently, some virus-like sequences related to totiviruses are found in the transcriptomes of haustoria of the common bean rust fungus (*Uromyces appendiculatus*) and the soybean rust fungus (*Phakopsora pachyrhizi*) (Link et al., 2014). Therefore, to explore the possible occurrence of similar uncharacterized totiviruses in fungal and other hosts, we conducted tblastn searches against transcriptome shotgun assembly (TSA) libraries of the fungi, plants and insects using totivirus-related sequences obtained from this study as queries. Interestingly, the proteins encoded by several TSA contigs from the rust fungus including above-mentioned sequences have significant similarities to the putative RPaTV coding proteins (CP and RdRP) (Table 3). RPaTV6-related TSAs from bean rust fungi such as *P. pachyrhizi* TSA1 (GACM01000569) and TSA2 (GACM01000768), *Ur. appendiculatus* TSA1 (GACI01004215) and TSA2 (GACI01000486), and an RPaTV3-related TSA from *Ur. appendiculatus* (TSA3, GACM01003374) are likely to be near-complete genome sequences (~4 kb) (Table 3). Some TSA sequences derived from plants and insects including the oriental tobacco budworm (*Helicoverpa assulta*), a springtail (*Tetrodontophora bielanensis*), a rockcress (*Boechera gunnisoniana*, relative of *Arabidopsis thaliana*), the tea plant (*Camellia taliensis*), and a legume

(*Lathyrus sativus*) also show modest levels of amino acid sequence identity to RPaTV CPs and/or RdRps (Table 3).

3.5. Phylogenetic relationships between RPaTVs and known totiviruses

To evaluate the phylogenetic relationships between the RPaTV sequences discovered in this study and known totiviruses (genus *Totivirus*), we constructed a maximum likelihood (ML) phylogenetic tree using amino acid sequences of the CP-RdRp (Gag-Pol-like) fusion and polyprotein sequences. As shown in Fig. 4, the constructed ML tree consisted of two major clades, named group I (most RPaTVs and totiviruses) and group II (RPaTV9 and UmV-H1). Group I was further separated into four sister clades (subgroups I-A to I-D, Fig. 4). RPaTVs 1 to 4 were distributed among the three sister clades along with known totiviruses. RPaTVs 1a and 1b were clustered with ScV-L-A, Tuber aestivum virus 1 (TaV1) and black raspberry virus F (BRV-F) in the subgroup I-A; RPaTV2 was clustered with ScV-L-BC and Scheffersomyces segobiensis virus L (SsV-L) in the subgroup I-B; and the subgroup I-C contained RPaTVs 3 and 4 together with Xanthophyllomyces dendrorhous virus L1B (XdV-LIB). The remaining four RPaTVs 5 to 8, were clustered together and independently formed the subgroup I-D (Fig. 4).

A similar tree topology was observed in ML trees based on CP (gag) or RdRp (pol) sequences of RPaTVs and known totiviruses (Figs. 5 and S6), further supporting the phylogenetic status of the RPaTVs. Notably, most of the totivirus-like TSAs derived from *P. pachyrhizi* and *Ur. appendiculatus* were placed within the forth sub-clade (the subgroup I-D) together with RPaTVs 5 to 8, whereas plant- and insect-derived TSAs were discretely placed in the other three sub-clades (Figs. 5 and S6). Recently, totivirus-related sequences have been

discovered as the fossil of viral elements in the genomes of some yeast strains (Frank and Wolfe, 2009; Liu et al., 2010; Taylor et al., 2013; Taylor and Bruenn, 2009). These endogenous totivirus-like sequences are closely related to totivirus CP or RdRps in the sub-clade I (Figs. 5 and S6), whereas a novel totivirus-CP-like sequence in the seaweed genome (*Chondrus crispus*, XP_005710022) (Query sequence=EaTV1a CP: Identity=24%, e-value =$2e^{-12}$), which was discovered during our data base search, is distantly related to all four other sub-clades (Fig. S6).

4. Discussion

Although the presence of mycoviruses has been recognized in all major groups of plant pathogenic fungi (Ghabrial and Suzuki, 2009), virus discovery in obligate biotrophs such as powdery mildews (ascomycetes) and rust fungi (basidiomycetes) that are unculturable on artificial media, is still very limited. The recent development of next generation sequencing technologies has allowed for the discovery of new viral sequences in environmentally derived biological samples; such new mycoviruses can be identified directly from field-collected fungal samples without a culturing process in the laboratory (Al Rwahnih et al., 2011; Feldman et al., 2012; Roossinck, 2015: Marzano and Domier, 2016). In this study, by conducting deep sequencing and RLM-RACE of the dsRNA-enriched sample recovered from field-collected conidia of red clover powdery mildew fungus (*E. trifoliorum*) (Fig. 1), we identified five-complete (referred to as RPaTVs 1a, 3 and 5 to 7) and five near-complete (RPaTVs 1b, 2, 4, 8 and 9) totivirus-like assembled sequences (Table 1). This finding is of great interest from several perspectives.

First, this study provides the first evidence of mycovirus infections in the powdery mildew fungus (Table 2). We also identified similar RPaTV-related sequences, in the powdery mildew fungus originating from the same field, but later propagated in the laboratory using healthy red clover plants (H. Kondo. and S. Hisano, unpublished results), confirming that RPaTV sequences indeed originated from the powdery mildew fungus populations infecting red clover plants in the field. Second, all of the newly identified totivirus-like sequences probably represent strains of novel totivirus species (genus *Totivirus*), although their putative encoded proteins (CP and RdRp or polyprotein) are related to those of known members of the genus *Totivirus* (Figs. 5 and S6, Table 2). The current species demarcations of the genus *Totivirus* share less than 50% sequence identity at the protein level with either CP or RdRp (Wickner et al., 2011). An official proposal of the newly identified sequences as species to the ICTV (International Committee of Taxonomy of Viruses) awaits thorough characterization of each of the totivirus-like sequences as a viral entity. For this purpose, virion transfection into a model fungus host, *Cryphonectria parasitica* (Chiba et al., 2013; Eusebio-Cope et al., 2015) will be performed in the near future. Third, many RPaTVs are possible unusual members of the family *Totiviridae* in terms of phylogeny and genome organization as mentioned below. The known members of the family *Totiviridae* mainly infect fungi and protozoans (Wickner et al., 2011) and have diverged into five distant evolutionary lineages: giardiavirus like (giardiaviruses), infectious myonecrosis virus like (artiviruses), victorivirus like (victori-, trichomonas- and leishmaniaviruses), ScV-L-A like (totiviruses) and UmV-H1 like (totivirus) (Liu et al., 2012) (see also Fig. 5). All viruses isolated thus far from filamentous ascomycetes and basideomycetes are classified into the genus *Victorivirus*, except for TaV (Stielow and Menzel, 2010), and the prototype of victorivirus

(Helminthosporium victoriae virus 190S) and possibly other members utilize a stop/restart mechanism for the translation of ORF2 (Li et al., 2015). However, RPaTV sequences show phylogenetic affinity for the genus *Totivirus* as mentioned above. In addition, nine RPaTV (RPaTV1–8, ScV-L-A like) sequences possess the characteristics typical of yeast totivirus genomes in that they contain two overlapping ORFs with the possible –1 ribosomal frameshift site. The RPaTV9 sequence, which is similar to the smut fungus totivirus (UmV-H1 like) genome sequence, contains a single ORF encoding a large polyprotein (Figs. 2, 3 and S3); the sequences of the RPaTV9 (>6.0 kbp) and UmV-H1 genomes (6.1 kbp) are slightly larger than those of other RPaTVs (4.8–5.1 kbp) and typical yeast totiviruses (see Fig. 2).

Further inspection of the RdRp-based phylogenic tree showed that the ScV-L-A-like RPaTVs and most other totiviruses (group I) are distantly related to quadripartite dsRNA viruses in the family *Quadriviridae* and their related viruses (Covelli et al., 2004; Lin et al., 2012; 2013) (Fig. 5). Likewise, RPaTV9 and UmV-H1 (group II) are distantly related to the novel multipartite dsRNA viruses in the proposed family *Botybirnaviridae*, and together form an evolutionary separated clade (Liu et al., 2015; Wu et al., 2012) (Fig. 5). It is therefore possible that these two independent bipartite (botybirnaviruses) and tetrapartite (quadriviruses) dsRNA viruses have evolved from different ancestral viruses with monopartite dsRNA genomes (Liu et al., 2012). The members of group II (UmV-H1 and RPaTV9) may form a new genus apart from the genus *Totivirus* due to above mentioned differences in coding strategies and phylogenetic relationships. For the ScV-L-A-like totiviruses (group I), these are further separated into four subgroups I-A to I-D (Figs. 4, 5 and S6). The RPaTV (ScV-L-A-like) sequences discovered in this study are distributed within all sub-clades, but interestingly, previously known fungal and yeast totiviruses

are not placed within the subgroup I-D (Figs. 5 and S6). In our BLAST searches, we found several totivirus-related sequences in the TSA libraries from bean rust fungi (Table 3). Most of them (except for *Ur. appendiculatus* TSA3, GACI01003374) are clustered together with RPaTVs 5 to 8 to constitute the subgroup I-D in the both CP and RdRp-based phylogenetic trees (Figs. 5, S6 and data not shown). Of these bean rust-derived TSAs, some have already been described, including the *P. pachyrhizi* TSA2 (GACM01000768, 5179 nt) that likely represents a complete ScV-L-A-like genomic sequence (Link et al., 2014) (Table 3). As we mentioned above, several RPaTV-related sequences were detected from the laboratory-propagated red clover powdery mildew fungus (*E. trifoliorum*) using originally fungus-free plants. Therefore, multiple totiviral infections may occur in field populations of this powdery mildew fungal species. It is important to note that mycoviral transmission is usually limited to intracellular transmission via anastomosis and sporulation, and horizontal movements to other fungal hosts are likely to be rare (Pearson et al., 2009). However, there are a few reports showing lateral mycovirus transfer across phyla, e.g., between *Sclerotinia homoeocarpa* and *Ophiostoma novo-ulmi* (Deng et al., 2003). Based on the phylogenetic topology of the forth sub-clade (the subgroup I-D) sequences (Figs. 5 and S6), we hypothesize that the horizontal transmission of ancestral totiviruses might have occurred across quite different fungal phyla, such as between powdery mildews (ascomycetes) and rust fungi (basidiomycetes) and later evolved within the new host fungus separately.

Some RPaTV (ScV-L-A like)-related TSAs derived from the rockcress, the tea and the legume plants (Table 3) show modest levels of amino acid sequence identity with the totiviral CPs or RdRps belonging to the subgroup I-A (Fig. S6). Several other ScV-L-A-related sequences including BRV-F (NC_009890, 5077 bp) and ribes virus F (EU495331, 1048 bp), have also

been detected in the dsRNA isolated from the plants (Al Rwahnih et al., 2011; Cox et al., 2000; Martin et al., 2006). Some of these ScV-L-A-like sequences are probably originating from viruses infecting fungi or yeast that propagated in those plants, because some mycoviruses (not in the case of totiviruses) discovered in the grapevine tree were found again in fungal cultures that were isolated from the same plant (Al Rwahnih et al., 2011). However, this notion is unlikely to account for the host origin of all of these viruses, because in some cases, the titer (read counts) of the putative virus is very high without any evidence of fungal infection (Roossinck, 2012). In fact, some viruses could be persistently detectable in both plants and fungi, such as the members of the family *Partitiviridae* and probably *Chrysoviridae* (bi- and multi-segmented dsRNA viruses, respectively) (Li et al., 2013; Roossinck, 2012). Interestingly, we found two insect TSA sequences from the tobacco budworm (2637 nt) and the springtail (1904 nt) as well as the novel endogenized torivirus-like sequence in the red alga genome (*C. crispu*s) that is likely a fossil record of ancient totivirus infection (Table 3, Figs. 5 and S6). Therefore, it is possible that many unseen totiviruses (*Totivirus* genus) infect a wide range of organisms across the different eukaryotic kingdoms such as fungi, plants, insects and red algae. Further wet and dry lab analyses (deep sequencing and more extensive database search) will be required to trace the deep evolution of totiviruses.

Acknowledgements

We thank Dr. Yu-Hsin Lin for technical assistance. We also thank two anonymous reviewers for their valuable suggestions. This research was partially supported by a Grant-in-Aid for Scientific Research KAKENHI of the Japan Society for Promotion of Science

(No.15K07312 to H.K., No.25252011 to N.S.) and part of Joint Research Program implemented at the Institute of Plant Science and Resources, Okayama University in Japan (F. F.).

References

Al Rwahnih, M., Daubert, S., Urbez-Torres, J.R., Cordero, F., Rowhani, A., 2011. Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. Arch. Virol. 156, 397–403.

Attanayake, R.N., Glawe, D.A., Dugan, F.M., Chen, W.D., 2009. *Erysiphe trifolii* causing powdery mildew of lentil (*Lens culinaris*). Plant Dis. 93, 797–803.

Attanayake, R.N., Glawe, D.A., McPhee, K.E., Dugan, F.M., Chen, W., 2010. *Erysiphe trifolii* - a newly recognized powdery mildew pathogen of pea. Plant Pathol. 59, 712–720.

Baeza, M., Bravo, N., Sanhueza, M., Flores, O., Villarreal, P., Cifuentes, V., 2012. Molecular characterization of totiviruses in *Xanthophyllomyces dendrorhous*. Virol. J. 9, 140. doi:10.1186/1743-422X-9-140.

Bao, Y.M., Chetvernin, V., Tatusova, T., 2014. Improvements to pairwise sequence comparison (PASC): a genome-based web tool for virus classification. Arch. Virol. 159, 3293–3304.

Bostian, K.A., Elliott, Q., Bussey, H., Burn, V., Smith, A., Tipper, D.J., 1984. Sequence of The preprotoxin dsRNA gene of type-I killer yeast - multiple processing events produce a 2-component toxin. Cell 36, 741–751.

Braun, U., Cook, R.T.A., 2012. Taxonomic manual of the *Erysiphales* (powdery mildews). Utrecht, the Netherlands: CBS-KNAW Fungal Bio-diversity Centre.

Brierley, I., 1995. Ribosomal frameshifting on viral RNAs. J. Gen. Virol. 76, 1885–1892.

Brierley, I., Digard, P., Inglis, S.C., 1989. characterization of an efficient coronavirus ribosomal frameshifting signal - requirement for an rna pseudoknot. Cell 57, 537–547.

Bruenn, J.A., 1993. A closely-related group of RNA-dependent RNA -polymerases from double-stranded-RNA viruses. Nucleic Acids Res. 21, 5667–5669.

Byun, Y., Han, K., 2009. PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. Bioinformatics 25, 1435–1437.

Chiba, S., Salaipeth, L., Lin, Y. H., Sasaki, A., Kanematsu, S., Suzuki, N., 2009. A novel bipartite double-stranded RNA mycovirus from the white root rot fungus *Rosellinia necatrix*: molecular and biological characterization, taxonomic considerations, and potential for biological control. J. Virol. 83, 12801–12812.

Chiba, S., Lin, Y.H., Kondo, H., Kanematsu, S., Suzuki, N., 2013. A novel victorivirus from a phytopathogenic fungus, *Rosellinia necatrix*, is infectious as particles and targeted by RNA silencing. J. Virol. 87, 6727–6738.

Covelli, L., Coutts, R.H., Di Serio, F., Citir, A., Acikgoz, S., Hernandez, C., Ragozzino, A., Flores, R., 2004. Cherry chlorotic rusty spot and Amasya cherry diseases are associated with a complex pattern of mycoviral-like double-stranded RNAs. I. Characterization of a new species in the genus *Chrysovirus.* J. Gen. Virol. 85, 3389–3397.

Cox, S., Mayo, M.A., Jones, A.T., 2000. The occurrence of dsRNA species in apparently healthy and virus-infected *Ribes* cultivars, and evidence that one such species originates from a member of the virus family *Totiviridae.* Eur. J. Plant Pathol. 106, 353–364.

Dantas, M.D.A, Cavalcante, G.H.O., Oliveira, R.A.C., Lanza, D.C.F, 2016. New insights about ORF1 coding regions support the proposition of a new genus comprising arthropod viruses in the family *Totiviridae*. Virus Res., in press. doi:10.1016/j.virusres.2015.10.020.

Deng, F., Xu, R., Boland, G. J., 2003. Hypovirulence-associated double-stranded RNA from *Sclerotinia homoeocarpa* is conspecific with *Ophiostoma novo-ulmi* mitovirus 3a-Ld. Phytopathology 93, 1407–1414.

Dinman, J.D., Icho, T., Wickner, R.B., 1991. A -1 ribosomal frameshift in a double-stranded-RNA virus of yeast forms a gag pol fusion protein. Proc. Natl. Acad. Sci. U. S. A. 88, 174–178.

Douchkov, D., Luck, S., Johrde, A., Nowara, D., Himmelbach, A., Rajaraman, J., Stein, N., Sharma, R., Kilian, B., Schweizer, P., 2014. Discovery of genes affecting resistance of barley to adapted and non-adapted powdery mildew fungi. Genome Biol. 15, 518. doi:10.1186/s13059-014-0518-8.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform 5, 1–19. doi:10.1186/1471-2105-5-113.

Eusebio-Cope, A., Sun, L.Y., Tanaka, T., Chiba, S., Kasahara, S., Suzuki, N., 2015. The chestnut blight fungus for studies on virus/host and virus/virus interactions: From a natural to a model host. Virology 477, 164–175.

Eusebio-Cope, A., Suzuki, N., 2015. Mycoreovirus genome rearrangements associated with RNA silencing deficiency. Nucleic Acids Res. 43, 3802–3813.

Feldman, T.S., Morsy, M.R., Roossinck, M.J., 2012. Are communities of microbial symbionts more diverse than communities of macrobial hosts? Fungal Biol. 116, 465–477.

Fondevilla, S., Chattopadhyay, C., Khare, N., Rubiales, D., 2013. *Erysiphe trifolii* is able to overcome *er1* and *Er3*, but not *er2*, resistance genes in pea. Eur. J. Plant Pathol. 136, 557–563.

Frank, A.C., Wolfe, K.H., 2009. Evolutionary capture of viral and plasmid DNA by yeast nuclear chromosomes. Eukaryot. Cell 8, 1521–1531.

Fujimura, T., Esteban, R., 2011. Cap-snatching mechanism in yeast L-A double-stranded RNA virus.

Proc. Natl. Acad. Sci. U. S. A. 108, 17667–17671.

Fujimura, T., Esteban, R., 2013. Cap snatching in yeast L-BC double-stranded RNA totivirus. J. Biol. Chem. 288, 23716–23724.

Ghabrial, S.A., Caston, J.R., Jiang, D.H., Nibert, M.L., Suzuki, N., 2015. 50-plus years of fungal viruses. Virology 479, 356–368.

Ghabrial, S.A., Suzuki, N., 2009. Viruses of plant pathogenic fungi. Annu. Rev. Phytopathol. 47, 353–384.

Goodman, R.P., Ghabrial, S.A., Fichorova, R.N., Nibert, M.L., 2011. *Trichomonasvirus*: a new genus of protozoan viruses in the family *Totiviridae*. Arch. Virol. 156, 171–179.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

Hamilton, R., Watanabe, C.K., Deboer, H.A., 1987. Compliation and comparison of the sequence context around the AUG startcodons in *Saccharomyces cerevisiae* messenger-RNAs. Nucleic Acids Res. 15, 3581–3593.

Haugland, O., Mikalsen, A.B., Nilsen, P., Lindmo, K., Thu, B.J., Eliassen, T.M., Roos, N., Rode, M., Evensen, O., 2011. Cardiomyopathy syndrome of Atlantic salmon (*Salmo salar* L.) is caused by a double-stranded RNA virus of the *Totiviridae* family. J. Virol. 85, 5275–5286.

Kang, J.G., Wu, J.C., Bruenn, J.A., Park, C.M., 2001. The H1 double-stranded RNA genome of Ustilago maydis virus-H1 encodes a polyprotein that contains structural motifs for capsid polypeptide, papain-like protease, and RNA-dependent RNA polymerase. Virus Res. 76, 183–189.

Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9, 286–298.

Koltin, Y., Day, P.R., 1976. Inheritance of killer phenotypes and double-stranded RNA in *Ustilago maydis*. Proc. Natl. Acad. Sci. U. S. A. 73, 594–598.

Kondo, H., Chiba, S., Suzuki, N., 2015. Detection and analysis of non-retroviral RNA virus-like elements in plant, fungal, and insect genomes. Methods Mol. Biol. 1236, 73–88.

Kondo, H., Chiba, S., Toyoda, K., Suzuki, N., 2013a. Evidence for negative-strand RNA virus infection in fungi. Virology 435, 201–209.

Kondo, H., Kanematsu, S., Suzuki, N., 2013b. Viruses of the white root rot fungus, *Rosellinia necatrix*. Adv. Virus Res. 86, 77–214.

Li, H., Havens, W.M., Nibert, M.L., Ghabrial, S.A., 2015. An RNA cassette from Helminthosporium

victoriae virus 190S necessary and sufficient for stop/restart translation. Virology 474, 131–143

Li, L.Q., Liu, J.N., Xu, A.X., Wang, T., Chen, J.S., Zhu, X.W., 2013. Molecular characterization of a trisegmented chrysovirus isolated from the radish *Raphanus sativus*. Virus Res. 176, 169–178.

Lin, Y.H., Chiba, S., Tani, A., Kondo, H., Sasaki, A., Kanematsu, S., Suzuki, N., 2012. A novel quadripartite dsRNA virus isolated from a phytopathogenic filamentous fungus, *Rosellinia necatrix*. Virology 426, 42–50.

Lin, Y.H., Hisano, S., Yaegashi, H., Kanematsu, S., Suzuki, N., 2013. A second quadrivirus strain from the phytopathogenic filamentous fungus *Rosellinia necatrix*. Arch. Virol. 158, 1093–1098.

Link, T.I., Lang, P., Scheffler, B.E., Duke, M.V., Graham, M.A., Cooper, B., Tucker, M.L., Van De Mortel, M., Voegele, R.T., Mendgen, K., Baum, T.J., Whitham, S.A., 2014. The haustorial transcriptomes of *Uromyces appendiculatus* and *Phakopsora pachyrhizi* and their candidate effector families. Mol. Plant Pathol. 15, 379–393.

Liu, H.Q., Fu, Y.P., Jiang, D.H., Li, G.Q., Xie, J.T., Cheng, J.S., Peng, Y.L., Ghabrial, S.A., Yi, X.H., 2010. Widespread horizontal gene transfer from double-stranded RNA viruses to eukaryotic nuclear genomes. J. Virol. 84, 11876–11887.

Liu, H.Q., Fu, Y.P., Xie, J.T., Cheng, J.S., Ghabrial, S.A., Li, G.Q., Peng, Y.L., Yi, X.H., Jiang, D.H., 2012. Evolutionary genomics of mycovirus-related dsRNA viruses reveals cross-family horizontal gene transfer and evolution of diverse viral lineages. BMC Evol. Biol. 12, 91 doi:10.1186/1471-2148–12-91.

Liu, L.J., Wang, Q.H., Cheng, J.S., Fu, Y.P., Jiang, D.H., Xie, J.T., 2015. Molecular characterization of a bipartite double-stranded RNA virus and its satellite like RNA co- infecting the phytopathogenic fungus *Sclerotinia sclerotiorum*. Front. Microbiol. 6, 406. doi: 10.3389/fmicb.2015.00406.

Lutcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F., Scheele, G.A., 1987. Selection of AUG initiation codons differs in plants and animals. EMBO J. 6, 43–48.

Martin, R.R., Tzanetakis, I.E., Sweeney, M., Wegener, L., 2006. A virus associated with blueberry fruit drop disease, in: L. Lopes da Fonseca, F.R.M. (Ed.), Acta horticulturae (Paper presented at the Eighth International Symposium on Vaccinium Culture, held May 3-8, 2004, Sevilla, Spain and Oeiras, Portugal), p. 497.

Marzano, S.Y.L., and Domier, L. L., 2016. Novel mycoviruses discovered from metatranscriptomics survey of soybean phyllosphere phytobiomes. Virus Res. (in press).

doi:10.1016/j.virusres.2015.11.002,

Muhire, B.M., Varsani, A., Martin, D.P., 2014. SDT: a virus classification tool based on pairwise sequence alignment and identity calculation. PLoS ONE 9, e108277. doi: 10.1371/journal.pone.0108277.

Pearson, M.N., Beever, R.E., Boine, B., Arthur, K., 2009. Mycoviruses of filamentous fungi and their relevance to plant pathology. Mol. Plant Pathol. 10, 115–128.

Potgieter, A.C., Page, N.A., Liebenberg, J., Wright, I.M., Landt, O., van Dijk, A.A., 2009. Improved strategies for sequence-independent amplification and sequencing of viral double-stranded RNA genomes. J. Gen. Virol. 90, 1423–1432.

Roossinck, M.J., 2012. Plant virus metagenomics: Biodiversity and ecology. Annu. Rev. Genet. 46, 359–369.

Roossinck, M.J., 2015. Plants, viruses and the environment: Ecology and mutualism. Virology 479, 271–277.

Routhier, E., Bruenn, J.A., 1998. Functions of conserved motifs in the RNA-dependent RNA polymerase of a yeast double-stranded RNA virus. J. Virol. 72, 4427–4429.

Spanu, P.D., Abbott, J.C., Amselem, J., Burgis, T.A., Soanes, D.M., Stuber, K., van Themaat, E.V.L., Brown, J.K.M., Butcher, S.A., Gurr, S.J., Lebrun, M.H., Ridout, C.J., Schulze-Lefert, P., Talbot, N.J., Ahmadinejad, N., Ametz, C., Barton, G.R., Benjdia, M., Bidzinski, P., Bindschedler, L.V., Both, M., Brewer, M.T., Cadle-Davidson, L., Cadle-Davidson, M.M., Collemare, J., Cramer, R., Frenkel, O., Godfrey, D., Harriman, J., Hoede, C., King, B.C., Klages, S., Kleemann, J., Knoll, D., Koti, P.S., Kreplak, J., Lopez-Ruiz, F.J., Lu, X.L., Maekawa, T., Mahanil, S., Micali, C., Milgroom, M.G., Montana, G., Noir, S., O'Connell, R.J., Oberhaensli, S., Parlange, F., Pedersen, C., Quesneville, H., Reinhardt, R., Rott, M., Sacristan, S., Schmidt, S.M., Schon, M., Skamnioti, P., Sommer, H., Stephens, A., Takahara, H., Thordal-Christensen, H., Vigouroux, M., Wessling, R., Wicker, T., Panstruga, R., 2010. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. Science 330, 1543–1546.

Sperschneider, J., Datta, A., 2010. DotKnot: pseudoknot prediction using the probability dot plot under a refined energy model. Nucleic Acids Res. 38, e103. doi: 10.1093/nar/gkq021

Stielow, B., Menzel, W., 2010. Complete nucleotide sequence of TaV1, a novel totivirus isolated from a black truffle ascocarp (*Tuber aestivum* Vittad.). Arch. Virol, 155, 2075–2078.

Sun, L.Y., Suzuki, N., 2008. Intragenic rearrangements of a mycoreovirus induced by the multifunctional protein p29 encoded by the prototypic hypovirus CHV1-EP713. RNA 14,

2557–2571.

Takamatsu, S., Ito, H., Shiroya, Y., Kiss, L., Heluta, V., 2015. First comprehensive phylogenetic analysis of the genus *Erysiphe* (Erysiphales, Erysiphaceae) I. The Microsphaera lineage. Mycologia 107, 475–489.

Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577.

Taylor, D.J., Ballinger, M.J., Bowman, S.M., Bruenn, J.A., 2013. Virus-host co-evolution under a modified nuclear genetic code. PeerJ 1, e50. https://dx.doi.org/10.7717/peerj.50.

Taylor, D.J., Bruenn, J., 2009. The evolution of novel fungal genes from non-retroviral RNA viruses. BMC Biol. 7, 88. doi:10.1186/1741-7007-7-88.

Wicker, T., Oberhaensli, S., Parlange, F., Buchmann, J.P., Shatalina, M., Roffler, S., Ben-David, R., Dolezel, J., Simkova, H., Schulze-Lefert, P., Spanu, P.D., Bruggmann, R., Amselem, J., Quesneville, H., van Themaat, E.V.L., Paape, T., Shimizu, K.K., Keller, B., 2013. The wheat powdery mildew genome shows the unique evolution of an obligate biotroph. Nature Genet. 45, 1092–1096.

Wickner, R.B., Fujimura, T., Esteban, R., 2013. Viruses and prions of *Saccharomyces cerevisiae*. Adv. Virus Res. 86, 1–36.

Wickner, R.B., Ghabrial, S.A., Nibert, M.L., Patterson, J.L., Wang, C.C., 2011. Family *Totiviridae*, in: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowits, E.J. (Eds.), Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, Tokyo, pp. 639–650.

Wu, M.D., Jin, F.Y., Zhang, J., Yang, L., Jiang, D.H., Li, G.Q., 2012. Characterization of a novel bipartite double-stranded RNA mycovirus conferring hypovirulence in the phytopathogenic fungus *Botrytis porri*. J. Virol. 86, 6605–6619.

Xie, J.T., Jiang, D.H., 2014. New insights into mycoviruses and exploration for the biological control of crop fungal diseases. Annu. Rev. Phytopathol. 52, 45–68.

Zhai, Y., Attoui, H., Jaafar, F.M., Wang, H.-q., Cao, Y.-x., Fan, S.-p., Sun, Y.-x., Liu, L.-d., Mertens, P.P.C., Meng, W.-s., Wang, D., Liang, G., 2010. Isolation and full-length sequence analysis of *Armigeres subalbatus* totivirus, the first totivirus isolate from Mosquitoes representing a proposed novel genus (*Artivirus*) of the family *Totiviridae*. J. Gen. Virol. 91, 2836–2845.

Figures

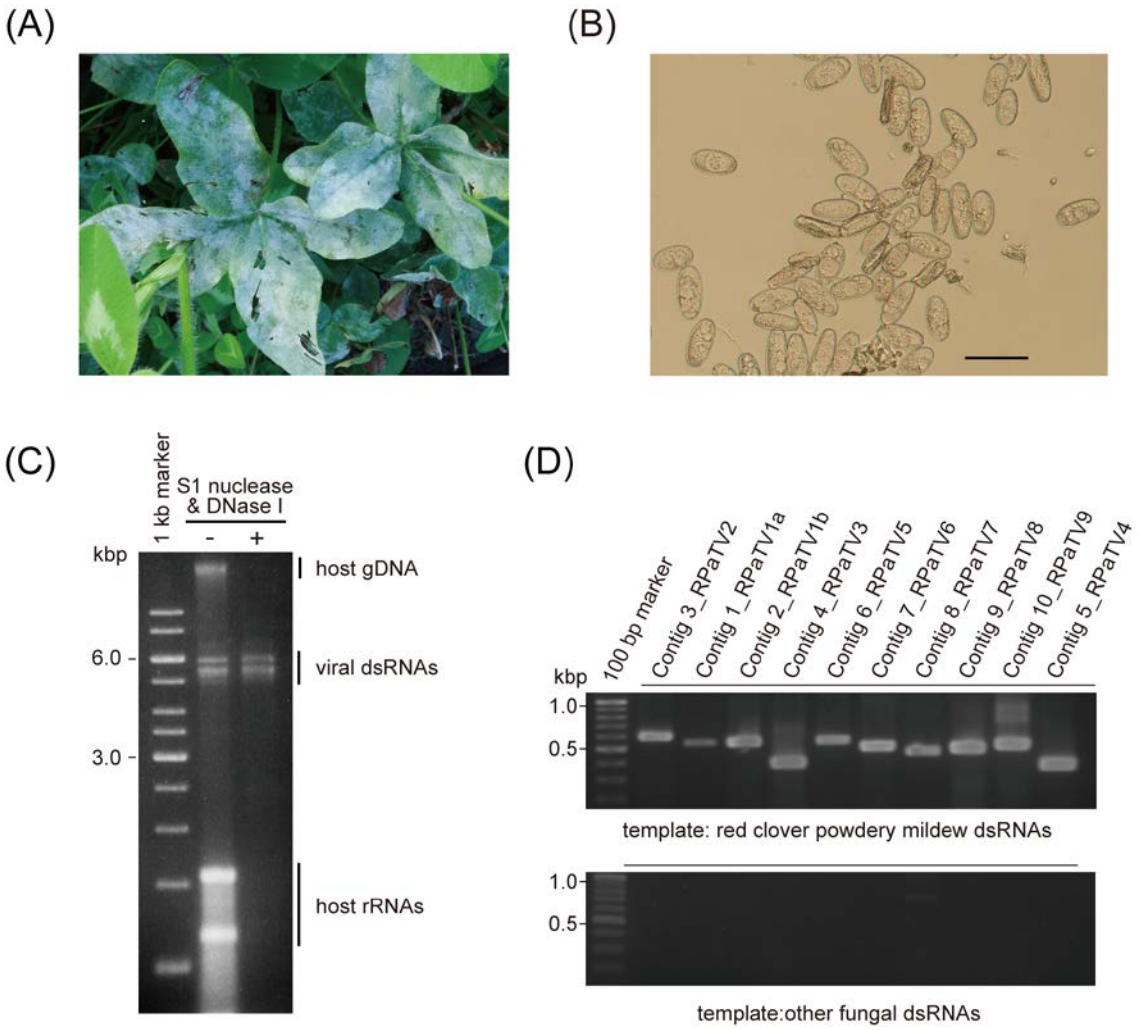

Fig. 1. The presence of totivirus-like sequences in powdery mildew fungus (*Erysiphe trifoliorum*). (A) Symptoms of powdery mildew disease on red clover leaves. (B) The conidia and/or conidiophores of *E. trifoliorum* collected from red clover leaves. Bar represents 50 µm. (C) Agarose gel electrophoresis of total nucleic acids extracted from the conidia of *E. trifoliorum*. Nuclease-treated and -untreated samples were analyzed in parallel. The nucleic acid types are denoted at the right. (D) RT-PCR detection of the totivirus-like sequences (referred to

as red clover powdery mildew-associated totiviruses, RPaTVs) (see Table 1) in the dsRNA samples derived from red clover powdery mildew (top panel) and other fungal materials (bottom panel). DNA was stained with ethidium bromide.

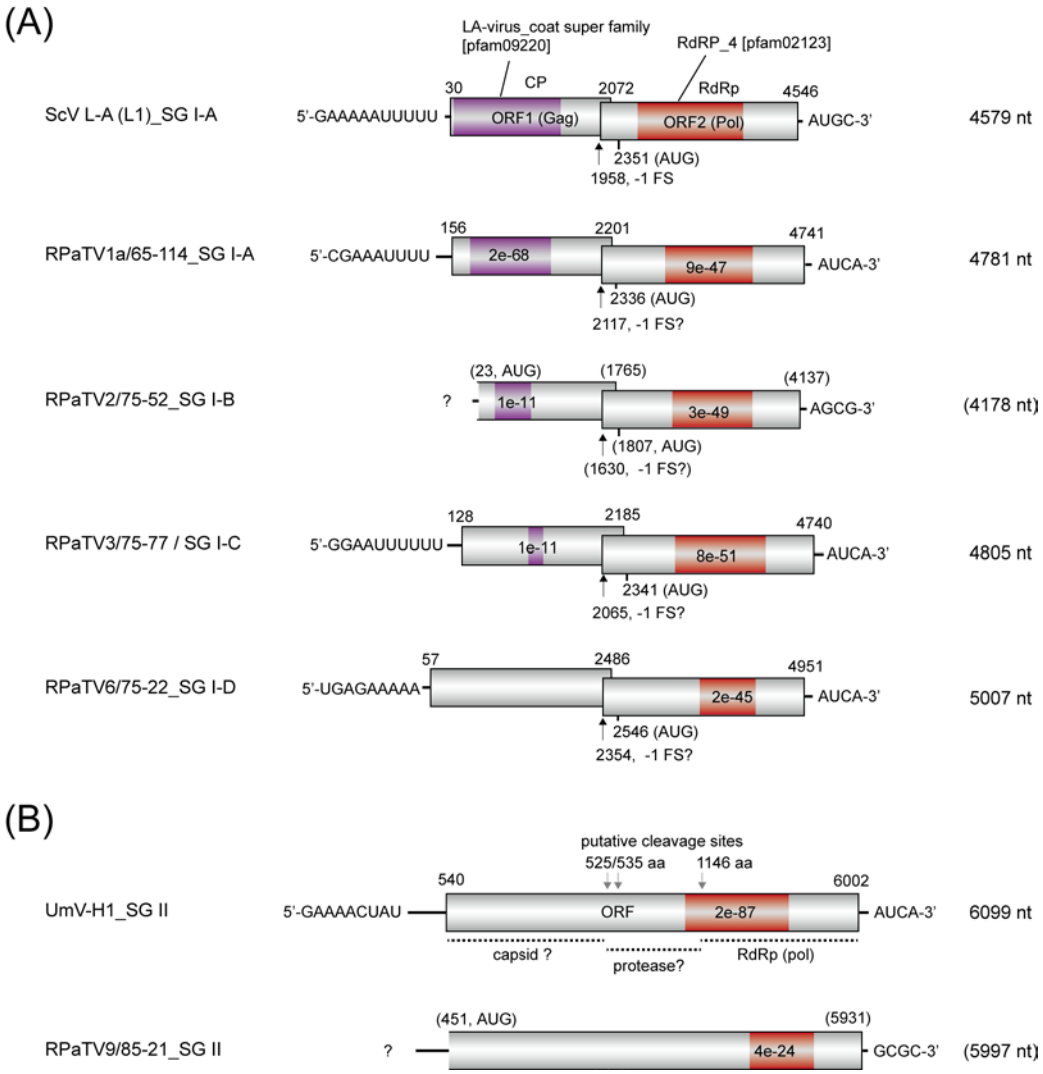

Fig. 2. Possible genomic organization of totivirus-like sequences. (A) The coding strand of the ScV-L-A genome and representative RPaTVs (RPaTV 1a, 2, 3 and 6, see other RPaTVs in Fig. S3) sequences. The sequences have two adjacent or overlapping open reading frames (ORFs) that encode major capsid protein (CP, Gag) and RNA dependent RNA polymerase (RdRp, Pol). (B) The genomic structures of UmV-H1 and its related RPaTV (RPaTV9) sequence. The sequences have a single large ORF that can encode a putative polyprotein. Predicted proteolytic processing sites for UmV-H1 and their putative products are indicated with arrows. Nucleotide

positions of ORFs and a putative heptameric slippery site for –1 frameshift (–1FS) are indicated. The dark colored boxes represent the totiviral conserved domains (pfam09220: LA-virus coat super family, pfam02123: RdRp4) with expected (E) values. The contig length and subgroup positions (SG-I-A–D, and SG-II) of each RPaTV sequence in the CP-RdRp-based phylogenetic tree (see Fig. 4) are indicated.

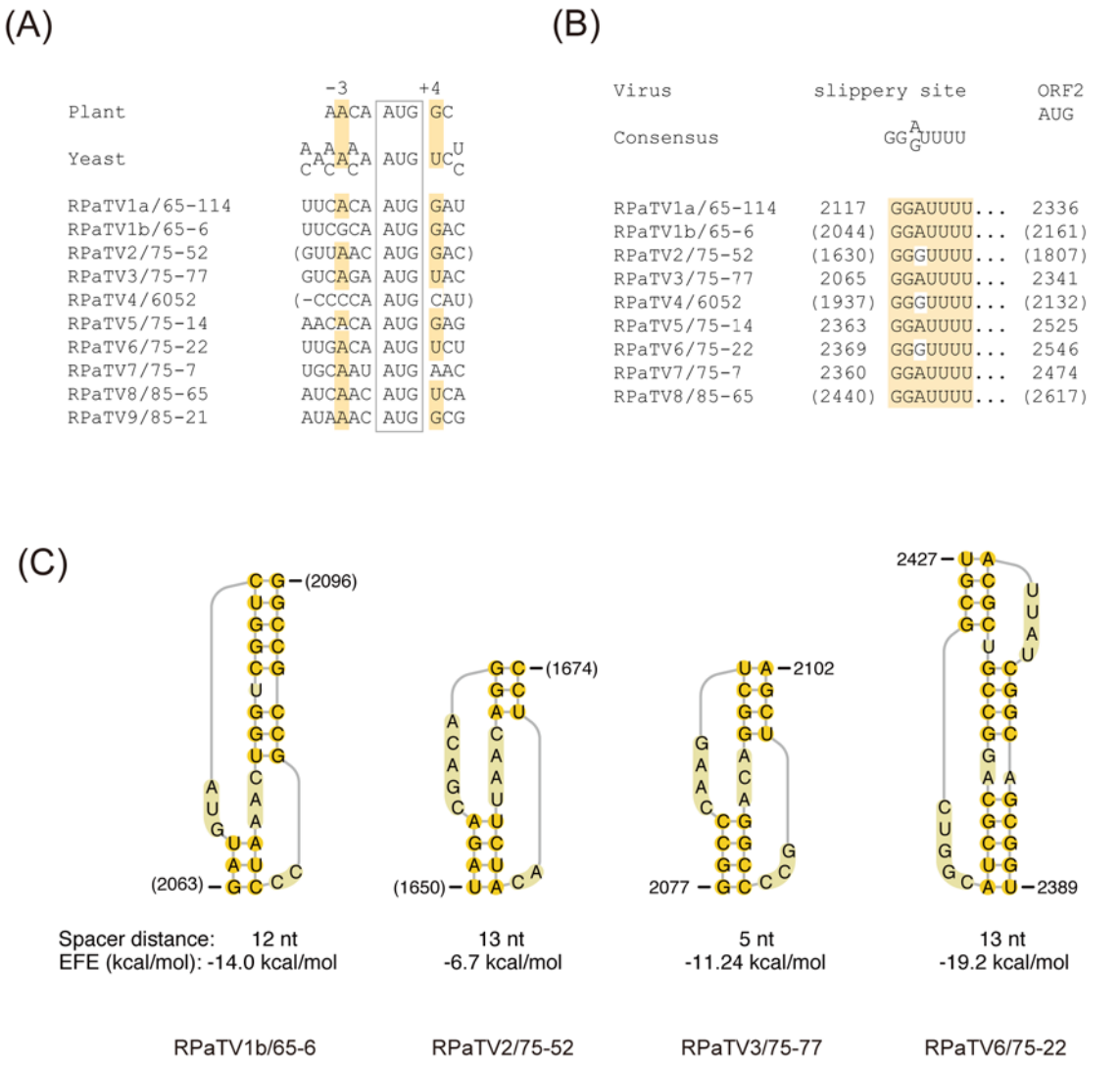

Fig. 3. Conserved sequences and predicted secondary structures found in totivirus-like sequences. (A) The sequence surrounding the predicted start codons (AUG) of ORF1 of totivirus-like sequences (RPaTVs). The most frequent nucleotide sequences at positions –3 and +4 relative to the AUG start codon (open box, +1~+3) in plants and yeasts are shown at the top. (B) The putative heptanucleotide (X XXY YYZ) slippery sequences for –1 translational

frameshift. (C) The predicted pseudoknot structures downstream of the potential slippery sequences for –1 frameshift. Spacer distance stands for the number of nucleotides (nt) between the slip site and the pseudoknot. EFE, estimated free energy (kcal/mol). Numbers (or numbers in parentheses) indicate nucleotide positions in the contigs.

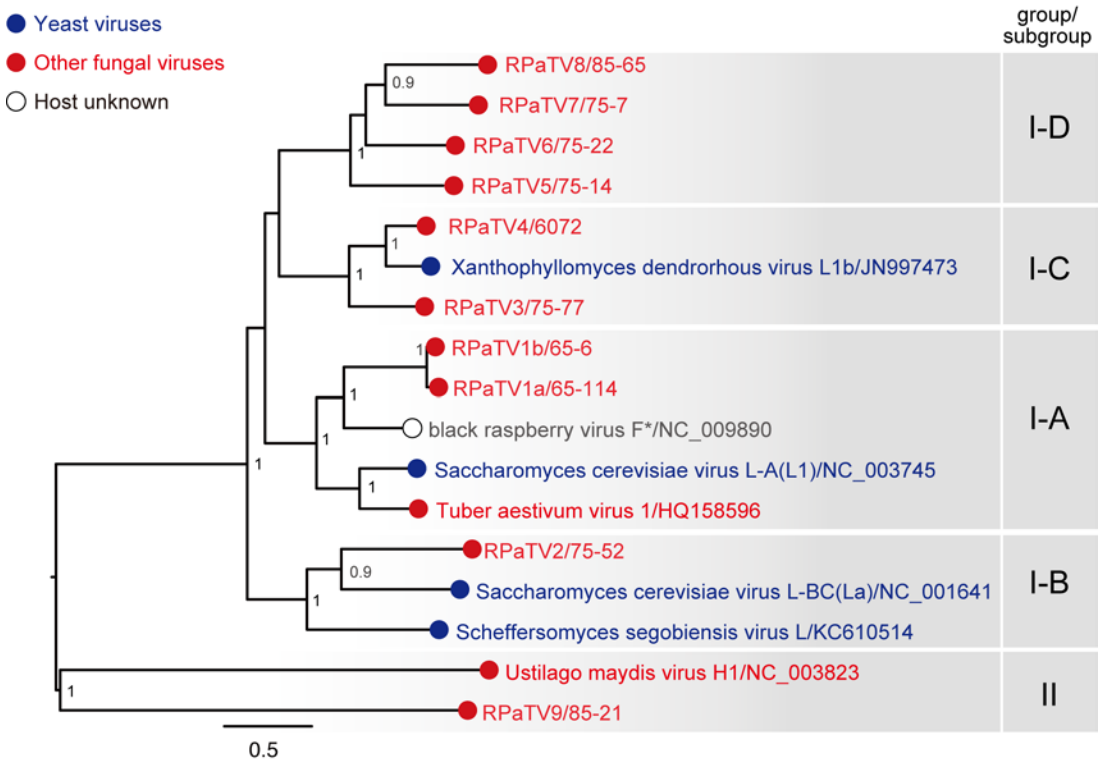

Fig. 4. Phylogenetic analysis of totivirus-like sequences. Maximum likelihood tree was calculated from the amino acid sequences of the CP-RdRp (Gag-Pol-like) fusion protein or polyproteins of totivirus-like sequences (RPaTVs) and known totiviruses. The phylogenetic tree was constructed using PhyML 3.0 based on the multiple amino acid sequence alignment (the cured regions are shown in Fig. S4) generated using MAFFT (ver 7.0). A model RtREV with + I + G + F was selected as the best fit model. Numbers at the nodes indicate aLRT values determined using an SH-like calculation (only values greater than 0.9 are shown). The scale bar represents amino acid distances. Virus names are followed by GenBank accession numbers. Two groups (I and II) and four subgroups (I-A to I-D) are defined according to the tree topology (clusters).

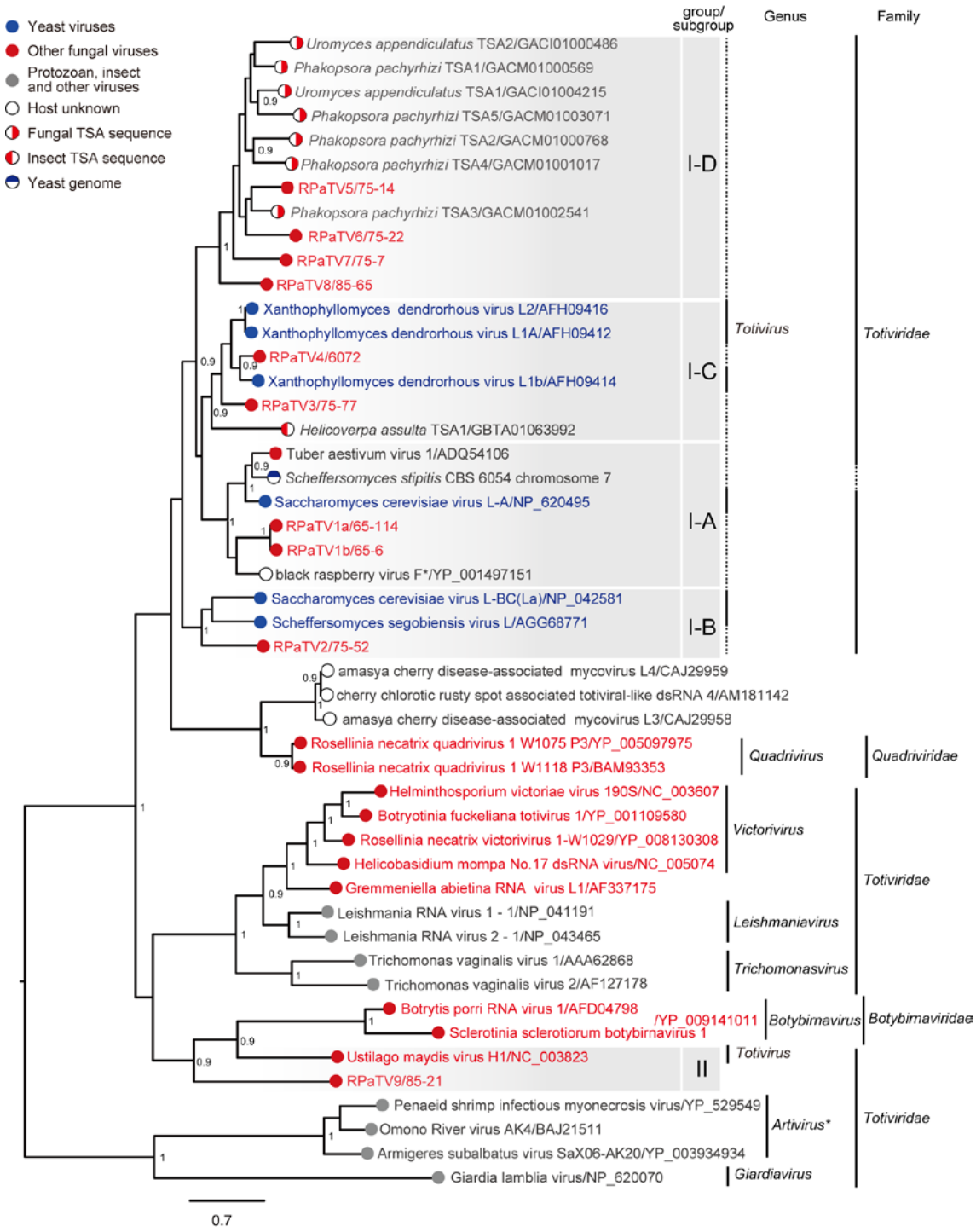

Fig. 5. Phylogenetic analysis of totivirus-like sequences. Maximum likelihood tree was calculated from the amino acid sequences of the RdRp (pol) protein of totivirus-like sequences (RPaTVs). Sequences of extant and probable members of the family *Totiviridae*, those related

viruses and virus-like elements were analyzed together with RPaTVs. A model LG with + I + G was selected as the best fit model. Numbers at the nodes represent aLRT values using an SH-like calculation (only values greater than 0.9 are shown). The scale bar represents the amino acid distances. An yeast endogenous totivirus-like sequence is also included. Totivirus-like sequences from the TSA databases of fungus and insects are listed in Table 3.

Supplementary Figure legends

Supplementary Fig. S1. Detection of dsRNA from fungal materials. (A) Agarose gel electrophoresis of total nucleic acids isolated from the conidia of red clover (*E. trifoliorum*) and pea (*E. pisi*) powdery mildew fungi. Two lanes for each species represent 1/10 dilution (left) and direct (right) loading of nucleic acid samples. (B) Enlarged image of the Fig. 1C. Multiple dsRNA bands are detected raging from 5.0 to 7.0 kbp. (C) Agarose gel electrophoresis of the dsRNA-enriched fraction isolated from the conidia of red clover powdery mildew fungus. Agarose gel was stained with ethidium bromide. (D) Agarose gel electrophoresis of RLM-RACE fragments corresponding to the 3'-terminal region of both strands of totivirus-like dsRNA. Primer sequences including an adaptor (PC3-T7 loop) (Potgieter et al., 2009) are shown in Table S1. Obtained fragments were subjected for sequencing. For the coding-strand of contig 6 (RPaTV5), an alternative adaptor (3RACE-adaptor, Table S1) (Chiba et al., 2009) was used for RLM-RACE (data not shown).

Supplementary Fig. S2. (A–J) Mappings of sequence reads to each totivirus-like sequence (RPaTV). The reads are color-differentiated by red (positive strand) and green (negative strand) horizontal lines. Unrepresented reads are shown by the graph with the scale indicating the numbers of reads. Locations of single-nucleotide polymorphisms (SNPs) within each RPaTVs sequences are indicated with vertical lines.

Supplementary Fig. S3. Possible genomic organization of totivirus-like sequences (RPaTVs). The coding strand of RPaTVs (RPaTVs 1b, 4, 5, 7 and 8) have two adjacent or overlapping

open reading frames (ORFs) that encode major capsid protein (CP, Gag) and RNA dependent RNA polymerase (RdRp, Pol). Nucleotide positions of ORFs and a putative heptameric slippery site for –1 frameshift (–1FS) are indicated. The dark colored boxes represent the totiviral conserved domains (pfam09220: LA-virus coat super family, pfam02123: RdRp4) with expected E-values. The contig length and subgroup positions (SG-I-A–D) of each RPaTV sequence in the CP-RdRP-based phylogenetic tree (see Fig. 4) are indicated.

Supplementary Fig. S4. Multiple alignment of amino-acid sequence of the gag-pol-like (CP-RdRp) fusion proteins encoded by totivirus-like contigs (RPaTVs) and known totiviruses. Gblocks 0.91b was used to identify ambiguous aligned regions. The residues in ScV-L-A CP that are important for the cap-snatching mechanism (Tyr150, Asp152, His154, Tyr452, Tyr538) are denoted by arrows. Eight RdRp motifs (I to VIII) are shown.

Supplementary Fig. S5. Pairwise comparison of ORF1 proteins (CPs, A) and ORF2 proteins (RdRps, B) encoded by RPaTVs. Each number represents the pairwise amino acid identities (%) between RPaTVs calculated using SDT version 1.2.

Supplementary Fig. S6. Phylogenetic analysis of totivirus-like sequences. Maximum likelihood tree was calculated from the amino acid sequences of the CP (gag) protein of totivirus-like sequences (RPaTVs), yeast and other related totiviruses and virus-like sequences. A model RtREV with + G + F was selected as the best fit model. Numbers at the nodes represent aLRT values using an SH-like calculation (only values greater than 0.9 are shown).

The scale bar represents the amino acid distances. Totivirus-like sequences from the TSA database of fungus, plants and insects are listed in Table 3. Several endogenus totivirus-like sequences from yeast and a seaweed (*Chondrus crispus*) genomes are also included.