Phylogenetic analysis reveals the emergence, evolution and dispersal of carnivore parvoviruses

Karin Hoelzer,1 Laura A. Shackelton,2 Colin R. Parrish1 and Edward C. Holmes2,3

Correspondence
Karin Hoelzer
kh294@cornell.edu

1Baker Institute for Animal Health, Department of Microbiology and Immunology, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA
2Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802, USA
3Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA

Canine parvovirus (CPV), first recognized as an emerging virus of dogs in 1978, resulted from a successful cross-species transmission. CPV emerged from the endemic feline panleukopenia virus (FPV), or from a closely related parvovirus of another host. Here we refine our current understanding of the evolution and population dynamics of FPV and CPV. By analysing nearly full-length viral sequences we show that the majority of substitutions distinguishing CPV from FPV are located in the capsid protein gene, and that this gene is under positive selection in CPV, resulting in a significantly elevated rate of molecular evolution. This provides strong phylogenetic evidence for a prominent role of the viral capsid in host adaptation. In addition, an analysis of the population dynamics of more recent CPV reveals, on a global scale, a strongly spatially subdivided CPV population with little viral movement among countries and a relatively constant population size. Such limited viral migration contrasts with the global spread of the virus observed during the early phase of the CPV pandemic, but corresponds to the more endemic nature of current CPV infections.

INTRODUCTION

Efficient transmission in a new host species is a major obstacle to the successful cross-species transmission and emergence of viruses. Many emergent viruses are likely to be maladapted to the new host species and therefore cause only short transmission chains (Kuiken et al., 2006). Hence, the process of viral emergence will often require multi-stage adaptations to a new host, an idea supported by the observation that RNA viruses, which often exhibit evolutionary plasticity, are the most common agents of emerging disease (Woolhouse et al., 2001). Indeed, there is growing evidence that, although replication fidelity is low in RNA viruses (and some ssDNA viruses), it has been fine-tuned in some cases (Sanjuan et al., 2005, 2007; Vignuzzi et al., 2006) and might be environment-specific (Pita et al., 2007).

Canine parvovirus (CPV) represents one of the few examples where the process of cross-species viral transmission has been observed in ‘real-time’. CPV emerged during the 1970s as a host-range variant of an endemic parvovirus of cats, mink or related hosts within the order Carnivora. This group of closely related viruses includes feline panleukopenia virus (FPV) and mink enteritis virus (MEV), both of which have long been endemic in their natural hosts. These viruses are globally distributed and infect a variety of domestic and feral hosts including cats, raccoons, foxes and mink (Barker et al., 1983; Parrish, 1990). Host range barriers restrict transmission of these viruses among dogs and related canids, in which viral replication is limited to the thymus and probably bone marrow (Truyen & Parrish, 1992).

CPV was first recognized in 1978 when it spread worldwide, but while this virus had probably circulated locally before that time, the exact time of emergence is uncertain. The first CPV identified is referred to as CPV-2 to distinguish it from the previously identified but distantly related parvovirus of dogs, minute virus of canines (canine minute virus). CPV-2 caused a severe pandemic with high mortality in dogs, but was unable to infect cats, raccoon or mink. A variant virus (referred to as CPV-2a) emerged in 1979 and replaced CPV-2 worldwide in approximately one year (Parrish et al., 1988). CPV-2a differed antigenically from CPV-2 and infected both dogs and cats. Another antigenic variant (referred to as CPV-2b) with an equivalent host-range was defined by a single substitution
(N426D) in the capsid protein gene and spread globally after 1984. Further mutations arose and spread widely, such as a mutation at residue 297 of the capsid protein. Currently, another mutation at the exposed capsid residue 426 (D426E), with a thus far uncharacterized impact on fitness, is spreading globally. This mutation was first described in Italy in 2000 (Buonavoglia et al., 2001), but has since been observed in other European countries (Decaro et al., 2006, 2007; Martella et al., 2004), South America (Perez et al., 2007), Asia (Nakamura et al., 2004) and most recently in the United States (Hong et al., 2007; Kapil et al., 2007).

Parvoviruses are small DNA viruses with a single-stranded linear genome of about 5 kb. Their replication is dependent upon the host cell replication machinery, but estimates of rates of nucleotide substitution for both CPV and FPV are more comparable to those observed in RNA viruses than in other DNA viruses (Shackelton et al., 2005). The CPV genome contains two open reading frames (ORFs), one encoding the two non-structural proteins NS1 and NS2, and the other encoding the two structural proteins VP1 and VP2 (Supplementary Fig. S1, available with the online version of this paper). The N-terminal ends of NS1 and NS2 are identical in sequence, but the C-terminal end of the NS2 protein is derived by differential splicing of the mRNA, being translated in a different reading frame which overlaps that of NS1. VP1 and VP2 are splice variants and are identical in sequence, except for a 143 aa N-terminal stretch unique to VP1. VP2 is the major capsid protein, and accounts for approximately 90% of the viral capsid. The capsid is the major determinant of host range (Hueffer et al., 2003b) and subject to antibody-mediated selection (Nelson et al., 2007). To date, most studies have focused on the evolution of the VP2 gene, with limited work on the non-structural genes.

Despite widespread vaccination of domestic carnivores, FPV and CPV have remained important pathogens of domestic and wild carnivores. In addition, both the patterns of CPV dispersal and the selection pressures acting across the viral genome are poorly understood. This non-enveloped virus causes only short infections with no persistent virus remaining after infection, but is highly resistant in the environment. The global spread of the virus could therefore occur through the movement of infected animals or mechanical vectors, although the spatial dynamics of CPV have not been examined in detail. Here we explore these ideas by performing the largest evolutionary analysis of CPV undertaken to date, incorporating 143 sequences from 13 countries sampled over 29 years.

**METHODS**

**Isolate collection and sequencing.** Viral DNA was amplified by PCR from clinical specimens collected at various points before or after the emergence of CPV. The DNA was either amplified directly from the clinical specimen, or after purification using the QIAamp viral DNA purification kit (Qiagen), according to the manufacturer's recommendations. PCR amplification was performed using Phusion hot start polymerase (New England Biolabs) and primers that flanked either the NS1 or VP1 genes, respectively, or that spanned the whole coding region. PCR products were either sequenced directly after purification with the QIAquick PCR purification kit, or after subcloning of the viral DNA into the bacterial vectors pJET 1.2 blunt (Fermentas) or pSMART GC HK (Lucigen) and each nucleotide was sequenced multiple times. Consensus sequences were constructed using the ‘Seqman’ program in the Lasergene 7.0 software package (DNASTAR). A total of five FPV and six CPV sequences that cover the complete viral coding region have been generated and were deposited in GenBank (accession numbers EU659111–EU659121).

**Sequence data.** The sequences of 27 near full-length genomes, 54 NS1 and 199 VP2 genes from CPV, FPV and closely related parvoviruses were either determined here or obtained from GenBank (Supplementary Tables S1–S3, available with the online version of this paper). Several CPV-2c sequences (GenBank accession numbers AY380577, AY742942, EF375479, EF375481) covered only smaller parts of the coding region and were therefore excluded from the analysis. Similarly, not all published isolations of CPV-2c strains were apparently accompanied by GenBank submission of the viral sequences (see for example Kapil et al., 2007; Hong et al., 2007 or Decaro et al., 2007). Of the sequences compiled, 23 near full-length sequences, 50 NS1 and 126 VP2 gene sequences contained information on the time of isolation, while the place of isolation was available for 52 NS1 and 178 VP2 gene sequences. The first four codons of the VP2 sequences were deleted from the VP2 gene alignment, as some of the sequences were incomplete, resulting in an alignment of 1721 nt. For the analysis of selection pressures acting on the VP1 unique region, a second alignment of 26 VP1 gene sequences was constructed with the intron within the VP1 unique N terminus removed, resulting in an alignment of 2181 nt. The complete 2001 nt coding region of the NS1 gene was also analysed, and for the analysis of selection pressures on the NS2 gene, the N-terminal 89 codons were combined with the 76 codon long C-terminal region in the correct reading frame, resulting in an alignment of 495 nt. The near full-length sequences were trimmed to the complete NS1 and VP1 coding regions, with the introns included, resulting in a 4269 nt alignment. Sequences were aligned by hand in the Se-Al program (http://tree.bio.ed.ac.uk/software/seal/).

**Inference of phylogenetic trees.** Maximum-likelihood (ML) phylogenetic trees were estimated using PAUP* version 4.0 (Swofford, 2002) with the best-fit model of nucleotide substitution determined using MODELTEST (Posada & Crandall, 1998). The HKY85 substitution model was the best-fit for the full-length dataset and the key parameters (i.e. transition and transversion rates) were estimated from the data. For the VP2 datasets, the GTR + I + G substitution model was optimal. To assess support for individual nodes, bootstrap resampling values were estimated with 1000 neighbour-joining trees, again employing PAUP* (Swofford, 2002).

**Estimation of substitution rate and population dynamics.** Rates of nucleotide substitution per site, per year (subs site⁻¹ year⁻¹) were estimated using the Bayesian Markov chain Monte Carlo (MCMC) method available in the BEAST package (http://beast.bio.ed.ac.uk/). Phylogenetic trees were estimated using the HKY85 substitution model, and either strict (constant) or relaxed (uncorrelated lognormal) molecular clocks were assumed. Four different demographic models were compared, assuming constant population size, exponential or logistic population growth or a Bayesian skyline model, which provides a piecewise depiction of changes in relative genetic diversity through time. Models were compared using Bayes Factors, with uncertainty in each parameter estimate depicted in values of the 95% highest probability density (HPD) interval. All analyses were run for sufficient time to ensure convergence as assessed.
using the TRACER program (http://tree.bio.ed.ac.uk/software/tracer/), with 10% of runs removed as burn-in. For those models supporting population growth, estimates of the population doubling time (λ) were obtained from the BEAST estimates of population growth (r) by using the formula λ=ln2/r.

Estimates of site-specific selection pressures. We used the single likelihood ancestor counting (SLAC), fixed effects likelihood (FEL) and random effects likelihood (REL) algorithms available in the Datamonkey web interface of the HYPHY software package (Kosakovsky Pond & Frost, 2005) to estimate the relative rates of synonymous (dS) and nonsynonymous (dN) substitutions per site (ratio dN/dS) in our sample of CPV and FPV isolates. The HKY85 substitution model was employed in each case.

Analysis of migration patterns. To determine the strength of phylogenetic clustering by country of virus isolation we employed a parsimony character mapping approach (Carrington et al., 2005). Each CPV sequence was first assigned a character state reflecting its country of origin. Given the ML phylogeny for these sequences (as determined above), the minimum number of state changes needed to produce the observed distribution of country character states was estimated using parsimony (excluding ambiguous changes). To determine the expected number of changes under the null hypothesis of complete mixing among countries, the states of all isolates were randomized 1000 times. The difference between the mean number of observed and expected state changes indicates the level of geographical isolation, with statistical significance assessed by comparing the total number of observed state changes to the number expected under random mixing. All analyses were performed using PAUP* (Swofford, 2002).

RESULTS

Phylogenetic analysis of the full-length genome

We analysed 27 near full-length genomes of carnivore parvoviruses, several of which were determined during the course of this study. The FPV sequences cover a time period of 42 years (1964 to 2006) and were collected in various geographical regions. The oldest CPV isolate was collected in 1978, the year in which CPV was first described, and the CPV sequences cover a total of 29 years. The topology of the ML tree of these data (Fig. 1) was similar to that of trees inferred from the VP2 region, with the longest internal branch separating the FPV and CPV clades. The FPV and CPV clades were separated by 16 nt substitutions, while 7 substitutions separated the CPV-2 from the CPV-2a clade. The majority of these substitutions (11 between FPV and CPV and 5 between CPV-2 and CPV-2a) were located in the capsid protein region (data not shown).

Substitution rates and population dynamics

Bayesian MCMC estimates of rates of nucleotide substitution were inferred for the full-length genome and separately for the NS1 and VP2 genes, either analysing all carnivore parvoviruses simultaneously or subdividing the sequences into FPV and CPV (i.e. all CPV-2, CPV-2a and derived viruses). For the VP2 sequence, two additional datasets were analysed, one consisting of the CPV-2a sequences and the other comprising all FPV and most CPV sequences, but excluding the original CPV-2 sequences. Very similar substitution rates were obtained for individual datasets under all molecular clock and demographic models, reflecting the robustness of this analysis (all results available from the authors on request). However, estimates of the substitution rate from the capsid protein gene were strikingly different between FPV and CPV based on the NS1 gene were 2.2 ± 0.6 substitutions per site (95% HPD 1.3–3.6), while that of CPV was significantly higher at 6.6 ± 2.2 substitutions per site (95% HPD 1.7–10.4), with overlapping HPD values.

Using the same approach, we were able to estimate the time to the most recent common ancestor (TMRCA) of both FPV and CPV. The estimated TMRCA for FPV was 98 years from the most recent analysed isolate collected in 2006 (95% HPD 53–168 years), while that for CPV was 32 years (95% HPD 29–37 years), which closely corresponds with the first CPV case recognized in 1978. The mean TMRCA estimate for CPV-2a (28 years) is lower...
than that for all CPVs, again matching the first observed CPV-2a collected in 1979. The TMRCA estimates based on the NS1 gene place the emergence of FPV, and particularly CPV, considerably earlier than those based on the capsid protein gene (128 years for FPV and 85 years for CPV). However, the spread of the HPD values is very wide (and overlaps with those inferred from VP2), revealing the inherent statistical uncertainty in this estimate (although, again, the 95% HPD values overlap). However, these estimates of growth rate should be treated with caution due to the potentially confounding effect of population subdivision. Both FPV and CPV NS1 sequences were best described by models of constant population size.

### Selection pressures in CPV and FPV

An analysis of selection pressures, manifest as $d_N/d_S$ ratios, across both viruses and genes revealed the sporadic action of positive selection, with mean $d_N/d_S$ values of 0.15 for NS1, 0.53 for NS2, 0.15 for the VP1 unique region and 0.12 for VP2 (Table 2). Notably, the short NS2 protein contained four positively selected sites, all within the 76 aa long C-terminal region where the encoded sequence overlaps that of NS1 in a different reading frame. Hence, the ‘positively selected’ sites in this region are difficult to interpret. While the possibility of positive selection cannot be ruled out, the results might represent false-positives caused by synonymous mutations in the alternate reading frame. In contrast, only a single NS1 site, at codon 597, appeared to be under positive selection. This amino acid is caused by synonymous mutations in the alternate reading frame. In contrast, only a single NS1 site, at codon 597, appeared to be under positive selection. This amino acid is

### Selection pressures in CPV and FPV

An analysis of selection pressures, manifest as $d_N/d_S$ ratios, across both viruses and genes revealed the sporadic action of positive selection, with mean $d_N/d_S$ values of 0.15 for NS1, 0.53 for NS2, 0.15 for the VP1 unique region and 0.12 for VP2 (Table 2). Notably, the short NS2 protein contained four positively selected sites, all within the 76 aa long C-terminal region where the encoded sequence overlaps that of NS1 in a different reading frame. Hence, the ‘positively selected’ sites in this region are difficult to interpret. While the possibility of positive selection cannot be ruled out, the results might represent false-positives caused by synonymous mutations in the alternate reading frame. In contrast, only a single NS1 site, at codon 597, appeared to be under positive selection. This amino acid is caused by synonymous mutations in the alternate reading frame. In contrast, only a single NS1 site, at codon 597, appeared to be under positive selection. This amino acid is

### Table 1. Nucleotide substitution rates, population dynamics and times to common ancestry for CPV and FPV

<table>
<thead>
<tr>
<th>Demographic model</th>
<th>VP2</th>
<th>All</th>
<th>NS1</th>
<th>Full-length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence length (bp)</td>
<td>1721</td>
<td>1721</td>
<td>1721</td>
<td>1721</td>
</tr>
<tr>
<td>Number of sequences</td>
<td>30</td>
<td>96</td>
<td>90</td>
<td>120</td>
</tr>
<tr>
<td>Mean substitution rate*</td>
<td>$2.1 \times 10^{-4}$</td>
<td>$1.6 \times 10^{-4}$</td>
<td>$1.1 \times 10^{-4}$</td>
<td>$1.7 \times 10^{-4}$</td>
</tr>
<tr>
<td>HPD substitution rate</td>
<td>$2.2 \times 10^{-4}$</td>
<td>$1.7 \times 10^{-4}$</td>
<td>$1.1 \times 10^{-4}$</td>
<td>$1.2 \times 10^{-4}$</td>
</tr>
<tr>
<td>TMRCA (years)†</td>
<td>98</td>
<td>32</td>
<td>28</td>
<td>71</td>
</tr>
<tr>
<td>Mean growth rate (years$^{-1}$)</td>
<td>NA</td>
<td>0.2</td>
<td>0.4</td>
<td>NA</td>
</tr>
<tr>
<td>Mean epidemic doubling time (j)</td>
<td>NA</td>
<td>3.7</td>
<td>1.8</td>
<td>NA</td>
</tr>
</tbody>
</table>

*sub site$^{-1}$ year$^{-1}$.

†Time to most recent common ancestor, age since most recent sequence included in analysis (i.e. 2006).
suggesting that this result might be a false positive. No positively selected sites were detected in the VP1 unique region. As expected, given the capsid’s known role in host range control, four capsid residues, at VP2 positions 101 (dS/dN=∞; dN=2.24), 300 (dS/dN=∞; dN=3.0), 324 (dS/dN=4.43) and 426 (dS/dN=∞; dN=4.56), were under positive selection across all sequences. Among the FPV sequences, only VP2 codon 562 in the VP2 gene appears to be under positive selection (dS/dN=∞; dN=13.05). Among the CPV sequences, VP2 residues 87 (dS/dN=∞; dN=5.03) and 426 (dS/dN=∞; dN=4.55) were under positive selection.

**Migration patterns of CPV**

The differences between the observed and expected numbers of migration events involving CPV isolates sampled from different countries are shown in Table 3, with negative values indicative of population subdivision and positive values characterizing directional migration between the specified geographical regions. Overall, we detected a strong signal of population subdivision, with only a few instances of migration between specific localities. Indeed, only 12 of the 182 possible directional migration pathways showed signs of CPV migration, and none had strong signals, suggesting only sporadic geographical movement of these viruses. Conversely, 95 of the 182 possible pathways were negative, indicative of population subdivision, sometimes strongly so. Interestingly, China represents a far weaker source of viruses for other regions than expected given the number of available sequences from this country. Similarly, Brazil is also a weaker source population than expected by chance (especially for movements to China and Taiwan), but appears to be a source for movement to Europe (Italy and Poland) and potentially the United States. Overall, the strongest evidence of viral movement was from the United States to Vietnam (observed−expected=1.54) and from Vietnam to China (observed−expected=1.36), but the signals were weak.

When we mapped the positively selected sites onto the phylogenetic tree depicted in Fig. 2 (data not shown), the VP2 N426D mutation appeared to be co-circulating in several countries, such as South Korea, the United States and Taiwan, but was absent in all but four sequences collected in China (CPV2b-BD1.ch, CPV2b-BK044/07.ch, CPV2b-HN-3.ch and CPV2b-ZD1.ch), of which the first three clustered with sequences from Vietnam. Interestingly, the N426D mutation was not monophyletic, indicating that it has arisen independently in these countries. Only a single Brazilian sequence (BR185.br_85) had the N426D mutation, and this sequence had signatures of migration, since it clustered with the European and US isolates instead.

### Table 2. Analysis of selection pressures in each genomic region

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mean dS/dN</th>
<th>Mutation</th>
<th>dS</th>
<th>dN</th>
<th>dS/dN</th>
<th>SLAC</th>
<th>FEL</th>
<th>IFEL</th>
<th>REL</th>
<th>Location*</th>
<th>Potential function*</th>
</tr>
</thead>
<tbody>
<tr>
<td>NS1</td>
<td>0.15</td>
<td>L597H</td>
<td>10.47</td>
<td>0</td>
<td>∞</td>
<td>0.23</td>
<td>0.089</td>
<td>0.014</td>
<td>77</td>
<td>C terminus</td>
<td>P38 promoter transactivation</td>
</tr>
<tr>
<td>NS2</td>
<td>0.53</td>
<td>T94R</td>
<td>2.42</td>
<td>0.91</td>
<td>2.66</td>
<td>0.445</td>
<td>0.25</td>
<td>0.195</td>
<td>331</td>
<td>C terminus</td>
<td>CRM1 interaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N107R</td>
<td>2.42</td>
<td>2.12</td>
<td>1.14</td>
<td>0.602</td>
<td>0.76</td>
<td>0.26</td>
<td>136</td>
<td>C terminus</td>
<td>CRM1 interaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D151N</td>
<td>2.42</td>
<td>0.92</td>
<td>2.63</td>
<td>0.615</td>
<td>0.296</td>
<td>0.09</td>
<td>301</td>
<td>C terminus</td>
<td>14-3-3 family interaction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E160K</td>
<td>2.42</td>
<td>0.87</td>
<td>2.78</td>
<td>0.52</td>
<td>0.38</td>
<td>1</td>
<td>325</td>
<td>C terminus</td>
<td>Unclear</td>
</tr>
<tr>
<td>VP1</td>
<td>0.15</td>
<td>I101T†</td>
<td>2.24</td>
<td>0</td>
<td>∞</td>
<td>0.353</td>
<td>0.10</td>
<td>0.34</td>
<td>5617†</td>
<td>Threefold spike</td>
<td>Possibly TfR binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A300G</td>
<td>3.0</td>
<td>0</td>
<td>∞</td>
<td>0.089</td>
<td>0.189</td>
<td>0.361</td>
<td>6†</td>
<td>Threefold spike</td>
<td>TfR binding, antigenic change</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Y324I†</td>
<td>4.43</td>
<td>0</td>
<td>∞</td>
<td>0.506</td>
<td>0.54</td>
<td>0.286</td>
<td>504†</td>
<td>Threefold spike</td>
<td>Likely TfR binding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N426D/E</td>
<td>4.56</td>
<td>0</td>
<td>∞</td>
<td>0.043</td>
<td>0.255</td>
<td>0.12</td>
<td>6†</td>
<td>Capsid surface</td>
<td>Antigenic escape</td>
</tr>
<tr>
<td>FPV</td>
<td>0.11</td>
<td>V562L</td>
<td>13.05</td>
<td>0</td>
<td>∞</td>
<td>0.408</td>
<td>0.33</td>
<td>0.097</td>
<td>4704</td>
<td>Capsid surface</td>
<td>Unclear</td>
</tr>
<tr>
<td>CPV</td>
<td>0.11</td>
<td>M87L</td>
<td>5.03</td>
<td>0</td>
<td>∞</td>
<td>0.465</td>
<td>0.219</td>
<td>0.069</td>
<td>‡</td>
<td>Shoulder</td>
<td>TfR binding; possibly antigenic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N426D/E</td>
<td>4.55</td>
<td>0</td>
<td>∞</td>
<td>0.045</td>
<td>0.231</td>
<td>0.104</td>
<td>‡</td>
<td>Capsid surface</td>
<td>Antigenic escape</td>
</tr>
</tbody>
</table>

*Location and function of the NS1 and NS2 mutations were inferred from the known functions of homologous regions in minute virus of mice. Residues were mapped to the NS2 gene of minute virus of mice using a previously published alignment (Wang et al., 1998) and the functions listed correspond to those known in that virus.
†Only detected in the full-length dataset.
‡Containing too many sequences for analysis.
of the other Brazilian isolates. Geographical structure was even more striking in the case of the Y324I mutation. This was only present in sequences from China and South Korea and again was not monophyletic. Conversely, the VP2 mutations at codons 87, 101 and 300 were present in all but five CPV-2a sequences. Three of the Vietnamese sequences collected in 1997 from feral cats (CPV-V139.vi_97, CPV-V203.vi_97 and LCPV_V140.vi_97) and one sequence from South Korea (CPV-2c-Pome) harboured a G300D mutation (Ikeda et al., 2000) and one Chinese sequence isolated from a red panda in 2004 (CPV-RPPV.ch_104) possessed a G300D mutation. Finally, only one sequence in the dataset (HNI-4.1.ha.vi_102) contained the alternative substitutions of residue 300, suggesting a role of this residue in the feline adaptation of CPV. We have previously shown that an Asp at residue 300 is selected during passage of CPV-2 in feline cells in culture, and also previously shown that an Asp at residue 300 is selected during passage of CPV-2 in feline cells in culture, and also during passage of CPV-2 in feline cells in culture, and also that this prevents the infection of dog cells (Parker & Parrish, 1997; Parrish & Carmichael, 1986). The two positively selected mutations at VP2 residues 324 and 426 arose later, and apparently independently, in different geographical locations. Thus, the early period of CPV emergence was characterized by strong viral spread and global strain replacement, while later mutations likely arose through parallel emergence of the same mutations in separate geographical regions. Indeed, by far the strongest phylogeographic signal in these data is that of population subdivision, with relatively little movement of viruses between geographical areas. In addition, the majority of migrations that do occur are between countries in close geographical proximity, such as Taiwan or Vietnam and China, or between countries with close economic connections such as the United States and Germany or Italy, indicating that CPV is likely spread among countries through the movement of mechanical vectors or, in some cases, infected animals. The movement of viruses from the

### DISCUSSION

This study provides novel insights into the population dynamics of the emerging CPV virus, contrasting the evolutionary pressures acting on CPV and FPV, and among the genes of CPV. The emergence of CPV was characterized by strong selection for specific mutations within VP2, but not within NS1. Optimal receptor binding and antigenic escape were the likely driving forces, as indicated by the location and function of the positively selected sites in the capsid protein (located at VP2 codons 101, 300, 324 and 426).

This study is the first to systematically explore CPV dynamics on a global scale. Notably, we found the CPV population to be spatially structured, with very limited gene flow between clusters. This is in contrast with a previous study which reported strong temporal, but mild or absent spatial clustering, among CPV isolates from Brazil, the US and Europe (Pereira et al., 2000), but is consistent with other reports of geographical clustering of isolates from Vietnam, Japan/Taiwan and the United States (Doki et al., 2006). Three of the mutations at positively selected sites (at VP2 residues 87, 101 and 300) arose soon after CPV emerged. The mutations at residues 87 and 101 have arisen once and all but five of the CPV-2a viruses harbour the A300G mutation. The remaining CPV-2a viruses were isolated from feral feline hosts and show alternative substitutions of residue 300, suggesting a role of this residue in the feline adaptation of CPV. We have previously shown that an Asp at residue 300 is selected during passage of CPV-2 in feline cells in culture, and also that this prevents the infection of dog cells (Parker & Parrish, 1997; Parrish & Carmichael, 1986). The two positively selected mutations at VP2 residues 324 and 426 arose later, and apparently independently, in different geographical locations. Thus, the early period of CPV emergence was characterized by strong viral spread and global strain replacement, while later mutations likely arose through parallel emergence of the same mutations in separate geographical regions. Indeed, by far the strongest phylogeographic signal in these data is that of population subdivision, with relatively little movement of viruses between geographical areas. In addition, the majority of migrations that do occur are between countries in close geographical proximity, such as Taiwan or Vietnam and China, or between countries with close economic connections such as the United States and Germany or Italy, indicating that CPV is likely spread among countries through the movement of mechanical vectors or, in some cases, infected animals. The movement of viruses from the

### Table 3. Migration analysis of CPV, based on VP2 gene

<table>
<thead>
<tr>
<th>To/from</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>F</th>
<th>G</th>
<th>I</th>
<th>J</th>
<th>N</th>
<th>O</th>
<th>P</th>
<th>S</th>
<th>T</th>
<th>U</th>
<th>V</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>−0.15</td>
<td>NA</td>
<td>−2.71</td>
<td>−0.13</td>
<td>−0.31</td>
<td>1.10</td>
<td>−0.46</td>
<td>−0.14</td>
<td>−0.15</td>
<td>0.86</td>
<td>−1.25</td>
<td>−1.04</td>
<td>0.07</td>
<td>−2.08</td>
<td>−6.37</td>
</tr>
<tr>
<td>C</td>
<td>−0.26</td>
<td>−3.79</td>
<td>NA</td>
<td>−0.27</td>
<td>−0.48</td>
<td>−1.68</td>
<td>0.24</td>
<td>−0.28</td>
<td>−0.25</td>
<td>−0.24</td>
<td>−2.29</td>
<td>−1.85</td>
<td>−3.49</td>
<td>−3.67</td>
<td>−18.3</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>I</td>
<td>−0.01</td>
<td>−0.08</td>
<td>−0.06</td>
<td>0</td>
<td>−0.01</td>
<td>NA</td>
<td>−0.02</td>
<td>−0.01</td>
<td>−0.01</td>
<td>0</td>
<td>−0.04</td>
<td>−0.04</td>
<td>−0.06</td>
<td>−0.06</td>
<td>−0.39</td>
</tr>
<tr>
<td>J</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>−0.01</td>
<td>0</td>
</tr>
<tr>
<td>N</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>O</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>−0.01</td>
<td>−0.13</td>
<td>−0.12</td>
<td>−0.01</td>
<td>−0.02</td>
<td>−0.05</td>
<td>0.02</td>
<td>−0.01</td>
<td>−0.01</td>
<td>−0.01</td>
<td>NA</td>
<td>−0.07</td>
<td>−0.1</td>
<td>−0.12</td>
<td>−0.67</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>−0.08</td>
<td>0.92</td>
<td>−0.01</td>
<td>−0.01</td>
<td>−0.04</td>
<td>0.98</td>
<td>−0.01</td>
<td>−0.01</td>
<td>−0.04</td>
<td>−0.05</td>
<td>NA</td>
<td>−0.08</td>
<td>−0.05</td>
<td>1.58</td>
</tr>
<tr>
<td>U</td>
<td>−0.03</td>
<td>−0.52</td>
<td>−0.48</td>
<td>−0.04</td>
<td>0.95</td>
<td>0.80</td>
<td>−0.09</td>
<td>−0.03</td>
<td>−0.03</td>
<td>−0.02</td>
<td>−0.29</td>
<td>−0.24</td>
<td>NA</td>
<td>1.54</td>
<td>1.53</td>
</tr>
<tr>
<td>V</td>
<td>−0.04</td>
<td>−0.65</td>
<td>1.36</td>
<td>−0.03</td>
<td>−0.09</td>
<td>−0.27</td>
<td>−0.1</td>
<td>−0.05</td>
<td>−0.04</td>
<td>−0.04</td>
<td>−0.33</td>
<td>0.73</td>
<td>−0.53</td>
<td>NA</td>
<td>−0.08</td>
</tr>
<tr>
<td>Σ</td>
<td>−0.49</td>
<td>−4.25</td>
<td>−1.11</td>
<td>−0.49</td>
<td>0.03</td>
<td>−0.14</td>
<td>0.54</td>
<td>−0.51</td>
<td>−0.48</td>
<td>0.54</td>
<td>−4.25</td>
<td>−2.5</td>
<td>−4.20</td>
<td>−4.45</td>
<td>0</td>
</tr>
</tbody>
</table>

Observed minus expected migration pathways are shown below, with directional movements from country indicated in row to country shown in column. Likely migration events (positive values) are indicated in bold. NA, Not applicable. Country codes: A, India; B, Brazil; C, China; F, Finland; G, Germany; I, Italy; J, Japan; N, New Zealand; O, outgroup (1964 FPV isolate, UK); P, Poland; S, South Korea; T, Taiwan; U, USA; V, Vietnam.
United States to Vietnam appears surprising, but all of these sequences were isolated in 1997, only 3 years after the embargo against Vietnam was lifted, indicating a potential role of the developing trade relations between the countries.

While recombination can potentially impact phylogenetic relationships, we found no clear signal of recombination among the FPV or CPV isolates studied here (unpublished data; available from the authors on request). Indeed, the presence and role of recombination among FPV or CPV sequences is uncertain. While phylogenetic evidence of recombination has been detected for related parvoviruses of mink (Aleutian mink disease virus), swine (porcine parvovirus) and rodents (several rodent parvoviruses including LuIII) (Shackelton et al., 2007), there is no clear evidence for the process among the FPV- and CPV-related parvoviruses studied to date. It is therefore unlikely that the analyses reported here were strongly impacted by recombination.

Our molecular clock estimates of emergence times for CPV-2 and CPV-2a are notable in that they are very close to the dates when the viruses were first detected in nature. In particular, we estimate that CPV-2 emerged around 1974 (i.e. 32 years before the most recently analysed sequence was collected in 2006), only four years before it spread globally in 1978. This is compatible with serological evidence that the virus circulated in Europe a few years before the CPV pandemic (Truyen et al., 1994b). Similarly, our estimate for the emergence date of CPV-2a is approximately four years later than CPV-2, which corresponds closely with the first detection of CPV-2a in nature in 1979. Notably, the pattern of logistic population growth for CPV-2 and CPV-2a indicates the initially rapid spread of the virus among animals. This corresponds to the expected population dynamics of a newly emerging pathogen, and has also been reported by previous studies (Pereira et al., 2006; Shackelton et al., 2005). In marked contrast, FPV evolution is characterized by constant population sizes, as expected for an endemic pathogen.

The change in evolutionary dynamics associated with the process of emergence is also apparent in the significantly higher rate of substitution in the capsid protein gene of CPV compared to the FPV sequences, with mean rates of 1.2 and 0.4 substitutions per genome per year, respectively. Although these rates are comparable to those inferred previously (Horiuchi et al., 1998; Pereira et al., 2006; Shackelton et al., 2005), this is the first time a statistically significant difference between FPV and CPV has been revealed. In theory, the elevated rate of evolutionary change
in CPV might be due to either lower replication fidelity in dogs or increased positive selection after the transmission to the new host. Since these viruses are all replicated by host cell DNA polymerases, significantly different fidelities appear unlikely. Moreover, substitution rates inferred from the NS1 gene were comparable between FPV and CPV sequences, and similar to those inferred for FPV based on the capsid gene, indicating that the elevated substitution rate observed among the CPV sequences was most likely due to strong positive selection, probably driven by host adaptation and immune escape.

There is clear evidence for strong positive selection on the capsid protein gene, and differences among the host species again appear to be of importance in this regard. Four VP2 residues, 101, 300, 324 and 426, are subject to positive selection among all carnivore parvoviruses. Residues 101 and 300 are located on a surface-exposed region of the viral capsid (Supplementary Fig. S2), interact with the transferrin receptor (which is the receptor mediating viral entry), and partially determine the reacquired feline host range of CPV-2a. The substitution at residue 324 has not been previously characterized, but is adjacent to residue 323, a known host-range mutation, which together with residue 93 controls canine transferrin receptor (TFR) binding and the canine host range (Hueffer et al., 2003a). The 324 mutation appears to be circulating only in China and South Korea (data not shown), and has appeared a number of times independently. Although clear information on time of sampling is available for only a few of the sequences containing this mutation, most appear to have been collected in 2006 or 2007, indicating that this mutation probably arose recently. Mutations at aa 426 change the antigenic profile of the virus (Nakamura et al., 2004; Parrish et al., 1991). This codon has undergone two mutations since CPV emerged, first from asparagine to aspartate, and more recently to glutamate, indicating complicated selection dynamics acting on this residue. The exact phenotypic consequences of the mutations are not clear, and the N426D mutation has been present in some CPV populations for decades without replacing the wild-type (Parrish et al., 1991).

Mapping the codon 426 variants onto the phylogenetic tree in Fig. 2 revealed both the Asn and Asp amino acids were co-circulating in the United States, Taiwan, Vietnam and Europe, while the Asp mutation was at very low frequency in China and Brazil (data not shown; available from the authors on request). Interestingly, the Asp codon-containing sequences from those countries showed signatures of migration and clustered with sequences from other countries that carried the respective mutation. The D426E mutation has recently been isolated from several countries around the world, but only a single VP2 sequence of known origin that harboured this mutation was available, preventing more detailed analysis. One additional sequence containing the D426E mutation (GenBank accession number AF401519) was excluded from the analysis since the place of isolation could not be clearly determined. Among the CPV sequences, residues 87 and 426 are under strong positive selection. Residue 87 is also surface exposed and, although its specific functions are unknown, appears to be associated with the regained feline host range of CPV-2a, together with residue 300.

Less positive selection was observed in FPV, and only VP2 codon 562 appeared to be under positive selection, although relatively strongly. Adaptive evolution at this site has not been described previously. However, two residues in close spatial proximity, 564 and 568, partially determine the in vivo host range of FPV (Truyen et al., 1994a).

Finally, our study is notable in that we detected putatively positively selected sites in the NS1 (codon 597, overlapping codon 107 in NS2) and NS2 (codons 94, 107,151 and 160) genes, although these are potentially false positives due to the overlapping reading frames. The role of NS2 in the CPV life cycle is unclear and knockout mutants appear to behave similarly to wild-type both in tissue culture and in dog infections (Wang et al., 1998). However, in the related rodent parvoviruses, NS2 is required for efficient translation and capsid assembly in cells from natural host species, but not in cells from other hosts (Choi et al., 2005; Cotmore et al., 1997; D’Abramo et al., 2005; Eichwald et al., 2002; Li & Rhode, 1991; Naeger et al., 1993). A possible role for NS2 in host adaptation therefore merits further study.

ACKNOWLEDGEMENTS

Leland E. Carmichael and Edward Dubovi provided clinical samples, and Wendy S. Weichert and Nell Bond provided excellent technical support. Christian Nelson provided valuable help with the three-dimensional representation of the positively selected capsid residues. This work was supported by NIH grants GM080533 to E.C.H. and AI028385 to C. R. P.

REFERENCES


amino acid change require additional silent mutations that regulate NS2 accumulation. Virology 340, 143–154.


