Relationships among the positive strand and double-strand RNA viruses as viewed through their RNA-dependent RNA polymerases

Jeremy A. Bruenn

Department of Biological Sciences, State University of New York at Buffalo, Buffalo, NY 14260, USA

Received August 17, 1990; Revised and Accepted November 28, 1990

EMBL accession no. X54405

ABSTRACT

The sequences of 50 RNA-dependent RNA polymerases (RDRPs) from 43 positive strand and 7 double strand RNA (dsRNA) viruses have been compared. The alignment permitted calculation of distances among the 50 viruses and a resultant dendrogram based on every amino acid, rather than just those amino acids in the conserved motifs. Remarkably, a large subgroup of these viruses, including vertebrate, plant, and insect viruses, forms a single cluster whose only common characteristic is exploitation of insect hosts or vectors. This similarity may be due to molecular constraints associated with a present and/or past ability to infect insects and/or to common descent from insect viruses. If common descent is important, as it appears to be, all the positive strand RNA viruses of eucaryotes except for the picornaviruses may have evolved from an ancestral dsRNA virus. Viral RDRPs appear to be inherited as modules rather than as portions of single RNA segments, implying that RNA recombination has played an important role in their dissemination.

INTRODUCTION

The evolution of RNA-dependent RNA polymerases (RDRPs) is of great interest, since these are probably among the first enzymes. They are also encoded by all known non-satellite RNA viruses, except the retroviruses, in which they appear in the metamorphosis of the reverse transcriptase (1). They are thus the best ruler for molecular taxonomy of the RNA viruses. Although the secondary and tertiary structure of none of these is known, many primary structures have been deduced from genomic sequences. The RDRPs of several positive strand viruses were first identified by sequence similarity to the known poliovirus polymerase by Kamer and Argos (2). One conserved motif, with a glycine-aspartic-aspartic (GDD) motif, was also noted in Q β replicase and in the reverse transcriptases of the retroviruses. The availability of many more sequences of RNA viral genomes has enabled a number of investigators to identify several conserved motifs in the RDRPs of positive strand, negative strand, and double-strand RNA viruses (1, 3, 4, 5, 6, 7, 8, 9) and in the reverse transcriptases (1, 5). Most of the RDRP motifs have, in fact, been identified by primary sequence similarity with the known poliovirus RDRP, although there is

genetic evidence identifying the RDRP coding sequences of the segmented positive strand and dsRNA viruses (10, 11, 12, 13) and biochemical evidence identifying some of the dsRNA RDRP coding motifs as well (14, 15).

In a general sense, the genomic organization and replication strategy of the RNA viruses is reflected in the similarities among their polymerases. The positive strand viruses form one easily recognized class, the dsRNA viruses another, the negative strand RNA viruses a third, and the retroviruses a fourth (1). Comfortingly, this follows the general practice of viral taxonomy, which is to assign viruses to classes based primarily on genome structure, as well as on morphology, host range, and mode of infection (16). The positive strand and the dsRNA viruses are recognizably similar in RDRP sequence, whereas the negative strand RNA viruses and retroviruses are both distinctly different (1; this work). This corresponds with suggestions that RNA viruses represent several independent lines of evolution (17).

Since Kamer and Argos (2), all of the sequence comparisons have been limited to several conserved motifs. In contrast, the present comparison is of the entire sequences of a set of the known RDRPs throughout their most conserved region, the carboxyterminal 200 amino acids. The sequences chosen are 50 representative sequences, including 43 positive strand RNA virus RDRPs and all 7 of the dsRNA virus RDRPs available. Mutational distances for each pairwise combination have been calculated and a dendrogram constructed for this group of 50 viruses. Alignment of the reverse transcriptases and the RDRPs of the negative strand RNA viruses is outside the scope of the present work.

While 'Virus taxonomy at its present stage has no evolutionary or phylogenetic implications', (16) describing evolutionary relationships should clearly be the goal of the field. The dendrogram constructed here reflects both the similarities among positive strand and dsRNA viruses due to homology (common descent) and homoplasy (convergent and parallel evolution). If shared ancestry accounts for the largest proportion of the observed similarity in the alignments, much of present taxonomy may be rooted in real evolutionary relationships. There are exceptions. For instance, hepatitis A virus is not an enterovirus; in fact, it probably belongs to a new genus of picornavirus. The flaviviruses and alphaviruses should not be in the same family. In general, the relationships among RDRPs follow the division of procaryotic and eucaryotic host, but not the separation of plant, animal, and fungal cells. One large cluster of viruses includes those that infect

plants, vertebrates, and/or insects. This may be the result of specialization of insect viruses to infect host species on which insects are pathogens. All the positive strand RNA viruses of this supergroup may have evolved from an ancestral dsRNA virus.

MATERIALS AND METHODS

Sequence of ScV-La

A partial sequence of the carboxy-terminal 241 amino acids of the ScV-La RDRP was deduced from the sequence of a 737 bp cDNA clone constructed and sequenced as described for ScV-L1 clones (7).

Sequence alignments

Closely similar sequences were aligned by GAP (18). Less similar groups of sequences were aligned by eye, using the program LINEUP (18). Alignments were evaluated by GAP and by RDF (19). GAP is similar to ALIGN; both use the Needleman and Wunsch algorithm (18, 19). The alignment of Fig. 1 assumes similarity classes of the following: G,P; I, M, C, L, A, V; Y, W, F, H; T, S; N, Q, D, E; R, K.

Dendrograms

Pairwise distances were calculated by DISTANCES from the PRETTY reconfiguration of the LINEUP alignments (18). DISTANCES calculates an average distance for each pair of sequences by counting the sum of matches at each aligned position in which the comparison value is greater than a match threshold and dividing by the total number of aligned residues. Distances to the ScV-La RDRP and to the HCV RDRP, for which only partial sequences are available, were calculated only within the sequenced regions and compared to the results for the entire data set; no significant differences were found. Dendrograms were constructed by the join algorithm of the cluster analysis program of SYSTAT, using the Euclidian distance option (20), from the pairwise distances provided by DISTANCES.

Published sequence data

Published sequences were obtained from Genbank, EMBL, or NBRF. The sequences compared and their origins, classification, and abbreviations are shown in Table 1. Sequence file names are given to facilitate similar comparisons by others.

RESULTS

Conserved motifs

The 50 viruses chosen and their abbreviations are listed in Table 1. The alignment of the 50 RDRPs is shown in Fig. 1. Alignment of very similar RDRPs, such as those of Yfv and Wnv, ScV-L1 and ScV-La, the picornaviruses, or Sinv and Tmv was performed by the program GAP (18). Less similar classes were then aligned by eye, preserving the alignment of the 21 most highly conserved residues and the 8 conserved motifs. The first four of these motifs: the acidic motif beginning at position 1 (motif 1), the SG.T motif about position 100 (motif 2), the GDD motif around position 160 (motif 3), and the basic motif around position 280 (motif 4) have all been noted previously (1), and the first three of these were also identified independently within the plant viruses (6). Motif 5 has a consensus sequence FCG at position 303 and is the same as motif IV of Habili and Symons (6). Motif 6 has the consensus LKR at position 381; motif 7 is a basic sequence preceded by an aromatic residue at position 480; and motif 8 is an aromatic

residue preceded by a basic sequence at position 514. These last three motifs have not been identified before, although some are evident in the original Kamer and Argos alignment (2). The conserved region of the viral RDRP corresponds well to the region of Qb replicase required for enzymatic activity as determined by in vitro mutagenesis and expression of cDNA clones (67). The conserved region begins 104 amino acids C-terminal to the most N-terminal lethal mutation in Qb replicase (outside its dispensable N-terminal region) and ends 3 amino acids after the most C-terminal lethal mutation mapped (67).

Alignment of proteins with little sequence similarity is bound to be somewhat arbitrary, but some of the conserved domains demonstrated in Fig. 1 are similar to those of other investigators. The first four motifs are those of Poch et al (1), in which 1 is A, 2 is B, etc. The alignment at motif 5 is the same as that of all the plant viruses at motif IV of Habili and Symons (6), except for that of Tbrv and Cpmv, which are aligned somewhat differently in this region in the scheme of Fig. 1. Since none of the viruses in the potyvirus-like group conserve motif 5, agreement on the alignment of the RDRPs of the remaining members of this group in this region (Tvmv, Ppv, and Tev) is surprising.

The dsRNA viruses

The alignment of the dsRNA virus RDRPs followed the alignment of the ScV-L1 RDRP, in which three of the first four conserved motifs were identified previously (4, 7). Although the original publications of the Phi6 segment L and reovirus segment L1 sequences did not find similarity to other viral RDRPs (13, 57), there is genetic evidence that these segments do encode the viral RDRPs (11, 13). For BTV (15) and ScV-L1 (14), there is biochemical evidence identifying the aligned regions as RDRPs. The alignments arrived at for the dsRNA viral RDRPs agree fairly well with those of Koonin et al (9). Within the conserved motifs, even these viral RDRPs, although distinct from those from other classes, are well conserved. For instance, of the 21 most highly conserved residues, reovirus RDRP retains 17, while poliovirus RDRP retains 19. Conserved residues are those in which similarity is preserved (see Materials and Methods). The only glaring non-similarity between the dsRNA viral RDRPs and those of the positive strand viral RDRPs is the S of phi6 and the I of Ibdv in the place of G in the GDD of motif 3. This position also varies among the negative strand RNA viruses (1). Five of the dsRNA viruses also have a region of about 50 amino acids between motifs 3 and 4 and another five have a region of 19 to 50 amino acids between motifs 5 and 6 that is missing in the positive strand RNA viruses. The fungal dsRNA viruses differ from the other dsRNA viruses in having a maximum length for both these unique regions.

The most highly conserved region of the RDRP is motif 2, beginning at the S residue at position 98 in the alignment of Fig. 1. This region is well conserved in all of the dsRNA viral RDRPs. Each of the conserved domains has a conserved predicted secondary structure (1). The secondary structure predicted for this motif among some of the RNA polymerases at large, including only two of the dsRNA viral RDRPs, is a turn at the SG both preceded and followed by beta sheet (1). This is precisely what is predicted for the 6 dsRNA viral RDRPs whose sequences are known in this region (Fig. 2), in which prediction of a turn at the SG position reaches 100% (versus 50% for viral polymerases at large), and beta sheets are predicted at 50% preceding and 83% following the turn (compared to 50% and 50% for RNA viruses at large).

1	50
Carmv	DMS.RF..DQ HVSVAALFE HSCYLACFE. ...GDAHLAN LLMQVLVNHG
Bydv	DAS.RF..DQ HVSQAALKE HGIYNGIF. ...GDSEMAI ALZHQITNNI
Plrv	DCS.GF..DW SVAYMLED MEVRNRLTF. ...NNTQTLER LRAAMKICIG
Bwv	DCS.GF..DW SVADNMLED MIVNRRLTI. ...DIMPATER LRSCLWLCIS
Cnv	DAS.RF..DQ HCSVEALQFE HGFYRAMYP. ...GNKLLSK LLENQHLHNG
Cyrv	DAS.RF..DQ HCSVEALQFE HSFYRALYP. ...GNKLLSK LLENQHLHNG
Mcmv	DAS.RF..DQ HVSVEALRWE HVSYSRIYG. ...YPELLTQ LHRVHLKTVS
Pvx	DYT.AF..DQ SQDGAMQFE VIKAKHCCI. ...PEEIIQ AYIDIKITNAQ
Tymv	DYT.AF..DQ SQDGAMQFE VIKAKHCCI. ...PEEIIQ AYIDIKITNAQ
Yfv	DTA.GW..DT RITEADLDE QELILNYM. ...SPHKKK LAQAVMENTY
Wnv	DTA.GW..DT RITEADLDE QELILNYM. ...SPHKKK LAQAVMENTY
Dengue	DTA.GW..DT RITEADLDE QELILNYM. ...SPHKKK LAQAVMENTY
Jev	DTA.GW..DT RITEADLDE QELILNYM. ...SPHKKK LAQAVMENTY
Hcv	DTK.AW..DT QVTTRDLELI RDQKFY. ...FKKK WHKFIDITLK
Bbv	EVI.ET..DF SNLDGRVSSW MORN.IAQ KAMVO.AFRP
Tev	DGS.QF..DS SLTPFLINAV LKVR.LAF MEEM..DIG
Ppv	DGS.QF..DS SLTPFLINAV LKVR.LAF MEEM..DIG
Tbrv	DYS.GF..DG LLNAQVIECI AKMINRL.YALSG ESEV..QQA
Cpmv	DYS.SF..DG LLSKQVMDVI ASMINEL.CGG EDQL..KNA
Tmv	DGS.QF..DS SLTPFLINAV LKVR.LSS MEEM..DVG
Trv	DMS.KF..DK SA...NRF HLO.LQL EIRY..LFG
Bnyvv	DAA.AC..DS GO...GVF TOL.IE RHIA..ALG
Tmv	DYS.LL..LEN VD...LF VKR.VE EIMR..RLG
Mhv	DYS.LL..LEN VD...LF VKR.VE EIMR..RLG
Bsmv	DYS.KF..DK SK...TGLH IKA.VIG LYKL..FG
Almv	DYS.KF..DK SK...TGLH IKA.VIG LYKL..FG
Bmv	DLS.KF..DK SQ...GEL HLE.FQR EILL..ALG
Cucmv	DLS.KF..DK SQ...GEL HLE.FQR EILL..ALG
Sinv	DIA.SF..DK SQ...DDS LAL.TGL MILE..DLG
Sfv	DIA.SF..DK SQ...DDS LAL.TGL MILE..DLG
Onv	DIA.SF..DK SQ...DDS LAL.TGL MILE..DLG
Rrv	DIA.SF..DK SQ...DDS LAL.TGL MILE..DLG
Rot	DVS.QW..DS SQHNTQ.PFR GIIIMGLDM ANMTDARV.QTLN..LYKO
Reo	DASITW..DF FISVIMAAIH EGVASSSIG PFMGVPSIV.NDES..VVG
Ph16	DLI..C..DE LLMQY..A PWW VKLFET SKLPLVYVGA.PAPE..QG
Ibdv	DLE.KG..EA NCTQHQMAA MYIL.TRG
Btv	DAYIRL..DE SE...RD QGSEKVPKGV LPVSSVDVAN.RIAV..DKG
Scv11	DGASSCFDY DD...FNS QHISASMYTV LCA.FRDTFS.RNMS..DEQ
Scv1a	DYT.GY..DA SLSPAMFEAL.KMY LEKI..G
Polio	DYS.NF..DS THSVAMFRLL.AEFF TPEN..G
Emc	DYS.NF..DS THSVAMFRLL.AEFF TPEN..G
Tmev	DYS.NF..DS THSVAMFRLL.AEFF TPEN..G
Hrv14	DYS.NF..DA SLSPVWFVCL.EKK LTKL..G
Fndv	DYS.AF..DA NHCS...DAM.NIM FEVFTDFG
Coxv	DYS.GY..DA SLSPVWFVCL.KML LE..KLG
Hav	DYS.AF..DA SLSPVWFVCL.GRIM SE..LSG
Qbeta	DLSAAS..D. SISLALCELL L.P
Sp	DLSAAS..D. SISLALCELL L.P
Ms2	DLSAAS..D. SISLALCELL L.P
Ga	DLSAAS..D. SISLALCELL L.P

1

101	150
Carmv	DMS.RF..DQ HVSVAALFE HSCYLACFE. ...GDAHLAN LLMQVLVNHG
Bydv	DAS.RF..DQ HVSQAALKE HGIYNGIF. ...GDSEMAI ALZHQITNNI
Plrv	DCS.GF..DW SVAYMLED MEVRNRLTF. ...NNTQTLER LRAAMKICIG
Bwv	DCS.GF..DW SVADNMLED MIVNRRLTI. ...DIMPATER LRSCLWLCIS
Cnv	DAS.RF..DQ HCSVEALQFE HGFYRAMYP. ...GNKLLSK LLENQHLHNG
Cyrv	DAS.RF..DQ HCSVEALQFE HSFYRALYP. ...GNKLLSK LLENQHLHNG
Mcmv	DAS.RF..DQ HVSVEALRWE HVSYSRIYG. ...YPELLTQ LHRVHLKTVS
Pvx	DYT.AF..DQ SQDGAMQFE VIKAKHCCI. ...PEEIIQ AYIDIKITNAQ
Tymv	DYT.AF..DQ SQDGAMQFE VIKAKHCCI. ...PEEIIQ AYIDIKITNAQ
Yfv	DTA.GW..DT RITEADLDE QELILNYM. ...SPHKKK LAQAVMENTY
Wnv	DTA.GW..DT RITEADLDE QELILNYM. ...SPHKKK LAQAVMENTY
Dengue	DTA.GW..DT RITEADLDE QELILNYM. ...SPHKKK LAQAVMENTY
Jev	DTA.GW..DT RITEADLDE QELILNYM. ...SPHKKK LAQAVMENTY
Hcv	DTK.AW..DT QVTTRDLELI RDQKFY. ...FKKK WHKFIDITLK
Bbv	EVI.ET..DF SNLDGRVSSW MORN.IAQ KAMVO.AFRP
Tev	DGS.QF..DS SLTPFLINAV LKVR.LAF MEEM..DIG
Ppv	DGS.QF..DS SLTPFLINAV LKVR.LAF MEEM..DIG
Tbrv	DYS.GF..DG LLNAQVIECI AKMINRL.YALSG ESEV..QQA
Cpmv	DYS.SF..DG LLSKQVMDVI ASMINEL.CGG EDQL..KNA
Tmv	DGS.QF..DS SLTPFLINAV LKVR.LSS MEEM..DVG
Trv	DMS.KF..DK SA...NRF HLO.LQL EIRY..LFG
Bnyvv	DAA.AC..DS GO...GVF TOL.IE RHIA..ALG
Tmv	DYS.LL..LEN VD...LF VKR.VE EIMR..RLG
Mhv	DYS.LL..LEN VD...LF VKR.VE EIMR..RLG
Bsmv	DYS.KF..DK SK...TGLH IKA.VIG LYKL..FG
Almv	DYS.KF..DK SK...TGLH IKA.VIG LYKL..FG
Bmv	DLS.KF..DK SQ...GEL HLE.FQR EILL..ALG
Cucmv	DLS.KF..DK SQ...GEL HLE.FQR EILL..ALG
Sinv	DIA.SF..DK SQ...DDS LAL.TGL MILE..DLG
Sfv	DIA.SF..DK SQ...DDS LAL.TGL MILE..DLG
Onv	DIA.SF..DK SQ...DDS LAL.TGL MILE..DLG
Rrv	DIA.SF..DK SQ...DDS LAL.TGL MILE..DLG
Rot	DVS.QW..DS SQHNTQ.PFR GIIIMGLDM ANMTDARV.QTLN..LYKO
Reo	DASITW..DF FISVIMAAIH EGVASSSIG PFMGVPSIV.NDES..VVG
Ph16	DLI..C..DE LLMQY..A PWW VKLFET SKLPLVYVGA.PAPE..QG
Ibdv	DLE.KG..EA NCTQHQMAA MYIL.TRG
Btv	DAYIRL..DE SE...RD QGSEKVPKGV LPVSSVDVAN.RIAV..DKG
Scv11	DGASSCFDY DD...FNS QHISASMYTV LCA.FRDTFS.RNMS..DEQ
Scv1a	DYT.GY..DA SLSPAMFEAL.KMY LEKI..G
Polio	DYS.NF..DS THSVAMFRLL.AEFF TPEN..G
Emc	DYS.NF..DS THSVAMFRLL.AEFF TPEN..G
Tmev	DYS.NF..DS THSVAMFRLL.AEFF TPEN..G
Hrv14	DYS.NF..DA SLSPVWFVCL.EKK LTKL..G
Fndv	DYS.AF..DA NHCS...DAM.NIM FEVFTDFG
Coxv	DYS.GY..DA SLSPVWFVCL.KML LE..KLG
Hav	DYS.AF..DA SLSPVWFVCL.GRIM SE..LSG
Qbeta	DLSAAS..D. SISLALCELL L.P
Sp	DLSAAS..D. SISLALCELL L.P
Ms2	DLSAAS..D. SISLALCELL L.P
Ga	DLSAAS..D. SISLALCELL L.P

2

51	100
Carmv	VGFA...S NGMLRYTKE.GCRMS.G
Bydv	KMFV...E DKMLRFKVR.GHRMS.G
Plrv	NSVLCI...S DGTLLAQTV.PGVQK.S.G
Bwv	NSVLCI...S DGTLLAQTV.PGVQK.S.G
Cnv	KGIV...P DGTITTYKE.GCRMS.G
Cyrv	KGIV...P DGTITTYKE.GCRMS.G
Mcmv	TAYA...S DGAFNVQVD.GCRMS.G
Pvx	IFL...G TLS.IMRIT.G
Tymv	TOF...G PLT.CMRIT.G
Yfv	KNKVVKVLR APGKAYMD.VI SRRDQGS.G
Wnv	RHKVVKVLR AADGKTVM.VI SRRDQGS.G
Dengue	QNKVVKVLR TPRG.TVMD.II SRRDQGS.G
Jev	RHKVVKVLR AAEKGTVM.VI SRRDQGS.G
Hcv	HMSEVPVISA DGEVYIRKG.ORGS.G
Bbv	...EYRD.MDTIINCSA KAKRF.GFRY EPGVGVKS.G
Tev	EQ...MLRN LYT.EIVYT.PI LTP.DGTIHK KHGN.NS.G
Ppv	EQ...MLSN LYT.EIVYT.PI ATP.DGTIVK KFGN.NS.G
Tbrv	QR...YMLMA LV.GRYAFV.GPEVYK VNCGL.PS.G
Cpmv	RR...NLLMA CCS.RLAIC.KNTVMR VECGT.PS.G
Tmv	QK...MLQN LYT.EIVYT.PI STP.DGTIVK KFGN.NS.G
Trv	LD...EMAA FLMEVSHST.QTTV RDI.QNGMMA HIMYQOKS.G
Bnyvv	IS...DFFL DMYFSFRE.KYVM Q..SRYVRA HMYVYKTS.G
Tmv	FE...DFFL EVWKQGRH.KTLT KDYTA.GIKT CIWYOKRS.C
Mhv	DG...LVEPLL LD...GLVP RSYLYLKS.G
Bsmv	LD...GLLKV LWEKSOYO.TYV KDRNF.GLEA YLLYQOKS.G
Almv	IPN...EPLT LWF.NAHR.KSRI SDS.KNGVFF NVDFORRT.G
Bmv	FP...APLT LWF.NAHR.DSYL SDP.HAKVGM SVSFORRT.G
Cucmv	CP...APIT KMWCDPFR.FSYI KDK.RAGVGM PISFORRT.G
Sinv	VD...QPLL DLIECAF.EISS THL.PGTTRF KFGAMMS.G
Sfv	VD...QPLL DLIECAF.EISS THL.PGTTRF KFGAMMS.G
Onv	VD...QPLL DLIECAF.EISS THL.PGTTRF KFGAMMS.G
Rrv	VD...QPLL DLIECAF.EISS THL.PGTTRF KFGAMMS.G
Rot	TOINLMS.YOI.PDGVNKK IQYGAVAS.G
Reo	VRARIPISGM QMIQHLISKL YKRGFSYRVN DSFSFGNDET HMTTTFPS.G
Ph16	HTLGDPS.N PDL.EVLSS.G
Ibdv	WSNDGDPH.F NCTWASFAMN IAPALVDSS CLI..MNIQ IKSYQOGS.G
Btv	FDLIAATDG SDA.LIDTHLS.G
Scv11	AE...MHWVC.ESVR HMVLDPT.KEWYR..LQGTLLS.G
Scv1a	FG...DRV DYIDYLNHS.HLYKNTYK VKGGM.PS.G
Polio	FG...PLTR EYLESIAIS.THAFEEKRFL ITGGL.PS.G
Emc	FD...RRIA EYLSLAIS.RHAYEDRRVL IRGGL.PS.G
Tmev	FD...G..SLIQSCINT.HHIFDEIYV VEGGM.PS.G
Hrv14	FA...PNAE WILKTLVNT.EHAYENKRIT VEGGM.PS.G
Fndv	YH...THKET NYIDYLCNS.HHLYRNKHYF VRGGM.PS.G
Coxv	TP...SHEGTALI NTIY..S.KHLYNCCYH VCGSM.PS.G
Hav	PGWFEVMD.LRSP.KGRLPDG.SV VYCKISSMG
Qbeta	PEWYDLITD.LRSD.EGILPDG.RV VYCKISSMG
Sp	PEWYDLITD.LRSD.EGILPDG.RV VYCKISSMG
Ms2	PEWYDLITD.LRSD.EGILPDG.RV VYCKISSMG
Ga	PEWYDLITD.LRSD.EGILPDG.RV VYCKISSMG

2

151	200
Carmv	...SRLI NNGDDCVLIC ERTDI.
Bydv	...GVAELC NNGDDCVIIT DRAN.
Plrv	...WAM AMGDDALEAP NSDL.
Bwv	...WAM AMGDDALEAP NSDL.
Cnv	...GINEYSLA NCGDDCVLIV ERRNL.
Cyrv	...GINEYSLA NCGDDCVLIV ERRNL.
Mcmv	...GI.PARLI NNGDDNVLIC PAVEV.
Pvx	...AGTAQV YAGDSDALDC.VPEV.
Tymv	...SCPIV VSGDDSLIDH PLPT.
Yfv	HGCDRLKRNA VSGDDCVVRP IDDR.
Wnv	NGEERLSRMA VSGDDCVVRP LDDR.
Dengue	VGRERLSRMA VSGDDCVVRP LDDR.
Jev	NGEERLSRMA VSGDDCVVRP LDDR.
Hcv	...DRVAKIH VCGDDGFLI.TER.
Bbv	EDLFRILG.P KCGDDGLSRA IIQK.SI NR.AAKCF.
Tev	EIVV...YY VNGDDLLIAI HPDKA.
Ppv	DTDCICRYF VNGDDLVAV HPAYE.
Tbrv	NRFNQVQVCLL VYGDNDLISV SPSIA.
Cpmv	QSFDKLIGLV TYGDNDLISV NAVVT.
Tmv	NSQEDVCKFF ANGDDLLIAI SPELE.
Trv	KAV...MVT YGDDSLIAF PRGT.
Bnyvv	GPF...CMA MKGDDGFKRQ ANLK.
Tmv	KIHK...GA FCGDDSLYF PKGC.
Mhv	...
Bsmv	DAH...FCV FGGDDSLIF DQGY.
Almv	PNVK...FVV ASGDDSLIGT VEEL.
Bmv	DCD...CAI FSGDDSLII SKVK.
Cucmv	QFD...RLL FSGDDSLAF SKLP.
Sinv	ERLKTSCAA FIGDDNIH.GVSDK.
Sfv	QRLTSCAA FIGDDNIH.GVSDK.
Onv	ERLKTSCAA FIGDDNIH.GVSDK.
Rrv	ERLKTSCAA FIGDDNIH.GVSDK.
Rot	KYSFATKIR VCGDDNYAVL QNTEVTEQM VQDVSDVRE TYARM.NAK
Reo	KSLTIQRMV CQDDGLMI D.GTTA.GKV NSETIONDLE LISKYGEFG
Ph16	WGHEEIRQT SKSDAILCM TKGRALV.G GHRIFEMLKE GKVNSPYMK
Ibdv	DKLGINKIE RSIDD IRGK LQVLPVLAQ FDKVSGVEPE QSSPTVELD
Btv	...EQ YGDDTFLYT KLHTD.ITV FYLSVAISFD TWAKC.
Scv11	...SV HNGDDVMISL NRSTAVRIM DAMHRINARA QPACNLFIS
Scv1a	...
Polio	...DHLKMI AYGDVIAVY PH.EVD.ASL LAQSGKDY.G
Emc	...DDVKVL SYGDDLLVAT NYQL.DFDKV RASLAKT.G
Tmev	...DDIKVL SYGDDLLIGT NYQI.DFNL VKERLAPP.G
Hrv14	...DKILK AYGDVIAVY PYEL.D.POV LATLGKNY.G
Fndv	...DTYMI SYGDDVIVAS DYDL.DFE.A LKPHFKSL.G
Coxv	...DQFRMI AYGDVIAVY PWP1.D.ASL LAEAGKY.G
Hav	VFFCQALRIL CYGDDVIVF SRDV.Q.IDN LDILGKQI.V
Qbeta	...VT VYGDVILIP CAPPALREVE EYVGT.T
Sp	...VS VYGDVILIP CAPPALREVE EYVGT.P
Ms2	...G IYGDVILIP EAPRVLAL EYVGT.P
Ga	...G IYGDVILIP EAPRVLAL EYVGT.P

3

201			250
Carmv			
Bydv			
Plrv			
Bwyv			
Cnv			
Cyrv			
Mcmv			
Pvx			
Tymv			
Yfv			
Wnv			
Dengue			
Jev			
Hcv			
Bbv			
Tev			
Ppv			
Tbrv			
Cpmv			
Tvmv			
Trv			
Bnyvv			
Tmv			
Mhv			
Bsmv			
Almv			
Bmv			
Cucmv			
Sinv			
Sfv			
Onv			
Rrv			
Rot	VKAL	.VSTVG.IEI	AKRYIAGGKI
Reo	WKYD	.IAYDGTAEY	LKLYFIFGCR
Phi6	ISYE	.HG.GAFIGDILL	YDSRREP GSA
Ibdv	LL		
Btv			
Scv11	SEFLRVEHGM	SGGDLGAQY	LSRSCATLVH
Scv1a	AEFLRVD.M	RAKNSTGSQY	LTRGIATPTH
Polio	LTMTPAD		
Emc	YKITPAN		
Tmev	YKITPAN		
Hrv14	LTITPPD		
Fmdv	QTITPAD		
Coxv	LIMTPAD		
Hav	DEFKKLG		
Qbeta	NT		
Sp	NR		
Ms2	NL		
Ga	NE		

251			300
Carmv			
Bydv			
Plrv			
Bwyv			
Cnv			
Cyrv			
Mcmv			
Pvx			
Tymv			
Yfv			
Wnv			
Dengue			
Jev			
Hcv			
Bbv			
Tev			
Ppv			
Tbrv			
Cpmv			
Tvmv			
Trv			
Bnyvv			
Tmv			
Mhv			
Bsmv			
Almv			
Bmv			
Cucmv			
Sinv			
Sfv			
Onv			
Rrv			
Rot	WDQAAYLYSN	YIVNRLRGEF	GLQWRWI
Reo	EPWFAILLQI	MGVFFNGVHD	GLQWRWI
Ibdv	QSGVDRDSKR	KRPFFGL	
Btv			
Scv11	ARLDLANRT	RQSAVTAIK	
Scv1a	TRYDEILARG	.AST.DNMK	FLYRKQLFFA
Polio			
Emc			
Tmev			
Hrv14			
Fmdv			
Coxv			
Hav			
Qbeta			
Sp			
Ms2			
Ga			

301			350
Carmv			
Bydv			
Plrv			
Bwyv			
Cnv			
Cyrv			
Mcmv			
Pvx			
Tymv			
Yfv			
Wnv			
Dengue			
Jev			
Hcv			
Bbv			
Tev			
Ppv			
Tbrv			
Cpmv			
Tvmv			
Trv			
Bnyvv			
Tmv			
Mhv			
Bsmv			
Almv			
Bmv			
Cucmv			
Sinv			
Sfv			
Onv			
Rrv			
Rot	RIFFPS	ER.VLTTNST	FKVDSSEDFI
Reo	CA.FSR	QR.TMIGESV	.GYLQY
Phi6	IERCW	.WNAFGEST	RAYRED
Ibdv	RLFCIS	.AAY.PKG	VENK
Btv	KGGCY	.VPQDRM	MII
Scv11	HRVCGGISTD	TWAPVETKII	TDNEAYEI.P
Scv1a	DISCGGL.QE	KGRVSEMVLQ	.EVDIENIDS
Polio			
Emc			
Tmev			
Hrv14			
Fmdv			
Coxv			
Hav			
Qbeta	RESCG		
Sp	RESCG		
Ms2	RESCG		
Ga	RESCG		

351			400
Carmv			
Bydv			
Plrv			
Bwyv			
Cnv			
Cyrv			
Mcmv			
Pvx			
Tymv			
Yfv			
Wnv			
Dengue			
Jev			
Hcv			
Bbv			
Tev			
Ppv			
Tbrv			
Cpmv			
Tvmv			
Trv			
Bnyvv			
Tmv			
Mhv			
Bsmv			
Almv			
Bmv			
Cucmv			
Sinv			
Sfv			
Onv			
Rrv			
Rot	IQRAFMSL	SQKSGIADEI	AASSTFKNYV
Reo			
Phi6			
Ibdv			
Btv			
Scv11			
Scv1a			
Polio			
Emc			
Tmev			
Hrv14			
Fmdv			
Coxv			
Hav			
Qbeta			
Sp			
Ms2			
Ga			

	401		450		501		520
Carmv	LKSVGMGCLR	IA...	GGV	PVV.QEFYQK	YVETA.....	Carmv	D RAKRGYS.AV
Bydv	MSAVAQCGLV	LN...	AGV	PIL.ESFYKC	LYRSS.....	Bydv	QGRRTYNET.
Plrv	ECGNP	EVI.QNYLAA	VFSVL.QELR	HDRELVAKLH		Plrv	
Bwv	GSGNA	EVV.SNYLAA	CFSVL.NELR	HDPASVELLY		Bwv	
Cnv	SNAQHHGGVA	LT...	SG.I	PVV.EKFY.S	RFTLY.....	Cnv	TGDEQALAE RLDL.WEMDL
Cyrv	SNAQHHGGVA	LS...	AG.I	PVV.ERFY.S	RFTLY.....	Cyrv	TGDEQVALED RLDL.WEMDL
Mcmv	MRAVGEGLS	LT...	GGI	PVK.SNEY.T	ALVKH.....	Mcmv	SNRK.WQEVQ
Pvx	KKCQDSYED	LSYA.YDHKD	SLH.DLFDEK	.QC.QAHT	LTCR.....	Pvx	P RL.RNF.....
Tymv	RLSYLTFETT	GH....	LLGE	SLW.HLLPET	HVQY..QSAC	Tymv	NAMYLLP AKRLRLAITSS
Yfv		PCR.EQDE				Yfv	LMYFHK RDMRLSLAV
Wnv		PCR.GQDE				Wnv	LLLYFHR RDLRLMANAC
Dengue		PCR.NQDE				Dengue	LMYFHR RDLRLMANAI
Jev		PCR.GQDE				Jev	LLLYFHR RDLRLMANAI
Hcv						Hcv	RICLLVLSTE LQVRPGKSTT
Bbv		PTIPLADA				Bbv	QVRNQRSSRN KEKPYWLTCT
Tev	DKLVEEIRNF	YAMVLEQA	PYSQLAEE			Tev	LDYDIPTT ENLY.FQ...
Ppv	KELLREIRKE	YSWVLEQA	PYNALSKD			Ppv	Y DNIN.DDGES
Tbrv	R.SAIFSSL	YMLTPDKSKF	HASQRASDFQ	GTVDVVEELI	LNVNVALMEL	Tbrv	FCEAF...HSA.QQ...
Cpmv	K.ASLMSQH	YVN.....		CMNCE		Cpmv	QLQEF.Y EYQR.QQ...
Tymv	TDLLQNIIRF	YKWTIEQE		PYRSIAEQ		Tymv	NNEFLR ETVR.FQ...
Trv	QHLAEIYISL			NDNRA		Trv	HALC ALWK.HIK...
Bnyvv	KHF.....	CEYQE	SLR.DWIKNL			Bnyvv	VQR.WLDAI
Tmv	EHLEEFRRSL		CDV	AVSLMNCAYY		Tmv	IDGSS
Mhv						Mhv	
Bsmv	EVLSIYISI			NDNYS	Y.KDFKVLDA	Bsmv	NCEGK FVDKKLRDF
Almv	IFDEWYQ.SW		IDIIG	GF..NDHVI	RCVAAM.....	Almv	LCKECL FNEK.HESNV
Bmv	QMLRAHFVSE		CD.RM	KFI.MQDSE	KMITTL.....	Bmv	GIRVYQMSD PVCK.FKRTT
Cucmv	DNDLFPAHF	MSF..VD.RL	KFM.DRMSQS	CIDQL.....		Cucmv	QCDLVNTFCI SEFRVIRRTT
Sinv	DEQDEDRRA	L....LDETK	AW..FRVGI	GTAVAV.....		Sinv	GEIK.HLYGP
Sfv	DEQDEDRRA	L....SDEVS	KW..FRTGLG	ALEVAL.....		Sfv	GPVI.HLYGP
Onv	DEQDEDRRA	L....ADEVT	RW..QRTGLV	TELEKA.....		Onv	GPVV.TLYGG
Rrv	DEQDEDRRA	L....KDETD	RW..ARVGLK	SELEIAL.....		Rrv	GPVH.HLYGG
Rot	ALTEKAKLN	..SYA.....	PISLEKRA			Rot	RDVKPFFTVN
Reo	PTGDLMYMS	ISLIR.....	PLMT.RMWVA	NGYVTRCS	TVFG.....	Reo	PKKS.GRAAS REVRQETQA
Phi6	GLAELTIDIL	EVLAD.....	PNKL.....	QYKW.TEADV	SA.....	Phi6	
Ibdv	PLDEFLEAWS	ELSEF.....	GETF.....	EGF.....		Ibdv	GLKAV.SNAL KTRG.YRNEA
Btv	SNDDEGFTLI	QIRN.....	PLAL.YVPIG	WNGYGAHPAA	L.....	Btv	IRRIV..HDI..PPCW.NET
Scvll	..WERTMYKA	YK.....	GLAVSYA..A	LSKFMSIPPM	A.....	Scvll	ALDSS DPLR.ALQVI
Scvla	AYHERAVRHA	WK.....	GMSGLHIV.N	RIRM.GVS..		Scvla	KSG DPTK.WLAVL
Polio			PVMPKEIHE	SIRW.....		Polio	RALLPEYST LYRR.WLDSF
Emc			PVMNREALEA	MSLY.....		Emc	GVVPSFES VEYR.WLSLF
Tmev			PQMDAVNLKA	MVSY.....		Tmev	GIVVPTYSS MYLR.WLSLF
Hrv14			PVMPMKDIHE	SIRW.....		Hrv14	KCLILPEYSV LRRR.WLDSF
Fmdv			PVMPMKDIHE	ILSF.....		Fmdv	PSYRS LYLR.WVNAV
Coxv			PAMPMKDIHE	SIRW.....		Coxv	RCLTLPFTT LGRR.WLDSF
Hav			PAISEKTIWS	LMAW.....		Hav	EMIEYRL KSYDWMRMRF
Qbeta	RWATIDGVWD		PAHSAVLYKY	RKLL.....		Qbeta	VL...INFP AKNRGWIRYV
Sp	RWGTVDGIWD		PRALTVEEYK	LKLL.....		Sp	AT...TNPF VIVKNYSALY
Ms2	GWGVVGGMSD		PRLYKVMVRL	SSQV.....		Ms2	FR...LARI ARERKFFSEK
Ga	GWQTVGGMSD		PRLFPIWKEF	ADMI.....		Ga	VS...LVRI AKVRSGFNHA
							..
							8
	451		500				
Carmv	GNVRE	NKNITEK...	SSSG	FFMMA.....			
Bydv	GYKKV	SEEFIKNVI..	SYG	TDERL.....			
Plrv	QWL.V.PSATT	KEH.....					
Bwv	SWLVDPVLPQ	KIPGE.....					
Cnv	..EVPKKHQ	RI.DTVTNVH	KWR..GSGGD	YVVT..EA	RASFWAAGFL		
Cyrv	..DMPVKHQ	RI.DTVTNVH	KWR..GSGGN	YHVT..ES	RASFWAAGFL		
Mcmv	..GLDPK...	..NIKQ.GKD	FD....SGLY	YLSKL.....			
Pvx	TLIKSGRG...	TVSL.....					
Tymv	..CRRCP.R	HE.KMLDDSTPA	LSLLERITSS	PRWTK.....			
Yfv	..LIGRGR...	..VSPGN.GNM	IKETACLSKA	Y.NMWS.....			
Wnv	..LIGRAR...	..ISPGA.GWN	VRDTACLAKS	YAQMW.....			
Dengue	..LIGRAR...	..ISQGA.GWS	LRETACLGKS	YAQMWS.....			
Jev	..LIGRAR...	..ISPGA.GWN	VKDTACLAKA	YAQMW.....			
Hcv	D..SSGE...	..RGTIAYEKA	VAF.....S	FILMYS.....	WNPLIR		
Bbv	CDRVEGHILCT	DALTPLI.SD	YCKMVL..RL	YGPTAS.....	TE		
Tev	..GKAPYIAE	TALKFLYTSQ	HGTNSEIE.E	YLVKL.....			
Ppv	..GKAPYIAE	TALKKLYTDT	EASETEI..R	YLEAF.....			
Tbrv	YLHNDPREFS	RV.RDFYIKA	LPLAT..GQ	FRTWA.....			
Cpmv	YMH.SPREAT	EF.RRKVLKK	VSWIT..SGD	LPTLA.....			
Tymv	..GLAPYLSE	VALRRLYTSQ	IATDNELTDY	YKELA.....			
Trv	..GN.YM	..VVSKLSESV	SDRYL.....	YKDSV.....			
Bnyvv	..PKDPA	..VYADF.LEC	NASLSR...	..NVDD.....			
Tmv	..KTAPPGSF	..VY.KSLVKYL	SDKVL.....	FRSLF.....			
Mhv							
Bsmv	..RSPYS...	..AISALV.SL	CY.HIFDFNK	FKLIF.....			
Almv	..RRPSLYL	EAALA...	SLGK1...	FAGKT.....			
Bmv	YGKEKFWIFE	EVRAALAAF.	SLYSE...N	FLRFS.....	DCYCTE..		
Cucmv	YKK.SGN..E	AAL.VLGAFK	KYTA...N	FNAYKEL...	YYSRQ		
Sinv	Y.EVDNI...	..TP.VLLALR	TFAQ...SKRA	FQAIR.....			
Sfv	Y.EVEGC...	..KS.ILIAMT	TLAR..DIKA	FKKLR.....			
Onv	Y.EVQGI...	..TA.VITSMA	TFT.N.SKEN	FKKLR.....			
Rrv	Y.EVNGT...	..GN.IVRAMA	TLAK...SLKN	FKKLR.....			
Rot	..QISALLTM	LQKP.....	VT	FKSSKITIND	IL.....		
Reo	..NADYR...	RC	FMELKLYQGY	YM..AQLPRN			
Phi6	..NIHEVLMH	GVSVEKT...	ER	FLRSVMPR...			
Ibdv	..NIKLTVTR	ENLAEL.....	..NK	PVPPKPPNVN	..RPVNTG		
Btv	..NIVMTE...	EM	YVDSIMISKL	DEIMAP			
Scvll	..NIEFGQ...	AR	YAMQA.....				
Scvla	..NLVMVSK	INP.....	..AK	ANVLA.....			
Polio	..TKDPRNTQ	DHVRSLCLLA	WHNGEEYKN	FLAKI.....	RSVPFG		
Emc	..YR.PGTLS	EKLTSITMLA	VHSGKQYDR	LFAPP.....	REV...		
Tmev	..CK.PGTLK	EKLMSIALIA	VHSGPDIDE	IFLPP.....	RNV...		
Hrv14	..TKDPKNTQ	DHVRSLCMLA	WHSGEKEYNE	FIQKI.....	RTTDIG		
Fmdv	..AR.RGTIQ	EKLISVAGLA	VHSGPDEYRR	LFEPF.....	QGLFEI..		
Coxv	..TKDPKNTQ	DHVRSLCLLA	WHNGEHEYEE	FIRKI.....	RSVPVG		
Hav	..QRSNAEFE	QNLNAQWFA	FMHGYEFYQK		QSCLEK.		
Qbeta	..PK.QL	QRNTIP..DG	YG.DGAL...	VGS			
Sp	..PR.NW	RRNRIP..DG	YG.DGAL...	VGL			
Ms2	..PS.MF	FGGTDLAADY	YVSPPTAVS	VYTKTPYGR	L..ADRTSG		
Ga	..PP.KF	KGCCNLDRTD	YLVSPDK...	PG			

Figure 1. Alignment of positive strand and dsRNA viral RDRPs. Alignments were made as described in Materials and Methods. Partial sequences of Hcv and ScV-La RDRPs are shown. All sequences are from data in the Genbank, EMBL, or NBRF databases as described in Table 1. Large numbers below the sequences indicate conserved motifs 1–8. Asterisks indicate highly conserved residues. Motif 1 is from position 1–9; motif 2 from 98–111; motif 3 from 160–168; motif 4 position 281; motif 5 from 303–305; motif 6 from 381–383; motif 7 from 481–484; motif 8 from 514–516.

Table 1: RDRP sources

Virus	Abbr.	Seg.	File*	Family or group	Genus	Major host	Ref
carnation mottle	Carmv		gb:cmvxx	Carmovirus		plant	21
barley yellow dwarf	Bydv		em:bydvpav	Luteovirus		plant	22
potato leaf roll	Plrv		gb:plrvvxx	Luteovirus		plant	23
beet western yellow	Bwyv		gb:bwyvfl1	Luteovirus		plant	24
curcumber necrosis	Cnv		gb:cnvgeno			plant	25
cymbidium ringspot	Cyrv		gb:x15511	Tombusvirus		plant	26
maize chlorotic mottle	Mcmv		em:mcmvxx	Bromoviridae		plant	27
potato virus x	Pvx		gb:pvxx3	Potexviridae		plant	28
turnip yellow mosaic	Tymv		gb:mtymvxx	Tymoviridae		plant	29
yellow fever	Yfv		gb:yfv	Togaviridae	Flavivirus	mammal	30
West Nile	Wnv		gb:wnfcg	Togaviridae	Flavivirus	mammal	31
Dengue fever	Dengue		gb:denrcg	Togaviridae	Flavivirus	mammal	32
Japanese encephalitis	Jev		gb:jevvcg	Togaviridae	Flavivirus	mammal	33
Hog cholera	Hcv		gb:hcvcgsga	Togaviridae	Pestivirus	mammal	34
black beetle	Bbv	1(A)	gb:bbv1g	Nodaviridae		insect	35
tobacco etch	Tev		em:tevgen	Potyvirus		plant	36
plum pox	Ppv		gb:ppv	Potyvirus		plant	37
tomato black ring	Tbrv		gb:tbrna1	Nepovirus		plant	38
cowpea mosaic	Cpmv	B	gb:mcpvgb	Comoviridae		plant	39
tobacco vein mottle	Tvmv		gb:tvmx	Potyvirus		plant	40
tobacco rattle	Trv	1	gb:mttrna1	Tobravirus		plant	41
beet necrotic yellow vein	Bnyvv	1	gb:bnyvv1	Furoviruses		plant	42
tobacco mosaic	Tmv		gb:mtvvcg	Tobamovirus		plant	43
mouse hepatitis	Mhv		gb:mhvp01	Coronaviridae		mammal	44
barley stripe mosaic	Bsmv	γ	gb:mbsrnagt	Hordeiviridae		plant	45
alfalfa mosaic	Almv	2	gb:maacg2z	Tricornaviridae		plant	12
brome mosaic	Bmv	2	gb:mbrcg2z	Tricornaviridae	Bromovirus	plant	46
cucumber mosaic	Cucmv	2	gb:mcvrna2c	Tricornaviridae	Cucumovirus	plant	47
sindbis	SinV		n:grsdbv	Togaviridae	Alphavirus	mammal	48,49
Semliki Forest	Sfv		em:alsfv42s	Togaviridae	Alphavirus	mammal	50- 53
O'nyong-nyong	Onv		em:togonpp	Togaviridae	Alphavirus	mammal	54
Ross River	Rrv		em:togrrvb	Togaviridae	Alphavirus	mammal	55
bovine rotavirus	Rot	L1	gb:robvp1	Reoviridae		mammal	56
reovirus	Reo	L1	gb:M24734	Reoviridae	Reovirus	mammal	57
bacteriophage phi6	Phi6	L(P2)	gb:M17461	Cystoviridae	Cystovirus	bacteria	13
infectious bursal disease	Ibdv	B	n:rrxsib	Birnaviridae		bird	58
bluetongue	Btv	L1	gb:btv10vp1	Reoviridae	Orbivirus	mammal	59
S. cerevisiae L1	Scvl1	L1	gb:M28353	Totiviridae	Totivirus	fungi	7
S. cerevisiae La	Scvla	La	em:X54405	Totiviridae	Totivirus	fungi	this work
poliovirus type 1	Polio		n:gnyy2p	Picornaviridae	Enterovirus	mammal	60
encephalomyocarditis	Emc		n:gnyy	Picornaviridae	Cardiovirus	mammal	61
Theiler's murine enceph.	Tmev		em:tmep	Picornaviridae	Cardiovirus	mammal	62
human rhinovirus 14	Hrv14		n:gnyh4	Picornaviridae	Rhinovirus	mammal	63
foot and mouth disease	Fmdv		n:gny2f	Picornaviridae	Aphthovirus	mammal	64
coxsackievirus	Coxv		n:gnyb3	Picornaviridae	Enterovirus	mammal	65
hepatitis A	Hav		gb:hpaacg	Picornaviridae	Enterovirus	mammal	66
bacteriophage Q β	Q β		em:lepqbrep	Leviviridae	Levivirus	bacteria	6
bacteriophage SP	Sp		em:mypxx	Leviviridae	Levivirus	bacteria	68
bacteriophage MS2	Ms2		gb:ms2cg	Leviviridae	Levivirus	bacteria	69
bacteriophage Ga	Ga		n:grbpga	Leviviridae	Levivirus	bacteria	70

*File names are the sequence names in the NBRF (n), Genbank (gb), or EMBL (em) databases. Accession numbers are given for those sequences without currently assigned names. The order of viruses listed is the same as that of Fig. 1. The 'Seg.' column indicates which segment encodes the RDRP in each of the segmented viruses.

Virus classification

A dendrogram of the 50 RDRPs was drawn by applying SYSTAT (20) to the table of distances calculated by the GCG program DISTANCES (18), as described in Materials and Methods. The result is shown as Fig. 3. In general, the clustering of similar viruses agrees with previous classification attempts using other protein sequence similarity criteria. For instance, Fmdv and Emc are closer to each other than either is to Hrv14 or polio, in agreement with a molecular taxonomy based of the viral capsid protein vp3 (71, 72, 73, see Discussion).

The dendrogram of Fig. 3 puts the plant viruses into three large groups. One of these corresponds to the 'Sindbis-like' group of

Goldbach and Wellink (74) and includes Bnyvv, Bsmv, Tmv, Trv, Almv, Bmv, and Cucmv, as well as the animal virus Mhv. This group has been labeled tobamovirus-like in Fig. 3 (see below). The second corresponds to the 'picornavirus-like' group of the same authors, and includes Tbrv, Cpmv, Ppv, Tvmv, and Tev. This group has been designated potyvirus-like in Fig. 3. Goldbach and Wellink (74) lumped the three groups of plant viruses into two and placed Carmv in the 'Sindbis-like' group. Habili and Symons (6), however, divided the plant viruses into three groups, placing Carmv (and Bydv) in a 'luteovirus-like' group, intermediate between the other two groups. This group includes Pvx, Tymv, Plrv, Bwyv, Cyrv, Cnv, Mcmv, Carmv, and Bydv, according to the dendrogram. The three groups of

Figure 2. Secondary structure predictions for motif 2 (31 amino acids). Chou–Fasman turns (CF turns) and beta sheet predictions (CF beta sheet) were performed by PEPTIDESTRUCTURE and PLOTSTRUCTURE (18). The sequence of the ScV-L1 motif 2 is shown as the last line aligned with predictions for each residue. Asterisks indicate highly conserved amino acids. Predictions of turns or beta sheet are visible as vertically displaced portions in each line.

plant viruses in the dendrogram are the same as those of Habili and Symons, except for Pvx and Tymv, which have been placed (for convenience) in the luteovirus-like group rather than the 'Sindbisvirus-like' group, although they appear as close to the flaviviruses as to the luteoviruses. This is remarkable agreement, since the dendrogram of Fig. 3 is based on RDRP similarities and the Habili and Symons classification scheme is based on the RNA helicases. Consideration of similarities solely within four conserved domains in the RDRP gives a similar result (6).

The designations 'picornavirus-like' and 'Sindbis-like' are inappropriate for any of the plant viruses. The picornaviruses are equally distant from all the plant viruses. This corresponds to alignments arrived at by computer algorithms as well as those derived by eye. Both the GCG program GAP (18) and the RDF program (19) support this conclusion. For instance, highly similar RDRPs appear as such by GAP: Yfv and Wnv are 63% identical; ScV-L1 and ScV-La are 30% identical. Both similarities are scored as highly significant by RDF: $z=115$ for the first and 18 for the second. That is, the optimum alignment of Yfv and Wnv RDRPs is matched with a value 115 times the standard deviation of the mean of the values of alignments of randomly

Figure 3. Dendrogram of dsRNA and positive strand RNA viruses based on RDRP similarity. The dendrogram was constructed as described in Materials and Methods from the alignment of Fig. 1. Horizontal distances within groups are proportional to distance as calculated by DISTANCES (18), but the scale is arbitrary. Thus Cyrv and Cnv are very closely clustered, while Cyrv and Bydv are more distantly clustered. Supergroups of viruses are indicated on the left and brackets enclose viruses within these supergroups.

scrambled sequences. A z value of 3–6 is considered possibly significant, 6–10 probably significant, and 10 or more highly significant. Although Sindbis has 27% identity with Tmv and a z value of 7.4, polio has only a 22% identity and a z value of less than 1 with Tev, a plant virus in the 'picornavirus-like' category. On the basis of these data, as well as the dendrogram, it is inappropriate to categorize any of the plant viruses as 'picornavirus-like.' Although the alphaviruses are more similar to the plant positive strand RNA viruses than are the picornaviruses, the dendrogram clearly places the alphaviruses equally distant from all the plant viruses. Therefore the terms potyvirus-like, in lieu of 'picornavirus-like,' and tobamovirus-like, in lieu of 'Sindbis-like,' are preferable.

All of the similarities of more distant RDRPs in the dendrogram are undetectable by computer algorithms. For instance, as pointed out by Weiner and Joklik (57), the reovirus RDRP has no detectable similarity with other proteins by simple application of search programs. The dendrogram places Reo and Phi6 in the same group, but these are only 15% identical by GAP and their z value is close to zero. Nevertheless, alignment of RDRPs preserves both primary and predicted secondary structure in the conserved regions (Figs. 1 and 2). The hydropathy plots of the

Reo and Phi6 RDRPs are quite similar (not shown), lending support to their structural similarity. GAP also fails to align the conserved motifs of Reo and Phi6, even though they are obvious by eye. Consequently, a calculation of potential evolutionary distance based on an alignment made at least partially by eye may be more sensitive to distant relationships. Among such distant relationships undetectable by computer algorithms is that of Bydv to Tev (20% identity, z value less than 1). Nevertheless, the dendrogram places these closer together than either is to the picornaviruses, which seems reasonable.

There are other relationships not consistent with current classification schemes (16). The togaviridae are not a well-defined group of viruses, since the alphaviruses are in one supergroup and the flaviviruses in another (Fig. 3). Within the family togaviridae, the one known pestivirus (Hcv) is more similar to the flaviviruses than to the alphaviruses. These results are not surprising, since the alphaviruses and flaviviruses are grouped together solely on the basis of their common possession of a lipid envelope, and none of the classifications based on viral morphology (including capsid morphology) are supported by this or other schemes based on protein sequence similarity. For instance, Tmv (like all the plant viruses) and Sindbis are clearly similar, although Tmv has a helical capsid and Sindbis an icosahedral one. The new proposed classification scheme of the International Committee on Taxonomy of Viruses makes the Togaviridae and Flaviviridae separate families, with the pestiviruses and flaviviruses genera within the Flaviviridae, which corresponds exactly to the dendrogram of Fig. 3. Within the picornaviruses, Hav, which has been classified as an enterovirus, like Polio or Coxv, is equally distant from all other picornaviruses and probably belongs in a new genus.

Relationship to genome structure

Clustering based on genome structure is more apparent among the more highly similar viruses. For instance, the RDRPs of the positive strand RNA viruses with three segments (Almv, Bmv, and Cucmv) are more similar to each other than they are to viruses in any other group. On the other hand, Yfv, which, like the picornaviruses, has a single RNA segment expressed as a polyprotein is clustered more closely to the luteoviruses than it is to the picornaviruses. The mode of expression of the viral RDRP is not consistent within the viral supergroups. For instance, within the luteovirus-like group, some RDRPs are thought to be produced by translational read-through of nonsense codons (Cyrv, Cnv, Mcmv, and Carmv), some by translational frame-shifts (Plrv and Bwyv), and some by processing of polyproteins (Pvx and Tymv). However, each of the subgroups based on RDRP similarity corresponds to a subgroup based on mode of expression of the RDRP (Fig. 3). Only Bydv RDRP, probably read as a translational frame-shift, is an exception. As might be expected, RNA viruses with segmented genomes express the RDRP as a single reading frame without translational read-through or frame-shifting. Tbrv is the only known exception to this rule. The close clustering of Bsmv and Trv and Tmv is a surprise, since Bsmv has a trisegmented genome, while Trv and Tmv have a single segment of RNA; and since Bsmv expresses the RDRP as a single open reading frame, while Trv and Tmv express it as a translational read-through. Similarly, Tbrv and Cpmv are clustered together by RDRP similarity, even though Cpmv produces its RDRP as do the potyviruses (as a polyprotein), while Tbrv utilizes translational read-through followed by processing. The mechanism of coding of the viral RDRP seems to have little relevance to classification within supergroups: frame-shifting

Figure 4. Comparison of the dendrograms of picornaviruses based on the viral capsid protein vp3 and the RDRP. The vp3 alignment was constructed as described for Fig. 1, starting with a previous alignment (72).

mechanisms are found among the dsRNA viruses (ScV-L1) as well as among the luteoviruses; read-through mechanisms are found among the luteovirus-like group, the potyvirus-like group and the tobamovirus-like group; polyproteins are found among the luteovirus-like group, the flavivirus-like group, the potyvirus-like group, the tobamovirus-like group (Bnyvv and Mhv), the alphaviruses, and the picornaviruses.

Similarly, classification solely on the basis of the number of genomic segments does not correspond well to classification by RDRP similarity. Reovirus is grouped more closely to the rotaviruses than it is to Btv by RDRP similarity, although Btv and Reo have 10 segments and Rot 11. Although Tbrv and Cpmv, both with 2 viral segments, are closely grouped, they are not closely grouped to Bbv, also with 2 segments. Bsmv, with 3 segments, is more closely grouped to Tmv (with 1 segment) than to the tricoronaviridae. Viruses with a single RNA segment fall into every supergroup.

Confidence in this dendrogram and the alignment on which it is based is increased by the fact that essentially the same distance matrix is generated by the carboxy-terminal 241 positions or the 104 amino terminal positions of the alignment as is generated by all 520 residues.

DISCUSSION

The most remarkable conclusion of this classification based on RDRP sequence is the grouping of all positive strand animal viruses except the picornaviruses with the plant viruses. In fact, the flaviviruses are more closely grouped to the luteoviruses than they are to the alphaviruses. The most obvious explanation for this grouping is that all six groups of viruses in this superfamily (the luteovirus-like, the flavivirus-like, the potyvirus-like, the tobamovirus-like, the alphavirus-like, and Bbv) either infect insects or have insect vectors. There are a number of exceptions to this rule: nothing is known of the transmission of Pvx and Mhv, and Hcv has no invertebrate vectors. Bnyvv, Tmv, and Trv are also known not to have insect vectors. However, both

Tmv and Trv can survive in a non-persistent manner in aphids and be transmitted to plants under laboratory conditions (75, 76). Bnyvv, Tmv, and Trv may have recently evolved minor coat protein modifications that interfere with insect transmission. All of the flaviviruses and alphaviruses are transmitted by or infect insects. Many of the plant viruses in this group transmitted by insects also multiply in their insect hosts (16). There are a number of possible explanations for the clustering of RDRPs within this diverse group of RNA viruses: convergent evolution of RDRP sequences, derivation from a genomic sequence of a common host species, divergence from a common viral ancestor, and/or interviral recombination. Convergent evolution seems unlikely, since a similar classification scheme (among the plant viruses) is dictated by the sequences of RNA helicases (6) and a similar clustering of the picornaviruses is dictated by the capsid polypeptide vp3. A comparison of dendrograms for the picornaviruses constructed from the RDRP and from vp3 is shown as Fig. 4. The two dendrograms are essentially identical, except that Hrv14 appears more similar to Polio and Coxv by vp3 similarity than by RDRP similarity. Since it is unlikely that two or more proteins would independently, convergently evolve in such a way as to produce the same network of similarities among different viruses, divergence from a common ancestor appears the more likely explanation for the observed similarities. The rapid rate of mutation and of fixation of mutations in RNA viruses (77, 78, 79, 80, 81, 82, 83) and the ancient divergence of plant and animal cells implies that the observed similarity among viruses with insect hosts and/or vectors is a fairly recent event. It can best be explained if all of the viruses in this group originally infected insect cells and subsequently evolved to infect the hosts of insect pathogens, either plant or animal. This is similar to the conclusion reached by Goldbach and Wellink (74) for a more limited data set of 18 positive strand RNA viruses.

There is also good evidence of a significant role for interviral recombination. The lack of correlation of genomic structure or mode of expression of the viral RDRP with RDRP sequence similarities is strong evidence for recombination among the RNA viruses. The data are best explained by inheritance of the RDRP as a module, and this can take place (in nonsegmented viruses) only by recombination. Recombination by means of template switching during replication is well documented in poliovirus (84) and there is no reason not to expect it in other RNA viruses. In fact, there is growing evidence for the universality of RNA recombination, including interviral recombination (85, 86, 87, 88). In summary, the extant sequence similarities among diverse groups of plant and animal viruses can be explained by divergence from a common ancestral insect virus and the spreading of RDRP modules among viruses by interviral recombination taking place in co-infected tissues, possibly the cells of insect vectors themselves. The absence of plant viruses from the animal virus supergroups and vice versa, with the exception of Mhv in the tobamovirus-like supergroup, may be the result of the inability of individual insect species to feed on both plants and animals. Similarly, Reaney (78) has proposed the spread of individual RNA segments in segmented RNA viruses by reassortment following co-infections in the same tissue by way of insect vectors.

The dendrogram of Fig. 3 also suggests that the positive strand RNA viruses other than the picornaviruses and the leviviruses arose from an ancestral dsRNA virus, and that viruses of procaryotes and eucaryotes diverged prior to the emergence of dsRNA viruses. This scheme differs considerably from suggestions that dsRNA viruses arose from the positive strand

RNA viruses, and on more than one occasion (9). A cladistic analysis to test this hypothesis is in progress. The sequences of the RDRPs of the plant dsRNA virus wound tumor virus, the insect dsRNA virus cytoplasmic polyhedrosis virus, the cryptic dsRNA viruses of plants (89), and the dsRNA viruses of protozoans (90) should soon be available to enlarge the necessary data set. These sequences may also help resolve the anomaly of Btv, the only virus with an insect vector not placed in the major grouping of viruses with insect hosts or vectors.

One major exception exists to this scheme: phi6. Phi6 is more similar to the dsRNA viruses of eucaryotes than to the leviviruses. This is the only well-characterized RNA bacteriophage that is not a levivirus, of which more than 30 are known (16). It is also unique in having a dsRNA genome and in being enveloped (91). Its host is *Pseudomonas phaseolicola*, a plant pathogen (92). Phi6 may have evolved from a plant virus by acquiring the ability to infect its bacterial pathogen, just as many positive strand RNA plant viruses and animal viruses may have evolved from the viruses of their insect pathogens. This is a suggestion made previously on the basis of the genome structure of Phi6, its uniqueness among RNA bacteriophages and its restricted distribution, with a single host species (78). Btv

ACKNOWLEDGEMENTS

I thank Ian Baldwin for numerous discussions and Ian Baldwin and Jim Berry for critical reading of the manuscript. This work was supported grant GM22200 from the National Institutes of Health.

REFERENCES

1. Poch, O., Sauvaget, I., Delarue, M., and Tordo, N. (1989). EMBO J. **8**, 3867-3874.
2. Kamer, G. and Argos, P. (1984). Nucl. Acids Res. **12**, 7269-7282.
3. Gorbalenya, A.E. and Koonin, E.V. (1988). Nucl. Acids Res. **16**, 7735.
4. Pietras, D.F., M.E. Diamond, and J.A. Bruenn. (1988). Nucl. Acids. Res. **16**, 6226.
5. Argos, P. (1988). Nucl. Acids Res. **16**, 9909-9916.
6. Habili, N. and Symons, R.H. (1989). Nucl. Acids Res. **17**, 9543-9555.
7. Diamond, M.E., J.J. Dowhanick, M.E. Nemeroff, D.F. Pietras, C. Tu, and J.A. Bruenn. (1989). J. Virol. **63**, 3983-3990.
8. Icho, T. and Wickner, R.B. (1989). J. Biol. Chem. **264**, 6716-6723.
9. Koonin, E.V., Gorbalenya, A.E., and Chumakov, K.M. (1989). FEBS Lett. **252**, 42-46.
10. Bruenn, J. (1980). Ann. Rev. Microbiol. **34**, 49-68.
11. Drayna, D. and Fields, B.N. (1982). J. Virol. **41**, 110-118.
12. 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.
13. Mindich, L., Nemhauser, I., Gottlieb, P., Romantschuk, M., Carton, J., Frucht, S., Strassman, J., Bamford, D.H., and Kalkkinen, N. (1988). J. Virol. **62**, 1180-1185.
14. Fujimura, T., and Wickner, R.B. (1988). Cell **55**, 663-671.
15. Urakawa, T., Ritter, D.G., and Roy, P. (1989). Nucleic Acids Res. **17**, 7395-401.
16. Matthews, R.E.F. (1982). Classification and nomenclature of viruses. Intervirology **17**, No. 1-3.
17. Baltimore, D. (1980). Ann. NY Acad. Sci. **354**, 491-497.
18. Devereux, J., P. Haerberli, and O. Smithies. (1984). Nucl. Acids Res. **12**, 387-395.
19. Lipman, D.J. and W.R. Pearson. (1985). Science **227**, 1435-1441.
20. Wilkinson, L. (1989). SYSTAT: The System for Statistics. Evanston, IL. SYSTAT, Inc.
21. Guillely, H., Carrington, J.C., Balazs, E., Jonard, G., Richards, K. and Morris, T.J. (1985). Nucleic Acids Res. **13**, 6663-6677.
22. Miller W.A., Waterhouse P.M., Gerlach W.L. (1988). Nucleic Acids Res. **16**, 6097-6111.
23. Van der Wilk, F., Huisman, M.J., Cornelissen, B.J.C., Huttinga, H. and Goldbach, R.W. (1989). FEBS Lett. **245**, 51-56.

24. Veidt, I., Lot, H., Leiser, M., Scheidecker, D., Guilley, H., Richards, K. and Jonard, G. (1988). *Nucleic Acids Res.* **16**, 9917-9932.
25. Rochon, D.M. and Tremaine, J.H. (1989). *Virology* **169**, 251-259.
26. Grieco, F., Burgyan, J. and Russo, M. (1989). *Nucleic Acids Res.* **17**, 6383-6383.
27. Nutter, R.C., Scheets, K., Panganiban, L.C., Lommel, S.A. (1989). *Nucleic Acids Res.* **17**, 3163-3177.
28. Huisman, M.J., Linthorst, H.J.M., Bol, J.F. and Cornelissen, B.J.C. (1988). *J. Gen. Virol.* **69**, 1789-1798.
29. Morch, M.D., Boyer, J.C. and Haenni, A.L. (1988). *Nucleic Acids Res.* **16**, 6157-6173.
30. Rice, C.M., Lench, E.M., Eddy, S.R., Shin, S.J., Sheets, R.L. and Strauss, J.H. (1985). *Science* **229**, 726-733.
31. Castle, E., Leidner, U., Nowak, T., Wengler, G. and Wengler, G. (1986). *Virology* **149**, 10-26.
32. Hahn, Y.S., Galler, R., Hunkapiller, T., Dalrymple, J.M., Strauss, J.H. and Strauss, E.G. (1988). *Virology* **162**, 167-180.
33. Sumiyoshi, H., Mori, C., Fuke, I., Morita, K., Kuhara, S., Kondou, J., Kikuchi, Y., Nagamatsu, H. and Igarashi, A. (1987). *Virology* **161**, 497-510.
34. Meyers, G., Ruemenapf, T. and Thiel, H.-J. (1989). *Virology* **171**, 555-567.
35. Dasmahapatra, B., Dasgupta, R., Ghosh, A. and Kaesberg, P. (1985). *J. Mol. Biol.* **182**, 183-189.
36. Allison R., Johnson R.E., Dougherty W.G. (1986). *Virology* **154**, 9-20.
37. Maiss, E., Timpe, U., Briske, A., Jelkmann, W., Casper, R., Himmler, G., Mattanovich, D. and Kattinger, H.W.D. (1989). *J. Gen. Virol.* **70**, 513-524.
38. Greif, C., Hemmer, O. and Fritsch, C. (1988). *J. Gen. Virol.* **69**, 1517-1529.
39. Lomonosoff, G.P. and Shanks, M. (1983). *EMBO J.* **2**, 2253-2258.
40. Domier, L.L., Franklin, K.M., Shahabuddin, M., Hellmann, G.M., Overmeyer, J.H., Hiremath, S.T., Siaw, M.F.E., Lomonosoff, G.P., Shaw, J.G. and Rhoads, R.E. (1986). *Nucleic Acids Res.* **14**, 5417-5430.
41. Hamilton, W.D.O., Boccara, M., Robinson, D.J. and Baulcombe, D.C. (1987). *J. Gen. Virol.* **68**, 2563-2575.
42. Bouzoubaa, S., Quillet, L., Guilley, H., Jonard, G. and Richards, K. (1987). *J. Gen. Virol.* **68**, 615-626.
43. Goelet, P., Lomonosoff, G.P., Butler, P.J.G., Akam, M.E., Gait, M.J. and Karn, J. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5818-5822.
44. Soe, L.H., Shieh, C.-K., Baker, S.C., Chang, M.-F. and Lai, M.M.C. (1987). *J. Virol.* **61**, 3968-3976.
45. Gustafson, G., Hunter, B., Hanau, R., Armour, S.L. and Jackson, A.O. (1987). *Virology* **158**, 394-406.
46. Ahlquist, P., Dasgupta, R. and Kaesberg, P. (1984). *J. Mol. Biol.* **172**, 369-383.
47. 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.
48. Rice, C.M., and Strauss, J.H. (1981). *Proc. Nat. Acad. Sci. USA* **78**, 2062-2066.
49. Strauss, E.G., Rice, C.M., and Strauss, J.H. (1984). *Virology* **133**, 92-110.
50. Garoff, H., Frischauf, A.M., Simons, K., Lehrach, H., Delius, H. (1980). *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6376-6380.
51. Garoff H., Frischauf A.M., Simons K., Lehrach H., Delius H. (1980). *Nature* **288**, 236-241.
52. Riedel H., Lehrach H., Garoff H. (1982). *J. Virol.* **42**, 725-729.
53. Takkinen, K. (1986). *Nucleic Acids Res.* **14**, 5667-5682.
54. Levinson, R.S., Strauss, J.H. and Strauss, E.G. (1990). *Virology* **175**, 110-23.
55. Faragher S.G., Meek A.D.J., Rice C.M., Dalgarno L. (1988). *Virology* **163**, 509-526.
56. Cohen, J., Charpilienne, A., Chilmonczyk, S. and Estes, M.K. (1989). *Virology* **171**, 131-140.
57. Wiener, J.R. and Joklik, W.K. (1989). *Virology* **169**, 194-203.
58. Morgan, M.M., Macreadie, I.G., Harley, V.R., Hudson, P.J., and Azad, A.A. (1988). *Virology* **163**, 240-242.
59. Roy, P., Fukusho, A., Ritter, G.D., and Lyon, D. (1988). *Nucleic Acids Res.* **16**, 11759-11767.
60. Racaniello, V.R., and Baltimore, D. (1981). *Proc. Nat. Acad. Sci. USA* **78**, 4887-4891.
61. Palmenberg, A.C., Kirby, E.M., Janda, M.R., Drake, N.L., Duke, G.M., Potratz, K.F., and Collett, M.S. (1984). *Nucl. Acids Res.* **12**, 2969-2985.
62. Pevear D.C., Calenoff M., Rozhon E., Lipton H.L. (1987). *J. Virol.* **61**, 1507-1516.
63. Stanway, G., Hughes, P.J., Mountford, R.C., Minor, P.D., and Almond, J.W. (1984). *Nucl. Acids Res.* **12**, 7859-7875.
64. Carroll, A.R., Rowlands, D.J., and Clarke, B.E. (1984). *Nucl. Acids Res.* **12**, 2461-2472.
65. Lindberg, A.M., Stalhandske, P.O.K., and Pettersson, U. (1987). *Virology* **156**, 50-63.
66. Najarian, R., Caput, D., Gee, W., Potter, S.J., Renard, A., Merryweather, J., Van Nest, G. and Dina, D. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2627-2631.
67. Mills, D.R., Priano, C., DiMauro, P., Binderow, B.D. (1988). *J. Mol. Biol.* **205**, 751-764.
68. Inokuchi, Y., Jacobson, A.B., Hirose, T., Inayama, S., Hirashima, A. (1988). *Nucleic Acids Res.* **16**, 6205-6221.
69. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min Jou, W., Molemans, F., Raeymaekers, A., Van den Berghe, A., Volckaert, G. and Ysebaert, M. (1976). *Nature* **260**, 500-507.
70. Inokuchi, Y., Takahashi, R., Hirose, T., Inayama, S., Jacobson, A.B., and Hirashima, A. (1986). *J. Biochem.* **99**, 1169-1180.
71. Luo, M., Vriend, G., Kamer, G., Minor, I., Arnold, E., Rossmann, M.G., Boege, U., Scraba, D.G., Duke, G.M., and Palmenberg, A.C. (1987). *Science* **235**, 182-191.
72. Bruenn, J.A., Diamond, M.E., and Dowhanick, J.J. (1989). *Nucl. Acids Res.* **17**, 7487-7493.
73. Acharya, R., Fry, E., Stuart, D., Fox, G., Rowlands, D., and Brown, F. (1989). *Nature* **337**, 709-716.
74. Goldbach, R. and Wellink, J. (1988). *Intervirology* **29**, 260-267.
75. Pirone, T.P. and Kassanis, B. (1975). *J. Gen. Virol.* **29**, 257-266.
76. Pirone, T.P. and Shaw, J.G. (1973). *Virology* **53**, 274-276.
77. Holland, J.J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and VandePol, S. (1982). *Science* **215**, 1577-1585.
78. Reaney, D.C. (1982). *Ann. Rev. Microbiol.* **36**, 47-73.
79. Holland, J.J. (1984). Continuum of change in RNA virus genomes, p. 137-143. In A.L. Notkins and M.B.A. Oldstone (ed.). *Concepts in viral pathogenesis*. Springer-Verlag. New York.
80. Domingo, E., Martinez-Salas, E., Sobrino, F., de la Torre, J.C., Portela, A., Ortin, J., Lopez-Galindez, C., Perez-Brefia, P., Villanueva, N., Najera, R., VandePol, S., Steinhauer, D., DePolo, N. and Holland, J.J. (1985). *Gene* **40**, 1-8.
81. Goldbach, R.W. (1986). *Ann. Rev. Phytopathol.* **24**, 289-310.
82. Smith, D.B. and Inglis, S.C. (1987). *J. Gen. Virol.* **68**, 2729-2740.
83. Rodriguez-Cerezo, E., Moya, A., and Garcia-Arenal, F. (1989). *J. Virol.* **63**, 2198-2203.
84. Kirkegaard, K. and Baltimore, D. (1986). *Cell* **47**, 433-443.
85. Huisman, M.J., Cornelissen, B.J., Groenendijk, C.F., Bol, J.F., and van Vloten-Doting, L. (1989). *Virology* **171**, 409-16.
86. Angenent, G.C., Posthumus, E., Brederode, F.T., and Bol, J.F. (1989). *Virology* **171**, 271-4.
87. Banner, L.R., Keck, J.G., and Lai, M.M. (1990). *Virology* **175**, 548-55.
88. Allison, R., Thompson, C., and Ahlquist, P. (1990). *Proc. Natl. Acad. Sci. USA* **87**, 1820-1824.
89. Boccardo, G. and Accotto, G. (1988). *Virology* **163**, 413-9.
90. White, T.C. and Wang, C.C. (1990). *Nucleic Acids Res.* **18**, 553-9.
91. Mindich, L. (1978). Bacteriophages that contain lipid. In H. Fraenkel-Conrat and R.R. Wagner (ed.), *Comprehensive Virology*, vol. 12. Plenum Press. NY. pp. 271-335.
92. Vidaver, A.K., Koski, R.K., and Van Etten, J.L. (1973). *J. Virol.* **11**, 799-805.