Comparative analysis reveals frequent recombination in the paroviruses

Laura A. Shackelton,1 Karin Hoelzer,2 Colin R. Parrish2 and Edward C. Holmes1,3

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

Paroviruses infect a diverse array of hosts. These small DNA viruses can be autonomous or depend upon a larger helper-virus for replication. Among the paroviruses that infect vertebrate hosts there are currently five recognized genera. Members of the Parovirus, Bocavirus, Amdovirus and Erythrovirus genera are all autonomous and can be found in humans, dogs, cats, cows, pigs, rodents and a number of other mammals. Indeed, the recent emergence of canine parovirus (CPV2) from feline parovirus (FPV) constitutes a valuable model of a successful cross-species transmission event (reviewed by Parrish & Kawaoka, 2005). Analyses of nucleotide substitution rates have also suggested that paroviruses evolve far more rapidly than other (larger) DNA viruses, although the reasons for this remain unclear (Shackelton et al., 2005; Pereira et al., 2007).

The severity of parovirus disease varies greatly. Some paroviruses, such as human B19 virus and the rodent paroviruses, usually cause mild disease (Young & Brown, 2004; Jacoby et al., 1995). Animals infected with Aleutian mink disease parovirus (ADV), however, may show a variety of symptoms, ranging from subclinical to chronic disease and death, with more severe disease in susceptible genotypes (Jackson et al., 1996). Porcine parovirus (PPV) causes little clinical disease in adults, but infection of pregnant females results in infection of the immunologically tolerant fetus, leading to reproductive complications such as fetal death and abortion (Brown et al., 1980; Mengeling & Cutlip, 1975).

Paroviral genomes are composed of a linear single-stranded segment of DNA approximately 5 kb in length. The two primary ORFs, the 3' nonstructural ‘NS’ ORF and the 5’ structural ‘VP’ ORF each encode at least two proteins. Hairpin structures, necessary for priming replication, are found in the non-translated regions at both ends of the genome, bordering the ORFs. Genomic replication takes place in the nucleus of actively dividing cells and utilizes cellular machinery, including subunits of the cellular polymerase. As the virus cannot induce mitosis, it replicates primarily in rapidly dividing cells, which may be in a variety of tissues depending on the age of the host and the tropism of the virus. Replication requires a double-stranded template and occurs through a rolling hairpin mechanism where the imperfect palindromes located in the 3’ and 5’ ends of the genome are used to assemble concatameric virus templates. Nicking by the viral NS1 protein and strand exchange are involved in the resolution of dimeric or tetrameric DNA replication intermediates (Cotmore & Tattersall, 2003).

The antibody-mediated immune response against paroviruses appears to be quite effective in many hosts and in the case of CPV2, FPV, PPV and B19 the virus is generally cleared within a few days of the response developing. However, other paroviruses, including ADV and probably many rodent paroviruses, are not cleared despite the
Fig. 1. PPV phylogenies. Separate phylogenetic trees were inferred for nucleotides (a) 1–562, (b) 563–1729 and (c) 1730–2620 of the alignments. Viruses isolated from Germany in 2001, which generally form a tight clade, are outlined in black. Viruses isolated from Germany in 2002, along with a vaccine strain from 1964, which consistently form a clade, are labelled in light grey. Isolate 225b is in dark grey. Nodes with \( \geq 70\% \) bootstrap support are shown and branch lengths are scaled according to nucleotide substitutions per site.
strong immune response that they elicit (Alexandersen et al., 1988; Porter, 1986). For example, some rodent parvoviruses persist in the kidneys and are shed in the urine over long periods (Jacoby et al., 1996). In the case of B19, persistent viral DNA can be detected by PCR years after the initial infection, although it is not clear whether the virus is still replicating over that period (Manning et al., 2007; Lefrere et al., 2005). Finally, persistent viral shedding has also been found in PPV infections (Guérin & Pozzi, 2005).

While there is a growing body of work on the mechanisms of replication and the dynamics of evolutionary change in parvoviruses, little is known about the occurrence or characteristics of parvoviral recombination in nature. However, parvovirus replication would appear to provide opportunities for recombination if a cell is co-infected by two different genomes. Indeed, the ubiquitous nature of the parvoviruses means that co-infection may be commonplace, and in many cases DNA from more than one virus strain is seen in an individual animal (Norja et al., 2006), allowing for the possibility of recombination.

Recombination is an important evolutionary mechanism in many virus families, having the potential to combine beneficial mutations within a single genome, and similarly, to decouple advantageous mutations from deleterious genomic baggage (Awadalla, 2003). Recombination has been observed for a number of DNA virus families, including the herpesviruses, the poxviruses, the single-stranded DNA geminiviruses and the anelloviruses (Fleischmann, 1996; Thiry et al., 2005; Bugert & Darai, 2000; Monci et al., 2002; Hino & Miyata, 2007). However, neither the frequency nor determinants of recombination among the parvoviruses are known. During mixed infection in cell culture recombination can be seen among parvovirus genomes. This can generate altered forms that may give rise to replicating genomes, as seen in studies using mixtures of DNAs for the production of gene therapy vectors derived from adeno-associated viruses (AAVs) or some autonomous parvoviruses (Allen et al., 1997; Brandenburger & Velu, 2004). Furthermore, extensive recombination was inferred as the explanation for the presence of mosaic genomes that arose during the passaging of a CPV2 strain in tissue culture (Badgett et al., 2002). There is, however, less evidence for recombination among parvovirus genomes recovered from natural infections, although an analysis of mice, hamster, and LuIII parvovirus genomes that examined the phylogenies of ORFs 1 and 2 separately revealed different evolutionary histories for each ORF, indicative of recombination (Lukashov & Goudsmit, 2001). Similarly, incongruencies were observed in the phylogenetic trees of ORFs 1 and 2 of the AAVs, with respect to AAV2 and AAV4 (Lukashov & Goudsmit, 2001).

Herein, we analyse three groups of autonomous parvoviruses for evidence of recombination either within or among species. We focus on PPV, ADV and the rodent parvoviruses as there is a substantial amount of information available about these viruses, along with relatively long and diverse nucleotide sequences which facilitates the in silico examination of recombination frequency. Viral sequences were compiled from GenBank and aligned by eye or with Muscle (Edgar, 2004; sequence alignments are available as supplementary information with the online version of this paper). Isolate names, year, and location of isolation, where available, were provided either on GenBank or by Zimmermann et al. (2006). GenBank accession numbers are as follows. PPV: VRI-1, AY390557; Challenge, AY684866; SR-1, DQ675456; Kresse, U44978; PPV_NADL-2, NC_001718; China-318, AY583318; 15a, AY684865; 27a, AY684871; 21a, AY684868; 225b, AY684864; PPV/Tornau, AY684869; 143a, AY684867; vaccine_IDT, AY684872; 106b, AY684870. ADV: V3, DQ630715; V9, DQ630716; M15, DQ630719; M19, DQ630721; M21, DQ630722; Far East, DQ371395; TH5, AF124791; TR, AMU39013; Pullman, AMU39014; Utah 1 kit, AMU39015; MDMV-T, M63044; MDMV-T7, M63045; Utah 1, M32981; SL-3, X97629; ADV-G, NC_001662. Rodent parvoviruses (using standardized nomenclature; Besselsen et al., 2006): LuIII virus (LuIII), NC_004713; Minute virus of mice `prototype' (MVMP), NC_001510; MVM immunosuppressive variant (MVMi), X02481; MVMm, DQ196317; MVMc, MVU34256; mouse parvovirus 1a (MPV-1a), NC_001630; MPV-1e, DQ989166; MPV-1c, MOU34254; MPV-1b, MOU34253; MPV-2, NC_008186; MPV-3, NC_008185; hamster parvovirus (HaPV), HOU34255.

To screen for recombination we employed two preliminary detection programs: RDP2 (Martin et al., 2005) and Genetic Algorithms for Recombination Detection (GARD; Kosakovsky Pond et al., 2006). The former includes six separate recombination detection programs—Bootscan, Chimeric, GENECONV, MaxChi, RDP, and SiScan—which were employed with their default parameters (and the following general options: window size of 20, linear sequences, Bonferroni correction, finding consensus daughter sequences, and polishing breakpoints). General parameter settings were used for GARD (GTR model of nucleotide substitution and Beta-Gamma rate variation with 3 rate classes). To exclude the possibility of false-positive recombination detection, putative recombinant regions were considered only if three or more different programs detected recombination within the same general region of the alignment. We then separated the alignments at a point within the region, estimated maximum-likelihood (ML) phylogenetic trees for the individual sections, and compared their evolutionary histories. Modeltest (Posada & Crandall, 1998) was used to determine the most appropriate models of nucleotide substitution, which were then used as the basis of phylogenetic inference using the ML method available in paup* (Swofford, 2003). The selected model (TVM + i in the case of the PPV and GTR + i + I in the case of ADV and the mouse/hamster/LuIII parvoviruses) was used in the ML
Fig. 2. ADV phylogenies. Separate phylogenetic trees were inferred for nucleotides (a) 1–542, (b) 543–1050, and (c) 1051–1533. The two groups that show significant changes in topology, with respect to each other, are shaded in light and dark grey. Nodes with $\geq 70\%$ bootstrap support are shown and branch lengths are scaled according to nucleotide substitutions per site.
analysis of all sections of the alignment. In each case, support for tree nodes was determined with neighbour-joining bootstrap resampling (based on 1000 replicates). Only nodes with $\geq 70\%$ bootstrap support were considered.

For PPV we constructed a 2620 bp alignment (corresponding to base pairs 1990–4609 of reference genome NC_001630 positions 144–4728), which includes a C-terminal part of the NS1 gene sequence and the entire VP1 ORF. Putative recombination regions were located around sites 563 and 1730 of our alignment. Phylogenetic trees of the sections bordered by these points did, indeed, show significant phylogenetic incongruence (Fig. 1). In particular, isolate 225b, collected by Zimmermann et al. (2006) in Germany, appears to be a recombinant between two distinct genetic groups circulating in that country. In addition, isolate 15a may have been involved in a recombination event.

An adequate alignment could only be obtained for a 1533 bp portion of the ADV VP2 gene (reference genome NC_001662 positions 2643–4175). Putative recombination regions were located near nucleotides 543 and 1051 of our alignment. Cutting the alignment at these points and constructing separate phylogenies for each section again resulted in incongruent tree topologies (Fig. 2). Although the bootstrap support for many of the nodes is below 70%, there is sufficient resolution to show significant incongruencies, especially with regard to clades TH5/SL-3/ADV-G/TR and Utah 1/MDPMVT6/Utah 1 kit/Far East.

Almost full-length rodent parvovirus genome sequences were analysed for recombination, and preliminary phylogenetic comparisons suggested frequent recombination among the mouse parvoviruses (MPVs), Minute viruses of mice (MVMs), LuIII virus and hamster parvovirus (HaPV). Therefore, we focused on these viruses and not the rat parvoviruses, which did not provide clear evidence of recombination. The former viruses were aligned (reference genome NC_001630 positions 144–4728) and analysed as described above. A number of putative breakpoints were estimated, some of which were located in regions around positions 855 and 2227 of our alignment. Phylogenies of the sections bordered by these positions show several topological inconsistencies (Fig. 3). For example, among other differences, HaPV and MPV-3 are very closely related in all but the first part of their genomes; most MPVs show relatively close relationships with some MVMs in the first part of their genomes but the two groups are clearly distinct in the second half of their genomes, and LuIII shows unique relationships to each of these groups in different genome regions.

While the complexity of recombination in these viruses means that we are unlikely to have identified exact breakpoints of recombination, we have demonstrated that different sections of the genomes have conflicting phylogenetic histories, indicating that recombination may be a relatively common event within and/or among these parvoviral species. Given the differences in epidemiology, cell tropism, pathogenicity and host species, many of which are not fully understood, the specifics of recombination and the evolutionary forces acting on any recombinants are likely to differ among these and other parvoviruses.

The initial requirements for recombination include coinfection of the same animal and host cell. High seroprevalences of MVMs and MPVs are reported for wild mice populations, and in many cases mice go on to develop persistent infections with chronic shedding of the virus from the kidneys (Becker et al., 2007; Jacoby et al., 2000). This suggests the potential for frequent co-infection, although the exact frequencies or likelihood of mixed infections are not known. High population densities and contact frequencies of rodents may explain, in part, why evidence of recombination has been readily observed both here and previously (Lukashov & Goudsmit, 2001) among different viral species infecting these populations.

In humans, the frequent detection of persistent erythroviral DNA and multiple genotypes also suggests co-infection and opportunities for recombination (Hokynar et al., 2002; Parsyan et al., 2007). Likewise, multiple strains of AAV infect humans and more than one genotype can be found in an individual (Gao et al., 2004). In experimental studies, the viral sequences recovered from AD-affected mink contained multiple genotypes, suggesting that mixed infections may be common (Gottschalck et al., 1994, 1991), although the degree of sequence variation and population structures of the viruses in natural infections have not been examined in detail. Little is known about the epidemiology of PPV, the likelihood of multiple infections, or the possible selection pressures acting on recombinant viruses. However, high densities and contact rates as well as crowding effects in intensively reared pig populations, along with partial herd immunity, might provide ready opportunities for recombination. The recombination of the PPV genome seen in this study involves viruses that differ antigenically, suggesting that immune pressure may have played a role in the emergence of the recombinant strain (Zimmermann et al., 2006; Zeeuw et al., 2007). It seems that all of the viruses examined here may, under certain circumstances, persist in the host. As this may increase the likelihood of recombination, it will be informative to examine the occurrence and relative frequency of recombination in those acute parvoviruses which are cleared by the host within a few days.

Parvoviruses have shown an ability to emerge in new hosts (Parrish & Kawaoka, 2005). Recombination or segmental reassortment has been reported for a number of other viruses during the processes of shifting host ranges; whether recombination would increase the likelihood of the emergence of parvoviruses in new hosts must be addressed in more depth. This uncertainty notwithstanding, our analysis suggests that, along with relatively rapid rates of sequence variation, co-infection and subsequent recombination may be important forces in the natural evolution of parvoviruses. As population densities and contact frequencies between and among humans and wild or agricultural animals continue to increase, it seems
Fig. 3. Rodent parvovirus phylogenies. Phylogenetic trees were inferred for nucleotides (a) 1–854, (b) 855–2226 and (c) 2227–4654. Groups of viruses are shaded or outlined for ease of visualization: MVMs are light grey; HaPV and MPV-3 are dark grey; all MPVs except MPV-3 are circled in black; and LuLL virus is unmarked. Nodes with ≥70 % bootstrap support are shown and branch lengths are scaled according to nucleotide substitutions per site.
inevitable that recombination will be an important factor, among others, in the emergence of new viral genotypes and species.

**Acknowledgements**

This work was supported by NIH grant number GM080533-01.

**References**


