Molecular Evolution of the Paramyxoviridae and Rhabdoviridae Multiple-Protein-Encoding *P* Gene

I. King Jordan, Ben A. Sutter IV, and Marcella A. McClure[1]

Department of Biological Sciences, University of Nevada at Las Vegas

Presented here is an analysis of the molecular evolutionary dynamics of the *P* gene among 76 representative sequences of the Paramyxoviridae and Rhabdoviridae RNA virus families. In a number of Paramyxoviridae taxa, as well as in vesicular stomatitis viruses of the Rhabdoviridae, the *P* gene encodes multiple proteins from a single genomic RNA sequence. These products include the phosphoprotein (P), as well as the C and V proteins. The complexity of the *P* gene makes it an intriguing locus to study from an evolutionary perspective. Amino acid sequence alignments of the proteins encoded at the *P* and *N* loci were used in independent phylogenetic reconstructions of the Paramyxoviridae and Rhabdoviridae families. *P*-gene-coding capacities were mapped onto the Paramyxoviridae phylogeny, and the most parsimonious path of multiple-coding-capacity evolution was determined. Levels of amino acid variation for Paramyxoviridae and Rhabdoviridae *P*-gene-encoded products were also analyzed. Proteins encoded in overlapping reading frames from the same nucleotides have different levels of amino acid variation. The nucleotide architecture that underlies the amino acid variation was determined in order to evaluate the role of selection in the evolution of the *P* gene overlapping reading frames. In every case, the evolution of one of the proteins encoded in the overlapping reading frames has been constrained by negative selection while the other has evolved more rapidly. The integrity of the overlapping reading frame that represents a derived state is generally maintained at the expense of the ancestral reading frame encoded by the same nucleotides. The evolution of such multicoding sequences is likely a response by RNA viruses to selective pressure to maximize genomic information content while maintaining small genome size. The ability to evolve such a complex genomic strategy is intimately related to the dynamics of the viral quasispecies, which allow enhanced exploration of the adaptive landscape.

Introduction

Paramyxoviridae and Rhabdoviridae are families of negative-strand RNA viruses belonging to the order Mononegavirales (Pringle 1997). RNA viruses, such as the Paramyxoviridae and the Rhabdoviridae, exist as large heterogeneous populations of related nonidentical genomes referred to as quasispecies (Domingo and Holland 1994; Domingo et al. 1995). RNA virus quasispecies are characterized by their potential for extremely rapid evolution (Domingo et al. 1996). The rapid evolutionary rate of RNA virus populations makes them a tractable system for studying molecular evolution in real time (Fitch 1996). In addition, the presence of overlapping reading frames in a number of viral taxa (Samuel 1989) provides the opportunity to examine de novo molecular evolution of unique coding sequences (Ohno 1984; Keese and Gibbs 1992).

A vast number of genetic, biochemical, and comparative sequence analyses have determined the precise functions of the various Paramyxoviridae and Rhabdoviridae gene products (Kingsbury 1991; De and Banerjee 1997). Despite the success of these efforts, the phosphoprotein-encoding, or *P,* gene, is still a relatively uncharacterized component of the genome. The role of the *P* gene has perhaps been most precisely defined in the Rhabdoviridae (Wagner 1991). In particular, work on vesicular stomatitis viruses (VSVs) has shown that the phosphoprotein, or P protein, interacts directly with the polymerase and nucleoproteins and plays a role in transcription and replication (Barik and Banerjee 1992; Canter and Perrault 1996; Spadafora et al. 1996; Das et al. 1997). However, knowledge of the exact function of the phosphoprotein remains elusive. In addition, the lack of obvious conserved amino acid sequence motifs resulting from the rapid evolution (Bilsel et al. 1990; Morgan 1991) of this locus contributes to the enigmatic status of the proteins encoded by the *P* gene.

The *P* gene is further distinguished by the fact that in a number of Paramyxoviridae taxa, it encodes multiple gene products from overlapping reading frames (Galinski and Wechsler 1991). In VSVs of the Rhabdoviridae family, the *P* gene also encodes multiple gene products from a single genomic template (Herman 1986; Hudson, Condra, and Lazzarini 1986). The primary product of the *P* gene is the polymerase-associated P protein. Indeed, P proteins are encoded from this locus in all Paramyxoviridae and Rhabdoviridae taxa. Within the Paramyxoviridae family, in both the Paramyxovirus and Morbillivirus genera, viral taxa typically encode three proteins from the *P* gene sequence (fig. 1). In addition to the P protein, a smaller C protein is encoded by the *P* gene. Translation of C is initiated from a downstream AUG and proceeds in the +1 frame relative to the P protein (Bellini et al. 1985; Galinski et al. 1986; Galinski and Wechsler 1991; Yamanaka et al. 1992). Another protein, V, is also encoded by the *P* gene sequence in these viruses. The V protein is translated from an edited mRNA template (Cattaneo et al. 1989; Galinski and Wechsler 1991; Curran and Rima 1992; Yamanaka et al. 1992). Translation of V begins at the same
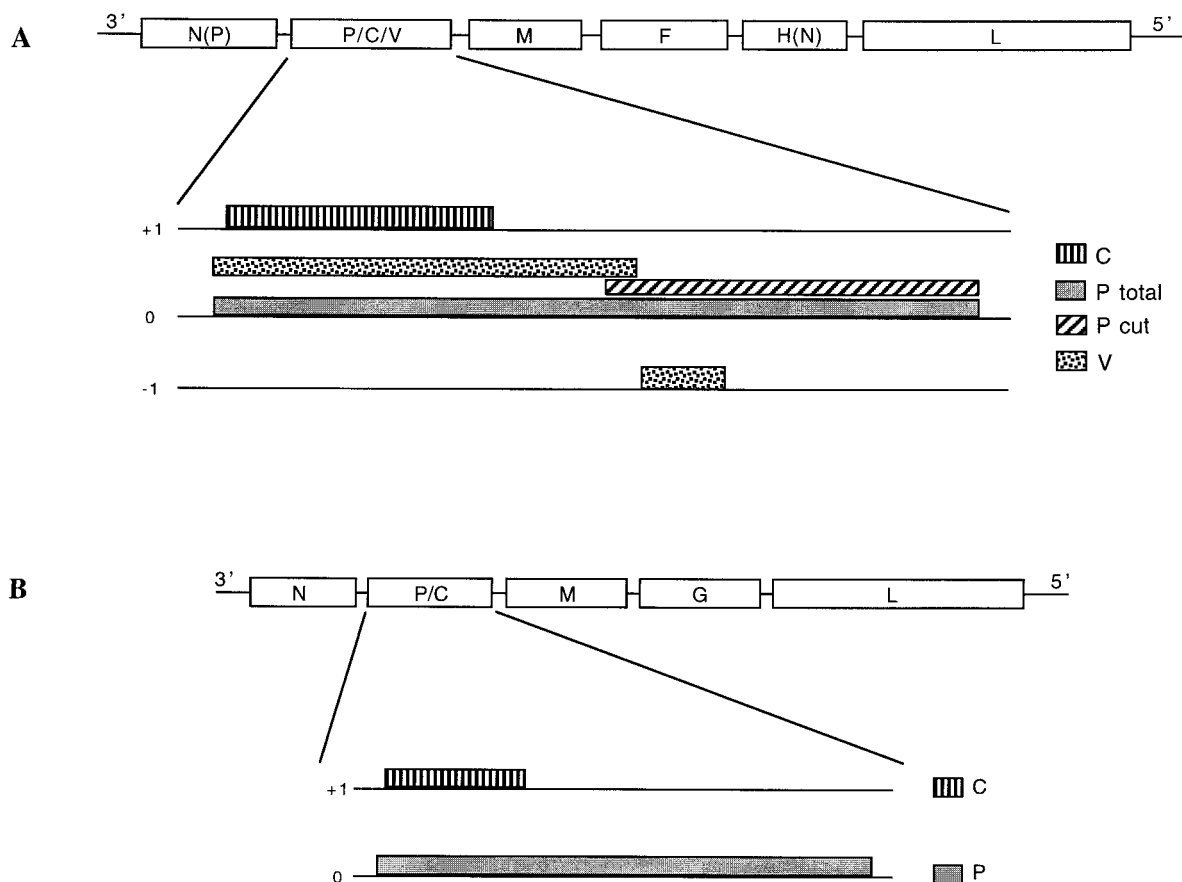
FIG. 1.—Negative-strand RNA genomes of the Paramyxoviridae and the Rhabdoviridae. *A,* The *P* gene is a multicoding sequence that encodes three proteins (P, C, and V) from overlapping reading frames in Paramyxovirus and Morbillivirus taxa of the Paramyxoviridae family. The C protein is encoded in the +1 frame relative to the P protein. The amino portion of the V protein is encoded in the same frame as the P protein, while the carboxyl cysteine-rich region of the V protein is encoded in the −1 frame relative to P. P-cut and P-total are defined in the text. *B,* In VSV taxa of the Rhabdoviridae family, the *P* gene encodes the C protein in the +1 frame relative to the P protein.

AUG site as P, and the amino portions of the two proteins are identical, but the carboxyl portion of V is translated in the −1 frame relative to P due to the insertion of nontemplated guanine residues into the mRNA. In the Rhabdoviridae family, the VSV *P* gene possesses an overlapping reading frame near the 5′ end of the gene. This reading frame encodes two polypeptides initiated from alternate AUGs, another so-called C protein and the shorter C′, both in the +1 frame relative to the P protein (Herman 1986; Spiropoulou and Nichol 1993; Kretzschmar et al. 1996). The Rhabdoviridae C protein bears no discernible sequence similarity to the Paramyxoviridae C protein.

The complex nature of the *P* gene, coupled with its high rate of evolution, makes it an intriguing locus to study from an evolutionary perspective. In this report, we present an analysis of the molecular evolutionary dynamics of the *P* gene. To evaluate the steps involved in the evolution of *P* gene overlapping reading frames, the coding capacities were mapped most-parsimoniously onto the phylogeny of the Paramyxoviridae taxa. In addition, relative levels of amino acid sequence identity for polypeptide sequences encoded by the overlapping reading frames were compared among all P-, C-, and V-containing Paramyxoviridae taxa, as well as with Rhab-

doviridae VSV taxa. The nucleotide changes that underlie this amino acid variation were also examined in an effort to evaluate the nature of selective forces that have acted on *P* gene sequences. The results of these analyses are discussed with respect to the impact of RNA virus quasispecies population dynamics on the evolution of viral genomic complexity.

Materials and Methods
Sequence Retrieval

Sequences for this study were obtained using the National Center for Biotechnology Information's (NCBI) ungapped advanced BLAST search engine on the NCBI's web site (http://www.ncbi.nlm.nih.gov/BLAST). The BLAST search was run using the tblastn algorithm and the PAM250 distance matrix. A total of 76 Paramyxoviridae (44) and Rhabdoviridae (32) sequences were transferred from GenBank and used for subsequent analyses. Paramyxoviridae *P* sequence accession numbers are as follows: Sendai virus—M30203, M30204, X17008, X0087; human parainfluenza virus-1—M74082; human parainfluenza virus-3—D00047, M14932, M14890, X04721, D10029, U51116, D84095; measles virus—S58435, M10456, K01711, Z66517, D63925, M88920,

X16569, X16566, X16567, D10635, X16568, L36044; rinderpest virus—S44819, X68311; dolphin morbilli virus—Z47758; phocine distemper virus—X75960, X65512, D10371; canine distemper—X51869; Newcastle disease virus—M20302; human parainfluenza virus-2—X57559; human parainfluenza virus-4—M55975; mumps virus—D86175; simian parainfluenza virus-4—X64275; simian parainfluenza virus-5—J03142; avian pneumovirus—U22110; mouse pneumovirus—U09649; bovine respiratory syncytial virus—M93127; ovine respiratory syncytial virus—U07232; human respiratory syncytial virus—AF013254. Rhabdoviridae *P* sequence accession numbers are as follows: vesicular stomatitis virus (NJ)—M31862–M31880, X04718, X04063, K03387; vesicular stomatitis virus (IN)—J02428, X04453, U12967, M15121, U13898, X04196; piry virus—D26175; chandipura virus—M16608; rabies virus—X55728; hemorrhagic septicemia virus—X73873.

Sequence Alignment

Amino acid sequences were aligned using the CLUSTAL W program (Thompson, Higgins, and Gibson 1994). CLUSTAL W was run with default gap penalty options and the PAM distance series. Amino acid sequence alignments were refined when necessary by visual inspection. Gaps were altered to minimize the number of mutational events in the alignment. Independent amino acid alignments were performed for the Paramyxoviridae and Rhabdoviridae families, as well as for each of the six intraclade analyses described in tables 2 and 3. Independent nucleotide sequence alignments were also performed for the intraclade comparisons. Nucleotide sequences were aligned manually to correspond with the intraclade amino acid sequence alignments.

Phylogenetic Analyses

Phylogenetic analyses were performed using the PAUP* 4.0b1 test version (Swofford 1998) and the PHYLIP package (Felsenstein 1991). Paramyxoviridae and Rhabdoviridae N and P amino acid sequence alignments were used to generate distance matrices with the Dayhoff PAM matrix implemented in the PROTDIST program of PHYLIP. The tree topologies reported here were then generated by the neighbor-joining algorithm (Saitou and Nei 1987) implemented in PAUP* 4.0b1. One thousand bootstrap replicates were each generated with the SEQBOOT program in PHYLIP and subsequently analyzed with the PROTDIST and CONSENSE programs.

Sequence Variation

Amino acid sequence identities were determined using PAUP* 4.0b1. Pairwise mean character differences were averaged and converted to percentages of identity. Nucleotide diversity was determined using the method of Lynch and Crease (1990) with the Jukes-Cantor correction (Jukes and Cantor 1969), implemented in the DnaSP program (Rozas and Rozas 1997). Nucleotide diversity was subsequently converted to percentage of identity. Levels of synonymous (d_s) and nonsynonymous (d_n) nucleotide diversity were calculated with DnaSP using the method of Nei and Gojobori (1986).

Hardware

PAUP* 4.0b1 was run on a Macintosh Quadra 950. DnaSP was run on a Gateway 2000 with the Windows 95 operating system. All other analyses were performed on a Sun SPARCstation 5 running SunOS Release 5.6.

Results

Phylogenetic Reconstruction

The Paramyxoviridae P amino acid sequence alignment was used to reconstruct the phylogeny of the Paramyxoviridae family (fig. 2). The topology of the resulting phylogeny is consistent with the existing taxonomic classification of the family (Murphy et al. 1995). This phylogeny is also consistent with an independent phylogenetic reconstruction performed using more conserved N amino acid sequences (not shown).

The distribution of multiple coding capacity of the *P* gene for Paramyxoviridae taxa is shown to the right of the tree (fig. 2). Taxa that represent similar states of *P* multiple coding capacity tend to group together in the tree. For example, all Pneumovirinae taxa encode a single P protein. Since most of the closely related Rhabdoviridae family viruses also lack multiple coding capacity in the *P* gene sequence, single-P-protein-coding capacity is assumed to be the ancestral state for the *P* gene.

The *P* gene of the Rubulavirus genus encodes both a P and a V protein. The V protein (V′ in fig. 2) is translated from an unedited mRNA template. In contrast, the P protein of these taxa is translated from an edited mRNA template (Thomas, Lamb, and Paterson 1988; Kondo et al. 1990; Southern, Precious, and Randall 1990; Galinski and Wechsler 1991). Newcastle disease virus (NDV) is the only exception to this pattern in the Rubulavirus genus. NDV encodes the P protein from the genomic mRNA and the V protein from an edited mRNA template (McGinnes, McQuain, and Morrison 1988; Galinski and Wechsler 1991) (fig. 2). Morbillivirus and Paramyxovirus taxa also encode the P protein from an unedited mRNA and the V protein from an edited mRNA (Thomas, Lamb, and Paterson 1988; Kondo et al. 1990; Southern, Precious, and Randall 1990; Galinski and Wechsler 1991; Galinski, Troy, and Banerjee 1992).

All taxa within the Morbillivirus and Paramyxovirus genera, with the exception of human parainfluenza-1, contain P, C, and V open reading frames (ORFs). The Paramyxovirus genus appears to be uniquely labile in terms of the ability of the viral taxa to gain and lose coding capacity within the *P* gene. Both Sendai and human parainfluenza-3 viruses not only contain P, C, and V ORFs, but also have gained coding capacity for a number of other small peptides (''+'' in fig. 2) from the C reading frame (Giorgi, Blumberg, and Kolakofsky 1983; Curran, Richardson, and Kolakofsky 1986; Galinski et al. 1986; Curran and Kolakofsky 1987; Galinski and Wechsler 1991). Human parainfluenza-1 viruses, on the other hand, do not possess an intact V ORF (Matsuoka et al. 1991; Power, Ryan, and Portner 1992). Rem-
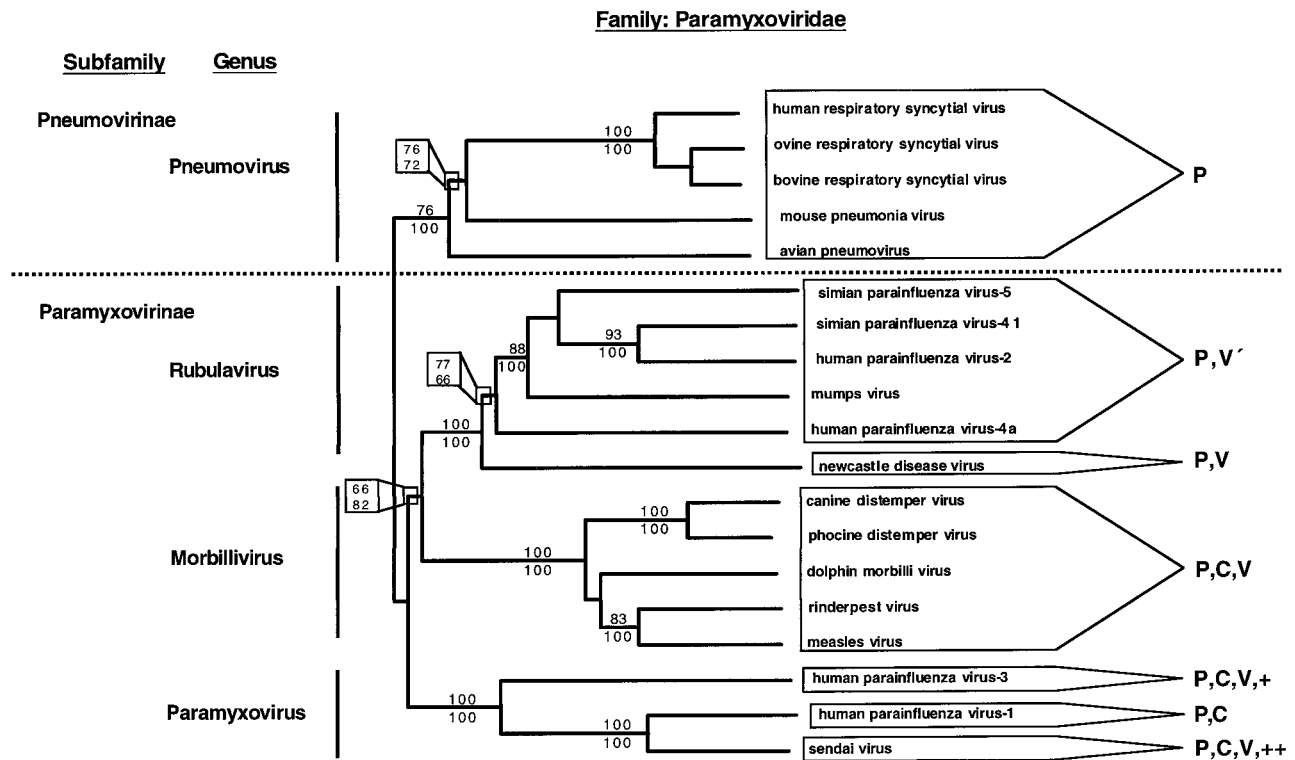
FIG. 2.—Phylogeny of the Paramyxoviridae based on amino acid sequences of the P protein. Phylogenetic reconstruction was performed as described in *Materials and Methods*. The tree shown here is the neighbor-joining tree found by PAUP* 4.0b1 and is rooted by midpoint. Parsimony and neighbor-joining bootstrap values are shown above and below the branches, respectively. Subfamily and genus designations are shown to the left of the tree. Viral-taxon *P* gene multiple coding capacity (P, C, and V) is indicated to the right of the tree. V′ indicates taxa that encode V from an unedited mRNA template. + and ++ indicate the presence of additional proteins encoded from an overlapping reading frame.

nants of an ancestral V ORF in these viruses suggest an independent loss of V-coding capacity within this genus.

Rhabdoviridae N and P amino acid sequence alignments were used to reconstruct the phylogeny of the Rhabdoviridae family. The neighbor-joining tree derived from the N proteins is shown in figure 3. The phylogeny derived from the more distantly related P proteins gave the same topology, with the exception of one weakly supported internal node (not shown). Vesiculovirus and Lyssavirus taxa group together in well-supported genus-specific clades consistent with their taxonomic statuses.

Paramyxoviridae Interclade Amino Acid Sequence Variation

Levels of amino acid variation were determined for the P, C, and V proteins within and between the Paramyxovirus and Morbillivirus genera (table 1). P proteins in Paramyxoviridae are larger (506–596 residues) than those in the Rhabdoviridae family (222–327 residues). The carboxyl region of the Paramyxoviridae P protein that is thought to be functionally analogous to the Rhabdoviridae P protein is referred to here as P-cut. The entire P amino acid sequence of the Paramyxoviridae is hereinafter referred to as P-total (fig. 1). The P-cut regions of the Paramyxoviridae P proteins are more conserved than the P-total sequences (table 1).

Published studies (Galinski and Wechsler 1991; Morgan 1991), as well as preliminary analyses in our lab, have indicated that different *P* gene products encoded by the same nucleotides in overlapping reading frames have different levels of amino acid sequence variation. Here, we examined this phenomenon in detail to infer the molecular evolutionary dynamics that gave rise to these patterns. To directly compare the amino acid sequence variation between different reading frames encoded by the same nucleotide sequences, we identified those amino acid sequences of the P proteins that are encoded by the same nucleotides that encode the C and V proteins in different frames (fig. 1). The C protein is encoded in a +1 frame relative to the P protein. The P-C region is defined as that region of the P protein that is encoded by the same nucleotides that encode the C protein. The V protein is encoded in the same frame as P starting from the amino terminus, but switches to the −1 frame, due to RNA editing, near the carboxyl end of the protein. The P-V region is defined here as the region of the P protein encoded by the same nucleotides that encode the region of V (after RNA edit) that is encoded in the −1 frame. The C and V (after RNA edit) protein sequences show similar patterns of amino acid identity relative to the P-C and P-V regions within and between the Paramyxovirus and Morbillivirus genera. In all cases, the P region is less conserved (i.e., shows less sequence identity) than the C and V protein sequences encoded by the same nucleotides in alternate reading frames (table 1).
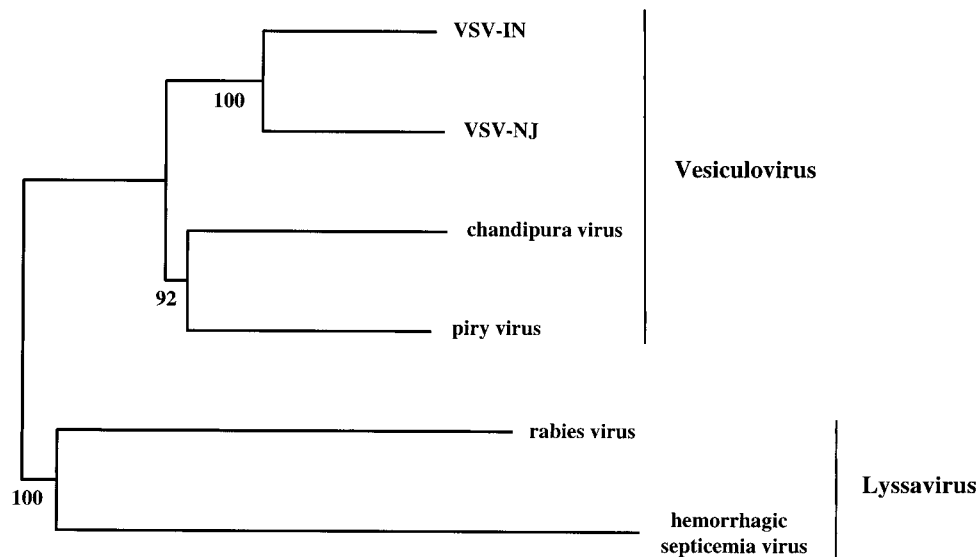
FIG. 3.—Neighbor-joining tree of the Rhabdoviridae based on amino acid sequences of the N protein. Phylogenetic reconstruction was performed as described in *Materials and Methods*. Numbers indicate neighbor-joining bootstrap values. Genus designations are shown to the right of the tree.

Paramyxoviridae Intraclade Amino Acid and Nucleotide Sequence Variation

To better understand the broad patterns of comparative amino acid identity for the six regions of the *P*-gene-encoded polypeptides described above, we analyzed the nucleotide changes that underlie this variation. Phylogenetic analysis of P proteins from the full complement of P-, C-, and V-containing Paramyxoviridae taxa (not shown) allowed us to identify well-supported clades of closely related P sequences. It was necessary to identify closely related sequences to avoid the problem of saturation when examining nucleotide differences. However, it was also necessary to choose clades with enough taxa and sequence diversity to obtain meaningful results. There are four clades that meet both of these criteria (74.2%–99.9% nucleotide identity): the Sendai clade, the human parainfluenza-3 clade, the measles clade, and the phocine/canine distemper clade.

For each of these clades, the same six measures of amino acid identity analyzed for the broader between-clades comparisons (table 1) were calculated. The nucleotide diversity for each of these regions was also calculated. To evaluate the effects of selection on these regions, the ratio of nonsynonymous (d_n) to synonymous (d_s) nucleotide diversity was calculated. A ratio below 1 (a relative excess of synonymous substitutions) is indicative of negative selection, presumably due to functional constraints on the protein sequence. The action of selection was also assessed by examining the distribution of variable nucleotide sites across the three codon positions for overlapping reading frames.

There are several patterns indicated by the intraclade analyses (table 2) that are the same in all four clades and are consistent with the broader patterns of amino acid identity shown in table 1. For example, the d_n/d_s ratios for the P-cut regions of all four of these clades suggest that a more stringent degree of negative selection is acting on P-cut than P-total. This is consistent with the higher levels of P-cut amino acid sequence identity in table 1.

For all four intraclade comparisons, amino acid identity is higher for the C protein than for the P-C region encoded by the same nucleotides. These data are also consistent with the patterns of interclade variation shown in table 1. The C protein is encoded in a +1 frame relative to the P protein. Thus, nucleotide changes that occur in the third positions of codons in the C reading frame change the first positions of codons in the P reading frame. Approximately 72% of third-codon-position nucleotide changes and 5% of first-position

Table 1
Average Percentages of Amino Acid Sequence Identity for *P*-Gene-Encoded Products from Paramyxoviridae Viruses that Encode the P, C, and V Proteins

	All P, C, V Taxa[a]	Paramyxo Clade[b]	Morbilli Clade[c]
P-total[d]	31.71	43.86	55.56
P-cut[e]	36.43*	51.06*	62.69*
P-C region	25.93	39.49	44.35
C.	31.35*	52.38*	52.37*
P-V region	33.51	48.86	53.21
V.	47.45*	53.95	67.08*

NOTE.—Average amino acid identity percentages calculated as described in *Materials and Methods*.

[a] Average amino acid identity percentages among all Paramyxoviridae taxa analyzed here that encode the P, C, and V proteins.

[b] Average amino acid identity percentages among proteins for all taxa of the Paramyxovirus genus analyzed here that encode the P, C, and V proteins.

[c] Average amino acid identity percentages among proteins for all taxa of the Morbillivirus genus analyzed here that encode the P, C, and V proteins.

[d] P-total refers to the entire Paramyxoviridae P protein sequence.

[e] P-cut refers to the carboxyl region of the Paramyxoviridae P protein sequence thought to be functionally analogous to the Rhabdoviridae P protein.

* Amino acid identity percentages differ at the *P* < 0.01 level by *t*-tests.

Table 2
Intraclade Amino Acid and Nucleotide Variation for Each of the Four Viral Taxa Analyzed in Detail Here that Encode the P, C, and V Proteins

CLADE	OPEN READING FRAME	AA % ID	NT % ID	d_n/d_s	CODON DIST[a]		
					1	2	3
Sendai	P-total	99.01	99.22	0.23			
	P-cut	98.99	99.03	0.16			
	P-C region	98.82	99.35	0.47	6	1	3
	C	99.02	99.35	0.29	1	3	6
	P-V region	99.41	99.51	0.22	1	0	1
	V	98.52	99.51	0.00	1	1	0
Human para3	P-total	88.28	89.87	0.30			
	P-cut	92.06**	90.47	0.11			
	P-C region	86.38	92.28	1.06	75	28	28
	C	92.13**	92.28	0.15	28	28	75
	P-V region	97.15	94.00	0.09	10	2	27
	V	87.22**	94.00	11.11	27	10	2
Measles	P-total	95.42	97.02	0.37			
	P-cut	95.98**	96.91	0.24			
	P-C region	94.17	97.09	0.79	29	14	17
	C	95.84**	97.09	0.30	14	17	29
	P-V region	94.76	97.41	0.99	1	15	7
	V	97.92**	97.41	0.16	7	1	15
Phocine/ canine . . .	P-total	87.51	87.59	0.18			
	P-cut	88.93	86.91	0.11			
	P-C region	84.94	89.52	0.48	41	18	37
	C	86.11	89.52	0.33	18	37	41
	P-V region	79.71	88.69	1.37	6	26	9
	V	89.13*	88.69	0.10	9	6	26

NOTE.—Amino acid identity percentages (AA % id), nucleotide identity percentages (Nt % id), and d_n/d_s ratios calculated as described in *Materials and Methods*.

[a] Distribution of variable nucleotide sites across the three codon positions.

* Amino acid identity percentages differ at the $P < 0.05$ level by *t*-tests.

** Amino acid identity percentages differ at the $P < 0.01$ level by *t*-tests.

changes are synonymous (Nei and Gojobori 1986). Therefore, synonymous third-position substitutions in the C reading frame that are hidden from the effects of selection will likely change the amino acid coded for in the P reading frame. There is a nonrandom distribution of intraclade nucleotide variable sites across the three codon positions of the C and P-C region overlapping reading frames (table 2). The excess of third-position variable sites in the C reading frame indicates that the evolution of this reading frame is being constrained by negative selection. The corresponding excess of first-position variable sites in the P frame indicates that the P-C region is less constrained by negative selection and is evolving rapidly relative to the C protein encoded by the same nucleotides. Consistent with the pattern of nonrandom distribution of variable sites across the codon positions, comparison of d_n/d_s ratios for these two reading frames indicates a higher degree of selective constraint acting on the C reading frame.

The V protein sequence encoded in the −1 frame relative to P is the most conserved of all the *P*-gene-encoded polypeptide sequences (table 1). However, comparative patterns of intraclade P-V region versus V sequence variation are not entirely consistent with the overall sequence conservation of the V protein. For ex-

ample, the Sendai and human parainfluenza-3 clades within the Paramyxovirus genus show the opposite pattern of relative variation between the P-V region and the V protein sequences. In both cases, the P-V region is more conserved than is the V protein sequence encoded by the same nucleotides (table 2). This difference is small in the Sendai clade due to the fact that this clade consists of the most closely related group of sequences in our analyses and there are only two nucleotide differences in the P-V/V coding regions. For the human parainfluenza-3 clade, however, there is a significant excess (χ^2, $P = 3.59 \times 10^{-6}$) of third-codon-position variable sites—indicative of negative selection—in the P reading frame (table 2). These same third-position nucleotide variations in the P frame map to the first codon positions of the V frame and are thus likely to change the V amino acid sequence. The d_n/d_s ratio also indicates the action of negative selection exclusively on the P-V reading frame (table 2). Therefore, V protein sequences are not highly constrained by selection in the human parainfluenza-3 clade and evolve more rapidly than the corresponding P-V region sequences, despite the fact that the broad overall patterns of variation within and between all Paramyxovirus and Morbillivirus genera indicate that the nucleotides that encode the V protein (after RNA edit) comprise the most selectively constrained reading frame of the *P* gene (table 1).

The measles and phocine/canine distemper clades of the Morbillivirus genus both encode V protein sequences that are more conserved than the P-V region sequences encoded by the same nucleotides. Examination of the distribution of variable nucleotide sites across the codon positions in this region of the *P* gene reveals that there is an excess of third-position variation in the V reading frame (table 2). These same variable sites map to the second codon positions of the P reading frame. All second-codon-position nucleotide changes are nonsynonymous. Therefore, the third-position variations that are mostly synonymous in the V frame are all nonsynonymous in the P frame. The d_n/d_s ratio also indicates a much higher degree of selective constraint acting on the V proteins than on the overlapping P-V region sequences within these clades. These data are consistent with the interclade patterns of amino acid variation (table 1), which indicate that the V protein is the most conserved reading frame of the *P* gene.

Rhabdoviridae Amino Acid and Nucleotide Sequence Variation

The *P* gene of VSV is known to encode multiple proteins from overlapping reading frames (Herman 1986; Hudson, Condra, and Lazzarini 1986; Spiropoulou and Nichol 1993; Kretzschmar et al. 1996). An ORF at the 5′ end of the *P* gene mRNA encodes a so-called C protein and a slightly shorter product C′ from a downstream AUG. Alignment of the Paramyxoviridae and Rhabdoviridae (VSV) C amino acid sequences reveals no homology between C proteins of the two families (not shown). The name C for both the Paramyxoviridae and the Rhabdoviridae proteins is due to a historical artifact and belies the fact that these two proteins have

Table 3
VSV *P* Gene Sequence Variation

Taxa		AA % ID	Nt % ID	d_n/d_s	Codon Dist[a]		
					1	2	3
VSV all....	P-total	59.64					
	P-C region	61.61*					
	C	52.45**					
VSV-IN....	P-total	96.73	98.46	0.22			
	P-C region	97.01	98.56	0.14	0	3	5
	C	94.76	98.56	4.17	3	5	0
VSV-NJ....	P-total	88.58	87.64	0.18			
	P-C region	96.30**	95.42	0.05	6	1	20
	C	86.85**	95.42	1.04	1	20	6

Note.—Amino acid identity percentages (AA % id), nucleotide identity percentages (Nt % id), and d_n/d_s ratios calculated as described in *Materials and Methods*.

[a] Distribution of variable nucleotide sites across the three codon positions.

* Amino acid identity percentages differ at the $P < 0.05$ level by *t*-tests.

** Amino acid identity percentages differ at the $P < 0.01$ level by *t*-tests.

likely evolved independently. VSV P protein amino acids encoded by the same nucleotides that encode the full-length C protein are defined here as the P-C region. Levels of amino acid identity for P-total, the P-C region, and C were determined for VSVs (table 3) in the same way as for the Paramyxoviridae overlapping *P* gene products. Amino acid identity in the P-C region is higher than that for P-total; however, C amino acid identity is lower than both the P-C region and the P-total identities. Analysis of the nucleotide sequences that underlie this variation indicates that negative selection is acting to conserve the P-C region and not the C amino acid sequences. Both distribution of variable nucleotide sites across the three codon positions and d_n/d_s ratios indicate that there is an excess of synonymous variation in the P reading frame that corresponds to nonsynonymous variation in the overlapping C frame (table 3).

Discussion

Multiple coding capacity, such as that exemplified by the *P* gene, is not entirely uncommon in viral genomes (Miyata and Yasunaga 1978; Keese and Gibbs 1992; Lartey, Voss, and Melcher 1996; Mizokami et al. 1997; Pavesi et al. 1997). The evolution of overlapping reading frames is one of several strategies by which viral genomes can increase information content while maintaining a small genome (Lamb, Mahy, and Choppin 1976; Clare and Farabaugh 1985; Yoshinaka et al. 1985; Cattaneo 1989; Brierley and Flavell 1990; Lamb and Horvath 1991; Shimotohno et al. 1995). While both DNA and RNA viruses have employed this evolutionary strategy, the ability of RNA viruses to evolve multicoding sequences is likely enhanced by their quasispecies population dynamics.

RNA virus quasispecies dynamics are characterized by accelerated rates of evolution due to high population numbers, rapid replication of RNA viruses, and the error-prone nature of RNA genome replication (Domingo and Holland 1997). The error-prone nature of RNA ge-

nome replication, in addition to the relative instability of the RNA molecule, puts a limit on the possible size of RNA genomes. Misincorporation rates for RNA polymerase are typically about 1 million times as high as those of DNA polymerase (Drake 1991, 1993; Domingo et al. 1996). This is mainly due to the fact that RNA polymerases lack the 3′→5′ exonuclease activity of DNA polymerases (Friedberg, Walker, and Siede 1995). The high mutation rates of RNA viruses put them in danger of error catastrophe, which occurs when the mutation rate is so high that genomic information can no longer be faithfully transmitted from one generation to the next. Small genomes allow RNA viruses to avoid error catastrophe in the face of high mutation rates. A small genome size lowers the possibility of incurring multiple deleterious mutations in a single genome during one round of replication. Because of their small genomes, when RNA viruses increase genomic content, they do so by maximizing the density of their encoded genetic information. The apparently independent evolution of overlapping reading frames in the Paramyxoviridae and Rhabdoviridae *P* genes, as well as in other RNA viruses, is likely a response to selective pressure to maximize genomic information content while maintaining a small genome.

The ability to adopt such a complex genomic strategy is also intimately related to quasispecies dynamics. RNA viruses in a quasispecies tend to replicate with a mutation rate that is just below the error catastrophe threshold. Such a mutation rate ensures faithful transmission of information while maximizing adaptability. The distribution of variants in the quasispecies that results from the high mutation rate includes a large reservoir of variants with novel and potentially successful phenotypes. The presence of these variants facilitates a succession of rapid evolutionary responses to various selective pressures. RNA viruses therefore may have an enhanced ability to explore adaptive landscapes. This enhanced adaptability can, in turn, lead to the evolution of novel and complex genomic strategies. The evolution of overlapping coding sequences is an example of such a complex genomic strategy. The evolution of viable overlapping coding sequences requires an extremely fine level of selective tuning, since changes in a particular codon position of one frame change a different codon position in the overlapping frame. Optimal exploration of sequence space that exists just below the error threshold makes such complex molecular sequence evolution possible in RNA virus genomes.

Evolutionary Path of Paramyxoviridae Overlapping Reading Frames

Examination of the phylogenetic distribution of *P* multiple coding capacity gives an idea of how overlapping reading frames might have evolved in the Paramyxoviridae family (fig. 4). The two most parsimonious paths of P-, C-, and V-coding capacity evolution along the Paramyxoviridae tree each involve seven steps (fig. 4). Each step corresponds to either a gain or a loss of an ORF in the *P* gene. A unique situation exists in the human parainfluenza-3 group (P, C, V*, + in fig. 4),
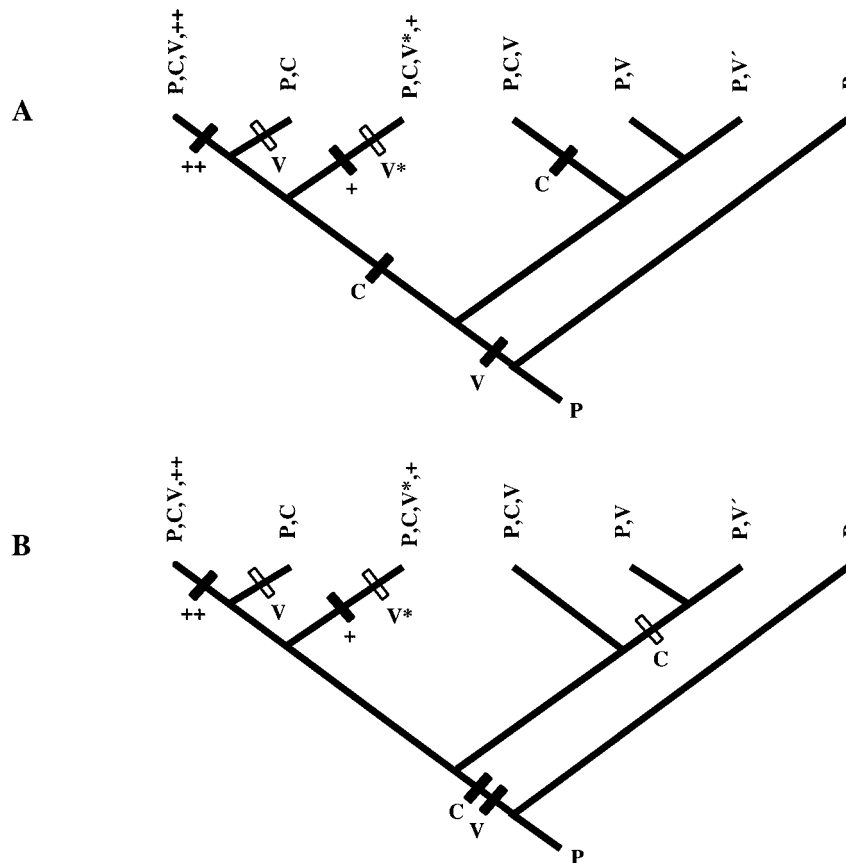
Fig. 4.—Reconstruction of the evolution of multiple coding capacity in the *P* gene of the P, C, and V proteins. Taxa names represent monophyletic groups of viruses with similar coding capacities. The tree topology is identical to that in figure 2. *P* gene multiple coding capacity evolves via the gain and loss of coding capacity for P, C, V, + and ++ (abbreviations defined in fig. 2). Black bars represent gains of coding capacity, and white bars represent losses. V* represents the recent loss of V-coding capacity as indicated by the presence of an intact but apparently nonfunctional (unexpressed) V open reading frame. Two equally parsimonious paths are shown, assuming that single P-protein-coding capacity is ancestral (see text) and that losses and gains are equally likely.

which encodes an intact but apparently nonfunctional (unexpressed) V ORF. This likely represents an extremely recent evolutionary loss of V-coding capacity. A change in V-coding capacity also occurred along the lineage from Newcastle disease virus (P, V) to the rest of the Rubulavirus taxa (P, V′). P, V′ taxa encode V from an unedited mRNA template and P from an edited template. This pattern is the opposite of what is seen in other P- and V-encoding taxa and could have evolved due to the gain or loss of only one or two nucleotides.

One most-parsimonious seven-step path (fig. 4*A*) suggests a single gain of V along the lineage leading to Paramyxovirinae subfamily, followed by independent gains of C in the Morbillivirus and Paramyxovirus genera. The alternate seven-step path (fig. 4*B*) requires single gains of both C and V coding capacity along the lineage leading to the Paramyxovirinae subfamily, followed by a loss of C in the Rubulavirus genus. The numbers of gains versus losses that constitute the seven steps distinguish these two paths. The path that involves a gain of both C- and V-coding capacity along the same lineage requires four gains and three losses of *P*-coding capacity (fig. 4*B*). The alternate path requires two independent gains of C-coding capacity and consists of five gains versus two losses (fig. 4*A*). If we hypothesize

that gains are less likely than losses, then the path which involves a gain of both C- and V-coding capacities along the lineage leading to the Paramyxovirinae lineage (four gains and three losses) is more parsimonious (i.e., fig. 4*B* is favored). The alternative path (fig. 4*A*) can also be considered less likely, because it involves two independent gains of C ORFs that would have created seemingly orthologous C proteins.

Paramyxoviridae P Protein Evolution

The P protein is the least conserved of all the Paramyxoviridae proteins studied thus far (Morgan 1991). This is unexpected when the presence of overlapping reading frames in the *P* gene is considered. One might expect that the presence of such overlapping reading frames would constrain the evolution of P amino acid sequences encoded by the same nucleotides (Miyata and Yasunaga 1978). There is evidence of such conservation of P due to the presence of an overlapping reading frame in the VSV *P* gene (Bilsel et al. 1990), and our results reveal that nucleotide variation in *P* gene overlap regions does tend to be slightly reduced. However, the high variability of the P protein is maintained despite the presence of overlapping reading frames, because negative selection acts predominantly to constrain the

evolution of one frame or the other in these regions (table 2). Overlapping reading frames may therefore evolve in the region of the *P* gene that is under the least functional constraint. This is suggested by the fact that P-C regions show the highest levels of P amino acid variation (tables 1 and 2). Thus, the evolution of *P* gene complexity appears to be an example of the exquisite adaptability allowed by the optimal exploration of sequence space of RNA virus quasispecies.

Considered in another light, the high variability of the P protein despite the complexity of the *P* gene locus may not be as surprising. In fact, high variability of the *P* gene may have been a prerequisite for the subsequent evolution of overlapping reading frames. The rapid evolution of the *P* gene may be due to a lack of functional constraints on its encoded products or a response to adaptive (positive) selection. The data presented here indicate that high variability of the *P* gene is due to a relative lack of functional constraints, as d_n/d_s ratios for both P-total and P-cut show no evidence of adaptive selection (table 2). This lack of constraint may lead to a unique *P* gene sequence distribution in the quasispecies. Despite the generally heterogeneous nature of RNA virus quasispecies, negative selection due to functional constraints is known to limit the movement of many other RNA viruses through narrow regions of available sequence space (Domingo et al. 1996). In the case of the *P* gene, a relative paucity of preexisting functional constraints on it may have allowed the exploration of a much broader sequence space and the subsequent evolution of overlapping reading frames.

Paramyxoviridae C Protein Evolution

The Paramyxoviridae C protein sequence is in all cases more conserved than the P protein sequence encoded by the same nucleotides (tables 1 and 2). Evidence from the d_n/d_s ratios and the distribution of nucleotide variation across the codons of these two overlapping reading frames invariably reveals that negative selection acts to constrain the evolution of the C reading frame more than that of the P frame (table 2). These data suggest that the C protein likely plays a critical role in the life cycle of the Paramyxoviridae taxa that encode it. Recent reports have suggested a role for the C protein in repression of both viral transcription and replication (Curran, Marq, and Kolakofsky 1992; Cadd et al. 1996; Horikami et al. 1997; Tapparel et al. 1997). The C reading frame is located at the 5′ end of the *P* gene that encodes the amino portion of the P protein. Evidence presented here suggests that the carboxyl portion of P contains the most important functional region of the protein (table 2). As alluded to above, the location of the Paramyxoviridae C reading frame in the relatively nonessential amino region of the *P* gene represents a commensalistic type of evolutionary strategy whereby the C ORF enhances its fitness by minimizing its potential harmful effects on the P ORF. The ability to maximally explore sequence space likely allows numerous attempts to evolve overlapping reading frames until a suitable location for one is found.

Paramyxoviridae V Protein Evolution

The post-RNA-edit region of the V protein (that is, the −1 frame in fig. 1) is the most conserved of the *P*-gene-encoded polypeptides (table 1). The high level of amino acid sequence conservation suggests that this cysteine-rich region of the V protein plays a critical role in Paramyxoviridae virus life cycles. A similar abundance of cysteine residues is found in zinc finger motifs. Taken together with the role of P in transcription, this suggests a possible role for the V protein in transcriptional regulation (Galinski and Wechsler 1991). While functional studies have shown the V protein to be nonessential in vitro (Schneider, Kaelin, and Billeter 1997), other data suggest that it plays an important role in pathogenesis in vivo (Kato et al. 1997; Tober et al. 1998).

Despite the high overall level of Paramyxoviridae V protein conservation (table 1), analysis of within-clade nucleotide variation revealed a lack of selection on the V region of the Paramyxovirus. Evidence for relaxed selective constraints on the V protein was particularly compelling for human parainfluenza-3 taxa. In this clade, unlike the Morbillivirus clades, selection acted to constrain the evolution of the P-V reading frame and not the V reading frame (table 2). This result is consistent with the finding that in human parainfluenza-3 the V reading frame is not accessed by RNA editing (Galinski, Troy, and Banerjee 1992). Therefore, the V reading frame, while intact in these taxa, likely represents a nonfunctional sequence artifact. The data for the Sendai virus also suggest relaxed selective constraints on V, although the lower levels of variation in this clade render this conclusion less certain. Furthermore, closely related human parainfluenza-1 taxa do not encode an intact V reading frame (Matsuoka et al. 1991; Power, Ryan, and Portner 1992). The Paramyxovirus genus therefore offers a glimpse of the evolutionary moment of loss of part of the multiple coding capacity of the *P* gene. The Paramyxovirus clade is also notable for the taxa that encode a number of novel peptides from the *P* gene (Giorgi, Blumberg, and Kolakofsky 1983; Curran, Richardson, and Kolakofsky 1986; Galinski et al. 1986; Curran and Kolakofsky 1987; Galinski and Wechsler 1991). It may be the case that in this genus these proteins are in the process of usurping the regulatory function of the V protein.

Rhabdoviridae C Protein Evolution

The VSV P-C region protein sequences are more conserved than the overall P-total amino acid sequences (table 3). A similar observation led to speculation that the presence of the overlapping C reading frame limited the nucleotide substitution rate of the *P* gene in the corresponding region (Bilsel et al. 1990). Implicit in this suggestion is the action of negative selection on the C reading frame. However, our data reveal that C amino acid sequences are less conserved than the P-C region sequences encoded by the same nucleotides within and among both VSV strains (table 3). Furthermore, the nucleotide data clearly indicate that negative selection is acting on the P-C region and not the C reading frame (table 3). It has been shown that C proteins are not need-

ed for VSV viability in tissue culture (Kretzschmar et al. 1996). Another study on C protein functions yielded conflicting results depending on the experimental system employed to assay C protein effects on transcription and replication (Peluso et al. 1996). Data reported here suggesting a lack of functional constraint on the C protein are consistent with the results from the former study, as well as with the results from the latter study, which found no effect of C proteins on VSV replication or transcription.

Conclusions

The large population size, rapid replication, and high mutation rate of the quasispecies allow RNA viruses to evolve effective and complex responses to a variety of selection pressures. This is exemplified by the ability of Paramyxoviridae and Rhabdoviridae taxa to evolve novel protein-coding sequences while maintaining small genome size. The evolution of overlapping reading frames in the least conserved locus of the Paramyxoviridae and Rhabdoviridae genomes is related to the broad distribution of *P* genes in sequence space. This unique *P* sequence distribution occupies a relatively large area of adaptive landscape and allows for selection to continuously evaluate many viable sequence combinations, including those with overlapping reading frames. The ability of the P protein to perform its original function despite the sequence changes wrought by emerging overlapping reading frames was likely necessary for the successful evolution of those novel reading frames. Elucidation of the molecular evolutionary dynamics of the Paramyxoviridae and Rhabdoviridae *P* gene multicoding sequences illustrates the profound effect of quasispecies population dynamics on RNA virus biology.

LITERATURE CITED

BARIK, S., and A. K. BANERJEE. 1992. Phosphorylation by cellular casein kinase II is essential for transcriptional activity of vesicular stomatitis virus phosphoprotein P. Proc. Natl. Acad. Sci. USA **89**:6570–6574.

BELLINI, W. J., G. ENGLUND, S. ROZENBLATT, H. ARNHEITER, and C. D. RICHARDSON. 1985. Measles virus P gene codes for two proteins. J. Virol. **53**:908–919.

BILSEL, P. A., J. E. ROWE, W. M. FITCH, and S. T. NICHOL. 1990. Phosphoprotein and nucleocapsid protein evolution of vesicular stomatitis virus New Jersey. J. Virol. **64**:2498–2504.

BRIERLEY, C., and A. J. FLAVELL. 1990. The retrotransposon copia controls the relative levels of its gene products posttranscriptionally by differential expression from its two major mRNAs. Nucleic Acids Res. **18**:2947–2951.

CADD, T., D. GARCIN, C. TAPPAREL, M. ITOH, M. HOMMA, L. ROUX, J. CURRAN, and D. KOLAKOFSKY. 1996. The Sendai paramyxovirus accessory C proteins inhibit viral genome amplification in a promoter-specific fashion. J. Virol. **70**:5067–5074.

CANTER, D. M., and J. PERRAULT. 1996. Stabilization of vesicular stomatitis virus L polymerase protein by P protein binding: a small deletion in the C-terminal domain of L abrogates binding. Virology **219**:376–386.

CATTANEO, R. 1989. How 'hidden' reading frames are expressed. Trends Biochem. Sci. **14**:165–167.

CATTANEO, R., K. KAELIN, K. BACZKO, and M. A. BILLETER. 1989. Measles virus editing provides an additional cysteine-rich protein. Cell **56**:759–764.

CLARE, J., and P. FARABAUGH. 1985. Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. Proc. Natl. Acad. Sci. USA **82**:2829–2833.

CURRAN, J. A., and D. KOLAKOFSKY. 1987. Identification of an additional Sendai virus non-structural protein encoded by the P/C mRNA. J. Gen. Virol. **68**:2515–2519.

CURRAN, J., J. B. MARQ, and D. KOLAKOFSKY. 1992. The Sendai virus nonstructural C proteins specifically inhibit viral mRNA synthesis. Virology **189**:647–656.

CURRAN, J. A., C. RICHARDSON, and D. KOLAKOFSKY. 1986. Ribosomal initiation at alternate AUGs on the Sendai virus P/C mRNA. J. Virol. **57**:684–687.

CURRAN, M. D., and B. K. RIMA. 1992. The genes encoding the phospho- and matrix proteins of phocine distemper virus. J. Gen. Virol. **73**:1587–1591.

DAS, T., A. K. PATTNAIK, A. M. TAKACS, T. LI, L. N. HWANG, and A. K. BANERJEE. 1997. Basic amino acid residues at the carboxy-terminal eleven amino acid region of the phosphoprotein (P) are required for transcription but not for replication of vesicular stomatitis virus genome RNA. Virology **238**:103–114.

DE, B. P., and A. K. BANERJEE. 1997. Role of host proteins in gene expression of nonsegmented negative strand RNA viruses. Adv. Virus Res. **48**:169–204.

DOMINGO, E., C. ESCARMIS, N. SEVILLA, A. MOYA, S. F. ELENA, J. QUER, I. S. NOVELLA, and J. J. HOLLAND. 1996. Basic concepts in RNA virus evolution. FASEB J. **10**:859–864.

DOMINGO, E., and J. J. HOLLAND. 1994. Mutation rates and rapid evolution of RNA viruses. Pp. 161–184 *in* S. S. MORSE, ed. The evolutionary biology of viruses. Raven Press, New York.

DOMINGO, E., J. HOLLAND, C. BIEBRICHER, and M. EIGEN. 1995. Quasispecies: the concept and the word. Pp. 181–191 *in* A. GIBBS, C. H. CALISHER, and F. GARCIA-ARENAL, eds. Molecular basis of virus evolution. Cambridge University Press, Cambridge, England.

———. 1997. RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. **51**:151–178.

DRAKE, J. W. 1991. A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA **88**:7160–1164.

———. 1993. Rates of spontaneous mutation among RNA viruses. Proc. Natl. Acad. Sci. USA **90**:4171–4175.

FELSENSTEIN, J. 1991. PHYLIP (phylogeny inference package). Distributed by the author, Department of Genetics, University of Washington, Seattle.

FITCH, W. M. 1996. The variety of human virus evolution. Mol. Phylogenet. Evol. **5**:247–258.

FRIEDBERG, E. C., G. C. WALKER, and W. SIEDE. 1995. DNA repair and mutagenesis. American Society for Microbiology, Washington, D.C.

GALINSKI, M. S., M. A. MINK, D. M. LAMBERT, S. L. WECHSLER, and M. W. PONS. 1986. Molecular cloning and sequence analysis of the human parainfluenza 3 virus mRNA encoding the P and C proteins. Virology **155**:46–60.

GALINSKI, M. S., R. M. TROY, and A. K. BANERJEE. 1992. RNA editing in the phosphoprotein gene of the human parainfluenza virus type 3. Virology **186**:543–550.

GALINSKI, M. S., and S. L. WECHSLER. 1991. The molecular biology of the Paramyxovirus genus. Pp. 41–82 *in* D. W.

KINGSBURY, ed. The Paramyxoviruses. Plenum Press, New York.

GIORGI, C., B. M. BLUMBERG, and D. KOLAKOFSKY. 1983. Sendai virus contains overlapping genes expressed from a single mRNA. Cell **35**:829–836.

HERMAN, R. C. 1986. Internal initiation of translation on the vesicular stomatitis virus phosphoprotein mRNA yields a second protein. J. Virol. **58**:797–804.

HORIKAMI, S. M., R. E. HECTOR, S. SMALLWOOD, and S. A. MOYER. 1997. The Sendai virus C protein binds the L polymerase protein to inhibit viral RNA synthesis. Virology **235**:261–270.

HUDSON, L. D., C. CONDRA, and R. A. LAZZARINI. 1986. Cloning and expression of a viral phosphoprotein: structure suggests vesicular stomatitis virus NS may function by mimicking an RNA template. J. Gen. Virol. **67**:1571–1579.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein evolution. Academic Press, New York.

KATO, A., K. KIYOTANI, Y. SAKAI, T. YOSHIDA, T. SHIODA, and Y. NAGAI. 1997. Importance of the cysteine-rich carboxyl terminal half of V protein for Sendai virus pathogenesis. J. Virol. **71**:7266–7272.

KEESE, P. K., and A. GIBBS. 1992. Origin of genes: ''big bang'' or continuous creation? Proc. Natl. Acad. Sci. USA **89**: 9489–9493.

KINGSBURY, D. W., ed. 1991. The Paramyxoviruses. Plenum Press, New York.

KONDO, K., H. BANDO, M. TSURUDOME, M. KAWANO, M. NISHIO, and Y. ITO. 1990. Sequence analysis of the phosphoprotein (P) genes of human parainfluenza type 4A and 4B viruses and RNA editing at transcript of the P genes: the number of G residues added is imprecise. Virology **178**: 321–326.

KRETZSCHMAR, E., R. PELUSO, M. J. SCHNELL, M. A. WHITT, and J. K. ROSE. 1996. Normal replication of vesicular stomatitis virus without C proteins. Virology **216**:309–316.

LAMB, R. A., and C. M. HORVATH. 1991. Diversity of coding strategies in influenza viruses. Trends Genet. **7**:261–266.

LAMB, R. A., B. W. MAHY, and P. W. CHOPPIN. 1976. The synthesis of Sendai virus polypeptides in infected cells. Virology **69**:116–131.

LARTEY, R. T., T. C. VOSS, and U. MELCHER. 1996. Tobamovirus evolution: gene overlaps, recombination and taxonomic implications. Mol. Biol. Evol. **13**:1327–1338.

LYNCH, M., and T. J. CREASE. 1990. The analysis of population survey data on DNA sequence variation. Mol. Biol. Evol. **7**:377–394.

MCGINNES, L., C. MCQUAIN, and T. MORRISON. 1988. The P protein and the nonstructural 38K and 29K proteins of Newcastle disease virus are derived from the same open reading frame. Virology **164**:256–264.

MATSUOKA, Y., J. CURRAN, T. PELET, D. KOLAKOFSKY, R. RAY, and R. W. COMPANS. 1991. The P gene of human parainfluenza virus type 1 encodes P and C proteins but not a cysteine-rich V protein. J. Virol. **65**:3406–3410.

MIYATA, T., and T. YASUNAGA. 1978. Evolution of overlapping genes. Nature **272**:532–535.

MIZOKAMI, M., E. ORITO, K. OHBA, K. IKEO, J. Y. LAU, and T. GOJOBORI. 1997. Constrained evolution with respect to gene overlap of hepatitus B virus. J. Mol. Evol. **44**:S83–S90.

MORGAN, E. M. 1991. Evolutionary relationships of Paramyxovirus nucleocapsid-associated proteins. Pp. 163–176 *in* D. W. KINGSBURY, ed. The Paramyxoviruses. Plenum Press, New York.

MURPHY, F. A., C. M. FAUQUET, D. H. L. BISHOP, S. A. GHABRIAL, A. W. JARVIS, G. P. MARTELLI, M. A. MAYO, and M. D. SUMMERS. 1995. Virus taxonomy. Springer-Verlag, New York.

NEI, M., and T. GOJOBORI. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.

OHNO, S. 1984. Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitious coding sequence. Proc. Natl. Acad. Sci. USA **81**:2421–2425.

PAVESI, A., B. D. IACO, M. I. GRANERO, and A. PORATI. 1997. On the informational content of overlapping genes in prokaryotic and eukaryotic viruses. J. Mol. Evol. **44**:625–631.

PELUSO, R. W., J. C. RICHARDSON, J. TALON, and M. LOCK. 1996. Identification of a set of proteins (C′ and C) encoded by the bicistronic P gene of the Indiana serotype of vesicular stomatitis virus and analysis of their effect on transcription by the viral RNA polymerase. Virology **218**:335–342.

POWER, U. F., K. W. RYAN, and A. PORTNER. 1992. The P genes of human parainfluenza virus type 1 clinical isolates are polycistronic and microheterogeneous. Virology **189**: 340–343.

PRINGLE, C. R. 1997. The order Mononegavirales—current status. Arch. Virol. **142**:2321–2326.

ROZAS, J., and R. ROZAS. 1997. DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Comput. Appl. Biosci. **13**:307–311.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SAMUEL, C. E. 1989. Polycistronic animal virus mRNAs. Prog. Nucleic Acid Res. Mol. Biol. **37**:127–153.

SCHNEIDER, H., K. KAELIN, and M. A. BILLETER. 1997. Recombinant measles viruses defective for RNA editing and V protein synthesis are viable in culture. Virology **227**:314–322.

SHIMOTOHNO, K., Y. TANJI, Y. HIROWATARI, Y. KOMODA, N. KATO, and M. HIJIKATA. 1995. Processing of the hepatitis C virus precursor protein. J. Hepatol. **22**:87–92.

SOUTHERN, J. A., B. PRECIOUS, and R. E. RANDALL. 1990. Two nontemplated nucleotide additions are required to generate the P mRNA of parainfluenza virus type 2 since the RNA genome encodes protein V. Virology **177**:388–390.

SPADAFORA, D., D. M. CANTER, R. L. JACKSON, and J. PERRAULT. 1996. Constitutive phosphorylation of the vesicular stomatitis virus P protein modulates polymerase complex formation but is not essential for transcription or replication. J. Virol. **70**:4538–4548.

SPIROPOULOU, C. F., and S. T. NICHOL. 1993. A small highly basic protein is encoded in overlapping frame within the P gene of vesicular stomatitis virus. J. Virol. **67**:3103–3110.

SWOFFORD, D. L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.

TAPPAREL, C., S. HAUSMANN, T. PELET, J. CURRAN, D. KOLAKOFSKY, and L. ROUX. 1997. Inhibition of Sendai virus genome replication due to promoter-increased selectivity: a possible role for the accessory C proteins. J. Virol. **71**: 9588–9599.

THOMAS, S. M., R. A. LAMB, and R. G. PATERSON. 1988. Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. Cell **54**:891–902.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, po-

sition-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

TOBER, C., M. SEUFERT, H. SCHNEIDER, M. A. BILLETER, I. C. JOHNSON, S. NIEWIESK, V. TER MEULEN, and S. SCHNEIDER-SCHAULIES. 1998. Expression of measles virus V protein is associated with pathogenicity and control of viral RNA synthesis. J. Virol. **72**:8124–8132.

WAGNER, R. R. 1991. Rhabdoviridae and their replication. Pp. 489–503 *in* B. N. FIELDS and D. M. KNIPE, eds. Fundamental virology. Raven Press, New York.

YAMANAKA, M., B. DALE, T. CRISP, B. CORDELL, M. GRUBMAN, and T. YILMA. 1992. Sequence analysis and editing of the phosphoprotein (P) gene of rinderpest virus. Virology **190**:553–556.

YOSHINAKA, Y., I. KATOH, T. D. COPELAND, and S. OROSZLAN. 1985. Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA **82**: 1618–1622.