Insights into the Evolutionary History of an Emerging Livestock Pathogen: Porcine Circovirus 2\(^{V}\)†

Cadhla Firth,1,† Michael A. Charleston,2 Siobain Duffy,1 Beth Shapiro,1 and Edward C. Holmes1,3

Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, Pennsylvania 168021; School of Information Technologies, Sydney Bioinformatics, and Centre for Mathematical Biology, University of Sydney, Sydney, New South Wales 2006, Australia2; and Fogarty International Center, National Institutes of Health, Bethesda, Maryland 208923

Received 16 August 2009/Accepted 29 September 2009

Porcine circovirus 2 (PCV2) is the primary etiological agent of postweaning multisystemic wasting syndrome (PMWS), one of the most economically important emerging swine diseases worldwide. Virulent PCV2 was first identified following nearly simultaneous outbreaks of PMWS in North America and Europe in the 1990s and has since achieved global distribution. However, the processes responsible for the emergence and spread of PCV2 remain poorly understood. Here, phylogenetic and cophylogenetic inferences were utilized to address key questions on the time scale, processes, and geographic diffusion of emerging PCV2. The results of these analyses suggest that the two genotypes of PCV2 (PCV2a and PCV2b) are likely to have emerged from a common ancestor approximately 100 years ago and have been on independent evolutionary trajectories since that time, despite cocirculating in the same host species and geographic regions. The patterns of geographic movement of PCV2 that we recovered appear to mimic those of the global pig trade and suggest that the movement of asymptomatic animals is likely to have facilitated the rapid spread of virulent PCV2 around the globe. We further estimated the rate of nucleotide substitution for PCV2 to be on the order of \(1.2 \times 10^{-3}\) substitutions/site/year, the highest yet recorded for a single-stranded DNA virus. This high rate of evolution may allow PCV2 to maintain evolutionary dynamics closer to those of single-stranded RNA viruses than to those of double-stranded DNA viruses, further facilitating the rapid emergence of PCV2 worldwide.

Livestock species are an increasingly important focus of emerging-disease research. Livestock are susceptible to infections from related wild species and can act as a conduit through which zoonotic diseases spill over into human populations. For example, contact between humans and livestock has been implicated in the emergence of pandemic influenza virus (25), Ebola-Reston virus (53), and *Chlamydia psittaci* (4). Pigs in particular are potentially important reservoirs for emerging human disease and have been implicated in the recent emergence of pandemic H1N1 influenza A virus (66), among others (3, 8, 53). Characteristics of the modern pig industry include high-density farming and expanding global trade, a combination that favors increased transmission and spread for many infectious agents. The role of swine in emerging infectious diseases may warrant further scrutiny, considering the expansive array of pathogens to which pigs are host, including a wide range of viruses: arboviruses, circoviruses, flaviviruses, herpesviruses, nidoviruses, orthomyxoviruses, paramyxoviruses, and picornaviruses all cause common infections in swine (52). Given the number of emergent pathogens currently identified in swine, it is essential to understand the circumstances under which these pathogens emerge and evolve. In this study, we examine the evolutionary dynamics of one of the most economically important emerging pig pathogens, porcine circovirus 2 (PCV2), the primary etiological agent of postweaning multisystemic wasting syndrome (PMWS).

PCV2 is a single-stranded, nonenveloped, circular DNA virus with an ambisense genome of only \(~1.76\) kb, making it the smallest autonomously replicating virus. The genome of PCV2 contains at least three open reading frames (ORFs) with known functions: ORF1 codes for two replicate proteins (rep gene products), ORF2 for the structural protein (cap gene product), and ORF3 for a protein implicated in viral pathogenesis (44, 45, 48). PCV2 is highly infectious, and transmission can occur through oronasal, fecal, urinary, sexual, and vertical routes (38, 40, 62, 64). The current prevalence of PCV2 in pigs has been estimated to be as high as 40 to 60%, with studies reporting up to 100% of farms infected in a given region (21, 58, 60, 65, 68, 73). PCV2 has been identified as the causative agent of PCV-associated diseases (PCVAD), a term recently used to describe all PCV2-associated clinical and subclinical manifestations, including those formerly considered to be PMWS (systemic infection) as well as porcine dermatitis and nephropathy syndrome and PCV2-associated pneumonia, enteritis, and reproductive failure (7, 55). However, clinical diagnosis of PCVAD remains rare, and the vast majority of research on PCV2 has centered on the causes and consequences of PMWS.

The majority of PCV2 infections (and subsequent PMWS disease symptoms) appear to occur shortly after weaning, when the protective effects of maternal immunity have waned (51, 59). The transition from infection with PCV2 (often asymptomatic) to a clinical diagnosis of PMWS is thought to depend on several factors, of which viral load and immune activation (potentially caused by early vaccination or infection with multiple pathogens) appear to be the most significant (35, 36, 37).
It has been suggested that there is a viral genetic component to the observed variation in virulence of PCV2, although this is still controversial (15, 18, 70). It is possible that the emergence of a new genotype (PCV2b) may be related to the appearance of highly virulent PCV2, although this link has yet to be confirmed (55). Symptoms of PMWS most commonly include wasting/weight loss, histologic lesions (particularly in the lymphatic system or lungs), and early mortality or abortion (67). Morbidity due to PMWS on an affected farm is often 5 to 15%, with a case fatality rate of close to 100% (18, 31).

PMWS was initially identified in Canada in 1991, with the first large outbreaks occurring in Europe in the late 1990s (9, 20, 30). Retrospective studies have now identified both antibodies to PCV2 and (in some cases) viral RNA dating back as far as 1969, indicating that PCV2 was present at least 22 years before the emergence of PCVAD (21, 47, 60). Outbreaks of PMWS and other PCVAD have now been confirmed in nearly every region with industrial pig farming (2, 11, 12, 17, 18, 21), and the incidences of all PCVAD have been steadily rising, particularly since 2004 (9, 55, 58). It remains unclear what event(s) precipitated the recent worldwide emergence of PCVAD (and PMWS in particular) from PCV2, despite the high level of scientific and economic interest directed toward this cause.

To address some of the uncertainty surrounding the recent emergence of PCVAD and PCV2, the present study employed phylogenetic inference to address the following key questions. (i) Is PCV2 a newly emergent pig virus, or is it much older? (ii) What is the rate and time scale of the evolution of PCV2? (iii) Are there viral genetic factors that are associated with virulence in PCV2? (iv) How did severe PCVAD emerge in both North America and Europe within a 10-year interval?

### MATERIALS AND METHODS

**Evolutionary relationships and time scale.** For the PCV2-specific analyses, all available PCV2 full-genome, rep, or cap gene sequences for which information on the sampling year was available were downloaded from GenBank (Table 1). Sequences were manually aligned by gene or genome by using Se-Al (version 2.0a11 Carbon) (http://tree.bio.ed.ac.uk/software/seal/) and examined for evidence of recombination using the RDP, GENECONV, Bootscan, MaxChi, and LARD methods with default parameters, as implemented in RDP3 (49). All evidence of recombination using the RDP, GENECONV, Bootscan, MaxChi, and LARD methods with default parameters, as implemented in RDP3 (49). All potential recombinants identified by two or more methods within RDP3 were removed from further analyses. Phylogenetic trees were created for both the PCV2 genomes and the cap genes by using a Bayesian Markov chain Monte Carlo (MCMC) method implemented in the BEAST package (version 1.4.8) (13), which incorporates time-of-sampling information and returns rooted trees. BEAST was also used to estimate the rate of evolution and the time to the most recent common ancestor (TMRCA) of PCV2 for both the full-genome data set and each ORF separately (as ORF3 is completely within ORF1, the substitution rates for ORF1 were analyzed both with and without inclusion of the ORF3 region). A relaxed molecular clock with an uncorrelated log-normal distribution

<table>
<thead>
<tr>
<th>Genomic region</th>
<th>No. of samples</th>
<th>Sequence length (nt)</th>
<th>Yr range</th>
<th>No. of nucleotide substitutions/site/yr (95% HPD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full genome</td>
<td>160</td>
<td>1,780</td>
<td>1980-2007</td>
<td>$1.21 \times 10^{-3}$ (8.23 $\times 10^{-4}$ to 1.61 $\times 10^{-3}$)</td>
</tr>
<tr>
<td>Coat protein</td>
<td>327</td>
<td>693</td>
<td>1980-2007</td>
<td>$1.44 \times 10^{-3}$ (1.01 $\times 10^{-3}$ to 1.90 $\times 10^{-3}$)</td>
</tr>
<tr>
<td>Replicase</td>
<td>168</td>
<td>942</td>
<td>1980-2007</td>
<td>$1.03 \times 10^{-3}$ (7.37 $\times 10^{-4}$ to 1.32 $\times 10^{-3}$)</td>
</tr>
<tr>
<td>Replicase-ORF3</td>
<td>168</td>
<td>627</td>
<td>1980-2007</td>
<td>$8.74 \times 10^{-4}$ (5.94 $\times 10^{-4}$ to 1.15 $\times 10^{-3}$)</td>
</tr>
<tr>
<td>ORF3</td>
<td>168</td>
<td>312</td>
<td>1980-2007</td>
<td>$1.32 \times 10^{-3}$ (7.94 $\times 10^{-4}$ to 2.16 $\times 10^{-3}$)</td>
</tr>
<tr>
<td>Genome (big clade)</td>
<td>122</td>
<td>1,780</td>
<td>1998-2007</td>
<td>$1.15 \times 10^{-3}$ (8.34 $\times 10^{-4}$ to 1.45 $\times 10^{-3}$)</td>
</tr>
<tr>
<td>Genome (small clade)</td>
<td>38</td>
<td>1,780</td>
<td>1980-2005</td>
<td>$1.64 \times 10^{-3}$ (9.43 $\times 10^{-4}$ to 2.40 $\times 10^{-3}$)</td>
</tr>
</tbody>
</table>
following genomic fragments (coding regions only): cap, rep, rep-ORF3, and ORF3. The HyPhy package was used to fit a codon model to each alignment by using the MG94xHKY85_3x4 substitution model, and the global dN/dS ratio and corresponding 95% confidence interval (CI) were estimated (34). If there was strong evidence that dN was >dS for any genomic region, the per-site dN/dS ratio was then estimated using the fixed-effect-likelihood algorithm implemented in HyPhy (32). A dN/dS ratio of >1 for any given site was considered significant if the P value derived from the likelihood ratio test was greater than 0.1.

A lineage-specific genetic algorithm approach was also used to infer the presence of positive selection along the branches of the Bayesian maximum clade credibility tree (full genome only) derived from BEAST, implemented in HyPhy (GABranch) (33). This approach assigns each branch to an incrementally estimated number of classes of dN/dS ratios without requiring a specification of the branches a priori and is less parameterized than a fully local model. Due to the computational intensity of this approach, the genome tree was pruned from 160 taxa down to 98 by arbitrarily removing all but one sequence from any monophyletic clade containing samples from any one outbreak/study.

PCVAD were discovered in both North America and Europe within a very short time frame, potentially as the result of high levels of migration of PCV2-infected pigs between disparate geographic regions. Here, we inferred the strength of PCV2 movement between geographic locations by using a geographically explicit Bayesian MCMC method implemented in BEAST (version 1.5) (43). This method simultaneously estimates a reversible diffusion rate matrix between previously defined locations along with the evolutionary and coalescent parameters. The diffusion pathways were estimated on two levels: by continent (Asia, Australia, Europe, North America, and South America; 10 possible diffusion pathways) and by country (Australia, Brazil, Canada, China, Taiwan, Denmark, France, Spain, Germany, Greece, Hungary, The Netherlands, and the United States; 55 possible diffusion pathways). Bayesian stochastic variable search selection was used to identify links between these locations that were necessary to explain the geographic diffusion process along the posterior sets of trees. Summary statistics based on a Bayes factor (BF) test were used to determine which diffusion links were statistically significant over the whole tree.

RESULTS

Evolutionary relationships and time scale. The phylogenetic trees inferred using tip-dated sequences resulted in very similar general topologies for the PCV2 genome and cap gene data sets, as previously noted (54). However, the increased sequence length in the genome data set (1,780 versus 693 bp) did result in a tree with substantially higher Bayesian posterior probabilities (BPPs) at the internal nodes (Fig. 1 for the genome tree; see Fig. S5 in the supplemental material for the cap tree). In both the cap and the genome trees, the two large clades (designated PCV2a and PCV2b) seen in previous phylogenetic analyses of PCV2 were recovered with high nodal support (7, 10, 42, 70). The majority of taxa fell within clade PCV2b, which may be replacing the older PCV2a genotype on a global scale (70). Notably, the PCV2b clade contained larger proportions of sequences both from recent outbreaks (post-1998) and from PMWS-affected animals (Fig. 1). Although both the PCV2a and the PCV2b clades contained taxa from Asia, Europe, and North America, the South American sequences were restricted to clade PCV2b and the Australian sequences to PCV2a. The majority of relationships within PCV2a had high nodal support, as did the deeper relationships within PCV2b (BPP > 0.9) (Fig. 1). The associations at the tips of the PCV2b clade were not significantly supported in the majority of cases, although small geographically structured clusters were resolved throughout the tree (Fig. 1).

The estimated age of the current diversity of PCV2 and the corresponding rate of nucleotide substitution were determined using a Bayesian MCMC analysis. The posterior distribution of the rate and TMRCA parameters from these analyses were compared with those from the data sets generated by randomizing the sequence-date relationships of all taxa as well as those from analyses using priors only. The parameter values estimated from the actual analyses had nonoverlapping 95% HPDs with all analyses from both the randomized data sets and the prior-only data sets. These results suggest that sufficient temporal structure was present in the data to estimate both the rates of nucleotide substitution and the TMRCA. After examining the distinct two-clade structure of the PCV2 phylogeny, rates were also estimated independently for the PCV2a and PCV2b clades. The mean substitution rate for the full-genome data set was 1.21 × 10^(-3) nucleotide substitutions per site per year, with a 95% HPD that ranged from 8.23 × 10^-5 to 1.61 to 10^-3 substitutions/site/year. This rate was consistent with those estimated independently for each genomic segment (Table 1). The TMRCA estimated for the current diversity of all PCV2 sampled viruses was 70 years before present (ybp) (95% HPD = 36 to 115 ybp), while the TMRCA values for the PCV2a and PCV2b genotypes were estimated at 41 ybp (95% HPD = 24 to 62 ybp) and 18 ybp (95% HPD = 12 to 27 ybp), respectively. Analyzing the population dynamics of PCV2a versus those of PCV2b also revealed genotype-specific differences, although neither Bayesian skyline plot demonstrated a significant departure from a constant level of genetic diversity (see Fig. S6 in the supplemental material). Under a model of neutral evolution, this is equivalent to a constant population size through time. However, there was a trend toward a recent increase in genetic diversity in PCV2b that was not mirrored in the Bayesian skyline plot of PCV2a. In addition, the relative genetic diversity estimated for both genotypes was extremely low (see Fig. S6 in the supplemental material).

Two competing hypotheses exist regarding the origin and emergence of PCV2 and PCVAD: either PCV2 is part of a family of viruses that have coevolved with their vertebrate hosts over millions of years (29) or it has recently emerged through cross-species transmission and/or host expansion. To distinguish between these hypotheses, we assessed the topological congruence between the circovirus and associated host phylogenies. The inferred phylogeny for both the circovirus and the host phylogenies were well resolved with high nodal support (Fig. 2). Phylogenetic reconciliation analysis of the host and virus trees returned 15 POpt solutions, containing three, four, or five codivergent nodes (6, 8 or 10 CEs) (see Fig. S7 in the supplemental material). Significance testing based on the maximum number of CEs revealed that 74% of randomly generated trees could be mapped onto the host tree with at least 10 CEs, revealing no statistically significant evidence for congruence by this measure. It was possible to generate significant congruence between the host and virus phylogenies by invoking a CE at the root of the vertebrate tree (see the supplemental material). However, the presence of a CE between circoviruses and the last common ancestor of birds and mammals is highly implausible, given the age of this split in the vertebrate tree, estimated to have occurred over 300 million years ago (6, 56). To test for consistency between the divergence times of each circovirus species with those of their hosts, the TMRCA values for a range of avian circoviruses and PCVs were estimated based on the rate of amino acid substitution (Fig. 2). The TMRCA for PCV2 was estimated to be more recent under the amino acid substitution model than when
Figure Legend

- diseased
- healthy
- unknown health status

- Australia
- Brazil
- Canada
- China
- Denmark
- France
- Germany
- Greece
- Hungary
- Netherlands
- South Africa
- Spain
- USA

* $d_n/d_s > 1$

FIG. 1. Bayesian maximum clade credibility phylogenetic tree of the full-genome sequences of PCV2. Tips are colored by the country of sampling, and the internal branches are colored based on the reconstructed ancestral state (country) as estimated by the reversible diffusion model. Tip times reflect the sampling time of the sequence in question (x axis). Numbers at the nodes indicate BPPs greater than 0.9. Disease status (when reported) is associated with each tip. The location of the branch leading to the primary PCV2b clade with a $d_n/d_s$ ratio of $>1$ is indicated with an asterisk (*).
nucleotides were used (38 ybp; 95% HPD = 14 to 69 ybp); and as tip-dated estimates of divergence times may be too recent as a whole (23), we considered only the upper 95% HPD of these TMRCA estimates for a conservative comparison with the host divergence times (Fig. 2). Despite this, our estimated divergence times for the circoviruses were very recent, with a mean TMRCA for PCV1 and PCV2 of 102 ybp and the mean divergence of pig and avian circoviruses estimated to have occurred 491 ybp.

To look for the presence of selection events that may have occurred only once in the evolutionary history of PCV2, $dN/dS$ was also calculated along individual branches of the genome tree. Four $dN/dS$ categories optimally described the selection pressure throughout the tree (based on Akaike Information Criterion scores; data not shown). Three of these had $dN/dS$ ratios of substantially less than 1, which accounted for 95% of the branches (see Fig. S8 in the supplemental material). The remaining $dN/dS$ category had a $dN/dS$ ratio of 15.4 and best fit 5% of the branches. Critically, while many of the branches associated with $dN/dS$ of >1 were near the tips of the tree (see Fig. S8 in the supplemental material) and therefore of little interest due to the influence of transient deleterious mutations, there was 99% support for $dN/dS$ of >1 on one of the primary branches leading to the PCV2b clade (Fig. 1). This branch was associated with an amino acid transition from threonine to proline at position 151 of the cap gene, present in the majority of taxa in the PCV2b clade.

Two phenomena may explain the nearly synchronous emergences of PCVAD in North America and Europe: either the emergence of PCVAD occurred independently in multiple regions or virulent PCV2 emerged only once and spread rapidly throughout the globe via the pig trade. When the diffusion pathways for PCV2 were estimated between continents, three were found to be significant: (in order of increasing significance) those between South America and Europe, Asia and North America, and Europe and Asia (BF > 12) (see Table S2).
in the supplemental material). When the diffusion pathways were estimated between countries, three pathways were again significant: those between France/Spain and Hungary, Canada and the United States, and France and The Netherlands (BF > 17) (see Table S2 in the supplemental material). The reconstruction of the ancestral locations (by country) of the internal nodes on the PCV2 genome tree is shown in Fig. 1.

DISCUSSION

Macropovolution of PCV2. This study provides a unique insight into the evolutionary context of the recent emergence of PCV2-associated diseases in swine. PMWS and other PCVAD are rapidly increasing in prevalence worldwide and are often associated with severe economic impacts. The origin and time scale of the emergence of PCV2 have remained unclear, as have questions surrounding a possible viral genetic component to recent reported increases in virulence of PCV2 (15, 18, 70). The outcome of our study indicates that PCV2 has recently emerged, potentially as the result of a cross-species jump from birds into swine, most likely through intermediate contact with wild boars (39, 50). Our results corroborate previous suggestions that a new genotype (PCV2b) may have contributed to the recent expansion of highly virulent PCV2, although it is clear from our analyses that PMWS is also associated with recent PCV2a infections (Fig. 1). We suggest that the high levels of movement of (asymptomatic) PCV2-infected pigs that occur as a result of the swine trade may have assisted in the rapid spread of the virulent PCV2 genotype around the globe.

Our PCV2 phylogeny shows the division of this virus into two genotypes, PCV2a and PCV2b, congruent with the results of previous studies (Fig. 1) (18, 54). The PCV2b clade contained the majority of recently sampled sequences and was more often associated with disease symptoms than PCV2a (e.g., all sequences from the PMWS-free country of Australia fell within the PCV2a clade). While previous work has indicated that the global emergence of PCVAD is coupled to recent increases in the prevalence of PCV2b, it is clear that this genotype is not exclusively associated with the presence of disease (Fig. 1) (1, 15, 18). However, our analysis of the population dynamics of PCV2 did reveal a trend toward increasing genetic diversity (and population size) in PCV2b that was not reflected in PCV2a, which could be the result of a recent population expansion that occurred in PCV2b only (see Fig. S6 in the supplemental material). As these results are not significant, a larger sample size is needed to confirm this trend. Notably, this analysis also indicates that the relative genetic diversity for both genotypes is extremely low, a feature that could be indicative of continuous selective sweeps over time or repeated population bottleneck events, both of which act to purge variation from the population.

The development of PMWS is often (but not exclusively) thought to be the result of multiple factors, one of which must be the presence of PCV2. Factors often linked to the development of a PCVAD from an asymptomatic infection include the presence of a coinfecting pathogen (such as porcine respiratory and reproductive syndrome virus or porcine parvovirus) and the stimulation of the neonatal immune system from early vaccination (46, 55). The combination of factors often associated with the development of PCVAD may help explain the presence of diseased animals in both the PCV2a and the PCV2b clades. If PCV2b is selectively favored over PCV2a due to properties such as increased replication rate, transmissibility, or virulence, we might expect to see evidence of selection along the lineage leading to the PCV2b clade. While the results of our branch-specific selection analysis did find evidence of positive selection (dN/dS > 1) along a branch leading to the primary PCV2b clade, no obvious directional evolution of multiple nucleotide or amino acid sites was identified (Fig. 1; see also Fig. S8 in the supplemental material). This contrasts with the clear signature of positive selection and directional evolution seen in the lineage leading to another newly emergent single-stranded DNA (ssDNA) virus (canine parvovirus) (63). However, the lack of a similarly strong signal in our data could be the result of an insufficient sample size.

The lack of strong evidence for recent positive selection could alternatively suggest that PCV2 is not a virus that has recently emerged in pigs but instead has perhaps only recently begun to cause disease. Based on similarities in the branching patterns of the Circoviridae phylogeny and those of their corresponding hosts, John et al. (29) suggested that circoviruses have been coevolving with vertebrates for millions of years. However, our analysis of the temporal and topological congruence between the host and circovirus phylogenies indicated that the circovirus phylogeny was no more topologically congruent with that of the host than would be expected by chance (Fig. 2). In addition, a comparison of the respective divergence times of host and circovirus species revealed substantial incompatibilities between the divergence times of these groups, which conflicts strongly with the hypothesis of codivergence. The last shared ancestor between mammals and birds is thought to have lived approximately 300 million years ago, whereas the results of our Bayesian MCMC analyses indicate that the common ancestor of all sampled circovirus diversity existed only 500 years ago (Fig. 2). In contrast to the results reported by John et al. (29), the lack of both temporal and topological congruence between the host and circovirus phylogenies suggests strongly that circoviruses have not shared a long history with their hosts.

Given that circoviruses are likely to be of relatively recent origin, we further estimated the TMRCA for the sampled genetic diversity of the PCV2 lineage. The results of our analyses, using both nucleotide and amino acid substitution rates, indicated that PCV2 originated within approximately the last 100 years. The TMRCA for PCV2b (18 ybp; 95% HPD, 12 to 27 ybp) was more recent than that for PCV2a (41 ypb; 95% HPD, 24 to 62 ybp), with almost nonoverlapping 95% HPDs. This result is consistent with the hypothesis that the PCV2b is associated with the emergence of PCVAD worldwide. The first PMWS case was identified in Canada in 1991, identical to our mean estimated date of origin for PCV2b. Interestingly, the deep bifurcation and long branch lengths leading to the PCV2a and PCV2b clades reveal that PCV2b, while more recent, did not arise from PCV2a. Rather, these two groups shared a common ancestor approximately 100 years ago and have been independently evolving, while cocirculating, since that time. How PCV2a and PCV2b have remained genetically distinct while coinfecting the same species in the same geographic area is unknown but is an interesting avenue for further research.

It has been a challenge to explain the cause of the nearly
simultaneous emergences of PCVAD in Europe and North America and the rapid global spread that followed. The results of the geographic diffusion model used in this study indicate that significant global movement of PCV2 exists, with links detected between countries in North America, Europe, and Asia as well as potentially between South America and Europe (Fig. 1; see also Table S2 in the supplemental material). France (Spain), Hungary, and The Netherlands were identified as being particularly important foci of PCV2 movement in Europe, while high levels of movement also appear to exist between Canada and the United States. Critically, the high levels of diffusion of PCV2 around the globe detected in this analysis are similar to those reported for the trade of live swine between nations (UN database, COMTRADE; available at http://comtrade.un.org/db/; accessed on 6 January 2009). Since the worldwide emergence of PCV2 was reported in 1997, more than 18 billion kilograms of live pigs have been exported from the countries for which PCV2 sequence data were available (based on reported data only). The largest exporters of swine since 1997 have been Canada, China, Denmark, and The Netherlands, while the majority of these exports went to France, Germany, Spain, and the United States. While both the COMTRADE database and the results of our geographic diffusion model are likely skewed by uneven reporting between countries, the similarities in movement patterns that were revealed through these two disparate types of data are striking. The high levels of swine trafficking that occur due to the globalization of the pig/pork industry, along with the high prevalence of asymptomatic PCV2 infections, combine to create an ideal circumstance for the emergence of a highly virulent strain of PCV2 (55). Epidemiological links between infected farms have been determined for outbreaks in the United Kingdom and New Zealand, among others, highlighting the potential for PCV2 transmission through the pig trade (41, 71).

**Microevolution of PCV2.** Virulence in PCV2 is likely determined by two genomic regions, those encoding the capsid protein (ORF2), potentially involved in tissue tropism (39), and ORF3, implicated in the apoptotic action of the virus (44, 45). While our selection analyses did not recover evidence for positive selection in the cap protein as a whole, we found tentative evidence for positive selection in ORF3, with individually selected sites throughout the protein. Previous in vitro and in vivo work has demonstrated that ORF3 is directly involved in viral pathogenesis and acts by inducing the apoptosis of T cells (44, 45). This destruction of T cells is thought to lead to immune suppression and, in turn, to a higher viral load in infected animals (44). Critically, we could not identify specific amino acids in ORF3 that were associated with the emergence of PCV2b or the presence of PCVAD in general; therefore, it is unlikely that strong directional selection is occurring within ORF3. Instead, the evolutionary pressures on this protein may be diversifying, indicative of a rough fitness surface favoring a range of phenotypes. However, that ORF3 is completely within ORF1 (and translated in the opposite direction) greatly complicates this analysis.

Another notable result of our analysis is that the estimated rate of nucleotide substitution for PCV2 (~1.2 × 10^−3 substitutions/site/year) (Table 1) is higher than that measured for any ssDNA virus to date. This places PCV2 within the range of evolutionary rates estimated for most ssRNA viruses and is particularly striking for a DNA virus that replicates using the host polymerase with its associated proofreading capabilities (19, 22, 28, 69). Although such a high substitution rate could have been driven by rapid, strong positive selection, we found little evidence for this process. A variety of factors may therefore explain our anomalously high rate estimates. First, as PCV2 is the smallest known autonomously replicating virus, it is possible that it is able to tolerate a higher mutation rate per site (in keeping with error threshold theory) (5). A rapid rate of viral replication may also contribute to a high error, as replication speed and mutational fidelity may be inversely correlated in RNA viruses (16). Unfortunately there are no available experimental measures of the replication rate of PCV2 or other ssDNA viruses by which to test this hypothesis. Alternatively, it is possible that our rate estimation is strongly skewed toward the mutation rate rather than the long-term substitution rate, a result of the bias in sampling toward the recent past (23, 24). However, this effect is unlikely to have inflated substitution rates by the orders of magnitude needed to reconcile it with those observed in double-stranded DNA viruses. More generally, the mechanism(s) by which ssDNA viruses can achieve high mutation rates while using host DNA polymerases to replicate is still unknown and represents a clear area for future study (14). Despite these uncertainties, our analysis further confirms that ssDNA viruses are closer to RNA than double-stranded DNA viruses in their evolutionary dynamics, helping to create the biological conditions necessary for rapid emergence (72).

Taken together, the results of our analyses strongly suggest that PCV2 is a virus of relatively recent origin in swine. We speculate that the emergence of this virus was likely the result of a recent host switch from birds into pigs, perhaps with an intermediate jump into wild boars (40, 50). High levels of pig movement due to the global trade market may have created a situation resembling a standing network, whereby any highly infectious pathogen, such as (but not limited to) PCV2, would be able to rapidly infect pigs around the world.

**ACKNOWLEDGMENTS**

Funding to C.F. was provided by the Natural Sciences and Engineering Research Council (Canada). E.C.H. was funded in part by NIH grant R01 GM080533.

We thank Andrew Read, Anton Nekrutenko, and Bryan Grenfell for their insight into PCV2 evolution and emergence; Sergei Kosakovsky Pond for his technical advice and support; and Maia Rabaa for her editorial assistance and invaluable contribution to all things swine.

**REFERENCES**


