Rapid Communication

Evidence for negative-strand RNA virus infection in fungi

Hideki Kondo a,∗, Sotaro Chiba a, Kazuhiro Toyoda b, 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

A R T I C L E   I N F O

Article history:
Received 11 August 2012
Returned to author for revisions 27 September 2012
Accepted 2 October 2012
Available online 23 October 2012

Keywords:
Endogenous viral element
Transcriptome shotgun assembly
Negative-strand RNA virus
Mononegavirales
Fungi
Erysiphe pisi
Sclerotinia homoeocarpa
Evolution

A B S T R A C T

Fungal viruses comprise two groups: a major group with double-stranded RNA genomes and a minor group with positive-sense single-stranded (ss)RNA genomes. Although many fungal viruses have been identified, no negative-stranded (−)ssRNA mycoviruses have been reported. Here we present two lines of evidence suggesting the presence of (−)ssRNA viruses in filamentous fungi based on an exhaustive search using extant (−)ssRNA viruses as queries. This revealed (−)ssRNA virus L protein-like sequences in the genome of a phytopathogenic ascomycete, Erysiphe pisi. A similar search for (−)ssRNA viruses in fungal transcriptome shotgun assembly libraries demonstrated that two independent libraries from Sclerotinia homoeocarpa, another phytopathogenic ascomycete, contained several sequences considered to correspond to the entire mononegavirus L gene and likely originating from an infecting (−)ssRNA virus. These results provide strong evidence for both ancient and extant (−)ssRNA virus infections in fungi.

© 2012 Elsevier Inc. All rights reserved.

Introduction

Fungal viruses (mycoviruses) are widespread in all major groups of fungi. Extensive searches for mycoviruses in endophytic, human-pathogenic, entomopathogenic and phytopathogenic fungi have revealed varying incidence rates. For example, approximately 21% of over 1000 field isolates of a soil-borne ascomycete pathogen, Rosellinia necatrix, were found to be infected with viruses (Arakawa et al., 2002; Ikeda et al., 2004) while approximately 68% of isolates of another basidiomycete phytopathogen, Helicobasidium mompa, carried viruses (Ikeda et al., 2004). As described by Ghabrial and Suzuki (2009), most fungal viruses have double-stranded (ds)RNA genomes, although some have single-stranded (ss)RNA genomes. This may be attributable to the method employed for dsRNA-based virus screening. Yu et al. (2010) recently reported an ssDNA virus infecting a plant pathogenic fungus, Sclerotinia sclerotiorum. According to the 9th report of the International Committee on Taxonomy of Viruses (ICTV), mycoviruses are now classified into five dsRNA families, five ssRNA families, two reverse-transcribing (RT) families, unassigned ssDNA and dsDNA viruses (King et al., 2011). It has long been unclear whether mycoviruses with negative-stranded (−)ssRNA (i.e., complementary to the mRNA) genomes exist in fungi, as in other eukaryotic organisms (Dolja and Koonin, 2011).

Nucleotide sequence databases for viral and cellular genomes and transcriptomes have been growing rapidly. Database searches have led to the discovery of non-retrovirus RNA viral sequences (NRVSs) integrated into the genomes of diverse eukaryotic organisms (Belyi et al., 2010; Chiba et al., 2011; Fort et al., 2012; Horie et al., 2010; Katzourakis and Gifford, 2010; Liu et al., 2010; Taylor et al., 2010, 2011), a phenomenon that was thought previously not to occur. Viruses integrated into host chromosomes include partitiviruses (members of the family Partitiviridae) with dsRNA genomes (plant), mononegaviruses (members of the order Mononegavirales) with non-segmented (−)ssRNA genomes, such as rhabdoviruses (plant, invertebrate, and vertebrate), bornaviruses (vertebrate), or filoviruses (vertebrate). In comparison with dsRNA and (−)ssRNA viruses, there are fewer examples of integration of positive-sense ssRNA viruses (plant and invertebrate) (Chiba et al., 2011; Crochu et al., 2004; Cui and Holmes, 2012). In fungi, Taylor and Bruenn (2009) and Liu et al. (2010) found totivirus (dsRNA virus)-like sequences in several fungal nuclear genomes, while Liu et al. (2011) reported the presence of ssDNA virus sequences in fungal chromosomes. A series of these studies has provided interesting insights into molecular evolution, virus/host interactions, and ancient virus genetic structures. It remains unknown how such sequences were endogenized or what their biological significance may be.
Negative-strand RNA viral sequences can be detected in all the major eukaryotic kingdoms except for fungi (Belyi et al., 2010; Chiba et al., 2011; Fort et al., 2012; Horie et al., 2010; Katzourakis and Gifford, 2010; Taylor et al., 2010). This fact prompted us to search available sequence databases for (−)ssRNA viruses in fungi. In the present study, we performed an extensive search of publicly available fungal genome and transcriptome shotgun assembly sequence databases, and found evidence strongly suggesting the presence of both ancient (−)ssRNA viruses and probably extant (−)ssRNA viruses in phytopathogenic filamentous fungi.

Results and discussion

Genome sequence database search

Because many mononegavirus-related sequences are present in plant and animal chromosomes, we searched the available NCBI databases using L-polymerase (RNA-dependent RNA polymerase, RdRp) sequences encoded by mononegaviruses or related viruses of different genera as queries (Table S1). We identified nine significant matches to (−)ssRNA viruses in the whole-genome shotgun (WGS) assemblies of the chromosomes of a plant pathogenic obligate ascomycete, *Erysiphe pisi* (pea powdery mildew fungus) (Table S2). These WGS sequences assembled into four independent contigs. According to our system of nomenclature for integrated NRVSs (Chiba et al., 2011), these sequences were termed *E. pisi* mono-
egavirus L protein-like sequences I–4 (EpMLLS1–EpMLLS4) (Fig. S4). A further BLAST search against the EMBL (WGS) database (http://www.ebi.ac.uk/Tools/sss/fasta/wgs.html) using mononegaviral L-proteins or EpMLLSs as queries resulted in identification of eight other *E. pisi* WGS sequences of candidate EpMLLSs (Table S2). These WGS sequences assembled into another four contigs (EpMLLS5–EpMLLS8), as shown in Fig. 1. Reverse BLAST analysis revealed that EpMLLSs were distantly related to members of a recently proposed novel genus Nyavirus (Midway and Nyamanini viruses) (Mihindukulasuriya et al., 2009), and several other mononegaviruses (Table S2 and data not shown).

EpMLLSs possess truncated L-like genes and some of them have ORF disruptions (Fig. 1). Their potentially encoded proteins differ from one another in the corresponding region of nyavirus L (Fig. 1) and show significant levels of amino acid sequence identity (20–80%) (Table S3). EpMLLS4 contains a possible rearranged virus sequence, and EpMLLS5 has a large internal deletion (Fig. 1). Some EpMLLSs have a domain found in RdRp (Pfam00946) of species belonging to the order Mononegavirales (Poch et al., 1990) (Table S2). Fig. S1 shows the multiple alignment of the hypothetical RdRp core module of EpMLLS2, EpMLLS3 and other mononegavirus L protein-like sequences (see below). The EpMLLS2 and EpMLLS3 deduced amino acid sequences contain the highly conserved motifs A to D and a typical catalytic motif GDNQ within C regions that are found in almost all mononegavirus L proteins, with the exception that the first residue in EpMLLS2 is Ser (S) instead of Gly (G) (Fig. S1).

EpMLLSs have distinguishable flanking sequences and some of them contain trace putative fungal transposable elements (Fig. 1). Together with the fact that the *E. pisi* shotgun library is derived from a single fungal strain, these results suggest that EpMLLS1–EpMLLS8 reside at different fungal genome positions. Notably, the flanking ORFs of EpMLLS6 show significant similarity to a TE1a (a class of LINE-1 retrotransposons, CgT1-like) retrotransposon in *Blumeria graminis* (a powdery mildew fungus in barley and wheat) and a powdery mildew–specific avirulence (AVR) effector gene, AVRk1. TE1a and AVRk1 are believed to be expressed as a single transcript and to be retrotransposed as a single unit (Sacristán et al., 2009). Therefore, we speculated that EpMLLS6 might have been integrated after template switching from this unique transcript to mRNA of (−)ssRNA virus (progenitor virus of EpMLLS6) during reverse transcription of the retrotransposon (Ballinger et al., 2012; Geuking et al., 2009).

Conversely, although the most numerous mononegavirus-like sequences in plant and animal genomes were related to the nucleoprotein (or nucleocapsid protein) N (*Belyi et al., 2010; Chiba et al., 2011; Fort et al., 2012; Horie et al., 2010; Katzourakis and Gifford, 2010; Taylor et al., 2010, 2011), we found no hit for N proteins and other viral proteins such as glycoproteins (G) in the fungal genome (data not shown), perhaps as a result of their low or moderate levels of conservation among mononegaviruses.

During the course of these analyses we noticed that WGS assembly sequences of another powdery mildew fungus, *Golovinomyces orontii* (formerly *Erysiphe orontii*, order Erysiphales), are available from the Max Planck Institute for Plant Breeding Research Powdery Mildew Genome Project site (http://www.mpiz-koeln.mpg.de/24322/Project_Description) (Spanu et al., 2010), but not from the NCBI website. Further BLAST analysis was therefore conducted against the *G. orontii* WGS library using EpMLLSs as queries. Consequently, in our preliminary database search, we found several significant matches, although most were short (data not shown). This finding strongly suggests that endogenization of mononegaviruses into the fungal genome might have occurred not only in the pea powdery mildew fungus (*E. pisi*), but also in another powdery mildew fungus (*G. orontii*), which has a broad host range including *Arabidopsis thaliana* (Spanu, 2012).

Detection of EpMLLSs by genomic PCR

In order to confirm the presence of EpMLLSs in fungal chromosomes, we harvested two samples of powdery mildew fungus from *P. sativum* L. and red clover (*Trifolium aestivum* L.) plants in the experimental field at the Institute of Plant Science and Resources (IPSR). The fungal genomic DNA was extracted and subjected to genomic PCR. Based on internal transcribed spacer (ITS) sequencing, the pea plants appeared to harbor a double infection with *E. pisi* and *E. trifolii*, whereas the red clover plants were infected only with *E. trifolii*. A total of 24 clones from pea samples were sequenced. While three clones showed over 99% sequence identity to the ITS sequences reported for *E. pisi*, the remaining clones were identical to those reported for *E. trifolii* (Fig. S2). The ITS sequence obtained by direct sequencing from red clover samples was identical to the reported ITS sequence for *E. trifolii* isolates (Fig. S2). Co-infection with two powdery mildew species on the same host plant had been reported previously (Clawe, 2008).

Genomic PCR using a different set of primers corresponding to each of the EpMLLS sequences (see Table S4 for the primer sequences) revealed that DNA fragments of the expected sizes were amplified on fungal genomic DNA from the pea sample (*P. sativum L.*), but not on those from the red clover sample (*E. trifolii*) (Fig. 2). No DNA fragment was amplified by PCR on pea or red clover genomic DNA (Fig. 2 and data not shown). The sequences of the resulting DNA fragments from the pea sample were sequenced. While three clones showed over 99% sequence identity to the ITS sequences reported for *E. pisi*, the remaining clones were identical to those reported for *E. trifolii* (Fig. S2). The ITS sequence obtained by direct sequencing from red clover samples was identical to the reported ITS sequence for *E. trifolii* isolates (Fig. S2). Co-infection with two powdery mildew species on the same host plant had been reported previously (Clawe, 2008).

Genomic PCR using a different set of primers corresponding to each of the EpMLLS sequences (see Table S4 for the primer sequences) revealed that DNA fragments of the expected sizes were amplified on fungal genomic DNA from the pea sample (*P. sativum L.*), but not on those from the red clover sample (*E. trifolii*) (Fig. 2). No DNA fragment was amplified by PCR on pea or red clover genomic DNA (Fig. 2 and data not shown). The sequences of the resulting DNA fragments from the pea sample were sequenced. While three clones showed over 99% sequence identity to the ITS sequences reported for *E. pisi*, the remaining clones were identical to those reported for *E. trifolii* (Fig. S2). The ITS sequence obtained by direct sequencing from red clover samples was identical to the reported ITS sequence for *E. trifolii* isolates (Fig. S2). Co-infection with two powdery mildew species on the same host plant had been reported previously (Clawe, 2008).

Genomic PCR using a different set of primers corresponding to each of the EpMLLS sequences (see Table S4 for the primer sequences) revealed that DNA fragments of the expected sizes were amplified on fungal genomic DNA from the pea sample (*P. sativum L.*), but not on those from the red clover sample (*E. trifolii*) (Fig. 2). No DNA fragment was amplified by PCR on pea or red clover genomic DNA (Fig. 2 and data not shown). The sequences of the resulting DNA fragments from the pea sample were sequenced. While three clones showed over 99% sequence identity to the ITS sequences reported for *E. pisi*, the remaining clones were identical to those reported for *E. trifolii* (Fig. S2). The ITS sequence obtained by direct sequencing from red clover samples was identical to the reported ITS sequence for *E. trifolii* isolates (Fig. S2). Co-infection with two powdery mildew species on the same host plant had been reported previously (Clawe, 2008).
Fig. 1. Schematic representation of *Erysiphe pisi* mononegavirus L protein-like sequences (EpMLLSs) and their flanking regions. EpMLLSs found in the whole-genome shotgun (WGS) database of *E. pisi* are shown to match L-polymerase (RNA-dependent RNA polymerase, RdRp; pfam00946) from an insect mononegavirus, Nyamanini virus (a prototypic member of Nyavirus). Top row shows a diagrammatic representation of the 3′-to-5′ arrangement of the ORFs (I, nucleoprotein N; II, putative phosphoprotein; III, unknown; IV, putative matrix protein; V, glycoprotein G; VI, large protein L or RNA-dependent RNA polymerase) of Nyamanini virus (virus complementary strand) (Mihindukulasuriya et al., 2009). Other conserved domains in the L-protein (mRNA-capping enzyme motif: pfam14318 and TIGR04198) found on the NCBI website (Marchler-Bauer et al., 2011) are also shown. The potential coding regions of EpMLLSs and flanking small ORFs are shown as boxes, and EpMLLSs are placed in the corresponding regions of the Nyamanini virus-L ORF. Retrotransposon-like sequences are shown by thick black lines. The positions of the primers used for genomic PCR (Table S4 and Fig. 2) are shown by arrows. The WGS assembled sequences and undetermined sequences are shown by solid and dashed lines, respectively. Symbols referring to mutations are also shown; kinked line, major deletion; asterisk, internal stop codon; F, frame-shift.
representative isolates: FJ378869, FJ378870 and FJ378872) and *E. trifolii* (representative isolates: AB015913, FJ378884 and FJ378878), calculated on the basis of the Kimura 2-parameter criterion using MEGA (Kimura, 1980), was 0.026 for the ITS region. Based on the molecular clock of the ITS (2.52 × 10⁻⁹ substitutions per site per year) (Takamatsu and Matsuda, 2004), the split between *E. pisi* and *E. trifolii* was roughly estimated to have occurred 5.2 million years ago (Fig. 2). Therefore, EpMLLSs integration may have occurred more recently than this. Further studies will be needed to examine this possibility. It was also suggested that NRVSs in fungal genomes can serve as a genetic marker for clarifying phylogenetic relationships that are otherwise unclear.

The discovery of these EpMLLSs provides the first example of (-)ssRNA virus sequences detected in the genomes of fungi. These sequences might have integrated into the genome of *E. pisi* following infection by ancestral fungal (-)ssRNA viruses related to possible extant fungal (-)ssRNA viruses (see below). Nevertheless, we cannot rule out the possibility that intimate interactions between obligate pathogens and plant hosts may have allowed integration of ancestral plant (-)ssRNA virus sequences into the fungal genome. Partitiviruses (dsRNA viruses) are assumed to be transferred laterally between fungi and plants, and these virus sequences might have integrated relatively frequently into plant host genomes as described above (Chiba et al., 2011; Liu et al., 2010).

Transcriptome shotgun library search

Several studies have demonstrated that pyrosequencing-based transcriptome analysis is a useful new research tool for discovery of (-)ssRNA viruses (Bekal et al., 2011; Honkavuori et al., 2008; Valles et al., 2012). Therefore, we further expanded our search to fungal transcriptome shotgun assembly (TSA) libraries in a search for unknown (-)ssRNA viruses. Interestingly, four TSA sequences showing moderate similarity to (-)ssRNA virus L protein sequences were found in the two independent libraries from *Sclerotinia homoeocarpa*, an ascomycete fungus causing dollar spot in grass plants (Fig. 3a and Table S5). Two TSAs, JU091017 and JU091016, originated from the same library (Orshinsky et al., 2012) and were almost identical to each other, apart from a short terminal heterologous sequence. Therefore, we termed their contig ShTSA1 (*S. homoeocarpa* TSA contig 1). The second library of TSAs (Hulvey et al., 2012), JW826636 (ShTSA2) and JW828891 (ShTSA3), showed moderate or low identity (73% and 53% nucleotide identity, respectively) over the whole of the ShTSA1 sequences. Each ShTSA has a long ORF (L-like ORF) considered to correspond to the entire region of the nyavirus L protein coding region (Table S5). The ShTSA L-like ORF proteins (ShTSA1-L–ShTSA3-L) show significant levels of amino acid sequence identity among these sequences (56–86% amino acid identity) (Table S6). These potentially encoded proteins have two typical domains, RdRp (pfam00946) and virus-capping methyltransferase (pfam14318), found in most mononegaviruses (Fig. 3a and Table S5). ShTSA has another two or three short ORFs (ORF1–3) (Fig. 3a). While the products of these small ORFs did not have any detectable similarity to the NCBI protein database, there was remarkable conservation among ShTSA in each of these small ORF proteins (ORF1 70%, ORF2 42–87 % and ORF3 30% amino acid identity) (Table S6).

In addition, a semi-consensus sequence (3'-AAUAAAAUUUUUUGAAUCUC-5') was found in the putative untranslated sequences between the ORFs of each ShTSA sequence (Figs. 3 and S3). These are probably 'gene-junction' sequences, which are broadly conserved among mononegaviruses (Conzelmann, 1998). Mononegaviral gene-junction sequences contain a gene-end sequence, which is required for transcription termination and polyadenylation of upstream genes, and a gene-start sequence, which plays an important role in initiating the transcription of downstream genes (Conzelmann, 1998). The AU-rich region (3'-AAUAAAAUUUUUUGAAUCUC-5') is reminiscent of the gene-end sequences found in plant rhabdoviruses and some other mononegaviruses, comprising an AU-rich region and a poly(U) tract (Jackson et al., 2008) (Fig. 3b). This sequence was followed by a moderate-consensus sequence GAAUCCUC which may be included in the untranslated intergenic spacer sequences (the first nucleotide sequence is G in many rhabdoviruses) and the initiation of transcription of downstream genes (the first three nucleotide sequences are CUN or CUA in plant cytorhabdoviruses and varicosaviruses) (Heim et al., 2008; Sasaya et al., 2004) (Fig. 3b).

Each of these ShTSA most likely originates from multiple transcripts corresponding to genes of an infecting virus that is not of host chromosomal origin, for the reasons stated below. The assembled sequences from two independent TSA libraries resemble the 5'-terminal half of a (-)ssRNA virus genome encoding the entire L protein, and each ShTSA ORF possesses putative gene-junction sequences as described above. These characteristics are typical of a mononega-like virus genome structure. Furthermore, we found no hit for ShTSA-Ls in a BLAST search against the *S. homoeocarpa* WGS database (8-fold redundancy) at the EMBL site (Hulvey et al., 2012). If ShTSA1 to ShTSA3 were transcripts from the same locus in undetermined regions of the *S. homoeocarpa* genome, they would be expected to show closer similarity to each other. In addition, NRVSs in host genomes frequently harbor ORFs that have been interrupted due to spontaneous mutations during the course of evolution, as is the case for EpMLLSs (Fig. 1). However, all L-like ORFs in the ShTSA appear to have remained intact. Taken together, our findings provide strong evidence for extant (-)ssRNA virus infection in fungi. Interestingly, reverse BLAST analysis using ShTSA3-L yielded a significant match (GeneBank accession GR225954; e-value=3×10⁻¹², identities=83/342=24%) in the EST database (8-fold redundancy) at the EMBL site (Hulvey et al., 2012). If ShTSA1 to ShTSA3 were transcripts from the same locus in undetermined regions of the *S. homoeocarpa* genome, they would be expected to show closer similarity to each other. In addition, NRVSs in host genomes frequently harbor ORFs that have been interrupted due to spontaneous mutations during the course of evolution, as is the case for EpMLLSs (Fig. 1). However, all L-like ORFs in the ShTSA appear to have remained intact. Taken together, our findings provide strong evidence for extant (-)ssRNA virus infection in fungi. Interestingly, reverse BLAST analysis using ShTSA3-L yielded a significant match (GeneBank accession GR225954; e-value=3×10⁻¹², identities=83/342=24%) in the EST database.
library of a fungal pathogen of caterpillars (medicinal fungus), *Cordyceps militaris* (Xiong et al., 2010), which belongs to an order different from that of *S. homoeocarpa*. This may be another similar example of a filamentous fungus infected with a (−)ssRNA virus.

It is noteworthy that searches for other mononegaviral genes such as N, which presumably would be most abundant among all the gene products, yielded no hits in the *S. homoeocarpa* TSA libraries (data not shown). Although the complete genome structure of *S. homoeocarpa* (−)ssRNA viruses is still unknown, this result suggests an unusual particle structure that would make any morphology-based identification of a subset of fungal (−)ssRNA viruses more difficult. In this connection, it should be noted that many fungal positive-sense ssRNA and dsRNA viruses form no particles. The candidate ORF2 gene of ShTSA1, whose position corresponds to that of a glycoprotein G for mononegavirus, has no putative signal peptide or C-terminal hydrophobic transmembrane anchor domain, which are common structural features of mononegavirus G protein (data not shown). Similarly, in lettuce big-vein associated virus (LBVaV, floating genus *Varicosavirus*), which has a rod-shaped virion (not enveloped) containing a bipartite (−)ssRNA with a plant rhabdovirus-like genome structure and is transmitted by a soil-inhabiting fungus (*Olpidium virulentus*), the LBVaV protein 5, whose position also corresponds to G protein, appeared to lack any direct (sequence) relatedness to the rhabdoviral G protein (Kormelink et al., 2011; Sasaya et al., 2002, 2004). In general, the G protein forms the membrane spikes of mononegaviral virions and is thought to play a critical role in virus membrane fusion during cell entry (Albertini et al., 2012). However, as mycoviruses have no known extracellular mode of transmission in nature (Ghabrial and Suzuki, 2009), (−)ssRNA viruses adapted to fungal hosts may lack the G protein gene. It is also speculated that differences among other virally encoded proteins and/or genomic structure may represent another unique history of co-evolution between (−)ssRNA virus and the fungus (and/or plant).
Fig. 4. Maximum likelihood phylogeny of the L protein sequences of mononegaviruses, EpMLLSs and ShTSA-Ls. A phylogenetic tree was constructed using PhyML 3.0 based on the multiple amino acid sequence alignment of the RdRp polymerase core module shown in Fig. S1. Mononegavirus L protein-like sequences from fungi, plants, insects, mammals and fishes (Tables S2, S7 and S8), and fungal TSA-derived sequences (Table S5) are shown in red, green, wine red, ochre, gray and blue boxes, respectively. L proteins from mononegaviruses (filoviruses, paramyxoviruses, rhabdoviruses and bornaviruses, order Mononegavirales) and other related viruses (nyaviruses, varicosavirus, dichorhabdovirus and others) (Table S1) are shown in black and gray letters, respectively. Asterisks show the viruses with the segmented (-)ssRNA genome. Shaded cartoons on the gray dashed circle indicate outlines of host species that encode mononegavirus L proteins-like sequences in their genomes. Numbers at the nodes represent aLRT values derived using an SH-like calculation (only values greater than 0.7 are shown).
Phylogenetic analysis of EpMLLSs and ShTSA-Ls

In order to assess the relationships among mononegaviruses and mononegavirus-like sequences (Figs. 1 to 3), we conducted phylogenetic analysis based on amino acid alignment of their L and L protein-like sequences. In this analysis, we included the type species for each genus in the Mononegavirales (four families Bornaviridae, Filoviridae, Paramyxoviridae and Rhabdoviridae and the proposed genus Nyavirus) and other related viruses with segmented (−)ssRNA genomes infecting plants (i.e., prototypic members of the genus Varicosavirus and the proposed genus Dicorhabdovirus) (King et al., 2011; Kondo et al., 2006; Mihindukulasuriya et al., 2009; Sasaya et al., 2002), as listed in Table S1, and the previously identified MLLSs (Belyi et al., 2010; Fort et al., 2012; Katzourakis and Gifford, 2010; Taylor et al., 2010), as listed in Table S7. We also included several novel MLLSs from plant and insect genomes identified in this study (Table S8 and Fig. S4). The multiple alignment of these sequences is shown in Fig. S1. Phylogenetic analysis of the most known MLLSs (rhabdovirus L protein-like sequences [RLLS], filovirus L protein-like sequence [FLLS] and bornavirus L protein-like sequences [BLLS]) in vertebrate and invertebrate genomes are placed into the modern mononegaviruses clades (e.g., rhabdoviruses, filoviruses and bornaviruses) (Belyi et al., 2010; Fort et al., 2012; Katzourakis and Gifford, 2010; Taylor et al., 2010) (Fig. 4). Similarly, the novel MLLSs (nyavirus L protein-like sequences: NLLSs) in lettuce plant (Lactuca sativa) (LnNLLS), silkworm (Bombyx mori) (BmNLLS1), leafcutter bee (Megachile rotundata) (MrNLLS), and carpenter and fire ants (Camponotus pennsylvanicus and Solenopsis invicta) (CNLLSs and SNLLS) form a cluster with the modern insect viruses (nyaviruses) (sister clade of the BLLSs and bornaviruses). In contrast, the EpMLLLs and ShTSA-L sequences demonstrated that they are the relatives closest to each other and distantly related to the modern mononegaviruses and related viruses (Fig. 4). Notably, two MLLSs in the genomes of the yellow fever mosquito, Aedes aegypti (AaMLLS), and the zebrafish, Danio rerio (DaMLLS), form a cluster distinct from the EpMLLLs and ShTSA-Ls (Belyi et al., 2010; Fort et al., 2012) (Fig. 4). It has been hypothesized by Hillman et al. (2004) that fungus-infecting mycoroenceptes (the genus Mycoroenceptes, family Reoviridae) and vertebrate-infecting coltiviruses (the genus Coltivirus, family Reo- viridae) might have evolved from a common ancestor native in the Acari. This hypothesis is based on the fact that some Acari members (mites) share their habitat with host fungi and others (ticks) transmit coltiviruses. Therefore, it is assumed that a reovirus progenitor may have been transferred laterally from Acari members to evolve into mycoroenceptes and coltiviruses. In this connection, a recent finding that horizontal transmission of mycoviruses by vector organisms may occur in nature is noteworthy (Yagashii et al., 2012). Likewise, it may be relevant to speculate that a common progenitor of (−)ssRNA viruses might have once infected fungi, mosquitoes and zebrafish. These organisms share an interesting connection. Some mosquitoes such as the anopheline species, Anopheles gambiae, feed on plants that could be infected by phytopathogenic fungi (Gouagna et al., 2010), while the diet of zebrafish can include mosquitoes (Spence et al., 2008).

Conclusion

Viral genomes are classified into seven groups based on nucleic acid composition and genome expression strategy (King et al., 2011). However, the genome type shows a trend that is dependent on the virus host kingdom. For example, many more positive-sense ssRNA viruses than dsDNA viruses are found in plants, whereas an inverse trend is observed in bacteria. In contrast to the large number of reports on fungal dsRNA viruses, no previous report has documented fungal (−)ssRNA viruses. Our findings strongly suggest that mononegaviruses infection might have occurred in fungi, as recorded in the form of endogenous virus-like sequences in their genomes, and also that novel mononegaviruses related to these virus-like elements are most likely present in fungi. Our discovery may provide novel insights into the origin and evolution of (−)ssRNA viruses.

Materials and methods

Database searches

To screen for (−)ssRNA virus-related DNA sequences, BLAST (tBLASTn) searches (Altschul et al., 1997) were conducted against genome sequence databases available from the NCBI (nucleotide collection, nr/nt; genome survey sequences, GSS; high-throughput genomic sequences, HTGS; whole-genome shotgun contigs, WGS; non-human, non-mouse expressed sequence tags, EST; transcriptome shotgun assembly, TSA, and others) (http://www.ncbi.nlm.nih.gov/). For this search, we used selected (−)ssRNA viruses from different families or genera as queries (as indicated in supplementary Table S1). Fungal genome sequences that matched viral peptide with e-values of < 0.01 were extracted (along with flanking sequences) and a putative ORF was restored by adding Ns as unknown sequences to obtain continuous amino acid sequences (edited residues are shown as Xs). These sequences were then used to screen the non-redundant (nr) database in a reciprocal tBLASTn search. (−)ssRNA virus-related sequences were also selected from S. homoeocarpa TSAs and the WGS sequences of other organisms (e.g., lettuce and several insect species). Transposable element sequences were identified using the Censor (http://www.girinst.org/censor/index.php) (Kohany et al., 2006).

Fungal materials, PCR amplification and DNA sequencing

Two samples of powdery mildew were collected from pea and red clover plants in the experimental fields at IPSR. Total genomic DNA was extracted from harvested powdery mildew (conidia and mycelia), pea plants, or red clover plants using the DNeasy® Blood and Tissue Kit (Qiagen) following the manufacturer's instructions. The resulting DNAs were stored at 4 °C until use. To amplify the ITS region of the ribosomal DNA, PCR was performed with total genomic DNA using the Erysipe-specific ITS primer pair (EryF and EryR) (Attanayake et al., 2009) or the ITS primer pair (EryF and EryR) (Chiba et al., 2011) (see Table S4). To amplify the candidate fungal DNA fragment by PCR, primer pairs were designed based on the virus-related sequences and/or their flanking sequences (see Table S4). All PCR reactions were carried out in a final volume of 50 μl containing 25 μl of Quick Taq HS DyeMix (TOYOBO) using the following conditions: an initial 94 °C for 3 min, followed by 35 cycles of 94 °C for 10 s, 58 °C for 30 s, and 68 °C for 1 min, then 68 °C for 10 min. The PCR products were fractionated by gel electrophoresis on 1% agarose gels and stained with ethidium bromide. Purified PCR fragments were sequenced using an ABI3100 DNA sequencer (Applied Biosystems, Foster City, CA, USA). To identify Erysipe spp. co-infecting pea, ITS fragments were cloned in pGEM T-easy (Promega) and sequenced. The sequences were assembled and analyzed using the DNA processing software packages, AutoAssembler (Applied Biosystems) and GENETYX (SDC, Tokyo, Japan).
Phylogenetic analysis

The deduced amino acid sequences of (−)ssRNA virus-related sequences in the genomes of fungus and other selected organisms were aligned with MAFFT version 6 under the default parameters (Katoh and Toh, 2008) (http://mafft.cbrc.jp/alignment/server). The alignments were manually refined to remove gaps using MEGA version 4.02 software (Tamura et al., 2007) (Fig. S1, highlighted region). The best-fit model of protein evolution (LG+G+F) was determined for maximum likelihood (ML) analyses using ProtTest (Abascal et al., 2005) (http://darwin.uvigo.es/software/prottest_server.html) based on the Akaike information criterion (AIC). Phylogenetic trees were created in PhyML 3.0 using the appropriate substitution mode (Guindon et al., 2010) (http://www.atgc-montpellier.fr/phyml/). The following four categories of rate variation were used. As starting trees, we selected the BIONJ tree, and the type of tree improvement was subtree pruning and regrafting (SPR) (Hordijk and Gascuel, 2005). The branch support values were estimated by the approximate likelihood ratio test (aLRT) with a Shimodaira–Hasegawa-like analysis (Anisimova and Gascuel, 2006). Trees were visualized using FigTree (version 1.3.1) (http://tree.bio.ed.ac.uk/software/figtree/).

For calculation of pairwise genetic distances, we used Kimura’s two-parameter model for nucleotide substitution (Kimura, 1980) with MEGA version 4.02. The molecular clock was based on a substitution rate of 2.52 × 10−9 per site per year in the rDNA ITS sequences of the order Erysiphales (Takamatsu and Matsuda, 2004). Estimate of the divergence time (T) was derived from the molecular clock (r) and genetic distance (K) using the simple equation T=r/K(2) (Fu and Li, 1997).

Acknowledgments

We thank Kazuyuki Maruyama, Sakae Hisano, Eri Umebayashi and Hideki Nishimura for technical assistance, Drs. Ida Bagus Andika, Tetsuo Tamada and Yuki Icinose for helpful advice, and Dr. Yuichiro Yamaai for providing a field fungal material. We are also grateful to the Max Planck Institute for Plant Breeding Research for use of the G. orontii BLAST Server. This work was supported in part by a Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology (KAKENHI to R.K., and Yomogi Inc. (to N. S.).

Appendix A. Supplementary information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.virol.2012.10.002.

References
