High rate of viral evolution associated with the emergence of carnivore parvovirus

Laura A. Shackelton*, Colin R. Parrish*, Uwe Truyen†, and Edward C. Holmes††

*Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom; †Department of Microbiology and Immunology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853; and ‡Institute for Animal Hygiene and Public Veterinary Health, Universität Leipzig, an den Tierkliniken 1, D-04103 Leipzig, Germany

Edited by Kenneth I. Berns, University of Florida College of Medicine, Gainesville, FL, and approved November 24, 2004 (received for review September 13, 2004)

Canine parvovirus (CPV) is an emerging DNA virus that was first observed to cause disease in canines in 1978 and has since become a ubiquitous pathogen worldwide. CPV emerged from feline panleukopenia parvovirus (FPLV) or a closely related virus, differing at several key amino acid residues. Here we characterize the evolutionary processes underlying the emergence of CPV. Although FPLV has remained an endemic infection in its host populations, we show that, since the 1970s, the newly emerged CPV has undergone an epidemic-like pattern of logistic/exponential growth, effectively doubling its population size every few years. This rapid population growth was associated with a lineage of CPV that acquired a broader host range and greater infectivity. Recombination played no role in the emergence of CPV. Rather, any preexisting variation in the donor species and the subsequent rapid adaptation of the virus to canines were likely dependent on a high rate of mutation and the positive selection of mutations in the major capsid gene. Strikingly, although these single-stranded viruses have a DNA genome and use cellular replication machinery, their rate of nucleotide substitution is closer to that of RNA viruses than to that of double-stranded DNA viruses.

Paroviruses (family Paroviridae) are small eukaryotic DNA viruses that infect a variety of animal species, including humans. Canine parvovirus (CPV), feline panleukopenia virus (FPLV), and a number of viruses similar to FPLV, such as blue fox parvovirus, the raccoon parvoviruses, and mink enteritis virus (MEV), are all host-range variants of the carnivore parvovirus subgroup (1, 2). Both disease and pathology differ depending on the age of the infected animal, because the viruses replicate only in cells in the S phase of the cell cycle. In neonatal animals, the virus replicates in a large number of tissues, and FPLV often causes cerebellar hypoplasia, whereas CPV causes myocarditis. In older animals, viral replication is limited to lymphoid and small intestinal cells, causing temporary panleukopenia or lymphopenia (3). FPLV is thought to have been endemic in felines since before the beginning of the 20th century (4). In contrast, it was not until 1978 that CPV (the first known strain was designated CPV2) was observed in canines, having emerged from FPLV or one of the closely related carnivore parvoviruses. Although CPV2 infected feline cells in culture, it did not infect cats. CPV2 was later replaced by a new lineage, designated CPV2a, which, along with its variants, can infect both dogs and cats (5, 6). There is also evidence that CPV2a is more effective at infecting canine cells than is CPV2 (7).

Although epidemiological studies of emerging viruses are commonplace, the evolutionary processes associated with cross-species virus transfer are poorly understood. Under some models, adaptation to the new host species is of fundamental importance, elevating the reproductive rate of the virus ($R_0$) above the critical value ($R_0 > 1$) needed for sustained transmission (8). However, most cross-species viral transfers result in “dead-end” infections in their new hosts, with no subsequent transmission. This lack of transmission implies that the key process in emergence is the transfer of a virus variant that, by chance, already has the ability to replicate in the new host species, rather than adaptation in the new hosts after transfer (9). Recombination also is often cited as a key process in viral emergence, most recently for severe acute respiratory syndrome (SARS) coronavirus (10), although this suggestion has been questioned (9). Recombination allows viruses to traverse the adaptive landscape faster than through mutation alone, although most recombinations, like most mutations, are likely to reduce fitness. Both the rate of mutation and the role of recombination in the evolution of paroviruses have yet to be fully elucidated. Nor is it clear whether the successful species jump of FPLV to CPV and subsequent transmission in the dog population was more strongly associated with preexisting genetic variation or posttransmission adaptation.

Paroviruses have a linear single-stranded DNA (ssDNA) genome ~5 kb in length with hairpin structures at both ends. The virus has two major ORFs, one encoding the two nonstructural proteins (NS1 and NS2) and the other encoding the capsid proteins VP1 and VP2 (11). The emergent canine parvoviruses are characterized by specific amino acid changes in and around a raised region of the capsid termed the threefold spike. Structural and mutational studies suggest that this region binds the cellular transferrin receptor and controls host range (7, 12). This capsid region is also highly antigenic, serving as the target of many neutralizing antibodies. The carnivore parvovirus group therefore provides a unique opportunity to document viral adaptation within and among hosts and to characterize the epidemiological dynamics of emerging viruses.

Because DNA viruses replicate by using DNA polymerases, it is generally assumed that their rate of nucleotide substitution and the underlying rate of mutation will not differ greatly from those of their hosts. Estimates of long-term rates of nucleotide substitution, primarily based on cases where virus and host are thought to have undergone co-speciation, show that this assumption is true for large double-stranded DNA (dsDNA) viruses such as the herpesviruses and possibly for small dsDNA viruses such as human papillomavirus and JC polyomavirus (13–16). However, few such rate estimates are available for ssDNA viruses. One experimental analysis found the number of mutations per base per genome replication for the ssDNA phage M13 to be at least 1 logarithm higher than the rates for the dsDNA phages λ, T2, and T4 (17). Although there has not been a similar study of eukaryotic ssDNA viruses, regression analyses of tem-

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: CPV, canine parvovirus; FPLV, feline panleukopenia virus; HPD, high-probability density; MEV, mink enteritis virus; ML, maximum likelihood; ssDNA, single-stranded DNA; dsDNA, double-stranded DNA.

Data deposition: The carnivore parvovirus sequences reported in this paper have been deposited in the GenBank database (accession nos. AY742932-AY742956 and AY787926-AY787930).

To whom correspondence should be addressed. E-mail: edward.holmes@zoo.ox.ac.uk.

© 2004 by The National Academy of Sciences of the USA
Table 1. Nucleotide substitution rates and population dynamics in carnivore paroviruses

<table>
<thead>
<tr>
<th>Data and ML estimates</th>
<th>VP2 FPLV clade</th>
<th>VP2 CPV clade</th>
<th>VP2 FPLV→CPV branch</th>
<th>VP2 CPV2a subclade</th>
<th>NS1 FPLV clade</th>
<th>NS1 CPV clade</th>
<th>NS1 FPLV→CPV branch</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best-fit population model</strong></td>
<td>Constant</td>
<td>Logistic</td>
<td>—</td>
<td>Exponential</td>
<td>Constant</td>
<td>Exponential</td>
<td>—</td>
</tr>
<tr>
<td><strong>Sequence length, bp</strong></td>
<td>1581</td>
<td>1581</td>
<td>1581</td>
<td>1581</td>
<td>2004</td>
<td>2004</td>
<td>2004</td>
</tr>
<tr>
<td><strong>No. of sequences</strong></td>
<td>28</td>
<td>63</td>
<td>—</td>
<td>56</td>
<td>16</td>
<td>19</td>
<td>—</td>
</tr>
<tr>
<td><strong>Chain length, millions</strong></td>
<td>16</td>
<td>20</td>
<td>—</td>
<td>20</td>
<td>20</td>
<td>40</td>
<td>—</td>
</tr>
<tr>
<td><strong>Burn-in length, millions</strong></td>
<td>1.6</td>
<td>2</td>
<td>—</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>—</td>
</tr>
<tr>
<td><strong>Mean substitution rate (substitutions per site per year)</strong></td>
<td>9.4 × 10^{-5}, 1.7 × 10^{-4}</td>
<td>7.1 × 10^{-3} to</td>
<td>1.7 × 10^{-4}</td>
<td>7.9 × 10^{-5}</td>
<td>7.9 × 10^{-5}</td>
<td>2.7 × 10^{-3} to</td>
<td>—</td>
</tr>
<tr>
<td><strong>HPD substitution rate</strong></td>
<td>2.1 × 10^{-5}, 1.2 × 10^{-4},</td>
<td>—</td>
<td>—</td>
<td>1.1 × 10^{-4}, 1.1 × 10^{-7}, 3.4 × 10^{-5},</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td><strong>Mean age, years</strong></td>
<td>116</td>
<td>36</td>
<td>—</td>
<td>28</td>
<td>212</td>
<td>46</td>
<td>—</td>
</tr>
<tr>
<td><strong>HPD age</strong></td>
<td>47, 161</td>
<td>30, 45</td>
<td>—</td>
<td>24, 33</td>
<td>37, 215</td>
<td>28, 72</td>
<td>—</td>
</tr>
<tr>
<td><strong>Mean growth rate, 1/years</strong></td>
<td>—</td>
<td>0.1</td>
<td>—</td>
<td>0.3</td>
<td>—</td>
<td>0.2</td>
<td>—</td>
</tr>
<tr>
<td><strong>Doubling time, years</strong></td>
<td>—</td>
<td>5.0</td>
<td>—</td>
<td>2.3</td>
<td>—</td>
<td>3.2</td>
<td>—</td>
</tr>
</tbody>
</table>

—, not applicable.

*Dates inferred, conservatively, from parameters estimated with CPV VP2 data. These dates were then used to estimate substitution rates along the FPLV→CPV branch.

By examining carnivore parovirus sequences sampled before and after the emergence of CPV from FPLV, we were able to examine in detail the evolutionary basis of viral emergence. In particular, we examine the roles of variation and adaptation in donor and recipient host species by asking whether the transition from felines to canines was associated with a burst of positive natural selection, as expected if the virus had to adapt to the new host. In addition, we test whether recombination had a role in viral emergence and determine the rates of nucleotide substitution and population growth of carnivore paroviruses before, during, and after cross-species transmission.

Materials and Methods

Isolate Collection and Sequencing. Viruses were obtained from clinical samples collected in various countries and stored at −80°C. The viral sequences were either from viruses propagated in feline cells for up to four passages before isolating the viral replicative form DNA, which was cloned into plasmids, or from samples of clinical material (feces or tissue) that were added directly to a PCR. Viral sequences were amplified by using primers that flanked the VP2 or NS1 genes. DNA was amplified by using TaqDNA polymerase, and the PCR products were purified by using the QIAQuick procedure (Qiagen, Valencia, CA) before sequencing. Viral sequence details are given in Tables 3–6, which are published as supporting information on the PNAS web site.

Sequence Data. The sequences of 91 VP2 and 35 NS1 genes from carnivore paroviruses were either determined here or obtained from GenBank. The first 162 nucleotides were deleted from the VP2 sequences because of incomplete sequencing of some strains, resulting in a 1,581-bp alignment. The entire 2,004-bp NS1 sequence was analyzed. We also examined the 495-bp NS2 sequence that is formed as a splice variant of NS1, constructed by combining 261 bp of the amino terminus of the NS1 sequence (in frame) with a 234-bp portion of the carboxyl terminus of NS1 (alternate reading frame) (24). Data sets were aligned by using the SE-AL program (http://evolve.zoo.ox.ac.uk).

Maximum likelihood (ML) phylogenetic trees were estimated by using PAUP* (25). The general time reversible (GTR) + I + Γ model of nucleotide substitution was used in all cases, with the key parameter values [the GTR substitution matrix, the proportion of invariant sites (I), and the Γ distribution of rate variation (with four rate categories, Γ)] estimated from the given data. (Parameter values are available on request.) Bootstrap values were estimated by using 1,000 replicate neighbor-joining trees on the ML substitution model with PAUP*. The program MACCLADE (26) was used to map nucleotide and amino acid changes to specific branches on the gene trees.

Estimating Substitution Rates and Population Dynamics. Overall rates of nucleotide substitution per site per year were estimated by using a Bayesian Markov chain Monte Carlo method, available in the BEAST package (http://evolve.zoo.ox.ac.uk). This method considers the differences in branch lengths between viruses sampled at different times and thoroughly samples the probability density function, thereby exploring different models with parameters that include tree topology, node dates, substitution rate, and substitution model. As before, phylogenetic trees were inferred by using the GTR + I + Γ model, with parameters optimized during multiple runs. Three different population dynamic models were used: constant population size, exponential population growth, and logistic population growth. The logistic model describes a curve with an exponential phase followed by a deceleration in growth rate. All models were compared by using Akaike’s information criterion; a single parameter distinguishes the constant and exponential models and the exponential and logistic models, whereas two parameters distinguish the constant and logistic models. Final chain and burn-in lengths are given in Table 1.
Recombination Tests. Two approaches were used to determine the extent of recombination in VP2 and NS1. First, each sequence alignment was split into four regions: nucleotides 1–399, 400–795, 799–1197, and 1198–1581 for VP2 and nucleotides 1–501, 502–1002, 1003–1503, and 1504–2004 for NS1. ML trees were inferred for each region by using the procedure described above. The trees from each region were then compared by using a simple visual test of incongruence. Second, Sawyer’s run test [GENECONV (27)] was used to search for gene conversion events between pairs of sequences.

Estimating Selection Pressures. We determined the ratio of non-synonymous to synonymous nucleotide changes per site \( (d_{NS}/d_S) \) for the VP2, NS1, and NS2 data sets, with \( d_{NS} > d_S \) indicative of positive selection. We used the ML method in the PAML package [CODEML (28, 29)], which involved the comparison of four evolutionary models that differ in how \( d_{NS} \) and \( d_S \) vary among codons or lineages. To analyze selection pressures at specific codons, we compared the M7 and M8 models. The former specifies that individual codons take 1 of 10 categories of \( d_{NS}/d_S \), all estimated from the data but where no category has \( d_{NS}/d_S > 1 \) so that the model allows only neutral evolution. The M8 model, however, allows positive selection by specifying an 11th category of codons in which \( d_{NS}/d_S \) can exceed 1.0. Evidence for positive selection is obtained if M8 rejects M7 in a likelihood ratio test and at least one category of codons in M8 has \( d_{NS}/d_S > 1 \). To analyze selection pressures along individual lineages, we compared (again using a likelihood ratio test) M0, in which each branch has the same \( d_{NS}/d_S \) ratio, with the free ratio model, in which branches are allowed to take on individual values of \( d_{NS}/d_S \).

Results

Phylogenetic Analysis. We collected 91 carnivore parvovirus VP2 sequences sampled from various geographical regions over a 42-year period. Although the first FPLV sequence is from 1962 isolate, there are CPV sequences from the first year in which this viral disease was observed (1978) until 2004. A ML phylogenetic tree of these data revealed two distinct clades separated by a long central branch (Fig. 1). One clade (from here forward referred to as the 28-member FPLV clade) included all FPLV, MEV, blue fox parvovirus, and raccoon parvoviruses; a second 63-member clade included all CPV sequences (those isolated from both canines and felines). Within the CPV clade, there was a long branch separating the CPV2a subclade from the ancestral CPV sequence. The phylogeny showed a general temporal structure, with the most recently sampled sequences generally falling farther from the root than viruses sampled in the more distant past.

We performed an equivalent phylogenetic analysis on 35 carnivore parvovirus NS1 sequences spanning 37 years (Fig. 3, which is published as supporting information on the PNAS web site). Although this phylogeny also showed the two distinct clades separated by a longer central branch, there was less temporal structure in the tree, with sequences from different sampling times often mixed.

Substitution Rates. The VP2 data set was divided into its FPLV and CPV clades, and substitution rates for each were estimated by using a Bayesian Markov chain Monte Carlo approach where a molecular clock is assumed but any rate variation is reflected in the high-probability density (HPD) interval for each estimated parameter. The mean substitution rate for the CPV clade was \( 1.7 \times 10^{-4} \) substitutions per site per year; the mean rate for the FPLV clade was \( 9.4 \times 10^{-5} \) substitutions per site per year, but with a larger HPD (Table 1). Hence, despite infecting different host species, these two clades are evolving at approximately the same rate. Most remarkable, however, was the greatly elevated rate on the central branch separating the FPLV and CPV clades. Assuming that the time interval over which CPV emerged was between 1 and 10 years (1968–1978; see below), the substitution rate on this branch ranges from \( 7.1 \times 10^{-5} \) to \( 0.7 \times 10^{-3} \) substitutions per site per year. These rates are within the range seen for a number of RNA viruses, including HIV-1 (30).

With these substitution rates, it was possible to estimate key dates in the evolutionary history of the canine parvoviruses. The mean age of the FPLV group was estimated to be \( \sim 116 \) years. Although this estimate is consistent with epidemiological observations, the large HPD associated with this date necessitates caution. In contrast, the mean age estimated for the CPV clade was 36 years, and the range was far tighter, suggesting that CPV had been in the canine population for \( \sim 10 \) years before it was first recognized in 1978. Indeed, a retrospective report from Greece cites evidence for CPV antibodies in dog sera from 1974. However, there was no accompanying disease, and the virus did not appear to be highly contagious, because most dogs tested were seronegative (31). Our analysis, therefore, supports the idea that a CPV-like virus was present in some localities a number of years before its association with large-scale canine disease.

After CPV diverged into CPV2 and CPV2a, the dominant CPV2a lineage gave rise to the variants N426D (designated CPV2b), S297A, G300D, and N/D426E, all of which now cocirculate among canines. This subclade, which is still designated CPV2a but includes the point variants, has a mean age of 28 years, implying that by 1976, \( \sim 8 \) years after CPV first emerged, the CPV population had already split into the CPV2 and the CPV2a subclades.

The same rate analysis was performed on the NS1 data. The more limited data set and weaker temporal structure of the phylogeny resulted in less precise rate estimates but with trends consistent with VP2 (Table 1). The FPLV and CPV clades both showed a mean substitution rate of \( 7.9 \times 10^{-5} \) substitutions per site per year, whereas the branch separating them showed an elevated rate of \( 2.7 \times 10^{-3} \) to \( 0.3 \times 10^{-3} \) substitutions per site per year.

Population Dynamics. Strikingly different models of growth were observed in the FPLV and CPV clades. Whereas the best-fit model for the FPLV clade was that of a constant population size, reflecting the more endemic nature of this disease, the CPV clade was characterized by logistic population growth. Hence, after the jump from felines to canines, the FPLV clade maintained its endemic nature, whereas the CPV viruses experienced a period of rapid population growth in the new host. The growth rate for the CPV clade results in a predicted population doubling time of 5.0 years. However, a separate analysis of the CPV2a subclade, which supported a model of exponential population growth, showed a population doubling time of 2.3 years. Hence, inclusion of the now-extinct CPV2b subclade in our previous analysis of CPV had lowered the initial estimates of growth rate. Clearly, the CPV2b subclade, which originally spread worldwide and caused the first observable outbreaks of disease, grew at a lower rate than its sister clade and was eventually out-competed by the CPV2a viruses.

Selection Pressures in FPLV and CPV. We measured \( d_{NS}/d_S \) across all VP2, NS1, and NS2 genes and lineages to determine the amount, location, and timing of positive selection. The overall \( d_{NS}/d_S \) ratio in VP2 was 0.121, indicating that most sites are subject to strong purifying selection. However, because the M8 model significantly rejects M7 (\( P = 0.0002 \)) and four sites (101, 232, 300, and 426) have \( d_{NS}/d_S \) values \( > 1 \), we conclude that positive selection is acting on specific amino acids in VP2. In the case of NS1, M8 does not significantly reject M7 (\( P = 0.0781 \)), so there is no evidence for positive selection within this gene.

Although informative, this analysis is unlikely to detect selec-
tion events that have occurred once in evolutionary history and perhaps are associated with host-switching. We therefore analyzed \( d_{NS}/d_{S} \) along individual branches of the tree. In both VP2 and NS1, the free ratio model rejects the M0 model (\( P = 2.72 \times 10^{-33} \) and \( P = 4.72 \times 10^{-12} \), respectively), indicating that \( d_{NS}/d_{S} \) significantly differs across lineages. Crucially, \( d_{NS}/d_{S} \) was greatly elevated on the FPLV \( \rightarrow \) CPV and CPV \( \rightarrow \) CPV2a branches of the VP2 tree. Along the former branch, which had a \( d_{NS}/d_{S} \) value of 0.382, there were seven nonsynonymous and four synonymous changes (Table 2). Although this \( d_{NS}/d_{S} \) value is still \(< 1.0\), it is significantly greater than the average values for internal branches in the two clades (\( \chi^2 \) test, \( P = 0.025 \) with one degree of freedom; observed and expected values were generated by using \( d_{N} \) and \( d_{S} \) multiplied by the number of sites in each class). Moreover, four of these seven nonsynonymous changes were at sites that are invariant in the FPLV and CPV clades, which suggests that they are associated with a selectively driven change of function rather than a relaxation of selective constraints.

Along the CPV \( \rightarrow \) CPV2a branch, three nonsynonymous and no synonymous changes occurred, resulting in a \( d_{NS}/d_{S} \) value of infinity, strongly indicative of positive selection. The NS1 branch separating the FPLV and CPV clades had a \( d_{NS}/d_{S} \) value of 0.803, and with three nonsynonymous and one synonymous changes, this branch also had a significantly greater \( d_{NS}/d_{S} \) ratio than the mean for the internal branches of the FPLV and CPV clades (\( P = 0.0004 \)). Interestingly, two of these changes, H247Q and T248I, are adjacent, suggesting coordinated substitution. In both the VP2 and NS1 genes, some substitutions that distinguish CPV from FPLV also are present in MEV and/or blue fox parvovirus. That these two FPLV-like viruses are the closest relatives of CPV in these phylogenies suggests that one may be the direct ancestor of CPV, although a larger sample of viruses is needed to confirm this.

Fig. 1. ML phylogenetic tree of 91 VP2 gene sequences from carnivore parvoviruses. The tree is rooted with the oldest sampled sequence. Bootstrap values are shown for relevant nodes. Other nodes with \( \sim 70\% \) support are marked with an asterisk. Horizontal branch lengths are drawn to scale. The name of each isolate is followed by the location and year of isolation. Locations are coded as follows: UK, United Kingdom; US, United States; JA, Japan; FR, France; AU, Australia; FI, Finland; GE, Germany; TA, Taiwan; NZ, New Zealand; VI, Vietnam; EU, Europe (no further information available); PO, Poland; IT, Italy; SA, South Africa. The FPLV clade is shown in blue, the CPV2 subclade is shown in yellow, and the CPV2a subclade is shown in red. BFPV, blue fox parvovirus.
Table 2. Selection analysis of FPLV→CPV→CPV2a branches

<table>
<thead>
<tr>
<th>Measure</th>
<th>FPLV→CPV</th>
<th>CPV→CPV2a</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_h/d_s )</td>
<td>0.382</td>
<td>( \infty )</td>
</tr>
<tr>
<td>( N_{ns}/N_{ds} )</td>
<td>2.353</td>
<td>( \infty )</td>
</tr>
<tr>
<td>Nonsynonymous</td>
<td>K80R (A239G)</td>
<td>M87L (A259T)</td>
</tr>
<tr>
<td></td>
<td>K93N (A279C)</td>
<td>A/G300G (G/C899C)</td>
</tr>
<tr>
<td></td>
<td>V103A (T308C)</td>
<td>D305Y (G913T)</td>
</tr>
<tr>
<td></td>
<td>V300A (T899G/C)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D323N (G967A)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N564S (A1691G)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A568G (C1703G)</td>
<td></td>
</tr>
<tr>
<td>Synonymous</td>
<td>V82V (A246G)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F212F (G694A)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y233Y (T699C)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>A541A (A1623C)</td>
<td></td>
</tr>
</tbody>
</table>

Amino acid changes along each branch are given, followed by the underlying nucleotide change.

*\( N_{ns}/N_{ds} \) = total nonsynonymous changes/total synonymous changes.

Finally, we examined selection pressures in NS2, a splice variant of NS1. In the parvovirus minute virus of mice (MVM), NS2 plays a role in capsid assembly and is required for replication in its normal host animal (32, 33). Because NS2 mutants in MVM influence replication in alternative hosts (34), it has been suggested that NS2 may also control host range in CPV and that mutations in this gene interact with those in the capsid, although experimental studies of mutated NS2 in CPV2 revealed no effect on virus replication in tissue culture or dogs (24). Although we saw one mutation (R105S) in the carboxyl terminus of NS2 along the FPLV→CPV branch, because isolated selection events do not give a strong signal in this analysis and M8 does not reject M7 (\( P = 0.118 \)), we provide no evidence for positive selection in this gene.

**Discussion**

Our study provides unique insight into the evolutionary events accompanying viral emergence. In particular, CPV is one of the few examples of host-switching in which ample sequence data are available from viruses collected at multiple time points before and after the host-switching event. Our analysis suggests that the FPLV clade has maintained a constant population size throughout its recent evolutionary history in cats. In contrast, CPV has experienced a great increase in growth rate, characteristic of an epidemic, as it adapted to its canine host, expanded its host range, and generated a number of antigenic and host-range variants. Notably, our molecular clock estimates indicate that the ancestral CPV likely emerged in the canine population up to 10 years before it was first described. This virus was maintained in the population for several years, seemingly accumulating beneficial mutations under strong positive selection. In particular, the CPV2 and CPV2a variants would have been selected during this period, coexisting in dogs during the 1970s, until finally emerging as the causes of major epidemics.

**Evolutionary Rates.** Perhaps our most striking observation was that the rate of nucleotide substitution in the carnivore parvoviruses is more similar to that seen in RNA viruses than to that in other DNA viruses and that the “emergent” branch separating FPLV from CPV had a substitution rate similar to that seen in rapidly evolving RNA viruses such as HIV-1 and human influenza A (30, 35, 36). There are two possible explanations for this high substitution rate: that the intrinsic rate of mutation in CPV is high or that persistent positive selection pressure has greatly elevated the rate of mutation fixation. From current data, it appears that both factors play a role. We found strong evidence for positive selection at a number of sites in VP2 and an evolutionary pattern suggestive of directional selection on the branch leading from FPLV to CPV. However, because this positive selection is restricted to a few sites and branches, most of the other substitutions observed are likely to be neutral and fixed by genetic drift. Furthermore, the NS1 gene, which is not exposed as an antigenic or receptor-binding site on the capsid surface, showed similar substitution rates, supporting the idea that nucleotide changes for immune escape and tropism shifts are not entirely responsible for the high substitution rates. Hence, although the VP2 substitution rate is likely to be elevated by natural selection, it generally reflects a high background mutation rate.

Why should mutation rates be elevated in parvoviruses and perhaps other ssDNA viruses? Despite the fact that cellular DNA machinery replicates these genomes, perhaps only a subset of this machinery is used for these viruses. The ssDNA template or mechanisms used for its replication may alter the efficiency or accuracy of the polymerases used. Also, proofreading and repair mechanisms may not be as active or efficient on these genomes or in an infected cell. Alternatively, lack of a double helix may leave them vulnerable to mutational processes such as deamination (37). Clearly, these questions cannot be resolved without further experimental analysis of the mutation process in ssDNA viruses.

**Natural Selection on VP2.** The overall low \( d_N/d_S \) ratio in the carnivore parvoviruses indicates that most amino acid residues are subject to purifying selection, with adaptive evolution restricted to specific residues within VP2. Most of the changes fixed along branches with elevated \( d_N/d_S \) ratios appear, in experimental studies, to be essential for determining host range, cell tropism, and/or antigenicity, further supporting the idea that they have been subject to positive selection (6, 38–40). For example, mutational analyses have shown that the FPLV→CPV changes K93N and D323N, acting in concert, can change the tropism of FPLV and allow it to bind and infect canine cells (38). Additional changes in the 300 region of the VP2 “shoulder region” (A300G, D305Y, and M87L) are also subject to positive selection (38). Together, these substitutions enhance the ability of CPV to infect canine cells (39). As shown in Fig. 2, the 300 region, together with residues 93 and 323, comprise the corners of a capsid surface spike, which may represent a broad...
Residues 80, 564, and 568 are also in the shoulder region, and changes at these sites are thought to be responsible for the loss of the feline host range in CPV2 (5). Whereas CPV2 can infect feline cells in tissue culture, it cannot infect cats unless residues 80, 564, and 568 revert to their FPLV identity. Intriguingly, the gain of feline host range on the CPV→CPV2a branch did not involve reversions at these sites, resulting instead from changes at nearby sites 87, 300, and 305 (42). The successful CPV2a strain, which has doubled in population size every few years, has generated at least four additional variants, which are now fixed in the population.

Although we know which amino acid substitutions occurred during the transition from FPLV to CPV, with a few exceptions it is uncertain which were directly involved in the host-range switch, which were hitchhikers present as variation in the direct ancestor of CPV, and which were compensatory changes. Allowing viruses with single or multiple mutations at the site or sites in mutation to compete in different tissue cultures will help distinguish adaptive from alternative explanations. The exact interactions between the threefold capsid region and both the transferrin receptor and antibodies may be visualized by structural analysis, and the extent to which mutations in this region result in antibody escape and/or affect transferrin receptor binding can also be determined.

The Evolutionary Genetics of Viral Emergence. Our study explores three key issues in viral emergence. (i) Does cross-species transmission and the establishment of emerging viruses require adaptation to the new host species? (ii) Does the genetic variation that facilitates such adaptive evolution arise in the donor or the recipient host? (iii) How many sequential or simultaneous changes are required for a successful host jump? In the case of the carnivore parvoviruses, we document a stepwise transfer and prolonged transmission. Crucially, some of these mutations have clearly arisen in the recipient host species. Support for this finding includes the following. First, the FPLV→CPV branch is associated with a large increase in the $d_s/d_d$ ratio, which is compatible with a punctuating episode of positive selection associated with adaptation to the new host species. Although the elevated $d_s/d_d$ ratio on this branch could also be explained by a relaxation of selective constraints, we consider this theory to be less plausible, given the functional importance of the sites involved and that most are invariant within the FPLV and CPV clades. Hence, it is likely that many changes along the FPLV→CPV branch are adaptive, although it remains to be determined whether all such changes are adaptive. Second, the first CPV variant that spread worldwide (CPV2) was unable to replicate in felines, so at least some of CPV2’s evolution must have been restricted to canines. At a later time, a new variant (CPV2a) acquired the ability to efficiently infect both cats and dogs and spread rapidly on a global scale, replacing CPV2. Finally, some of the first viruses to infect canines must have had the ability to cause small transmission chains rather than dead-end infections. It is likely that these “stuttering chains of transmission” allowed the virus to gradually adapt to canine transmission. During this period of evolution, CPV was sufficiently rare to evade veterinary detection, and there were no overt signs of disease in dog populations. However, it is likely that strong positive selection led to an accumulation of additional beneficial mutations, resulting in the emergence of the highly successful CPV2 and CPV2a variants.

The emergence of CPV demonstrates how a virus can successfully cross species barriers and become established as an epidemiologically spreading pathogen in a new host. In the case of CPV, this process probably involved rapid and sequential events of molecular adaptation, facilitated by an intrinsically high rate of mutation without recombination. Furthermore, at least some of these key changes occurred in the recipient host species. Given such rapid evolutionary dynamics, it is likely that CPV will continue to improve its capacity to spread among its carnivore hosts.

This work was supported by the Rhodes Trust and National Institutes of Health Grant R01AI20385 (to C.R.P.).