Journal of General Virology (2011), **92**, 2674–2678 **DOI** 10.1099/vir.0.034702-0

Short Communication

The remarkable evolutionary history of endornaviruses

Marilyn J. Roossinck,[1]† Sead Sabanadzovic,[2] Ryo Okada[3] and Rodrigo A. Valverde[4]

[1]The Samuel Roberts Noble Foundation, Plant Biology Division, Ardmore, OK 73402, USA

[2]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA

[3]Laboratory of Molecular and Cellular Biology, Faculty of Agriculture, Tokyo University of Agriculture and Technology, Tokyo 183-8509, Japan

[4]Department of Plant Pathology and Crop Physiology, Louisiana State University Agricultural Center, Baton Rouge, LA 70803, USA

Correspondence
Marilyn J. Roossinck
mjr25@psu.edu

Received 1 June 2011
Accepted 19 July 2011

The family *Endornaviridae* contains several members from diverse hosts, including plants, fungi and oomycetes. They are found as large dsRNA elements with a nick in the coding strand. All members encode a conserved RNA-dependent RNA polymerase, but no other domain that is conserved among all members. Based on the conserved domain database comparison the various domains have different origins, indicating a highly modular evolutionary history. In some cases, domains with similar putative functions are found that are derived from different protein families, indicating convergent evolution for a required function.

The endornaviruses are large dsRNA viruses that were first described from plants in the early 1990s (Valverde *et al.*, 1990; Wakarchuk & Hamilton, 1990) and later found in fungi and oomycetes (Fukuhara *et al.*, 2006), although a large dsRNA from *Vicia faba* was described in the early 1980s (Grill & Garger, 1981), which later proved to be an endornavirus. The first completed sequence was from a virus in rice in the mid 1990s (Moriyama *et al.*, 1995). Endornaviruses have recently been awarded family status (*Endornaviridae*) by the International Committee for the Taxonomy of Viruses (http://www.ictvonline.org/virusTaxonomy.asp?version=2009). They have a number of unique properties, including no evidence of encapsidation, a single very long ORF with a nick in the plus strand and the presence of a poly-C 3′ end in some, but not all members of the group (Fukuhara *et al.*, 2006). Endornaviruses have a persistent lifestyle in their hosts (Roossinck, 2010), and no evidence of horizontal transmission in plants or fungi, or of cell-to-cell movement in plants. They share a number of similar domains in their ORFs, but only the RNA-dependent RNA polymerase (RdRp) is clearly homologous among all species. The other domains are remarkable in that they are found in some, but not all members of the family, and they are not homologous. In addition, many endornaviruses contain a glycosyltransferase (GT) domain, a highly unusual protein for RNA viruses.

The sequences of the 11 completed genomes of members and putative members of the family *Endornaviridae* were analysed by phylogenetic analysis and for conserved domains using the conserved domain database (CDD) through the NCBI website (Marchler-Bauer *et al.*, 2011). When two closely related isolates had been sequenced, only one was used for the CDD analysis (nine in total), while all 11 were used for the phylogenetic analysis of the RdRp. Virus names, abbreviations and GenBank accession numbers are listed in Supplementary Table S1 (available in JGV Online). Domain hits with expect values larger than 10^{-2} were not considered. The various domains are shown in Table 1, with their protein families shown in different colours. Viruses are ordered by the size of their genomes. Two of the nine viruses, Bell pepper endornavirus (BPEV) and *Gremmeniella abietina* type B RNA virus XL1 (GaBRV-XL), contain a methyltransferase (MeTr) domain that is related to postive-sense ssRNA viruses. However, in the remaining seven viruses this domain is not detected, although we cannot rule out a highly diverged and unrecognizable MeTr domain. The MeTr in ssRNA viruses is used to add an m^7G cap structure to the 5′ end of the viral RNA, but experimental proof of a cap structure has not been shown for any endornavirus. Both BPEV and GaBRV-XL MeTr domains are most similar to *Sunn-hemp mosaic virus* (SHMV), a tobamovirus. However, while there is some overlap in the conserved domains, the similarity is more

---

†Present address: Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA 16802, USA.

A supplementary table is available with the online version of this paper.

Downloaded from www.microbiologyresearch.org by
IP: 130.18.13.93
On: Thu, 10 Aug 2017 20:31:13

034702 © 2011 SGM *Printed in Great Britain*

Table 1. Superfamilies of protein domains in endornavirus polyproteins

Domains are in the same order as found in the polyprotein ORF.

Virus	Host	MeTr*	Hel	Hel 2	GT	GT 2	RdRp	Length (bp)†
TaEV	Fungus	None	cl14882		None		cl03049	9760
GABrV-XL1	Fungus	cl03298	cl14882	cl14126	None		cl03049	10374
CeEV-1	Fungus	None	cl14126		cl10013		cl03049	11602
PEV-1	Oomycete	None	cl14126		cl10013		cl03049	13883
ORV	Plant	None	None		cl12292	cl10013	cl03049	13936
OSV	Plant	None	None		cl12292	cl10013	cl03049	13952
BPEV-KS	Plant	cl03298	cl14126		cl10013		cl03049	14727
HmEV-1	Fungus	None	None		cl07328		cl03049	16614
VFV	Plant	None	cl14126		None		cl03049	17635

*Protein domains are shown in different colours based on the protein superfamily they belong to. cl14882 is the DEAD-like helicase superfamily; cl03049 is the RdRp superfamily that includes most ssRNA virus RdRps; cl103298 is the methyl-transferase family that includes the ssRNA alpha virus MeTr, involved in capping viral RNA; cl14126 is the UvrD/REP helicase family that catalyses the ATP-dependent unwinding of DNA; cl10013 is a GT superfamily that shares a common GTB topology for nucleotide-sugar-dependentglycosyltransferases; cl12292 is another GT superfamily that contains a DXD motif, and uses nucleoside diphosphate sugars as donors; cl107328 is a third GT superfamily that is membrane associated, and is also known as the 28 N-terminal.

†Length in bp of the viral genomic RNA.

extensive between GaBRV-XL and SHMV than between BPEV and SHMV. In addition, many of the other viral MeTr domains most similar to BPEV are of mycovirus origin, while the GaBRV-XL MeTr is closer to plant viruses, even though GaBRV-XL is a fungal virus, while BPEV is a plant virus (Fig. 1). These differences suggest that the MeTr was acquired independently by the two viruses.

Most of the endornaviruses also contain a conserved helicase domain, although the two rice viruses, *Oryza sativa endornavirus* (OSV) and its closely related sister species *Oryza rufipogon endornavirus* (ORV) *Helicobasidium mompa endornavirus* 1 (HmEV1), and the root rot fungus virus do not by the criteria used here. Hence, if they do have a functional helicase it is too diverged from known helicases to be readily recognized. The origin of the helicase domains in the remaining six viruses is clearly different. Two of the fungal viruses, GaBRV-XL and *Tuber aestivum* endornavirus (TaEV) contain DEAD-like helicase domains, a diverse family of ATP-dependent helicases involved in unwinding RNA or DNA, and the remainder of the endornaviruses contain Uvr-B-like domains that are found in a number of ssRNA virus replicases. GaBRV-XL has both types of helicase (Table 1).

Six of the nine endornaviruses contain GT domains, but these are also from diverse origins (Table 1). Three of the four plant endornaviruses contain GTs from different sources. BPEV contains a GT domain that is most similar to bacterial GTs involved in the synthesis of antibiotics. This domain is also found in *Phytophthora* endornavirus 1 (PEV1), an oomycete virus, and *Chalara elegans* endornavirus 1 (CeEV1), a fungal virus. OSV and ORV, the rice viruses, contain a different type of GT with a DXD motif. This is a very diverse family of GTs found in many organisms. They also contain an additional GT motif that is in the same family as the BPEV GT domain. The fourth plant endornavirus, *Vicia faba endornavirus* (VFV) from broad bean, does not contain a GT domain. HmEV1 contains yet another GT motif from the '28 N family'. These GTs are characterized by an N-terminal domain that is probably membrane associated. It is highly unusual for RNA viruses to encode GTs. The only other GTs described in RNA viruses are found in some, but not all, of the hypoviruses of *Cryphonectria* (Linder-Basso *et al.*, 2005; Smart *et al.*, 1999). These GTs belong to the same superfamily as the GT in BPEV. Interestingly, the hypoviruses, like the endornaviruses, are found as dsRNAs in their infected hosts, lack any evidence of encapsidation, and have RdRps that are ssRNA RdRps. The function of the GTs in any of these viruses is unknown. They could play roles in protecting the viral RNA from degradation by modifying the RNA, or cellular membranes surrounding the RNA.

All of the endornaviruses encode an RdRp at the C terminus of the long ORF. These are all similar, and are the only highly conserved domain in members and putative members of the family *Endornaviridae*. They are related to the RdRps of ssRNA viruses, rather than dsRNA RdRps, and are closest to the RdRps of closteroviruses. The RdRp amino acid sequences of the nine viruses, plus an additional strain of

Downloaded from www.microbiologyresearch.org by
IP: 130.18.13.93
On: Thu, 10 Aug 2017 20:31:13

(a)

```
BPEV-KS 244 .[89].PCLDWRDVERY.[26].LISVDAMYDIDPTDVISLM.[1].RTG.[2].QL.[1].FCLST     AEVDFENT 410
SHMV     47 .[78].PALDARDLARN.[74].ATSLHSLYDIPYQNLGPAL.[1].RKR     IK     VLHAA.[4].EDLLLGAS 251
OMSV     38 .[79].EIITPADHTRY.[15].ILVDDALQHWSRERIDAAF     RHF     GG.[1].RLIGT.[4].DEVRSGHA 184
BVF      40 .[81].YAITARDHVRY.[15].LFMHDALQYMTPLDVYTLF.[1].TSP     EM.[1].SLVAT.[4].PESVDRLP 189
BVX      39 .[80].YAREAKDLSRF.[15].ALLHDTLHFLSPRQLAQLF.[1].HNP     NL.[1].LLYAT.[4].VEALHNLP 187
GVE      39 .[77].YCHEPKDVLRF.[15].AVISDTLHFMSTHQVWTLF.[1].NNP     KL.[1].RLYGT.[4].VEALHKLP 184
BmMLV    38 .[78].FRHTPKDITRY.[12].AFMHDALMFITPSQILGLF.[1].DSP     SM.[1].SLYCS.[4].AEAAYGVP 181
EMV      38 .[78].FRLTAADTTRY.[12].CFMHDALMYFSPAQIVDLF.[1].QSP     AL.[1].TLYCS.[4].PESHFTDL 181
PhyMV    38 .[78].YRLTSADTCRY.[12].CFMHDALMYYHPSQILDLF.[1].SCE     QL.[1].TLYCS.[4].PESHFTNL 181
GRVFV   169 .[78].YSLVARDITRY.[12].WFMHDSIMYLSAEEVAGCF.[1].ACP     HL.[1].RLYAS.[4].PESSFTDL 312

BPEV-KS 411 NGKLAHGQGEW.[2].HDDK     LVTVLRGD     DRPYVNNW     KLTK.[1].WSTADLIQ      AGHV 460
SHMV    252 EGLLTQIGGTF     QRNG.[2].LTFSFLDE.[1].SLIYTHSF.[1].NVFE     YVTRTFFV      ACNR 302
OMSV    185 SRYPDLYTIEY     LPNR.[1].YGLLSHRH.[1].SASYEASI     DEAW     MLDCGDFT      VDDR 233
BVF     190 AFWPELYQLAY     YEDH     LCYAPDGN.[1].ADAYNQPL     AAHQ     WMTMKSLH      GPDF 237
BVX     188 SLNPHAYTLEY     FSNG.[1].FAYMPGGH.[1].GSGYYHSA     ETLH     WLRAGQIK      LGHF 236
GVE     185 SLFPEIYKLEY     YERH     FAYMPGGY.[1].GGAYIHSY     GTLK     WLSTAQIG.[15].PLLD 247
BmMLV   182 SLFPDLYSYTI     KDDQ     LVYTLEGN.[1].TGNYTQPL     RSLD     WLRRSGIS      SGDL 229
EMV     182 SLFPEIYTYKI     SGQT     LHYIPENH.[1].SGSYNQPL     QAPS     WLKISSIL      SPSL 229
PhyMV   182 SLYPSVYTYQI     HGRT     LHYIPESH.[1].AGSYDQPL     DALS     WLKIHSIP      HPSL 229
GRVFV   313 SLNPSLYRYRF     HGDQ     LIYRLEEN.[1].SHSYQQPK     AALK     WLQTNSIQ      GNDI 360

BPEV-KS 461      ELSVQTVKTIGNH     IIR.[2].T.[151]. 630
SHMV    303 .[1].AYMKEFRSRRVDT.[4].FIR.[2].T.[151]. 477
OMSV    234      AYDVEFLLSYGPY.[4].VVP.[2].G.[ 64]. 320
BVF     238      TLSVDVPASRYSH.[4].ISK.[2].G.[ 64]. 324
BVX     237      SLSLNKEDSFGAH.[4].VTR.[3].P.[ 63]. 323
GVE     248      HLSIEKIETKAAH.[4].IQR     R.[ 70]. 338
BmMLV   230      HLSVTLLESFVSV.[4].ITR.[2].Q.[ 65]. 317
EMV     230      ALSVTKLESWGPV.[4].IQR     G.[107]. 357
PhyMV   230      TLSVTRLESWGPC.[4].IQR     G.[107]. 357
GRVFV   361      FLAITRLESWGPV.[4].ITR     G.[ 68]. 449
```

BPEV (BPEV-YW) and an additional strain of GaBRV-XL (GaBRV-XL2) were aligned along with the RdRp of *Pineapple mealybug wilt-associated virus 1* (PMWaV-1; family *Closteroviridae*, genus *Ampelovirus*), which was used as an outgroup for phylogenetic analysis. Initial alignments were done in MAFFT and were edited manually using Mesquite 2.74 (Maddison & Maddison, 2010). Aligned sequences were imported into Geneious (www.geneious.com) and analysed via a MrBayes (Ronquist & Huelsenbeck, 2003) plugin. The rate matrix was set to a Poisson distribution with a gamma rate variation. Burn-in was 100 000 and total chain length was 1 100 000. Branch lengths were unconstrained. Tree topologies were confirmed using PhyML (Guindon *et al.*, 2009) and PAUP 4.0 Beta 4b10 (Swofford, 2002), which gave very similar trees (not shown).

The topology of the tree does not follow the relationships of the hosts. For example, VFV, a plant virus, is most closely related to CeEV-1, a fungal virus, and the oomycete virus PEV-1 is a sister virus to these two (Fig. 2). The two rice viruses form a separate clade that is closer to the root rot virus (HmEV-1) than it is to the Bell pepper virus. One explanation is that these viruses have a common origin, most likely in fungi, and have been transmitted horizontally at some time in their history. The accompanying paper on BPEV suggests that short-term evolution may be congruent with a host group, but long-term evolution clearly is not.

The relationships of the viruses inferred from their RdRps follow the relationships based on shared domains in several cases, but not always. For example, CeEV-1, VFV and PEV-1 all share helicase domains, but while CeEV-1 and PEV-1 contain GT domains from the same superfamily, VFV lacks a GT domain altogether (Table 1). BPEV and GaBRV-XL are the only viruses with a MeTr domain, but they are found in distal portions of the tree (Fig. 2). TaEV and GaBRV-XL form a separate clade by RdRp analysis and share a Hel domain from the same superfamily, but TaEV does not contain the MeTr domain. In addition, GaBRV-XL contains a second Hel domain (a UVR-D Hel domain) that is shared with several other endornaviruses. HmEV-1, closest to the rice viruses by RdRp, shares the lack of a Hel domain with them, but has a different GT domain that is not found in any of the other viruses.

This interesting and highly modular arrangement of domains in the endornaviruses suggests that they have acquired various functional domains from different sources and/or at different times during their evolution. The lack of domains in some viruses could mean that these domains were not acquired or that the domains were in some ancestral virus and subsequently lost. The occurrence of domains with similar functions from obviously different sources, such as the Hel and GT domains from different protein superfamilies indicates a convergent pattern of

Downloaded from www.microbiologyresearch.org by
IP: 130.18.13.93
On: Thu, 10 Aug 2017 20:31:13

(b)

```
GaBRV-XL1 233 .[29].AAHS     HIYHHA     NRL.[1].LTEMLAS.[4].NALLYDIGGNINRHLDHGR     LNVHCVYTTS 315
SHMV       47 .[29].TSHA.[1].AAGLRG.[2].LEY     LYTLVPY     GAVSYDIGGNFPAHMMKGR     SYVHCCNPAL 127
AltMV      39 .[25].RAHT     HAAAKS.[2].LDM     YKITGYS.[5].PITFLFMKRSKLQYFHRGP.[1].HGDLFLNAWI 120
AMV        68 .[29].SSHC.[1].AAAHRL.[2].TDF     VYRCFGN.[1].VDSIIDLGGNFVSHMKVKR     HNVHCCCPIL 149
CiLRV      59 .[29].SSHA.[1].AAAHRV.[2].TDY     IYSRFQT.[1].NTTIIDIGGNFSTHAKMGR     SNVHSCCPIL 140
TVCV       50 .[29].AVHS.[1].AGGLRS.[2].LEY     LMMQVPF     GSLTYDIGGNFSAHLFKGR     DYVHCCMPNL 130
ORSV       50 .[29].AVHS.[1].AGGLRA.[2].LEY     LMMQIPF     GSITYDIGGNFSAHLYKGR     DYVHCCMPNL 130
TMGMV      50 .[29].AVHS.[1].AGGLRA.[2].LEY     LMLQVPY     GSPTYDIGGNFAAHLFKGR     DYVHCCMPNL 130
PMMV       50 .[29].AVHS.[1].AGGLRS.[2].LEY     LMMQIPY     GSTTYDIGGNFAAHMFKGR     DYVHCCMPNM 130
PEBV       63 .[29].MVHG.[1].AAAERK.[2].ALL     LMARVPK     LEPVDDIGGQWSFWLSRGD     KRVHSSCPLI 143

GaBRV-XL1 316 TPADLSRH.[29].ASMIKEDNS.[19].AMSIDTLVHLDPSDLIKFY.[1].DN.[1].VAHASH.[1].MTIPDNYAY 419
SHMV      128 DARDLARN.[ 8].NYLSRFEDK.[57].ATSLHSLYDIPYQNLGPAL.[1].RK.[3].VLHAAF     HFSEDLLLG 249
AltMV     121 EPKDVTRY.[ 8].SITPEIQTR.[ 1].AFIGDTLHFLPLGAIREIF.[1].NS.[4].TLYATM     VLPPEAMHR 187
AMV       150 DARDGARL.[ 8].SYVRKHPEI.[20].AFAIHSTSDLDVGELACSL.[1].QK.[3].KFICTM     MVDADMLIH 234
CiLRV     141 DVHDGERY.[ 8].GALEKQPDR.[20].AMAIHSISDIPITTVVKHC.[1].RR.[3].KLIASI     MDGSNDVNR 225
TVCV      131 DVRDIARH.[ 8].SYVNRLKRQ.[42].AVALHSIYDIPVEEFGSAL.[1].RK.[3].TCFAAF     HFHENMLLD 237
ORSV      131 DIRDVARH.[ 8].TYLARLERS.[42].AVGLHSIYDIPADEFGAAL.[1].RK.[3].ICYAAF     HFSENLLLE 237
TMGMV     131 DIRDIMRH.[ 8].MYLSRLSRS.[43].AVALHSLYDIPVHEFGAAL.[1].SK.[3].VCYAAS     ILAEALLLD 238
PMMV      131 DLRDVMRH.[ 8].LYLSKLAQK.[42].AVALHSLYDIPADEFGAAL.[1].RR.[3].VCYAAF     HFSENLLLE 237
PEBV      144 DMRDKQRE.[ 8].VFRDNATTS.[50].AIALHSLYDFKLDDVADAM.[1].EK.[3].FLHAAM     LFAPEAEIE 258

GaBRV-XL1 420 A.[1].KGVLRHNEGHWYRTN.[5].IEFVGE     SLSYEQKV     SYTD.[1].YLQTLIFT     AGEM 472
SHMV      250 A.[1].EGLLTQIGGTFQRNG.[2].LTFSFL.[3].SLIYTHSF.[1].NVFE     YVTRTFFV     ACNR 302
AltMV     188 M.[1].SIHPSIYELEFHERN     FIYKPG.[3].GASYCHEY     SQLQ     WLKVGKFE.[5].YQKH 242
AMV       235 N     EGEIPNFNVRWEIDR.[4].IHFDFI.[3].NLGYSHRF.[1].LLKH     YLTYNAVD     LGHA 288
CiLRV     226 H     RSFIPRLNVRWEVET.[6].ISFHFV.[3].GLSYTHNF.[1].VLMQ     YMTCNQVI     VTGK 281
TVCV      238 C.[1].TVTLDEIGATFQKSG.[2].LSFFFH.[3].TLNYTHSF.[1].NIIK     YVCKTFFP     ASQR 290
ORSV      238 T.[1].SAPLDEIGATFYKSG.[2].LSFFFQ.[3].TLNYEHSY.[1].NVIK     YVCKTFFP     ASNR 290
TMGMV     239 Q.[1].EVTLNEIGATFKREG.[2].VSFFFA.[3].TLNYSHKY.[1].NILH     YVVKSYFP     ASSR 291
PMMV      238 D.[1].YVSLDDIGAFFSREG.[2].LNFSFV.[3].TLNYTHSY.[1].NVLK     YVCKTYFP     ASSR 290
PEBV      259 K     EGPLPSVDGYYERKE.[8].IFFGFN.[3].SYAYIHDW.[1].EYKK     YLRGEPFS     RRGH 316

GaBRV-XL1 473      AVYNKVVGRKGPHML.[3].FLITRAKL.[14].MDEIFVM      I     PEI     DFDSAVTLMK 533
SHMV      303 .[1].AYMKEFRSRRVDTVF     CSFIRIDT.[52].KDKVCLP.[11].V.[1].TRH.[5].DFYWTALNHI 416
AltMV     243      YVTSQILETKGANHL     FVFQRGNF.[ 2].PTYRTFG      V.[1].TKF.[5].IFLPKKYNAR 294
AMV       289      AYRIERKQDFGGVMV     IDLTYSLG.[19].KGQMVVH.[13].A.[1].RRK.[5].KVLTRVTEVA 370
CiLRV     282 .[1].AYRVERVADLSGVFI     VEITLAST.[21].RRKTLVR.[ 9].W.[1].IKH.[5].DFVRRVAEVS 362
TVCV      291 .[1].VYHKEFLVTRVNTWY     CKFTRVDT.[54].RDMVIVP.[10].M.[1].RRE.[5].DFVYTVLNHI 405
ORSV      291 .[1].VYHKEFMCTRVNTWF     CKFTKVDT.[54].KDMVIVP.[10].M.[1].RSE.[5].DFVYTVLNHI 405
TMGMV     292 .[1].VYFKEFLVTRVNTWF     CKFTKVDT.[54].KDMVIVP.[10].M.[1].RSE.[5].DFVYTVLNHI 406
PMMV      291 .[1].VYMKEFLVTRVNTWF     CKFSRLDT.[54].KDMVIVP.[12].L.[1].RKE.[5].DFVYTVLNHI 407
PEBV      317      VFMFEPWQARGDTMF     FTLYRMTG.[20].EGMVVVP.[10].L.[1].KSS.[5].AYMDKCLDYV 396

GaBRV-XL1 534 .[1].EPFTMKTVSINLRFYERLLNRLLQ.[35]. 593
SHMV      417 .[2].YPDGKADFRGVMSFLESIRSRVVI.[35]. 477
AltMV     295      YPIKKTVAQQLFLYIKSVKTVTER.[12]. 330
AMV       371 .[6].NADAHSAIQSIATMLSSSTNHTII.[35]. 435
CiLRV     363 .[6].DTPLENLVQSVATMISSASNHCVI.[35]. 427
TVCV      406 .[2].YQAKALTYANVLSFVESIRSRVII.[35]. 466
ORSV      406 .[2].YQDKALTYKNVLSFVESIRSRVII.[35]. 466
TMGMV     407 .[2].YQAKALTYQNVLSFVESIRSRVII.[35]. 467
PMMV      408 .[2].YQSKALTYANVLSFVESIRSRVII.[35]. 468
PEBV      397 .[2].LSDQQLTINNVKSFMSSNNWVLFI.[35]. 457
```

Fig. 1. The MeTr domains of BPEV (a) and GaBRV-XL1 (b). Alignments were generated by the NCBI CDD tool. Numbers refer to the position of the domains in their respective proteins/polyproteins. Viruses not mentioned in the text are: Oyster mushroom sperical virus (OMSV), *Botrytis virus F* (BFV), *Botrytis virus X* (BVX), *Garlic virus E* (GVE), *Bombyx mori Macula-like latent virus* (BmMLC), *Eggplant mosaic virus* (EMV), *Physalis mottle virus* (PhyMV), Grapevine rupestis vein feathering virus (GRVFV), Alternanthera mosaic virus (AltMV), *Alfalfa mosaic virus* (AlMV), *Citrus leaf rugose virus* (CiLRV), *Turnip vein-clearing virus* (TVCV), *Odontoglossum ringspot virus* (ORSV), *Tobacco mild green mosaic virus* (TMGMV), *Pepper mild mottle virus* (PMMV) and *Pea early browning virus* (PEBV).

Downloaded from www.microbiologyresearch.org by
IP: 130.18.13.93
On: Thu, 10 Aug 2017 20:31:13

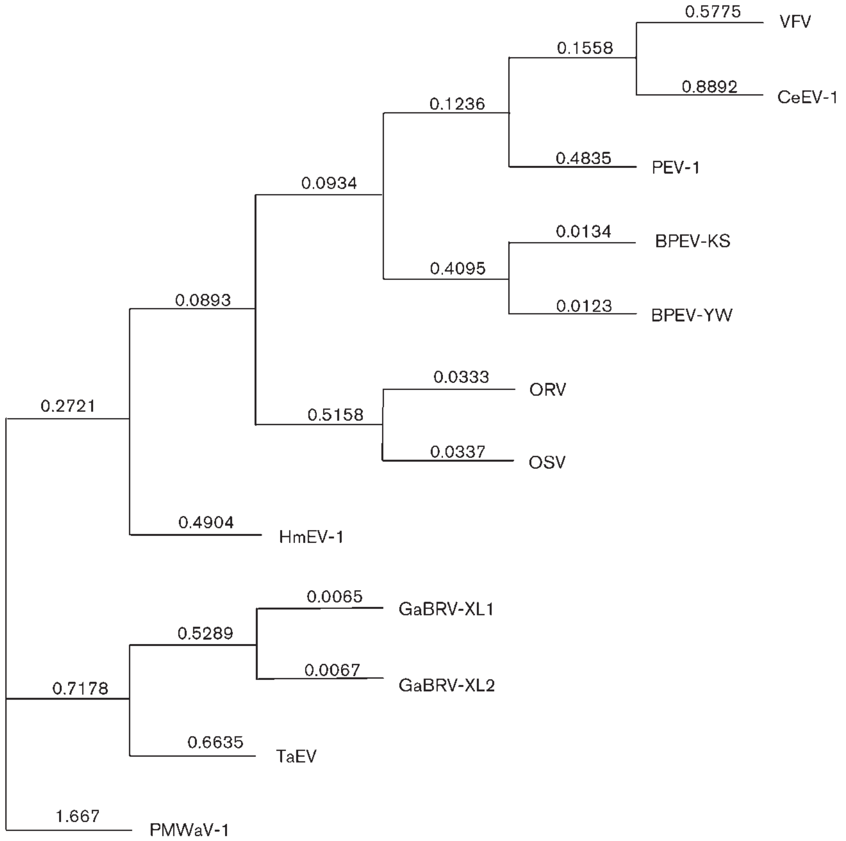

Fig. 2. Phylogeny of endornavirus RdRps. Bayesian analysis of the relationships among aligned amino acid sequences of the RdRps. PMWaV-1 was used as an outgroup. Relative branch lengths are indicated by numbers above branches. Virus GenBank accession numbers and abbreviations are given in Supplementary Table S1.

evolution. Although the method of acquisition of these domains is not known, it suggests a very complex and dynamic evolutionary history that is unprecedented in other families of plant or fungal viruses.

Since the endornaviruses do not have any known effects on their hosts, they have been understudied. However, sequence analysis of more endornavirus genomes will certainly shed new light on this interesting virus family.

Acknowledgements

The authors thank Dr Ping Xu for carefully reading of the manuscript. This work was supported in part by the Samuel Roberts Noble Foundation, and by the National Science Foundation (grant EF-067108 to M. R.). This manuscript is approved for publication as journal article no. 12039 of the Mississippi Agricultural and Forestry Experiment Station, MSU.

References

Fukuhara, T., Koga, R., Aoki, N., Yuki, C., Yamamoto, N., Oyama, N., Udagawa, T., Horiuchi, H., Miyazaki, S. & other authors (2006). The wide distribution of endornaviruses, large double-stranded RNA replicons with plasmid-like properties. *Arch Virol* **151**, 995–1002.

Grill, L. K. & Garger, S. J. (1981). Identification and characterization of double-stranded RNA associated with cytoplasmic male sterility in *Vicia faba*. *Proc Natl Acad Sci U S A* **78**, 7043–7046.

Guindon, S., Delsuc, F., Dufayard, J. F. & Gascuel, O. (2009). Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol* **537**, 113–137.

Linder-Basso, D., Dynek, J. N. & Hillman, B. I. (2005). Genome analysis of *Cryphonectria hypovirus 4*, the most common hypovirus species in North America. *Virology* **337**, 192–203.

Maddison, W. P. & Maddison, D. R. (2010). Mesquite: a modular system for evolutionary analysis, 2.74 edn. http://mesquiteproject.org/mesquite/mesquite.html

Marchler-Bauer, A., Lu, S., Anderson, J. B., Chitsaz, F., Derbyshire, M. K., DeWeese-Scott, C., Fong, J. H., Geer, L. Y., Geer, R. C. & other authors (2011). CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res* **39** (Database issue), D225–D229.

Moriyama, H., Nitta, T. & Fukuhara, T. (1995). Double-stranded RNA in rice: a novel RNA replicon in plants. *Mol Gen Genet* **248**, 364–369.

Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Roossinck, M. J. (2010). Lifestyles of plant viruses. *Philos Trans R Soc Lond B Biol Sci* **365**, 1899–1905.

Smart, C. D., Yuan, W., Foglia, R., Nuss, D. L., Fulbright, D. W. & Hillman, B. I. (1999). *Cryphonectria hypovirus* 3, a virus species in the family *Hypoviridae* with a single open reading frame. *Virology* **265**, 66–73.

Swofford, D. L. (2002). PAUP*: Phylogenetic analysis using parsimony (and other methods), version 4. Sunderland, MA: Sinauer Associates.

Valverde, R. A., Nameth, S., Abdallha, O., Al-Musa, O., Desjardins, P. & Dodds, A. (1990). Indigenous double-stranded RNA from pepper (*Capsicum annuum*). *Plant Sci* **67**, 195–201.

Wakarchuk, D. A. & Hamilton, R. I. (1990). Partial nucleotide sequence from enigmatic dsRNAs in *Phaseolus vulgaris*. *Plant Mol Biol* **14**, 637–639.

Downloaded from www.microbiologyresearch.org by
IP: 130.18.13.93
On: Thu, 10 Aug 2017 20:31:13