Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 10647–10651, December 1991
Evolution

Evidence for common ancestry of a chestnut blight hypovirulence-associated double-stranded RNA and a group of positive-strand RNA plant viruses

EUGENE V. KOONIN*†, GIL H. CHOI‡, DONALD L. NUSS‡, RONI SHAPIRA‡, AND JAMES C. CARRINGTON*§

*Department of Biology, Texas A&M University, College Station, TX 77843; and ‡Department of Molecular Oncology and Virology, Roche Institute of Molecular Biology, Roche Research Center, Nutley, NJ 07110

Communicated by Myron K. Brakke, August 1, 1991 (received for review June 1, 1991)

ABSTRACT **Computer-assisted analysis of the putative polypeptide products encoded by the two open reading frames present in a large virus-like double-stranded RNA, L-dsRNA, associated with hypovirulence of the chestnut blight fungus, *Cryphonectria parasitica*, revealed five distinct domains with significant sequence similarity to previously described conserved domains within plant potyvirus-encoded polyproteins. These included the putative RNA-dependent RNA polymerase, RNA helicase, two papain-like cysteine proteases related to the potyvirus helper-component protease, and a cysteine-rich domain of unknown function similar to the N-terminal portion of the potyvirus helper-component protein. Phylogenetic trees derived from the alignment of the polymerase domains of L-dsRNA, a subset of positive-stranded RNA viruses, and double-stranded RNA viruses, using three independent algorithms, suggested that the hypovirulence-associated dsRNA and potyvirus genomes share a common ancestry. However, comparison of the organization of the conserved domains within the encoded polyproteins of the respective viruses indicated that the proposed subsequent evolution involved extensive genome rearrangement.**

The phenomenon of transmissible hypovirulence represents a natural form of biological control in which the virulence of *Cryphonectria parasitica*, the chestnut blight fungus, is modulated by the presence of virus-like genetic elements composed of double-stranded RNA (dsRNA) (reviewed in refs. 1 to 4). Efforts to understand the molecular mechanisms responsible for transmissible hypovirulence have provided an emerging view of the structural and functional properties of hypovirulence-associated dsRNA genetic elements (reviewed in ref. 5). The largest dsRNA present in *C. parasitica* strain Ep713, L-dsRNA (12,712 base pairs), was shown recently to contain two contiguous coding domains, designated open reading frame (ORF) A and ORF B, consisting of 622 and 3165 codons, respectively (6). Both ORFs encode polyproteins that undergo autocatalytic processing during or immediately after translation (6, 7). On the basis of the similarity of the L-dsRNA genetic organization and expression strategy to those of several viral genomes, it was suggested that L-dsRNA should be considered the equivalent of a viral genome or replicative form and the descriptive term hypovirulence-associated virus (HAV) was proposed (6).

Similarities between one of the HAV-encoded proteases, p29, and the potyvirus-encoded protease, HC-Pro, were noted previously (7). In addition, computer-assisted analysis of the C-terminal portion of the ORF B-encoded polyprotein revealed a domain that was clearly related to the RNA helicase of potyviruses (6). We now extend these observations by demonstrating that three additional domains are conserved between the gene products of HAV and those of potyviruses, including the putative RNA polymerase. A detailed analysis of the sequence and organization of the conserved domains within the HAV-encoded and potyvirus-encoded polyproteins suggests that Ep713 HAV dsRNA and the positive-stranded RNA genomes of the potyviruses share a common ancestry and that their evolution included extensive gene shuffling.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Polyproteins encoded by HAV RNA, barley yellow mosaic virus (BaYMV) RNA 1, BaYMV RNA 2, and pea seed-borne mosaic virus (PSBMV) RNA were from refs. 6, 8, 9, and 10, respectively. All other sequences were taken from the Swissprot data base (release 16). Methods used for multiple sequence alignment, cluster dendrogram formation, and parsimony tree analysis were described in detail by Dolja *et al.* (11) and in references therein.

RESULTS AND DISCUSSION

Hypovirulence-Associated dsRNA Encodes a Putative RNA-Dependent RNA Polymerase Related to the Polymerases of Positive-Strand RNA-Containing Potyviruses and BaYMV. Considering the similarity between the putative helicase of HAV and the potyvirus-encoded helicase (6), we first compared the sequences of HAV gene products with those of potyviruses and BaYMV. The extensive sequence similarities, despite differences in genome organization, between the proteins encoded by potyviruses and BaYMV have been described (8, 9, 12). Inspection of the local similarity plots revealed one of the conserved motifs of the RNA polymerases of the positive-strand RNA viruses (designated motif IV in Fig. 1) as the region of highest similarity over the entire lengths of the HAV ORF B polyprotein and the polyproteins encoded by potyviruses and BaYMV, while motif V scored

Abbreviations: dsRNA, double-stranded RNA; ORF, open reading frame; HAV, chestnut blight fungus hypovirulence-associated virus; HC, helper component. Potyviruses: PPV, plum pox virus; PSBMV, pea seed-borne mosaic virus; PVY, potato virus Y; TEV, tobacco etch virus; TVMV, tobacco vein mottling virus. Potyvirus-like virus: BaYMV, barley yellow mosaic virus. dsRNA viruses: ScV, *Saccharomyces cerevisiae* virus; BTV, bluetongue virus; ROTA, human rotavirus SA11; REO, reovirus type 3; IBDV, infectious bursa disease virus. Togaviruses: SFV, Semliki forest virus; SNBV, Sindbis virus; RRV, Ross river virus; ONNV, O'Nyong-Nyong virus; VEEV, Venezuelan equine encephalitis virus. Picornaviruses: PV1, poliovirus type 1; EMCV, encephalomyocarditis virus; FMDV, foot-and-mouth disease virus type A10; HA, hepatitis A virus strain LA. Comovirus: CPMV, cowpea mosaic virus. Nepovirus: GCMV, Hungarian grapevine mosaic virus. Nodavirus: BBV, black beetle virus.

†Permanent address: Institute of Microbiology, Academy of Sciences, 117811 Moscow, Russia.
§To whom reprint requests should be addressed.

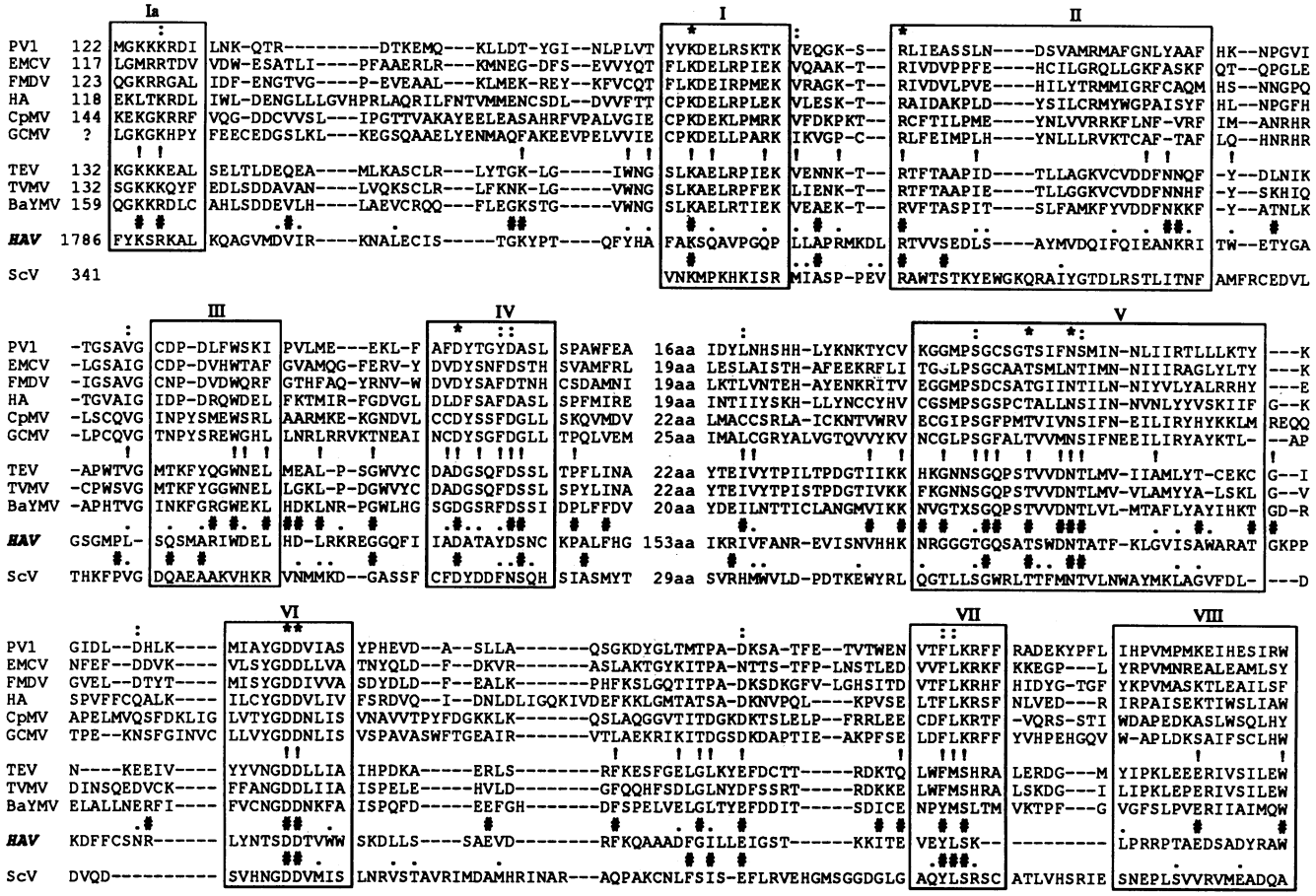

FIG. 1. Putative RNA-dependent RNA polymerase encoded by the chestnut blight hypovirulence-associated dsRNA. The alignment of the putative HAV polymerase domain with the conserved domains of polymerases from supergroup I positive-strand RNA viruses and the *Saccharomyces cerevisiae* dsRNA virus ScV is shown. (See abbreviation footnote for names of viruses.) The conserved motifs found in all positive-strand RNA viral polymerases (I-VIII), as well as a motif specific for the polymerase supergroup I (Ia), are boxed. For the ScV polymerase, no definite alignment in the region upstream of motif I could be found. For the polymerases of HAV and ScV, the spacers between segments IV and V were not aligned; their lengths are indicated. *, Identical amino acid residues in all sequences; :, identical or similar residues in all sequences; !, identical or similar residues in the sequences of HAV, the potyviruses and BaYMV; #, identical residues in HAV and BaYMV, or in HAV and ScV; ., similar residues in HAV and BaYMV, or in HAV and ScV. The following groups of similar amino acid residues were considered: G and A; S and T; D, E, N, and Q; K and R; L, I, V, and M; and F, Y, and W. The distances from the protein N-termini to the aligned segments are indicated. Additional sequence, extending past the C-termini of the alignments, is found in each protein.

second highest. Analysis of the surrounding portion of the ORF B protein using the SITE program and visual inspection also identified the counterparts to six other conserved motifs typical of positive-strand RNA viral RNA polymerases (13–16). Comparisons of the HAV polymerase-like sequence with all known sequences of the RNA polymerases of positive-strand RNA and dsRNA viruses showed that the closest similarity was indeed with potyviruses and BaYMV, followed by picorna-, como-, and nepoviruses. Multiple alignment of these supergroup I (picornavirus-like) positive-strand RNA virus polymerases (16) with the HAV sequence revealed two segments (motifs Ia–IV and motifs V–VII) that aligned with convincing alignment score (AS) values of 7.2 SD and 9.6 SD, respectively. Generally, scores above 7 SD are considered a solid indication of relatedness among sequences, while scores between 4 and 7 SD are significant, provided that additional evidence to support the suspected relationship is available.

In view of the recent suggestion that dsRNA viruses may constitute a monophyletic group (17), the possible relationship between the putative RNA polymerase of HAV and the RNA polymerases of dsRNA viruses was investigated. Initial local similarity searches did not reveal specific relationships between the putative HAV polymerase sequence and any of the six known sequences of dsRNA viral polymerases (not shown). Alignments of the polymerases of dsRNA viruses, supergroup I positive-strand RNA viral polymerases, and the HAV polymerase mostly had AS values of less than 5 SD. These results are exemplified in Fig. 1 by the alignment of the yeast dsRNA virus (ScV) polymerase. The ScV sequence was selected among the dsRNA virus polymerases for inclusion in Fig. 1 only on the basis that both ScV and HAV replicate in fungal hosts. It is clear that the overall similarity between HAV and ScV was less pronounced than that between HAV, the potyviruses, and BaYMV. Within the region compared, HAV shared 45 identical and 29 similar residues with BaYMV (263 residues aligned) versus 24 and 28 residues with ScV (259 residues aligned). A unique feature of the putative polymerase of HAV was the insertion of ca. 130 amino acid residues between motifs IV and V. Inserts of somewhat smaller size have been found previously at similar locations in the putative polymerases of two dsRNA viruses, bluetongue virus and reovirus (13). Another interesting feature of the putative HAV polymerase was the substitution of serine for glycine in the highly conserved GDD tripeptide (motif VI in Fig. 1). Analogous substitutions occur in the putative polymerases of coronaviruses, toroviruses, several negative-strand RNA viruses, and the dsRNA bacteriophage $\phi 6$ (13, 15, 18–20), although no particular sequence similarities could be detected between the putative polymerases of these viruses and HAV despite an extensive search.

The alignment shown in Fig. 1 was used for the generation of tentative phylogenetic trees using three independent algorithms. Additionally, the polymerase of black beetle virus (BBV), demonstrated to be a peripheral member of the polymerase supergroup I (16), and the polymerases of dsRNA viruses were included in the analysis. The cluster dendrogram shown in Fig. 2 suggests grouping of the HAV polymerase with the polymerases of the potyviruses and BaYMV. A similar branching order was obtained when the maximum topological similarity algorithm was used. The protein parsimony algorithm produced a somewhat different topology, suggesting separation of HAV before the division of the potyvirus/BaYMV and picorna/como/nepovirus groups (not shown). Despite this apparent difference, each method demonstrated the grouping of the HAV polymerase with the polymerases of positive-strand RNA viruses, and not with those of dsRNA viruses (Fig. 2).

The Putative Helicase of HAV Belongs to the Helicase Superfamily II and Is Related to its Positive-Strand RNA Viral Members. Previous analysis detected motifs typical of helicases in the C-terminal portion of the ORF B product of HAV (6) and showed some similarity to the putative RNA helicase (CI protein) of tobacco vein mottling potyvirus (TVMV). The more detailed analysis performed here confirms that the HAV helicase-like sequence largely conformed to the consensus pattern of conserved amino acid residues typical of the so-called helicase superfamily II (21), though it contained some notable deviations (Fig. 3). This superfamily brought together the (putative) RNA helicases of poty-, flavi- and pestiviruses, that of hepatitis C virus (22), the "DEAD" family of cellular RNA helicases (with translation initiation factor eIF-4A as the prototype), and a number of DNA helicases. We generated the alignment of these helicases with the putative helicase of HAV, yielding a convincing AS of over 10 SD (shown in excerpts in Fig. 3). Tentative phylogenetic trees revealed the grouping of the putative helicase of HAV with those of positive-strand RNA viruses, though it seemed to be roughly equidistant from the helicases of potyviruses and BaYMV, on the one hand, and those of

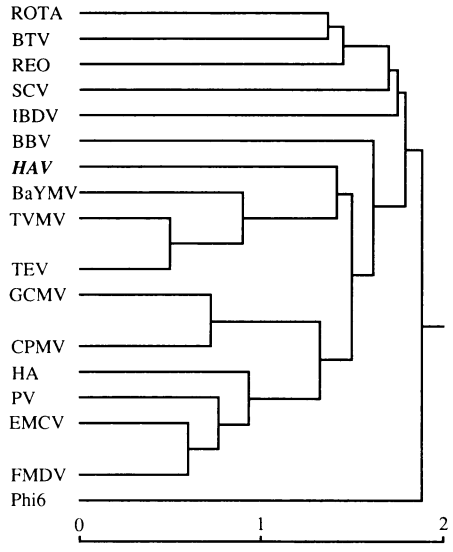

FIG. 2. Cluster dendrogram for the conserved regions of the RNA polymerases of HAV, positive-strand RNA viruses, and dsRNA viruses. The branch lengths are proportional to the distances calculated. A logarithmic scale showing these distances (in relative units) is at the bottom. The branching order in this dendrogram does not necessarily imply that the polymerases of the majority of dsRNA viruses are monophyletic. Upon further analysis, some may join other groups of positive-strand RNA viruses that were not analyzed here.

```
                          I                                  Ia
JEV    187aa QMTVLDLHPGSGKTR-KILPQ 10aa RTAVLAPTRVV 50aa
KUN    187aa QITVLLDLHGAGKTR-RILPQ 10aa RTAVLAPTRVV 50aa
WNV    187aa QITVLDLHPGAGKTR-KILPQ 10aa RTAVLAPTRVV 50aa
YFV    191aa MTTVLDFHPGAGKTR-RFLPQ 10aa RTLVLAPTRVV 50aa
TBEV   191aa QITVLDMHPGSGKTH-RVLPE 10aa RTLVLAPTRVV 50aa
BVDV  1898aa DFKQITLATGAGKTT-E-LPK 10aa RVLVLIPLRAA 54aa
HCV   1223aa QVAHLHAPTGSGKST-K-VPA  7aa KVLVLNPSVAT 49aa
TEV     77aa RDFLVRGAVGSGKST-G-LPY  6aa RVLMLEPTRPL 53aa
PPV     77aa QDILIRGAVGSGKST-G-LPF  6aa HVLLIEPTRPL 53aa
PVY     78aa LDFLVRGAVGSGKST-G-LPV  6aa SVLLIEPTRPL 53aa
TVMV    78aa KDIILMGAVGSGKST-G-LPT  6aa GVLLLEPTRPL 53aa
BaYMV   87aa WSMVV-GHTGSGKST-Y-LPV 14aa QILICEPTQAA 53aa
Consensus      ++    G GKS    LP         +++ PTR +
                                T
HAV   2657aa GHVtVAAKTASGKST-F-fPA 12aa KLWIVMPrKIL 45aa
eIF-4A  69aa YDVIAQAQSGTGKTATFAIsI 11aa QALVLAPTREL 63aa

          II                  III                 IV
JEV     YNLFVMDEAHFT 18aa EAAAIFM-TATPPG 29aa VWFVASVKMGNEI
KUN     YNLFVMDEAHFT 18aa EAAAIFM-TATPPG 29aa VWFVPSVKMGNEI
WNV     YNLFIMDEAHFT 18aa EAAAIFM-TATPPG 29aa VWFVPSVKMGNEI
YFV     WEVIIMDEAHFL 18aa ESATILM-TATPPG 29aa AWFLPSIRAANVM
TBEV    WEVAIMDEAHWT 18aa KCALVLM-TATPPG 29aa AWFVPSSAKGGII
BVDV    YSYIFLDEYHCA 16aa SIRVVAM-TATPAG 32aa LVFVPTRNMAVEV
HCV     YDIIICDECHST 18aa GARLVVLATATPPG 28aa LIFCHSKKKCDEL
TEV     YDFVIIDECHVN 16aa EGKVLKV-SATPPG 32aa LVYVASYNDVDSL
PPV     YKCIIFDECHVH 16aa PGKILKV-SATPPG 32aa LVYVASYNEVDTL
PVY     FNFVIFDECHVL 16aa ACKVLKV-SATPVG 32aa LVYVSSYNEVDTL
TVMV    YQFIIFDEFHVL 16aa NGKIIKV-SATPPG 32aa LVYVASYNEVDQL
BaYMV   FDAIFLDEAHDV 15aa SVRKFYV-SATPRD 31aa LVFLAGRPECIKA
Con     Y ++++DE H           + + TATP G      ++F+ S
                                    S           Y
HAV     dNLVFFDEFHEM 11aa KGPTIFM-SATPVA 36aa MIiVPTYNELKKT
eIF-4A  iKMFVLDEAdEM 18aa NTQVVLL-SATmPs 46aa VIFINTRRKVDWL

                        V                          VI
JEV     18aa VITTDISEMGANF-GAS-RVID 28aa SAAQRRGRVGRNP  153aa
KUN     18aa VVTTDISEMGANF-KAS-RVID 28aa SAAQRRGRTGRNP  153aa
WNV     18aa VYTTDISEMGANF-KAS-RVID 28aa SAAQRRGRIGRNP  153aa
YFV     18aa ILATDIAEMGANL-CVE-RVLD 27aa SAAQRRGRIGRNP  154aa
TBEV    18aa VVTTDISEMGANL-DVS-RVID 25aa SAAQRRGRVGRQE  153aa
BVDV    19aa IVATNAIESGVTLPDLD-TVID 30aa EQAQRRGRVGRVK 1797aa
HCV     17aa VVATDALMTGYTG-DFD-SVID 29aa SRTQRRGRTGRGK 1515aa
TEV     22aa IVATNIIENGVTI-DID-VVVD 26aa ERIQKLGRVGRHK  274aa
PPV     22aa VVATNIIENGVTL-DID-VVVD 26aa ERIQRLGRVGRNK  275aa
PVY     22aa VVATNIIENGVTL-DID-VVVD 26aa ERIQRLGRVGRFK  274aa
TVMV    22aa IVATNIIENGVTL-DVD-VVVD 26aa ERIQRFGRVGRNK  275aa
BaYMV   24aa IFTTNIIETGVTL-SVD-CVVD 26aa ERQQRIGRAGRLK  282aa
Con          ++ TDI E G  +       V+D         QR GR+GR
                    N
HAV     17aa LVCTpYVQTGIDIKPAPSILID 22aa TNEQRVnRVGRTM  136aa
eIF-4A  22aa LITTDLLaRGIDVQQVS-LVIN  7aa NYIhRIGRgGRFG   39aa
```

FIG. 3. Conserved motifs typical of the helicase superfamily II in the putative helicase of HAV. The consensus pattern of conserved amino acid (aa) residues for the positive-strand RNA viral members of the helicase superfamily II (after ref. 21 with some modifications), the putative HAV helicase, and eukaryotic translation initiation factor eIF-4A is shown. The HAV-encoded and eIF-4A residues that do not conform to the consensus are shown in lowercase. Residues that are structurally conserved (see Fig. 1 legend) between the HAV-encoded sequence and most of the other sequences are indicated by +.

flaviviruses, pestiviruses, and hepatitis C virus, on the other hand (not shown). Thus, our analysis demonstrated the possibility of a common origin for the HAV helicase and the RNA helicases of positive-strand RNA viruses, though it failed to reveal a close association of the former with a specific group of the latter. One could argue that the HAV helicase has diverged far from its positive-strand RNA viral relatives because of different functional requirements associated with replication and/or transcription of the HAV dsRNA.

HAV Encodes Two Cysteine Proteases and an Additional Domain Related to the Helper-Component Protein of Potyviruses. Recent studies have revealed that the two HAV-encoded polyproteins undergo autoproteolysis. The product of ORF A is processed into p29 and p40, with the proteolytic activity residing in p29 (7). A 48-kDa protein, p48, is autocatalytically cleaved from the N terminus of the ORF B product (6). Site-directed mutagenesis (23) has identified Cys-162 and His-215 of p29 as residues essential for autoproteolysis, indicating that this enzyme, like the potyvirus-encoded helper-component (HC) protease (24), resembles papain-like proteases. Additional similarities between p29 and HC protease have been noted previously (7) in the form of conserved amino acid sequences around the essential cysteine and histidine residues, the nature of the cleavage dipeptides, and the distances between the essential residues and the cleavage sites. Recent analysis of p48 has also

revealed a single cysteine (Cys-341) and a single histidine (His-388) as essential for autoproteolysis (25). A more detailed comparison of the two proteases of HAV with the potyvirus and BaYMV HC protease domains yielded an alignment including quite a number of conserved residues, despite its relatively modest statistical significance, with AS values of 5.1 and 3.9 SD for p29 and p48, respectively (Fig. 4).

The presence of two related proteases among HAV genome products led Shapira and Nuss (25) to speculate that the respective genes could have evolved by duplication. Tentative phylogenetic trees were generated to assess the relatedness of these proteases to one another and to potyviral HC proteases. The sequences of the C-terminal domains of alphavirus nsP2 proteins, which are cysteine proteases (26) that display significant similarity to the potyvirus HC proteases (A. E. Gorbalenya, E.V.K., and M. M. C. Lai, unpublished observations), were included as an outgroup. All three tree-generating algorithms mentioned above agreed in suggesting the grouping between the p29 and p48 proteases, as exemplified by the cluster dendrogram in Fig. 5. Thus the two protease genes of HAV could have evolved by intragenomic duplication followed by rapid divergence due possibly to the functional diversification of the proteases. Despite an extensive computer-assisted search, no similarity was detected between any additional HAV-encoded proteins and serine, cysteine, acid, or picornavirus 3C-type proteases.

Comparison of the sequences of the potyvirus HCs and p29 of HAV revealed an additional domain of moderate similarity (AS of 5.8 SD) at the N termini of these proteins, marked by the conservation of three Cys residues (Fig. 4). Potyvirus HC proteins contain approximately 200 amino acid residues between the N-terminal conserved region and the C-terminal protease domain, for which there is no counterpart in p29. HC proteins, therefore, may consist of three distinct domains, with the homologs of only the C-terminal protease and the N-terminal domain present in p29. In contrast, the respective protein of BaYMV contains only the protease domain and a truncated form of the central domain of HC (Fig. 4).

Comparison of the Genetic Organizations and Expression Strategies of HAV, Potyviruses, and BaYMV: A Possible Case of Evolution by Extensive Genome Rearrangement. The comparative amino acid sequence analysis described above revealed four conserved domains among the gene products of HAV, potyviruses, and BaYMV (one of which is duplicated in HAV). Tartaglia *et al.* (27) speculated previously that hypovirulence-associated dsRNAs are analogous to the replicative form of an ancestral single-stranded RNA virus. Considering the relative organization of the conserved domains within the HAV-encoded, potyvirus-encoded, and BaYMV-encoded polyproteins (Fig. 6), it is tempting to suggest that HAV-dsRNA could have evolved by rearrangement of a positive-strand RNA potyvirus-like genome. The scenario for this rearrangement could include the following major events (their indicated order is arbitrary): (*i*) transposition of the helicase gene; (*ii*) duplication of the sequence encoding the protease domain of HC; (*iii*) deletion of the sequence encoding the protease domain of NI_a; (*iv*) deletion of the capsid protein gene; and (*v*) emergence of the termination codon separating ORFs A and B.

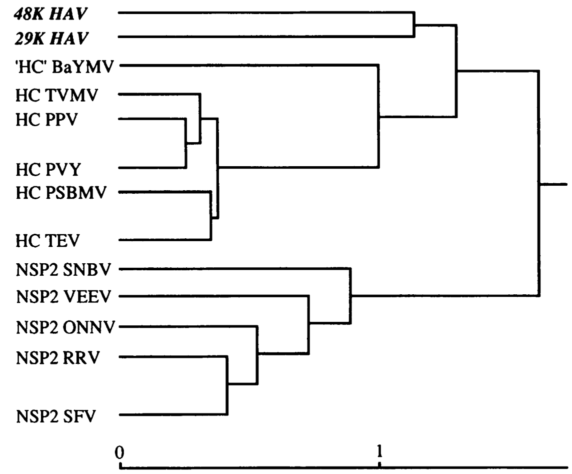

FIG. 5. Cluster dendrogram for the papain-like proteases of HAV (48- and 29-kDa proteases), potyviruses, BaYMV, and alphaviruses. NSP2, nsP2 cysteine proteases. For details see legend to Fig. 2.

For steps *iii* and *iv*, the alternative is the loss of the respective function followed by extensive divergence. In view of the potential for movement of the ancestral virus within the mycelium and the frequency of anastomosis, it is conceivable that an extracellular route of infection, and the required packaging function, would have been dispensable. Similarly, in the absence of a capsid protein, the ratio of the single-stranded RNA genome to the replicative form could have been altered so that the double-stranded form predominated.

```
                :   :                  ** : : :        :     *  *
TEV    HC 320aa FLELRPDGI-SHECTRGVS-VERCGEVAAILTQALSPCGKI---TCKRC-MV-ETPDIVEGESGESVTNQGK-LLAML-KEQYPDF
PPV    HC 322aa FMQCKLRET-DHQSTSDLD-VKECGDVAALVCQAIIPCGKI---TCLQC-AQ-KYSYMSQQEIRDRFSTVIEQHEKTA-MDNYPQF
PVY    HC 297aa WARMRYPS--DHTCVAGLP-VEDCGRVAALMAHSILPCYKI---TCPTC-AQ-QYASLPVSDLFKLLHKHARDGLNRL-GADKDRF
TVMV   HC 269aa FQEQKAIGL-DHTCTSDLP-VEACGHVAALMCQSLFPCGKI---TCKRC-IA-NLSNLDFDTFSELQGDRAMRILDVM-RARFPSF
PSBMV  HC 411aa YVKLRDESVSDHDCVGGIT-PEECGILAAQILRVFYPCWRI---TCTKC-IS-NWLSKPTSEQIEHIYERGNLAIQDL-NKRIPSA
HAV    P29 24aa FVSVRTEEVVPAGCITLWEYRDSCGDVPGPLSHGDLRRLRTPDGVC-KCQVHFELPTVLKSGSTGTVPEHPAVLAAFIGRPRRCSL

                                                                           *
                   :                        ::                           * *:      :
TEV    HC    PMAEKLLT-RFLQQKSLV 199aa HPTKRHLVIGNSGDSKYLD--LPVLNEEKMYIANEGYCYMNIFFALLVNVKEEDAKDFTKFIRDTI
PVY    HC    SHVLAFLK-RYRELMRVE 199aa SPTKNHLVVGNTGDSKYVD--LPTAKGGAMFIAKAGYCYINIFLAMLININEDEAKSFTKTVRDTL
PPV    HC    IHVNKFLI-ALEHLETPV 200aa PPTKKHLVIGNSGDQKFVD--LPKGDSEMLYIAKQGYCYINVFLAMLINISEEDAKDFTKKVRDMC
TVMV   HC    THTIRFLH-DLFTQRRVT 199aa MPTKNHLVIGNSGDPKYLD--LPGEISNLMYIAKEGYCYINIFLAMLVNVDEANAKDFTKRVRDES
PSBMV  HC    HHVTQMVE-LLRQRIKNT 197aa LPTRNHLVIGNTGDPKLVD--LPKTETGRMWIAKEGYCYINIXFAMLVNVSEKDAKDFTKFVRDEI
BaYMV'HC'                       107aa SHAPWLYMA-NNVCAYEAT-HLKPVQTFIAFDFAHGYCYLSLFIPLSFRITPENARSFSRFL-EQL
HAV    P29   EQRTKELDSRFLQLVHGG   0aa LPARPSYMIARPPRP-VRG--LCSSRNGSLAQFGQGYCYLS---AIVDSARWRVARTTGWCVRVAD
HAV    P48                      306aa AP-RQQLVPRRST---FVDNHEEVEIDTLRVPVEEGRCF-----ELLFN-NQVTPAIFDK--KPLL
                                                         *
                *  :                        :            *:                             *        :*
TEV    HC    VPKLGAWPTMQDVATACYLLSILYPDVLRAELPRILVDHDNKTMHVLDSYGSRTTGYHMLKMNTTSQLIEFVHSGLESEMKTYNVG
PVY    HC    VPKLGTWPSMMDLATACHFLAVLYPETRNAELPRILVDHEAKIFHVVDSFGSLSTGMHVLKANTINQLISFASDTLDSNMKTYLVG
PPV    HC    VPKLGTWPTMMDLATTCAQMRIFYPDVHDAELPRILVDHDTQTCHVVDSFGSQTTGYHILKASSVSQLILFANDELESDIKHYRVG
TVMV   HC    VQKLGKWPSLIDVATECALLSTYYPAAASAELPRLLVDHAQKTIHVVDSYGSLNTGYHILKANTVSQLEKFASNTLESPMAQYKVG
PSBMV  HC    MPQLGKWPTMMDVATACYKLAIIYPDVRDAQLPRILVDHSEQIFHVIDSYGSMTTGYHILKAGTVSQLISFXHGALLGEMKMYRVG
BaYMV'HC'    PDILGAYPTLAAIYKTMLFAIRLFPEVLQAPIPIIAKRPGVLQFHVSDARG-LPPSWFPMKCGSVRSFVALITNNLNSDLLDGIVG
HAV    P29   YLRLLQWVGRRSFGS--------F------QIEKSAVDH---VYHVV-----VDAEYQSEQDGALFYQAIL---GLAEKDPLARIG
HAV    P48   KDVLGVFEEV------C-------------TMDSLEISHSDQCVHIVA--GETFRNYDEIK--AVLEVI------LENE-PDILVG
```

FIG. 4. Two papain-like proteases and the HC-related domain of HAV. The alignments with two domains of potyvirus-encoded HC proteins are shown. The distances between the two HC domains, and the distances between the N terminus of each polyprotein and the indicated sequence, are given. The C-terminal boundaries of the aligned segments correspond to the cleavage sites. The N-terminal sequences of the HC-related protein of BaYMV and of p48 are not shown, as they displayed no detectable similarity to any portion of the potyvirus proteins or to each other. *, Identical residues in all sequences; :, similar residues in all sequences; ⁑, the (putative) catalytic residues of the proteases; boldface, individual residues that are similar or identical to HAV residues.

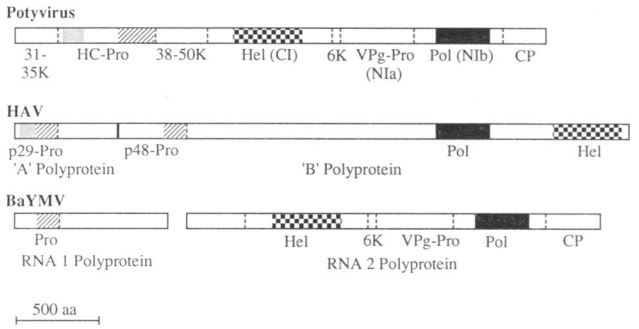

FIG. 6. Schematic comparison of the organization of the polyproteins of HAV, potyviruses, and BaYMV. Related domains are highlighted by identical shading. The termination codon of the HAV ORF A is indicated by a solid vertical line, and the cleavage sites are denoted by dashed lines. Pol, RNA polymerase; Hel, helicase; Pro, protease; CP, capsid protein. The BaYMV scheme was designed by combining data from refs. 8, 9, and 12.

Step *v* is parallel to the possible evolution of BaYMV, which appears to have involved the isolation of the 5'-terminal genes in a separate RNA segment. It should be noted that BaYMV is transmitted in nature by a fungal vector (28), although the ability of BaYMV to replicate in the vector has not been demonstrated. The order of the helicase and polymerase domains observed in the potyviruses and BaYMV (N-. . . helicase-. . . -polymerase-. . . -C) is typical of a wide range of positive-strand RNA viruses (21, 29). Interestingly, the reversed order observed in HAV has also been found in corona- and toroviruses (18–20). Moreover, the overall organization of the HAV L-dsRNA is surprisingly similar to that of coronaviruses, with coding sequences for two papain-like proteases in the 5'-proximal part of the RNA and sequences encoding the polymerase-helicase array in the 3'-proximal region (30). However, upon careful comparison, we have not found any specific sequence similarities between proteins encoded by HAV and the coronaviruses. Thus we believe that the striking analogy in the genetic organizations of these viruses is a peculiar case of evolutionary convergence.

If HAV and potyviruses had a common origin, was the host of the ancestral virus a plant or a fungus? Perhaps the most plausible series of events involves the acquisition of a potyvirus-like virus by the fungus during saprophytic or pathogenic interactions with a plant host. Since the HAV dsRNA confers a hypovirulent phenotype to *C. parasitica*, acquisition of an ancestral virus may have provided a survival advantage to both the plant host and fungal pathogen. Alternatively, a fungal host could have acted as a vector, transmitting an ancestral dsRNA virus to a plant host followed by subsequent evolution to a single-stranded RNA virus as suggested by Bruenn (17). These considerations are complicated because of the apparent lack of an encapsidated from of the HAV RNA, prohibiting a clear distinction between genomic and nongenomic RNA in the classical sense. Nevertheless, in considering the evolutionary relationship between single-stranded and double-stranded RNA viruses, it is interesting to note that HAV dsRNA frequently undergoes internal deletion events and that recombination with cellular RNA may also occur (31). This propensity for recombination is compatible with the proposed genome reorganization in the course of evolution of HAV dsRNA from a potyvirus-like ancestor. Moreover, one can envision that additional internal deletion events could lead to the formation of segmented dsRNA genomes from HAV-like genetic elements. Previously, gene module shuffling has been highlighted as a major trend in the evolution of positive-strand RNA viruses (32, 33). Here, we demonstrate that a similar process may account for the evolution of a dsRNA virus-like genetic element from a positive-strand RNA virus.

E.V.K. is most grateful to Dr. A. E. Gorbalenya for helpful discussions and for permission to use his unpublished observations. Thanks are also due to Dr. V. V. Dolja for critical reading of the manuscript and to Drs. A. Davidson and S. Kashiwazaki for providing preprints of their work. The excellent assistance of Ms. Ula Dodson and Ms. Ashley Henderson in the preparation of this manuscript is appreciated. This work was supported in part by grants from the National Institutes of Health (AI27832), the U.S. Department of Agriculture (89-37263-4503), and the Texas Advanced Technology Program, and by funds provided by the Department of Biology, College of Science, and Provost's Office of Texas A&M University.

1. Anagnostakis, S. L. (1982) *Science* **215,** 466–471.
2. Van Alfen, N. K. (1982) *Annu. Rev. Phytopathol.* **20,** 349–362.
3. Roane, M. K., Griffin, G. J. & Elkins, J. (1986) *Am. Phytopathol. Soc. Monogr.*, pp. 1–53.
4. Griffin, G. J. (1986) *Hort. Rev.* **8,** 291–335.
5. Nuss, D. L. & Koltin, Y. (1990) *Annu. Rev. Phytopathol.* **28,** 37–58.
6. Shapira, R., Choi, G. H. & Nuss, D. L. (1991) *EMBO J.* **10,** 731–739.
7. Choi, G. H., Shapira, R. & Nuss, D. L. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 1167–1171.
8. Kashiwazaki, S., Minobe, Y., Omura, T. & Hibino, H. (1990) *J. Gen. Virol.* **71,** 2781–2790.
9. Kashiwazaki, S., Minobe, Y. & Hibino, H. (1991) *J. Gen. Virol.* **72,** 995–999.
10. Johansen, E. (1990) PhD thesis (Institute for Ecology and Molecular Biology, Copenhagen).
11. Dolja, V. V., Boyko, V. P., Agranovsky, A. A. & Koonin, E. V. (1991) *Virology* **184,** 79–86.
12. Davidson, A. D., Prols, M., Schell, J. & Steinbiss, H.-H. (1991) *J. Gen. Virol.* **72,** 989–993.
13. Koonin, E. V., Chumakov, K. M. & Gorbalenya, A. E. (1989) *FEBS Lett.* **252,** 42–46.
14. Koonin, E. V., Gorbalenya, A. E., Chumakov, K. M., Donchenko, A. P. & Blinov, V. M. (1987) *Mol. Genet. Mikrobiol. Virusol.* **7,** 27–39.
15. Pock, O., Sauvageut, I., Delarue, M. & Tordo, N. (1989) *EMBO J.* **8,** 3867–3874.
16. Koonin, E. V. (1991) *J. Gen. Virol.* **72,** 2197–2206.
17. Bruenn, J. A. (1991) *Nucleic Acids Res.* **19,** 217–226.
18. Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. & Blinov, V. M. (1989) *Nucleic Acids Res.* **17,** 4456–4469.
19. Bredenbeek, P. J., Pachuk, C. J., Noten, A. F. H., Charite, J., Luytjes, W., Weiss, S. R. & Spaan, W. J. M. (1990) *Nucleic Acids Res.* **18,** 1825–1832.
20. Snijder, E. J., Den Boon, J. A., Brendenbeek, P. J., Horzinek, M. C., Rijnbrand, R. & Spaan, W. J. M. (1990) *Nucleic Acids Res.* **18,** 4535–4542.
21. Gorbalenya, A. E., Koonin, E. v., Donchenko, A. P. & Blinov, V. M. (1989) *Nucleic Acids Res.* **17,** 4713–4730.
22. Miller, R. M. & Purcell, R. M. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2057–2061.
23. Choi, G. H., Pawlyk, D. M. & Nuss, D. L. (1991) *Virology* **183,** 747–752.
24. Oh, C.-S. & Carrington, J. C. (1989) *Virology* **171,** 692–699.
25. Shapira, R. & Nuss, D. L. (1991) *J. Biol. Chem.* **266,** 19419–19425.
26. Hardy, W. R. & Strauss, J. H. (1989) *J. Virol.* **63,** 4653–4664.
27. Tartaglia, J., Paul, C. P., Fulbright, D. W. & Nuss, D. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 9109–9113.
28. Mathews, R. E. F. (1991) *Plant Virology* (Academic, San Diego), 3rd Ed.
29. Gorbalenya, A. E., Blinov, V. M., Donchenko, A. P. & Koonin, E. V. (1989) *J. Mol. Evol.* **28,** 256–268.
30. Lee, C.-J., Shieh, C.-K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhadzhyan, A. & Lai, M. M. C. (1991) *Virology* **180,** 567–582.
31. Shapira, R., Choi, G. H., Hillman, B. I. & Nuss, D. L. (1991) *EMBO J.* **10,** 741–746.
32. Zimmern, D. (1988) in *RNA Genetics*, eds. Holland, J. J., Domingo, E. R. & Ahlquist, P. (CRC, Boca Raton, FL), pp. 211–240.
33. Morozov, S. Yu., Dolja, V. V. & Atabekov, J. G. (1989) *J. Mol. Evol.* **29,** 52–62.