Volume 17 Number 23 1989 Nucleic Acids Research

Evolutionary relationship between luteoviruses and other RNA plant viruses based on sequence motifs in their putative RNA polymerases and nucleic acid helicases

Nuredin Habili and Robert H.Symons*

Centre for Gene Technology, Department of Biochemistry, University of Adelaide, Adelaide, SA 5000, Australia

Received October 5, 1989; Accepted November 7, 1989

ABSTRACT

Comparative studies of sequence motifs in the RNA polymerases and nucleic acid helicases of positive-sense RNA plant viruses have provided a new scheme for the classification of these pathogens. We propose a new luteovirus supergroup which should be added to the already described Sindbisvirus-like and picornavirus-like supergroups. Sequence motifs of nucleic acid helicases and RNA polymerases which previously were considered to be specific for each of the two supergroups now occur together within this new supergroup. We propose that this new viral supergroup provides an evolutionary link between the other two supergroups.

INTRODUCTION

Positive-strand RNA plant viruses consist of many economically important groups with wide host ranges, morphologies and modes of transmission (1−4). In recent years the RNAs of a number of these viruses have been sequenced and some functional sequence motifs in their structural and non-structural proteins have been described (3,5−11).

The recent discovery of a series of amino acid motifs from a wide range of prokaryotic and eukaryotic organisms led to the definition of a large group of proteins, the nucleic acid helicases, which are also found in the positive-strand RNA viruses of plants (5−7,10,11). Their specific functions include nucleic acid unwinding, and acting in recombination, transcription, translation, and possibly also in RNA splicing (7,10−13).

In the comparative analysis presented here, we have used sequence motifs in nucleic acid helicases (5−7,11) and RNA polymerases (14) to gain an insight on the presence and location of the relevant putative genes in different groups of plus-sense RNA plant viruses. In particular we are concerned with luteoviruses, carmoviruses and sobemoviruses which show partial homology in the sequence motifs of their non-structural proteins. We propose here a new luteovirus-like supergroup which includes these viruses and which links together the two previously described Sindbisvirus-like and picornavirus-like supergroups.

RESULTS AND DISCUSSION

Recently, the positive-sense RNA plant viruses were divided into two large supergroups called Sindbisvirus-like and picornavirus-like by Goldbach and Wellink (3). Their classification scheme was based on the presence or absence of VPg (*V*iral *P*rotein *g*enome) at the 5′-end of the genomic RNA, the use of a polyprotein as an intermediate in the expression of viral genome and the presence or absence of subgenomic RNAs *in vivo*.

In this work, we present a new basis for the classification of these viruses, which is now extended to include a new supergroup called luteovirus-like (Supergroup B), which

Table 1. The presence (+) or absence (−) of various putative nucleic acid helicase motifs in plant viral RNAs.

SUPERGROUP A: Sindbisvirus-like	Virus	Helicase Motifs I	Ia	II	III	IV	V	VI
A1 Potexvirus	Potato virus X PVX[a]	+	+	+	+	+	+	+
Potexvirus	Wild clover WClMV[a] mottle virus	+	+	+	+	+	+	+
Tobravirus	Tobacco rattle TRV virus	+	+	+	+	+	+	+
Tymovirus	Turnip yellow TYMV mosaic virus	+	+	+	+	+	+	+
A2 Tricornaviridae	Alfalfa mosaic AlMV virus	+	−	+	+	+	+	+
Tricornaviridae	Cucumber CMV mosaic virus	+	−	+	+	+	+	+
Tobamovirus	Tobacco mosaic TMV virus	+	−	+	+	+	+	+
Furovirus	Beet necrotic BNYVV[a] yellow vein virus	+	−	+	+	+	+	+
Hordeivirus	Barley stripe BSMV[a] mosaic virus	+	−	+	+	+	+	+
SUPERGROUP B: Luteovirus-like								
B1 Carmovirus	Carnation CarMV mottle virus	−	−	−	−	+	−	+
Carmovirus	Turnip crinkle TCV virus	−	−	−	−	+	−	+
Carmovirus	Maize chloroic MCMV mottle virus	−	−	−	−	+	−	+
Tombusvirus	Cucumber CNV necrosis virus	−	−	−	−	+	−	+
B2 Luteovirus: Subgroup 1	Barley yellow dwarf virus BYDV(PAV)	−	+	−	+	+	−	+
	Soybean dwarf virus SDV	+	−	+	+	+	−	+
Luteovirus: Subgroup 2	Potato leaf-roll virus PLRV	−	−	−	−	+	−	+
	Beet western yellow virus BWYV	−	−	−	−	+	−	+
B3 Sobemovirus	Southern bean mosaic virus SBMV	+	+	−	−	−	−	+
SUPERGROUP C: Picornavirus-like								
C1 Nepovirus	Tomato black ring virus TBRV	+	−	+	−	−	−	−
Comovirus	Cowpea mosaic virus CPMV	+	−	+	−	−	−	−
C2 Potyvirus	Plum pox virus PPV	+	+	+	+	+	+	+
Potyvirus	Tobacco etch virus TEV	+	+	+	+	+	+	+
Potyvirus	Tobacco vein mottling virus TVMV	+	+	+	+	+	+	+

[a] In each of these viruses the sequence motifs of the putative helicases appear twice (see Figs. 1,3).

is intermediate between the Sindbisvirus-like supergroup (Supergroup A) and the picornavirus-like supergroup (Supergroup C). The features of the amino acid sequences in putative nucleic acid helicases and RNA polymerases of the viruses listed in Table 1 which were used as a basis for our classification are as follows:

1. Putative nucleic acid helicases: a, type of sequence motif present; b, the amino acid gap length between each sequence motif and c, the degree of homology between corresponding sequence motifs in different viral helicases.
2. Putative RNA polymerases: a, the homology in the sequence motifs of various viral RNA polymerases and b, the gap length between each sequence motif.

Gorbalenya *et al.* (1989) (11) classified the nucleic acid helicases of prokaryotic and eukaryotic organisms into two large superfamilies (SF1 and SF2) based on the homology between their amino acid sequence motifs. In the study of the sequence motifs of putative nucleic acid helicases of the positive-sense RNA plant viruses, we observed that the Sindbisvirus-like Supergroup A carries sequence motifs of SF1 helicases and the picornavirus-like Supergroup C contains helicases of the SF2 type. On the other hand, members of the luteovirus group contain sequence motifs present in both Supergroups A and C (Fig. 1). Furthermore, the amino acid sequence motifs present in the RNA polymerases of each supergroup appear to be distinct (Fig. 2). These studies led us to conclude that the luteoviruses and related viruses (see below) can be grouped together into a new luteovirus-like Supergroup B.

Amino Acid Sequence Motifs of Viral Nucleic Acid Helicases

Seven sequence motifs derived from the predicted amino acid sequence of nucleic acid helicases (5−7,11) from 21 plant viruses are presented in Fig. 1. A consensus pattern for each supergroup is also shown.

Not all the plant viruses studied here possess all the seven types of sequence motifs in their nucleic acid helicases. Based on the presence or absence of a sequence motif (numbered I to VI, Table 1) and the other features mentioned above (see also Fig. 1), the classification scheme summarized in Table 1 was designed. Each of the three supergroups (A, B and C) can be sub-divided into two or more subdivisions (Table 1; A1, A2, B1−B3 and C1, C2) comprising one or more similar groups of plant viruses.

Although Supergroups A and C show some homology in sequence motifs I, Ia, II and V, motifs III, IV and VI are distinct. On the other hand, sequence motif III of Supergroup B, where present, appears related to that of Supergroup A, while motifs IV and VI are more closely related to those of Supergroup C (Fig. 1). Therefore, Supergroup B, which includes carmoviruses, tombusviruses, luteoviruses and sobemoviruses (Table 1), is distinct from the other two supergroups.

Our studies also show that the luteovirus group (Table 1; B2) can be divided into two subgroups. Subgroup 1, which includes the PAV isolate of BYDV and SDV (see Table 1 for the list of abbreviations), shows significant similarity to the carmovirus and tombusvirus groups in motifs IV and VI (Fig. 1; Table 1). Further, in contrast to the limited presence of sequence motifs I, Ia and II of the putative helicases of SDV and BYDV, motif IV of these viruses is 100% homologous while the homology in motif VI is 81%. Conversely, sequence motif VI of Subgroup 2 (which includes BWYV and PLRV) matches with that of SBMV, a sobemovirus (Fig. 1). A similar pattern of interhomology is observed when sequence motifs of the putative RNA polymerases in these viruses are compared (see below).

The gap length between sequence motifs III and IV of Subgroup 1 of the luteoviruses, and between IV and VI of Subgroup 2 is much larger than the corresponding gaps in other

			Motifs:	I		Ia		II
Supergroup A								
1.	AlMV RNA 1	(p126)	836	VDGVAGCGKTTNIK			55	RLIFDECFLQHA
2.	CMV RNA 1	(p111)	710	VDGVAGCGKTTAIK			53	RVLVDEVVLLHF
3.	TMV	(p126)	830	VDGVPGCGKTKEIL			57	RLFIDEGLMLHT
4.	TRV RNA 1	(p134)	901	VDGVPGCGKSTMIV	7	DVVLSTGRAA	39	VLHFDEALMAHA
5.	BSMV RNA 1	(p130)	835	IDGVPGCGKSTMIL			58	RFHFDEALKVHY
6.	BSMV RNA 2	(p 58)	266	ISGVPGSGKSTIVR			46	LLIIDEYTLAES
7.	BNYVV RNA 1	(p237)	939	VKGGPGTGKSFLIR			48	IIFVDEFTAYDW
8.	BNYVV RNA 2	(p 42)	120	VLGAPGVGKSTSIK			49	TMLVDEVTRVHM
9.	TYMV	(p206)	973	FAGFAGCGKTYPIQ	11	RVSCPTTELR	29	ILVIDEIYKMPR
10.	PVX	(p166)	732	IHGAGGSGKSHAIQ	12	TVVLPTNELR	28	IVIFDDYSKLPP
11.	PVX	(p 25)	26	VHAVAGAGKSTALR	2	ILRHPTFTVH	25	FAILDEYTLDNT
12.	*E.coli* rec B		20	IEASAGTGKTFTIA	25	LVVTFTEAAT	310	VAMIDEFQDTDP
Consensus A				o G G GKS o A T		o PT		oDE
Supergroup B								
13.	MCMV	(p111)	501					
14.	CarMV	(p 86)	308					
15.	CNV	(p 92)	366					
16.	BYDV-PAV	(p 99)	115			LSCEPTRQEL	154	
17.	SDV	(?)	?	HIGEAPFGKWFNTL			42	VVQVDTGEKLPP
18.	BWYV	(p116)	395					
19.	PLRV	(p119)	394					
20.	SBMV	(p105)	354	NRGKVKLGKREFAW	64	PVDPPTKQVA	184	
Consensus B				G GK		PT+Q		
Supergroup C								
21.	TBRV RNA 1	(p 72)	212	LFGQRHCGKSNFMA			94	GVRDIEAYRSRK
22.	CPMV RNA 1	(p 58)	165	FQGKSRTGKSLLMS			95	KVRDDEAFKNRR
23.	PPV CI	(p 70)	82	IRGAVGSGKSTGLP	9	LLIEPTRPLA	54	CIIFDECHVHDA
24.	TEV CI	(p 54)	81	VRGAVGSGKSTGLP	9	LMLEPTRPLT	54	FVIIDECHVNDA
25.	PVY CI	(p 70)	82	VRGAVGSGKSTGLP	9	LLIEPTRPLA	54	FVIFDECHVLDP
26.	vasa(Drosophila)		286	ACAQTGSGKTAAFL	21	VIVSPTRELA	63	FVVLDEADRMLD
27.	p68 (Human)		115	GVAQTGSGKTLSYL	21	LVLAPTRELA	63	YLVLDEADRMLD
Consensus C				G GKS o A T		ooo PTR LA		o DE

Supergroup A

		III		IV		V		VI	REF
1.	14	VIGFGDTEQIPF	22	ITWRSPADA	66	IFTTHEAQGKTFDNVYFCR	19	NGLVALSRHKKSFKYF	(47)
2.	14	ALCFGDSEQIAF	23	TTFRSPQDV	79	IKTVHEAQGISEDHVTLVR	13	YCLVAVTRHKVTFRYE	(48)
3.	14	AYVYGDTQQIPY	25	TTLRCPADV	62	VHTVHEVQGETYSDVSLVR	14	HVLVALSRHTCSLKYY	(49)
4.	14	CICQGDQNQISF	25	ETYRSPADV	64	VSTVHESQGETFKDVVLVR	13	YLIVALSRHTQSLVYE	(50)
5.	14	ILAQGDRAQLPM	25	RSYRIPGDV	78	ISTIHEAQGGTYENVILVR	17	YIVVGTSRHTKTFTYC	(19)
6.	14	VLLVGDVAQGKA	18	TTYRLGQET	52	CALAIDVQGREFDSVTLFL	12	LRLVALSRHKSKLIIR	(51)
7.	14	IYLVGDEQQTGI	25	MNFRNPVHD	61	KTTVRANQGSTYDNVVLPV	12	LNLVALSRHRNKLTIL	(52)
8.	14	VICFGDPAQGLN	18	ASRRFGKAT	56	SILYSDAHGQTYDVVTIIL	13	VRAVLLTRARKGGMIK	(53)
9.	15	VIILGDPLQGEY	28	WSYRIPQCI	45	SCTISSSQGLTFCDPAIIV	11	NGLVALTRSRSGVQFM	(54)
10.	15	IILTGDSRQSVY	28	ATHRNKKDL	46	TFTYAGCQGLTKPKVQIVL	10	VMYTALSRATDRIHFV	(55)
11.	5	QAIIADPYQAPE	10	TSFRVPRKV	52	FVKPCQVTGLELKVVTIVS	11	AFYNAITRSKGLTYVR	(55)
12.	17	LLLIGDPKQAIY	24	TNWRSAPGM	284	IVTIHKSKGLEYPLVWLPF	44	LLYVALTRSVWHCSLG	(56)
Consensus A		o GD Q		R o		T QG V o		oVAoSR o o T	

Supergroup B

				*				**	
				*				**	
13.				ERVFYVKDD	99			RYNVELGRYLRPVEHP	(40)
14.				ERVFYVERN	97			RYNVELGRYLKKYEHH	(57)
15.		**		ERVFCVERN	96			RYNVELGRYLRHMESK	(41)
16.		VENVGDKFKLDI	112	RRVFTVGKG	96			RYNIILGTRLKFNEKK	(15)
17.	22	FEHYGDARKYIR	147	RRVFTVGKG	95			RYNVELGRRLKFNEKK	(58)
18.				TYVFETTAP	346			QAKLDEGRYRLIMSVS	(42)
19.				NYVFESTAV	376			QSKLDEGRYRLIMSVS	(43,44)
20.								SRKLKEGRYRLISSVS	(9)
Consensus B		E GD K		+RVF Y				+ +o+ GRY	

Supergroup C

21.									(36)
22.									(59)
23.	17	ILKVSATPPGHE	40	VLVYVASYNE	42	VATNIIENGVTLDIDVVVD	26	ERIQRLGRVGRNKPGA	(60)
24.	17	VLKVSATPPGRE	40	ILVYVASYND	42	VATNIIENGVTIDIDVVVD	26	ERIQKLGRVGRHKEGV	(61)
25.	17	VLKVSATPVGRE	40	VLVYVSSYNE	42	VATNIIENGVTLDIDVVVD	26	ERIQRLGRVGRFKKGV	(62)
26.	21	TLMFSATFPEEI	51	TIVFVETKRG	42	IATSVASRGLDIKNIKHVI	9	DYVHRIGRTGCVGNNG	(22)
27.	19	TLMWSATWPKEV	57	TIVFVETKRR	42	IATDVASRGLDVEDVKFVI	9	DYIHRIGRTARSTKTG	(23)
Consensus C		o SAT +		VoV +		oAT o +Go o V		+ I++oGR R	

nucleic acid helicases (Fig. 1; Table 1). There is no explanation for the presence of such a large gap length, although it may be related to the presence of a translational frameshift in this region (ref. 15; F/S, Fig. 3). Large gap distances are also present between sequence motifs Ia and II and also between IV and V of the *E. coli* Rec B helicase (Fig. 1).

This translational frameshift which appears to be unique in the luteovirus RNAs (15) splits the helicase sequence motifs into two parts (pHEL, Fig. 3). In ORF1 of Subgroup 1, only motifs I to III are present, while in the portion of protein downstream from the translational frameshift site in ORF2, motifs IV and VI are found. In Subgroup 2, motif IV is found in ORF2 and motif VI in ORF3 (Table 1; Fig. 3). Although four and five sequence motifs are present in the putative helicases of BYDV (PAV) and SDV, respectively, only two are present in BWYV and PLRV (Subgroup 2), all being members of the luteovirus group (Fig. 1; Table 1, B2). The absence of most of the helicase motifs in BWYV, PLRV and some other plant viruses (Table 1) indicates that either an appropriate host protein may be sequestered to fulfil this function (10), a situation similar to that found in the replication of phage Qβ RNA (10,16), or viral proteins with similar functions but with a different set of sequence motifs may exist.

Sequence motif VI is the most abundant motif, being present in 22 out of a total of 24 putative plant viral helicases studied here (Fig. 1). It is believed that this motif, enriched with basic amino acid residues (Fig. 1), provides the nucleic acid-binding site (11).

Sequence motifs I and II, corresponding to the so-called 'A' and 'B' sites of the NTP-binding domain (10,11), are absent in most members of Supergroup B (Table 1). Some members of this supergroup (Table 1; B1) contain the smallest size genomes among the plant viruses (about 4 kb) (4) which implies that the function(s) provided by these motifs may be supplied by the host (10,16).

Motif I ('A' site motif) was the first to be described, and was shown to bind to ATP or GTP (17). In plant viruses this binding activity has only been studied in TMV RNA *in vitro* (18). The sequence of this motif has been quoted by many workers in their studies of the non-structural proteins present in the RNA viruses and was designated as a part of the replicative complex of these viruses (3,8).

Motif II ('B' site motif) is believed to interact with Mg^{2+} of the Mg-NTP complex via the conserved aspartic acid (D) residue (11). No putative role has as yet been stipulated for the sequence motifs Ia, III, IV and V.

Occasionally, a sequence motif in a certain viral helicase reveals a high degree of homology with a corresponding motif from a totally unrelated virus, indicating a significant evolutionary relationship. For example, by comparing type Ia motif of BYDV (a luteovirus) with that of the potyviruses (PPV, TEV and PVY) and also type II motif of SDV (a luteovirus) with that of the p166 of PVX (a potexvirus) a 50% amino acid sequence

Fig. 1. Seven amino acid sequence motifs (I−VI) in the putative nucleic acid helicases coded by the RNAs of three Supergroups (A,B and C) of positive-sense RNA plant viruses. For comparison, corresponding motifs present in the helicases of *E. coli* (rec B) (Supergroup A), *Drosophila* (vasa) and human (p68) proteins (Supergroup C) are also shown. The estimated M_r ($\times 10^{-3}$) of each putative protein (p) coded by viral RNAs is given in brackets. Numbers in the gaps between motifs are the number of amino acids contained within each gap. Each consensus line indicates residues which most commonly occur in that position. o, hydrophobic residues; +, polar residues. (?), partial sequence only obtained for SDV (58). The gap in which a translational frameshift occurs in the viral RNAs of Supergroup B is underlined. Single and double asterisks above amino acids in motifs in Supergroup B indicate those conserved in the same motifs of Supergroup A and Supergroup C, respectively. See Table 1 for the list of virus abbreviations.

homology is present (Fig. 1). This implies that these motifs originated from a common source.

In the monopartite potexvirus (PVX and WC1MV), and in the multipartite furovirus (BNYVV) and hordeivirus (BSMV), the sequence motifs of the putative helicase appear twice, either in the same RNA species (PVX, Figs. 1 and 3), or in two separate RNAs (BNYVV and BSMV, Figs. 1 and 3). It has been proposed that in these viruses, two oligomeric subunits of the nucleic acid helicase are coded by each viral genome (7). The same virus groups are furnished with a poly(A) tail structure at the 3′-end of their RNAs (Fig. 3). In the case of BSMV, the stretch of the poly(A) (about 20 nucleotides) is located 238 nucleotides upstream from the 3′-end (ref. 19; Fig. 3).

In the third position of the sequence motif I of the plant and animal viral helicases there is a glycine instead of an alanine, the latter being typical for the prokaryotic and eukaryotic helicases (refs. 12, 20–23; Fig. 1). Alanine is also present in the same position in the p25 and p26 of PVX and WCLMV helicases, respectively, (Fig. 1; ref. 24). For comparison, helicase motifs derived from *E. coli* recB, representing Supergroup A (11) and from *Drosophila* (vasa) and human (p68) representing Supergroup C are displayed at the bottom of each set in Fig. 1.

Role of Nucleic Acid Helicases in Virus Replication

It is clear that the plant viral coded proteins listed in Fig. 1 share some similarity with the true nucleic acid helicases in both the order and content of their sequence motifs. Further, it has long been known that in the tricornaviridae RNA 1 and RNA 2, which code for the motifs of nucleic acid helicase and RNA polymerase, respectively, are sufficient together for their successful replication in protoplasts (25,26). In the case of CPMV, only the RNA 1 (B RNA) which carries both putative nucleic acid helicase and RNA polymerase genes (see Fig. 3) is required to infect cowpea protoplasts (27). This indicates that both proteins are required for the replication of viral RNA.

Several functions for nucleic acid helicases in prokaryotic and eukaryotic cells have been reported (21–23,28,29). Of these, most emphasis has been placed in the unwinding of double-stranded nucleic acids during replication. The appearance of a double-stranded RNA in the life cycle of RNA viruses has been the subject of controversy for a long time (26,30). Accumulated data, at least in the viral and subviral pathogens of plants, favour the natural occurrence of dsRNA during the life cycle of these pathogens (7,29–32). In several systems replication complexes that have been isolated from cells infected with positive-sense RNA viruses resulted in the formation of a dsRNA as the end product (26,33,34). This implies that the helicase activity was not present in these fractions to provide unwinding of the duplex RNA and hence allow the synthesis of detectable amounts of single-stranded genomic RNA.

Sequence motifs of the nucleic acid helicases are not found in the negative-stranded RNA viruses, double-stranded RNA viruses, retroviruses and small DNA viruses of animals (7, and P. Keese, personal communication). This is surprising in the case of the first two groups and indicates that a helicase function is not required in the replication of these viral RNAs or that a helicase exists with a different set of amino acid motifs.

Amino Acid Sequence Motifs of RNA Polymerases

Four types of sequence motifs (I to IV) present in the putative RNA polymerases of different plant viruses are shown in Fig. 2. Type III motif, which was found by Kamer and Argos in 1984 (35), and more recently the type II motif, have been used frequently to denote that the gene under study was a putative RNA-dependent RNA polymerase (15,36,37).

		Motifs: I		II		III		IV	REF
Supergroup A									
AlMV RNA 2	(p 90)	KEADFSKFDKS	47	FQRRTGDALTYLGNTIV	17	FVVASGDDSLIG	29	ICSKFL	(63)
CMV RNA 2	(p 94)	LEIDLSKFDKS	47	FQRRTGDAFTYFGNTIV	15	RLLFSGDDSLAF	29	ICSKFY	(38)
TMV	(p183)	LELDISKYDKS	47	YQRKSGDVTTFIGNTVI	15	KGAFCGDDSLLY	31	FCGRYV	(49)
TRV RNA 1	(p194)	VEIDMSKFDKS	47	YQQKSGDADTYNANSDR	15	MVTYGGDDSLIA	31	FCGKFL	(50)
BSMV RNA 3	(p 87)	LEIDFSKFDKS	47	YQQKSGNCDTYGSNTWS	15	FCVFGGDDSLIL	31	FCGKFL	(64)
BNYVV RNA 1	(p237)	GVIDAAACDSG	45	YVKTSGEPGTLLGNTIL	15	CMAMKGDDGFKR	31	FCGYAL	(52)
TYMV	(p206)	IANDYTAFDQS	41	CMRLTGEPGTYDDNTDY	14	PIMVSGDDSLID	29	FCGYYV	(54)
PVX	(p166)	LANDYTAFDQS	41	IMRLTGEGPTFDANTEC	15	AQVYAGDDSALD	34	FCGWLI	(55)
Consensus A		Do oD+S		o + G+ To NT		o GDDSo		FCG oo I S	
Supergroup B									
MCMV	(p111)	IGFDASRFDQH	49	GKRMSGDMNISLGNCIL	16	RLINNGDDNVLI	37	FCQMRP	(41)
CarMV	(p 86)	IGFDMSRFDQH	49	GCRMSGDMNTALGNCLL	13	RLINNGDDCVLI	37	FCQMAP	(57)
CNV	(p 92)	IGLDASRFDQH	49	GCRMSGDINTSLGNYLL	17	SLANCGDDCVLI	37	FCQAHP	(40)
TCV	(p 87)	IGFDMKRFDQH	48	GCRMSGDVNTALGNCLL	14	KLINNGDDCVLF	36	FCQMAP	(41)
BYDV	(p 99)	IGVDASRFDQH	48	GHRMSGDINTSMGNKLI	16	ELCNNGDDCVII	36	FCQSKP	(15)
SDV	(?)	IGVDASRFDQH	48	GHRMSGDINTSSGNKLI	16	ELCNNGDDCVII	36	FCQSKP	(58)
BWYV	(p116)	TPTDCSGFDWS	53	GVQKSGSYNTSSSNSRI	11	WAMAMGDDALES	22	FCSHIF	(42)
PLRV[a]	(p119)	RPTDCSGFDWS	53	GVQKSGSYNTSSSNSRI	11	WMAMAGDDALES	22	FCSHIF	(43,44)
SBMV	(p105)	AEADISGFDWS	52	GIMKSGSYCTSSTNSRI	11	WCIAMGDDSVEG	34	FCSHIV	(9)
Consensus B		D SRFDQH G WS		G +MSGDoNTS N Lo K S A R		GDD oo E		FCQ S	
Supergroup C									
TBRV RNA 1	(p 92)	INCDYSGFDGL	52	CGLPSGFALTVVVNSVF	25	CLLVYGDDNLIS	45	FLKRKF	(36)
CPMV RNA 1	(p 87)	LCCDYSSFDGL	49	CGIPSGFPMTVIVNSIF	30	GLVTYGDDNLIS	45	FLKRTF	(59)
TVMV (NIb)	(p 57)	CDADGSQFDSS	50	KGNNSGQPSTVVDNTLM	23	KFFANGDDLIIA	41	ALSKDG	(65)
PPV (NIb)	(p 57)	CDADGSQFDSS	50	KGNNSGQPSTVVDNTLM	23	RYFVNGDDLVLA	42	VLYDDM	(60)
TEV (NIb)	(p 54)	SGWVYCDADGS	55	KGNNSGQPSTVVDNTLM	19	VYYVNGDDLLIA	41	ALERDG	(61)
Consensus C		D S FD		G SG P TVV N oo		oo GDD oo		L +	

Type I and type IV motifs were introduced (14) to propose that the amino acid sequence motifs of the polymerase of a double-stranded RNA birnavirus was related to the motifs of positive-strand RNA viruses.

Based on the amino acid sequence homology in the motif regions, putative RNA polymerases of the positive-sense RNA plant viruses were assigned into three different supergroups (A,B and C, Fig. 2), matching precisely with the classification scheme also proposed here for the nucleic acid helicases (Fig. 1). A consensus sequence was derived for each supergroup which is shown in the bottom line of each set in Fig. 2. The members of each supergroup not only show a high degree of sequence homology between their respective motifs, but also show a similar length in the amino acid sequence gap between their consecutive motifs.

In Supergroup A the high homology between the sequence motifs of the putative RNA polymerases of TMV and the tricornaviridae (AlMV and CMV) has been reported before (38,39). TRV and BSMV also enjoy a high homology in the sequence motifs of their putative RNA polymerases as do the pair of PVX and TYMV (Fig. 2).

As was observed with the sequence motifs of nucleic acid helicases (Fig. 1), two members of Subgroup 1 of the luteovirus group, i.e. SDV and BYDV (PAV serotype), show a high homology in the sequence motifs of their putative RNA polymerases with the members of subdivision B1, i.e. the carmoviruses and tombusviruses (Fig. 2, see also refs. 40 and 41). Of further interest is that, although BYDV infects monocots and SDV infects dicots, their amino acid sequence motifs, except for one change in motif II, are the same (Fig. 2).

On the other hand, the sequence motifs of the putative RNA polymerases of Subgroup 2 of the luteoviruses, i.e. BWYV and PLRV, show a high homology with the sequence motifs of the putative RNA polymerase of SBMV (42–44), a member of the sobemovirus group (Fig. 2, Table 1, B3). In addition, a significant homology occurs between the amino acid sequence motif I of the RNA polymerases of SBMV and those of the members of Supergroup C (Fig. 2). This implies that the evolutionary link between Supergroups B and C is, in fact, established via SBMV (see also Table 1 and Fig. 3). SBMV, on one hand, is related to Subgroup 2 of the luteoviruses (Table 1, B2; Figs. 1 and 2), and on the other to Supergroup C (Table 1, C1; Figs. 1–3).

A glycine residue is conserved in the first position of the sequence motif II of the RNA polymerase of the members of Supergroup B, while the same residue is conserved in the second position of this motif in Supergroup C (Fig. 2). Also, in Supergroup C, the conserved cysteine residue in position 2 of motif IV has been replaced by a leucine (Fig. 2). This holds true for the animal picornaviruses as well (14). The functional significance of these substitutions is not known.

Relative Position of Sequence Motifs of Nucleic Acid Helicases and RNA Polymerases in Plant RNA Virus Genomes

Helicase motifs are found in the RNA 1 of the tricornaviridae (Supergroup A, Table 1; AlMV and CMV), while the RNA polymerase motifs can be located in RNA 2 of these viruses (Fig. 3). In PVX the RNA polymerase motifs are flanked by the motifs of p166

Fig. 2. Conserved amino acid sequence motifs (I to IV) in the putative RNA polymerases coded by the RNAs of three supergroups (A, B and C) of plus-sense RNA plant viruses. The estimated M_r ($\times 10^{-3}$) of each putative protein (p) coded by viral RNAs is given in brackets. Numbers in the gaps between motifs are the number of amino acids contained within each gap. Each consensus line indicates residues which most commonly occur in that position. o, hydrophobic residues; +, polar residues. (?), partial sequence only obtained for SDV (58). See Table 1 for the list of virus abbreviations. [a] The C-terminal residue in motif III of the Scottish isolate of PLRV is A instead of S (44).

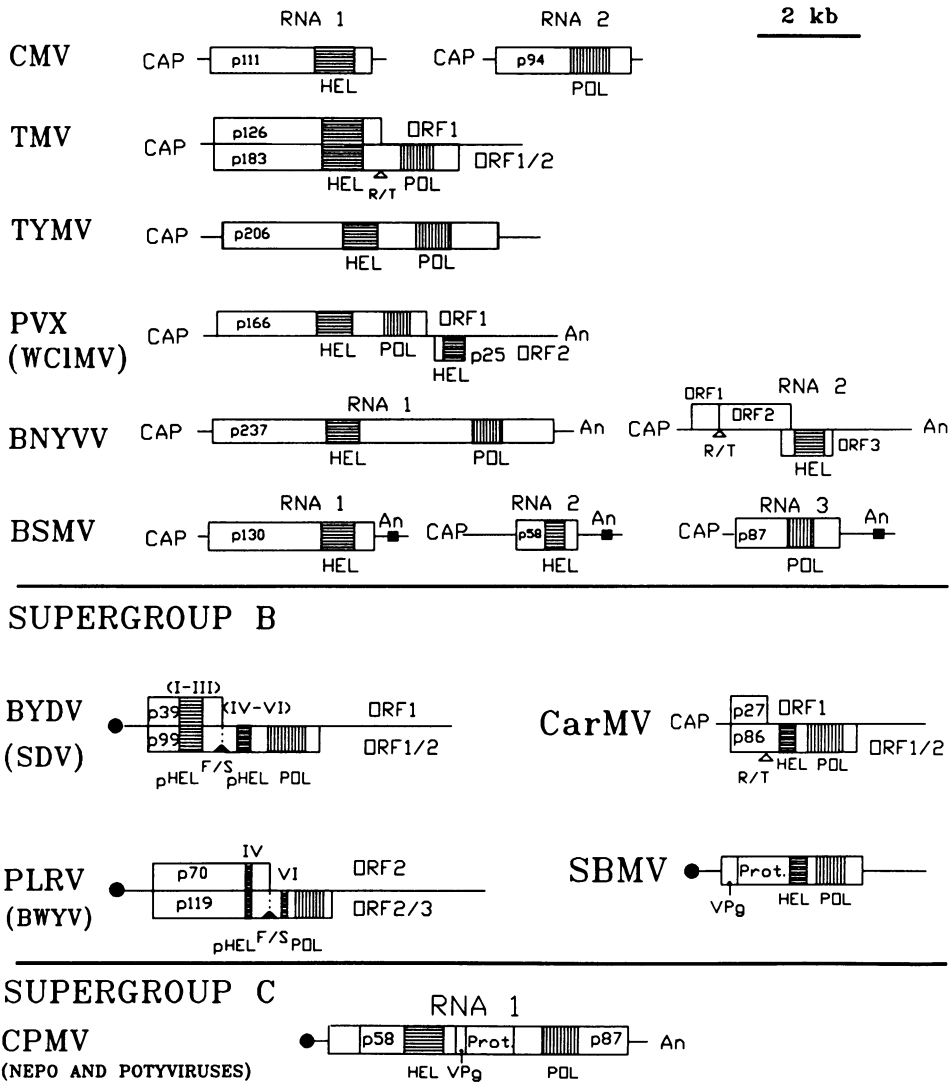

Fig. 3. Arrangement of sequence motifs of the putative nucleic acid helicases and RNA polymerases in the genomes of three supergroups (A, B and C) of positive-sense RNA plant viruses. Only ORFs displaying the conserved sequence motifs are depicted (open boxes). M_r ($\times 10^{-3}$) of each protein is given in each ORF. Positions of helicase (HEL) and polymerase (POL) motifs are indicated by horizontal and vertical hatches, respectively. CAP, 5′ terminal m^7 G-CAP structure; An, poly(A); where internal, indicated by a solid box; closed circles represent the position

and p25 helicases. In BSMV RNA 1 and RNA 2 each has a set of helicase motifs, while the RNA polymerase motifs are located in RNA 3 (Fig. 3). In the tobamoviruses (TMV; Fig. 3) and tobraviruses as in Sindbis virus (8), the nucleic acid helicase motifs are placed upstream from a readthrough termination codon and the RNA polymerase motifs are located after this codon (Table 1A; and Fig. 3, R/T).

In TYMV, PVX (p166 open-reading frame), BNYVV (RNA 1) and in the members of Supergroup B (Fig. 3), the nucleic acid helicase motifs occur upstream from the RNA polymerase motifs in the same open-reading frame in the absence of a readthrough termination codon (Fig. 3). Members of the luteovirus group are exceptional in that only the C-terminal motifs of the nucleic acid helicase (pHEL, Fig. 3) are in-frame with the RNA polymerase motifs.

The arrangement of sequence motifs of the putative helicases and RNA polymerases in Supergroup C, which is represented by CPMV in Fig. 3, is unique in that the protease and the VPg are located between these two sets of motifs. The VPg and the protease genes in the genome of SBMV (a member of Supergroup B, Table 1, Fig. 3) are located upstream of the nucleic acid helicase motifs (45,46).

Evolutionary Standpoint of the Luteoviruses

Based on the data presented here, members of the luteovirus group can be divided into two subgroups: Subgroup 1 which is evolutionarily linked to Supergroup A via the viruses in subdivision B1 (Table 1) and Subgroup 2 which is linked to Supergroup C via the viruses in subdivision B3 (Table 1, sobemovirus group). Members of the subdivision B1 were originally placed in the Sindbisvirus-like Supergroup (Table 1, Supergroup A) based on the presence of a 5'-terminal cap structure in their genome (3).

However, the following evidence supports that the carmoviruses and tombusviruses of subdivision B1 are related to Subgroup 1 of the luteoviruses:

(a) amino acid sequence motifs of their nucleic acid helicases are significantly homologous (Fig. 1, viruses listed 13−17, motifs IV and VI; Table 1).
(b) amino acid sequence motifs of their RNA polymerases are significantly homologous (Fig. 2, the top six viruses listed under Supergroup B).
(c) amino acid gap lengths between the sequence motifs of their RNA polymerases are similar (Fig. 2, the top six viruses listed under Supergroup B).

The following evidence supports the view that members of Subgroup 2 of the luteoviruses (Table 1, B2) are closely related to the sobemoviruses in subdivision B3.

(a) the amino acid sequence in motif VI of their nucleic acid helicases is significantly homologous (Fig. 1, BWYV, PLRV and SBMV).
(b) significant homology occurs between the amino acid sequence motifs of their RNA polymerases (Fig. 2, BWYV, PLRV and SBMV).
(c) the amino acid gap lengths between the sequence motifs of their RNA polymerases are similar (Fig. 2, BWYV, PLRV and SBMV). The only exception is the presence of a larger gap length between motifs III and IV of SBMV (Fig. 2).

of VPg. Positions in the RNAs at translational readthrough (R/T) and translational frameshift (F/S) are indicated by open and closed triangles, respectively. Part of the split helicase motif in the luteoviruses is shown by pHEL. Roman numbers indicate the position of each motif type in the helicase. Prot., protease; ORF, open reading frame. See Table 1 for the list of virus abbreviations.

In turn, based on the following evidence, the sobemoviruses of Supergroup B are related to Supergroup C, although nucleotide sequence data from other members of the sobemoviruses, when available, will further support these views.

(a) a basic amino acid residue (replacing the glycine) is present in the sixth position of sequence motif I of the nucleic acid helicases of SBMV, CPMV and TBRV, the latter two being members of Supergroup C (Fig. 1).
(b) sequence motif I of the RNA polymerase of SBMV is similar to motif I of the members of Supergroup C (Fig. 2).
(c) SBMV has a viral coded protease for processing of the polyprotein product (45,46), similar to the members of Supergroup C (Fig. 3).

The evidence listed above indicates that the luteoviruses provide an important evolutionary link between the Sindbisvirus-like Supergroup A and picornavirus-like Supergroup C.

ACKNOWLEDGEMENTS

This work was supported by a joint grant under the National Biotechnology Program Research Grants Scheme to the Department of Biochemistry, University of Adelaide, and the CSIRO Division of Plant Industry, Canberra. We thank Dr. W. Gerlach, Dr. A. Miller, and Dr. P. Waterhouse for cDNA clones of SDV and for plant material, Dr. P. Keese for many helpful discussions and Mrs. Ros Murrell for typing the manuscript.

*To whom correspondence should be addressed

REFERENCES

1. Matthews, R.E.F. (1982) Intervirology 17, 1–200.
2. Franssen, H., Leunissen, J., Goldbach, R., Lomonossoff, G. and Zimmern, D. (1984) EMBO Journal 3, 855–861.
3. Goldbach, R. and Wellink, J. (1988) Intervirology 29, 260–267.
4. Koenig, R. (1988) in: The Plant Viruses, vol. 3, Polyhedral Virions with Monopartite RNA Genomes (Koenig, R. Ed.) pp. 1–22. Plenum Press. New York and London.
5. Hodgman, T.C. (1988) Nature 333, 22–23.
6. Hodgman, T.C. (1988) Nature 333, 578.
7. Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P and Blino, V.M. (1988) FEBS Lett. 235, 16–24.
8. Strauss, J.H. and Strauss, E.G. (1988) Ann. Rev. Microbiol. 42, 657–683.
9. Wu, S., Rinehart, C.A. and Kaesberg, P. (1987) Virology 161, 73–80.
10. Gorbalenya, A.E., Blinov, V.M., Donchenko, A.P. and Koonin, E.V. (1989) J. Mol. Evol. 28, 256–268.
11. Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1989) Nucleic Acids Res. 17, 4713–4730.
12. Linder, P., Lasko, P.F., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J. and Slonimski, P.P. (1989) Nature 337, 121–122.
13. Lane, D. (1988) Nature 334, 478.
14. Gorbalenya, A.E. and Koonin, E.V. (1988) Nucleic Acids Res. 16, 7735.
15. Miller, W.A., Waterhouse, P.M. and Gerlach, W.L. (1988) Nucleic Acids Res. 16, 6097–6111.
16. Blumenthal, T. and Carmichael, G.G. (1979) Ann. Rev. Biochem. 48, 528–548.
17. Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) EMBO Journal 1, 945–951.
18. Evans, R.K., Haley, B.E. and Roth, D.A. (1985) J. Biol. Chem. 260, 7800–7804.
19. Gustafson, G., Armour, S.L., Gamboa, G.C., Burgett, S.G. and Shepherd, J.W. (1989) Virology 170, 370–377.
20. Nishi, K., Morel-Deville, F., Hershey, J.W.B., Leighton, T. and Schnier, J. (1988) Nature 336, 496–498.
21. Hirling, H., Scheffner, M., Restle, T. and Stahl, H. (1989) Nature 339, 562–564.
22. Lasko, P.F. and Ashburner, M. (1988) Nature 335, 611–617.
23. Ford, M.J., Anton, I.A. and Lane, D.P. (1988) Nature 332, 736–738.
24. Forster, R.L.S., Bevan, M.W., Harbison, S.A. and Gardner, R.C. (1988) Nucleic Acids Res. 16, 291–303.
25. Nassuth, A., Alblas, F. and Bol, J.F. (1981) J. Gen. Virol. 53, 207–214.

26. Hall, T.C., Miller, W.A. and Bujarski, J.J. (1982) in: Advances in Plant Pathology (Ingram, D.S. and Williams, P.H. eds). vol. 1, pp. 179–211.
27. Goldbach, R., Rezelman, G. and van Kammen, A. (1980) Nature 286, 297–300.
28. Bramhill, D. and Kornberg, A. (1988) Cell 54, 915–918.
29. Lamb, R.A. and Dreyfuss, G. (1989) Nature 337, 19–20.
30. Bove, J.M., Bove, C. and Mocquot, B. (1968) Biochem. and Biophys. Res. Commun. 32, 480–486.
31. Habili, N. and Kaper, J.M. (1981) Virology 112, 250–261.
32. Gargouri, R., Joshi, R.L., Bol, J.F., Astier-Manifacier, S. and Haenni, A-L. (1989) Virology 171, 386–393.
33. Jaspars, E.M.J., Gill, D.S. and Symons, R.H. (1985) Virology 144, 410–425.
34. Nitta, N., Takanami, Y., Kuwata, S. and Kubo, S. (1988) J. Gen. Virol. 69, 2695–2700.
35. Kamer, G. and Argos, P. (1984) Nucleic Acids Res. 12, 7269–7282.
36. Greif, C., Hemmer, O. and Fritsch, C. (1988) J. Gen. Virol. 69, 1517–1529.
37. Argos, P. (1988) Nucleic Acids Res. 16, 9909–9916.
38. Rezaian, M.A., Williams, R.H.V., Gordon, K.H.J., Gould, A.R. and Symons, R.H. (1984) Eur. J. Biochem. 143, 277–284.
39. Haseloff, J., Goelet, P., Zimmern, D., Ahlquist, P., Dasgupta, R. and Kaesberg, P. (1984) Proc. Natl. Acad. Sci. USA 81, 4358–4362.
40. Nutter, R.C., Scheets, K., Panganiban, L.C. and Lommel, S.A. (1989) Nucleic Acids Res. 17, 3163–3177.
41. Rochon, D. and Tremaine, J.H. (1989) Virology 169, 251–259.
42. Veidt, I., Lot, H., Leiser, M., Scheidecker, D., Guilley, H., Richards, K. and Jonard, G. (1988) Nucleic Acids Res. 16, 9917–9932.
43. van der Wilk, F., Huisman, M.J., Cornelissen, B.J.C., Huttinga, H. and Goldbach, R. (1989) FEBS Lett 245, 51–56.
44. Mayo, M.A., Robinson, D.J., Jolly, C.A. and Hyman, L. (1989) J. Gen. Virol. 70, 1037–1051.
45. Gorbalenya, A.E., Koonin, E.V., Blinov, V.M. and Donchenko, A.P., (1988) FEBS Lett. 236, 287–290.
46. Gorbalenya, A.E., Donchenko, A.P., Blinov, V.M. and Koonin, E.V. (1989) FEBS Lett. 243, 103–114.
47. Cornelissen, B.J.C., Brederode, F.T., Moormann, J.M.R. and Bol, J.F. (1983) Nucleic Acids Res. 11, 1253–1265.
48. Rezaian, M.A., Williams, R.H.V. and Symons, R.H. (1985) Eur. J. Biochem. 150, 331–339.
49. Goelet, P., Lomonossoff, G.P., Butler, P.J.G., Akam, M.E., Gait, M.J. and Karn, J. (1982) Proc. Natl. Acad. Sci. USA 79, 5818–5822.
50. Hamilton, W.D.O., Boccara, M., Robinson, D.J. and Baulcombe, D.C. (1987) J. Gen. Virol. 68, 2563–2575.
51. Gustafson, G. and Armour, S.L. (1986) Nucleic Acids Res. 14, 3895–3909.
52. Bouzoubaa, S., Quillet, L., Guilley, H., Jonard, G. and Richards, K. (1987) J. Gen. Virol. 68, 615–626.
53. Bouzoubaa, S., Ziegler, V., Beck, D., Guilley, H., Richards, K. and Jonard, G. (1986) J. Gen. Virol. 67, 1689–1700.
54. Morch, M., Boyer, J. and Haenni, A. (1988) Nucleic Acids Res. 16, 6157–6173.
55. Huisman, M.J., Linthorst, H.J.M., Bol, J.F. and Cornelissen, B.J.C. (1988) J. Gen. Virol. 69, 1789–1798.
56. Finch, P.W., Storey, A., Chapman, K.E., Brown, K., Hickson, I.D. and Emmerson, P.T. (1986) Nucleic Acids Res. 14, 8573–8582.
57. Guilley, H., Carrington, J.C., Balazs, E., Jonard, G., Richards, K., and Morris, T.J. (1985) Nucleic Acids Res. 13, 6663–6677.
58. Habili, N., Karageorgos, L.E. and Symons, R.H., unpublished.
59. Lomonossoff, G.P. and Shanks, M. (1983) EMBO Journal 2, 2253–2258
60. Maiss, E., Timpe, U., Brisske, A., Jelkmann, W., Casper, R., Himmler, G., Mattanovich, D. and Katinger, H.W.D. (1989) J. Gen. Virol. 70, 513–524.
61. Allison, R., Johnston, R.E. and Dougherty, W.G. (1986) Virology 154, 9–20.
62. Robaglia, C., Durand-Tardif, M., Tronchet, M., Boudazin, G., Astier-Manfacier, S. and Casse-Delbart, F. (1989) J. Gen. Virol. 70, 935–947.
63. Cornelissen, B.J.C., Brederode, F.T., Veeneman, G.H., van Boom, J.H. and Bol, J.F. (1983) Nucleic Acids Res. 11, 3019–3025.
64. Gustafson, G., Hunter, B., Hanau, R., Armour, S.L. and Jackson, A.O. (1987) Virology, 158, 394–406.
65. Domier, L.L., Franklin, K.M., Shahabuddin, M., Hellmann, G.M., Overmeyer, J.H., Hiremath, S.T., Siaw, M.F.E., Lomonossoff, G.P., Shaw, J.G. and Rhoads, R.E. (1986) Nucleic Acids Res. 14, 5417–5430.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.