



# Diagnostic phylogenetics reveals a new Porcine circovirus 2 cluster



Brendan Davies<sup>a</sup>, Xiong Wang<sup>a</sup>, Cheryl M.T. Dvorak<sup>a</sup>, Douglas Marthaler<sup>b</sup>,  
Michael P. Murtaugh<sup>a,\*</sup>

<sup>a</sup> Departments of Veterinary and Biomedical Sciences, University of Minnesota, St. Paul, MN 55108, United States

<sup>b</sup> Departments of Veterinary Population Medicine, University of Minnesota, St. Paul, MN 55108, United States

## ARTICLE INFO

### Article history:

Received 18 December 2015

Received in revised form 3 February 2016

Accepted 3 February 2016

Available online 3 March 2016

### Keywords:

PCV2

Swine

Virology

Phylogenetics

Diagnostics

## ABSTRACT

Porcine circovirus 2 (PCV2) was prevalent in swine in the United States before PCV2-associated disease (PCVAD) appeared in 2006. Limited nucleotide sequencing of open reading frame 2 (ORF2) encoding capsid, the only structural protein, revealed the presence of two genotypes, PCV2a and PCV2b. Later, PCV2c and mutant PCV2b, or PCV2d, were also described. However, extensive PCV2 ORF2 sequence databases in veterinary diagnostic laboratories have not been analyzed systematically to determine the genetic diversity of field isolates. Here, we interrogated >1100 PCV2 ORF2 nucleotide sequences to assess population diversity and genetic variation. We detected a novel PCV2 genotype that is substantially different, primarily in ORF2, from all known PCV2. Notably, ORF2 contains a unique carboxyl terminal amino acid insertion resulting in a 238 amino acid ORF2. All other PCV2 ORF2 proteins are 233 or 234 aa in length. Phylogenetic analysis indicates that it is more ancient than other PCV2 genotypes. The findings demonstrate the value of analyzing routine diagnostic laboratory sequence databases in population genetic analyses of animal pathogens.

© 2016 Elsevier B.V. All rights reserved.

## 1. Introduction

Porcine circovirus 2 (PCV2) is a small, circular, single-stranded DNA virus of swine. PCV2 was discovered in association with outbreaks of a postweaning, multisystemic wasting syndrome (PMWS) in the 1990's in Europe and Canada (Allan et al., 1998; Ellis et al., 1998). Retrospective analysis of archived diagnostic specimens showed it was present in multiple countries in Europe, Asia and North America in previous decades (Dupont et al., 2008; Patterson and Opriessnig, 2010). A national survey of the U.S. swine herd prior to disease outbreaks showed that PCV2 infection was present in nearly all herds at a prevalence greater than 80% of pigs before diseases associated with PCV2 were observed (Haley et al., 2011; Puvanendiran et al., 2011).

Diagnostic sequencing of PCV2 open reading frame 2 (ORF2), encoding the viral capsid protein, was widely adopted in North America and other swine growing regions to assist in understanding the etiology of diseases, including PMWS, porcine dermatitis and nephropathy syndrome, porcine respiratory disease complex, and reproductive diseases, associated with the presence of PCV2

(Segales, 2012; Segales et al., 2005). In the following years, extensive phylogenetic analysis showed that PCV2 consisted of several genetic clusters, without a clear consensus relating genotypic variation to biological characteristics or to disease causation (Chae, 2015; Franzo et al., 2015; Wang et al., 2009; Xiao et al., 2015).

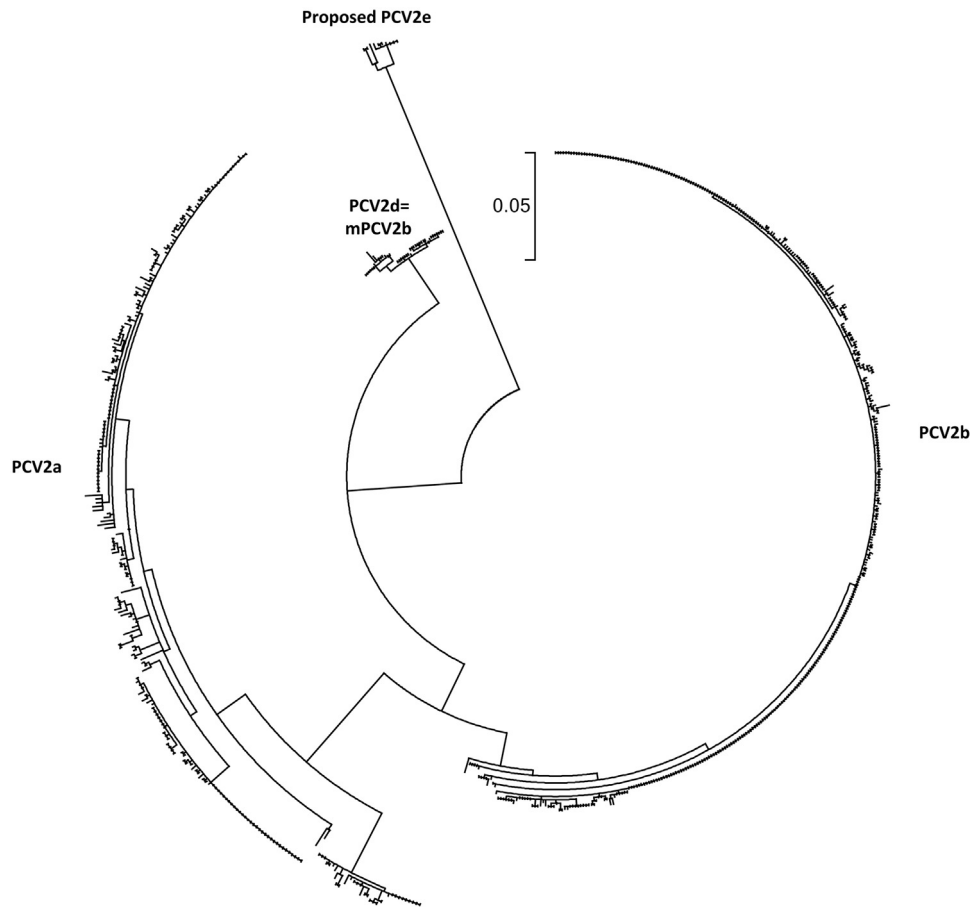
Numerous studies have described PCV2 field isolates and their phylogenetic relationships in swine growing regions. However, a systematic analysis of unpublished ORF2 sequences collected in veterinary diagnostic laboratories has not been conducted. We examined a repository of >1100 PCV2 ORF2 sequences dating back to 2001 in the Minnesota, USA, Veterinary Diagnostic Laboratory (MVDL). A cluster of novel sequences was identified that, by phylogenetic analysis, is ancestral to all known PCV2. Whole genome sequencing confirmed that these viruses are unique and form a new genetic group which we refer to as PCV2e.

## 2. Materials and methods

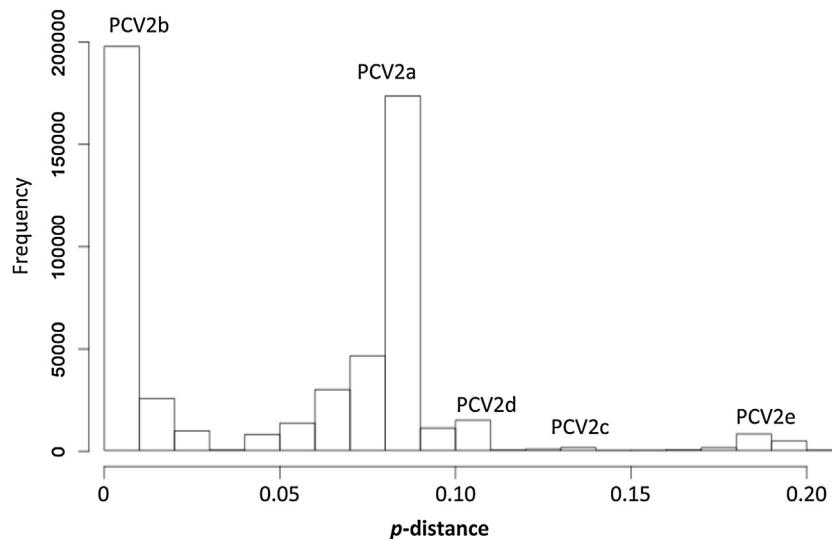
The MVDL PCV2 ORF2 database in August 2015 contained 1289 sequences generated by Sanger sequencing and linked by case number to date of isolation, geographic location, and descriptive reports. The files were merged and organized by geographic region and year of isolation then subjected to data quality analysis. Wild-type field origin was determined by accessing MVDL case files for

\* Corresponding author.

E-mail address: [murta001@umn.edu](mailto:murta001@umn.edu) (M.P. Murtaugh).



**Fig. 1.** Maximum likelihood (ML) phylogenetic analysis of 729 PCV2 ORF2 sequences in the Minnesota Veterinary Diagnostic Laboratory database. Substitution pattern and rates were estimated under the Tamura-Nei model (+G) (Tamura and Nei, 1993), and a discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories [+G]). For estimating ML values, a tree topology was automatically computed. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013).



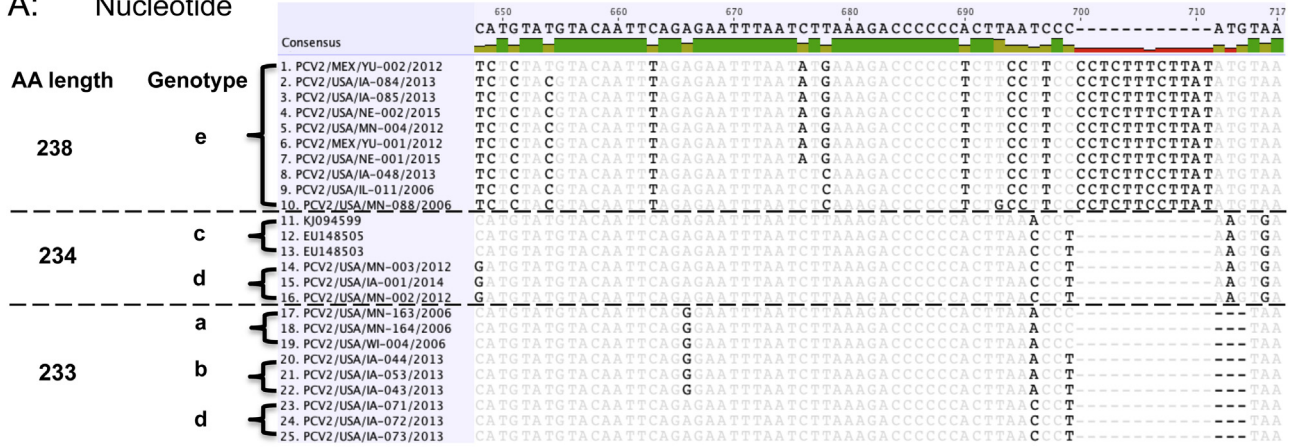
**Fig. 2.** Pairwise similarity analysis of 729 Minnesota Veterinary Diagnostic Laboratory PCV2 ORF2 sequences. Histogram bins are in 1% increments. Genotypes are shown relative to PCV2b. Analysis was performed in MEGA6 (Tamura et al., 2004, 2013). The total number of comparisons was 531,448.

missing or suspect information and to remove sequences generated in research studies. Sequences that could not be verified as field isolates were discarded. A fasta file was generated with the resulting list of 905 sequences, renamed to show virus name (PCV2),

geographic location (country, state or province), a unique identifier, and year of isolation.

The fasta file was imported into Geneious v6 (Biomatters Limited, Auckland NZ) and the “Find ORFs” tool was used to search for intact ORFs over 650 nucleotides in length since PCV2 ORF2

A: Nucleotide



B: Amino acid

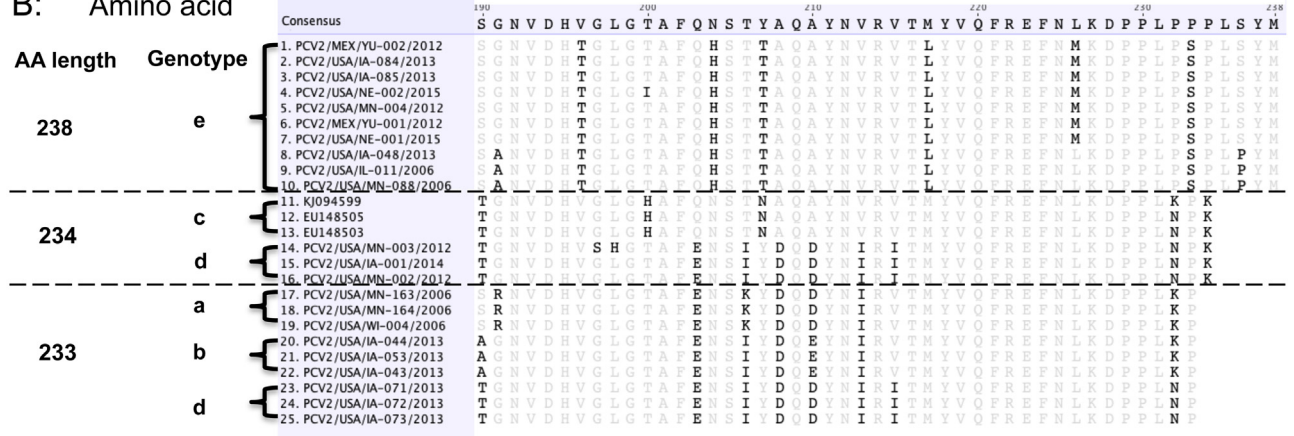


Fig. 3. Amino acid sequence alignment of representative PCV1 and PCV2 genotype clusters showing regions of similarity and dissimilarity.

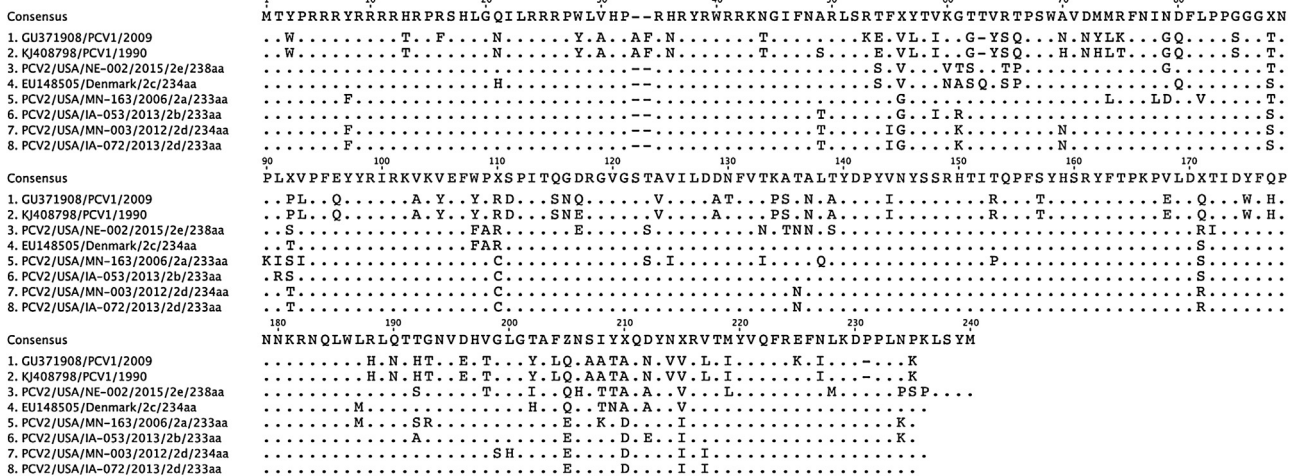
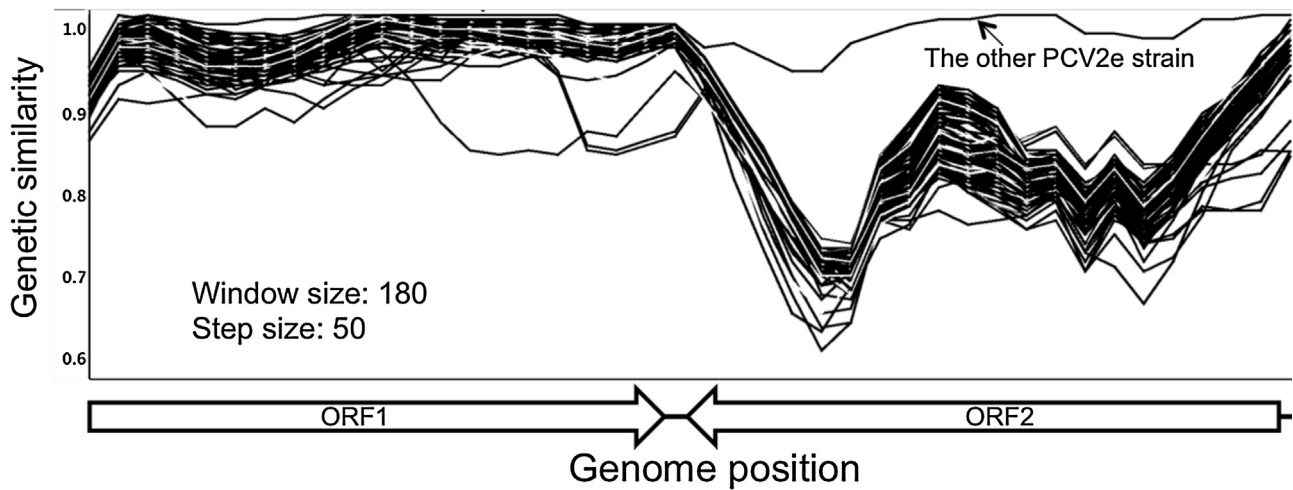


Fig. 4. Variation in PCV2 at the ORF2 3' gene terminus (A) and the carboxyl terminus of the protein (B). Representative strains are ordered by amino acid length of the protein, and by genotype as indicated.

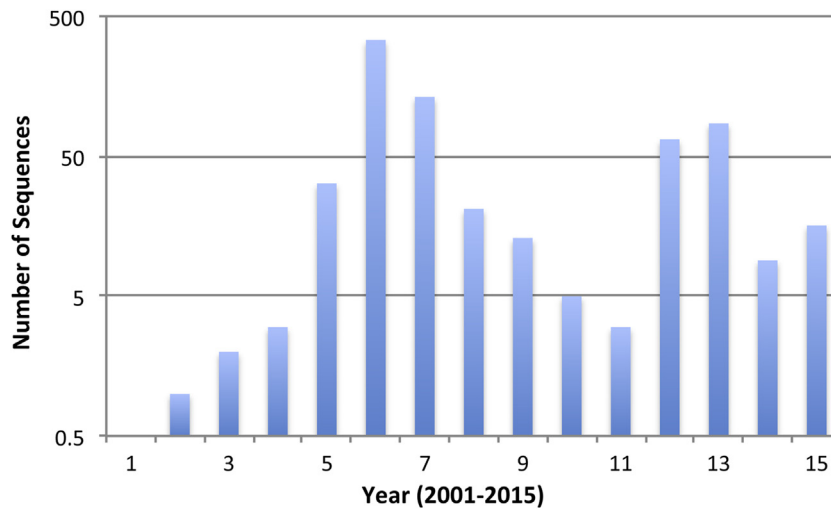
sequences are  $\geq 702$  bases. Sequences with an ORF less than 650 were deleted, and there were none between 650 and 702 bases. Sequences with >5 ambiguous bases were deleted, and for sequences with  $\leq 5$  ambiguities, original trace files were examined manually. Sequences with one or more ambiguities that could not be resolved were deleted. The resulting 729 validated sequences were subjected to phylogenetic analysis.

Nucleotide sequences were aligned in MUSCLE with default parameters and clustered with Geneious Tree Builder using default parameters. A maximum likelihood tree was constructed in MEGA 6.06 using GTR and default parameters without bootstrapping. Independent landmark genotypes to orient the tree were added from Genbank as described (Franzo et al., 2015).

Whole genome sequence analysis was performed on 1416 PCV2 Genbank accessions whose ORF1 start was at position 51, and two



**Fig. 5.** Recombination analysis comparing PCV2e whole genomes to all PCV2 genomes in Genbank. Analysis was performed using the Recombination Analysis Tool (RAT) (Etherington et al., 2005).



**Fig. 6.** Frequency pattern of PCV2 ORF2 sequence acquisition in the Minnesota Veterinary Diagnostic Laboratory from 2001 to 2015.

novel genomes identified in this study. Recombination analysis was performed using Recombination Analysis Tool (RAT) (Etherington et al., 2005).

PCV2 ORF2 sequences were deposited in Genbank with accession numbers KT867794–KT868522. Full length genome sequences PCV2/USA/NE-001/2015 and PCV2/USA/NE-002/2015 were deposited as KT870146 and KT870147, respectively.

### 3. Results

The genetic structure of 729 validated PCV2 ORF2 sequences from field cases submitted to the MVDL is shown in Fig. 1. Notable features include the limited genetic variation in PCV2b, especially the presence of 126 identical, independently acquired sequences in the years 2005–2013, and a second set of 88 identical sequences from 2005 to 2012; a more variable PCV2a genotype with a cluster of 33 identical sequences from 2006 to 2012; a conserved cluster of mPCV2b (=PCV2d) isolated in 2012–2015; and a novel, distant cluster of 10 sequences from 4 midwestern U.S. states and Mexico, collected from 2006 to 2015, proposed here as PCV2e (Fig. 1). Pair-wise sequence comparison of the 729 sequences also showed that the novel PCV2e cluster was unique (Fig. 2, Table 1). Amino acid sequence alignments of representative isolates, as shown in Fig.

3, and nucleotide sequence comparisons (not shown) also showed that PCV2e, as well as other PCV2 genotypes, were strikingly different from PCV1.

The novel PCV2e group of sequences was 1777 nt in genome length, and ORF 2 was 15–18% different in nucleotide sequence from the nearest neighbors in the MVDL database (15% between PCV2/MEX/YU-002/2012 and PCV2/USA/TX-003/2009, and 18% between PCV2/USA/MN-088/2006 and PCV2/USA/WI-004/2006). The group members encoded a capsid protein of 238 amino acids, which is markedly different from all known capsid proteins of 233 or 234 amino acids (Chae, 2015). Nucleotide and amino acid sequence comparisons showed that the marked variation is at the carboxyl end of the encoded protein, including extra amino acids at the carboxyl terminus (Figs. 3 and 4). This cluster of sequences shares a common ancestor with PCV2c on the ML tree and it is uniquely different from any other sequence in Genbank. The closest ORF2 identity by BLASTn comparison was 87% to five PCV2c isolates from Denmark, Romania and India (Table 2). Comparison of PCV1 and PCV2 genotype ORF2 amino acid sequences shows that the carboxyl terminus is relatively conserved across all porcine circoviruses in spite of the length variations that have been demonstrated (Fig. 3).



**Table 1**  
Pairwise comparison of ORF2 nucleotide sequences of PCV1 and PCV2a–e.

Genbank accession/Name	GU371908	KJ408798	PCV2/USA/NE-002/2015	EU148505	PCV2/USA/MN-163/2006	PCV2/USA/IA-053/2013	PCV2/USA/MN-003/2012	PCV2/USA/IA-072/2013
Genotype	PCV1	PCV1	PCV2e	PCV2c	PCV2a	PCV2b	PCV2d	PCV2d
AA length	233	233	238	234	233	233	234	233
Year of Isolation	2009	1990	2015	1990	2006	2013	2012	2013
GU371908		98.3	65.6	66.3	65.9	67.1	65.2	66.0
KJ408798	98.3		66.6	67.3	66.6	68.0	66.2	67.0
PCV2/USA/NE-002/2015	65.6	66.6		84.4	79.5	81.6	82.8	83.0
EU148505	66.3	67.3	84.4		85.7	89.9	89.1	88.9
PCV2/USA/MN-163/2006	65.9	66.6	79.5	85.7		91.6	88.9	89.9
PCV2/USA/IA-053/2013	67.1	68.0	81.6	89.9	91.6		93.3	94.3
PCV2/USA/MN-003/2012	65.2	66.2	82.8	89.1	88.9	93.3		98.7
PCV2/USA/IA-072/2013	66.0	67.0	83.0	88.9	89.9	94.3	98.7	

**Table 2**  
PCV2e gene and genome nucleotide sequence identities to nearest neighbors.

Isolate	Genbank	Year isolated	Country	(%) identity <sup>a</sup>
<b>A. ORF2</b>				
PCV2/USA/NE-002/2015	KT870147	2015	USA	100
DK1980PMWSfree	EU148503	1980	Denmark	87
DK1987PMWSfree	EU148504	1987	Denmark	87
DK1990PMWSfree	EU148505	1990	Denmark	87
India-GN-07	GU808525	2007	India	87
EU-RO-WB2010-794	JN382183	2011	Romania	87
<b>B. ORF1</b>				
PCV2/USA/NE-001/2015	KT870146	2015	USA	100
SPA1	AF201308	1999	Spain	99
Yantai-1	KP313251	2014	China	99
P2425NT	JX099786	2008	Vietnam	99
(blank)	AY713470	2004	Germany	99
YL5	HQ202972	2010	Taiwan	99
<b>C. Whole genome</b>				
PCV2/USA/NE-001/2015	KT870146	2015	USA	100
P2425NT	JX099786	2008	Vietnam	93
TZ60607	FJ158607	2006	China	93
HaiNan	AY556476	2004	China	93
10BJ-2	HaiNan	2010	China	93
LZ	DQ363860	2006	China	93

<sup>a</sup> Percent identity was taken from BLASTn analysis results performed 7 Sept. 2015 on the National Center for Biological Information (NCBI) BLAST website.

Whole genome sequence analysis of two PCV2e isolates confirmed the unique identity of this group of viruses. BLASTn analysis of the non-redundant Genbank database identified 19 genomes with 93% identity. Comparison of one PCV2e isolate to five representative genomes, with a range of isolation dates and geographical locations, shows that the genetic variation is localized predominately in ORF2 (Table 2). In addition, BLAST analysis of ORF1, encoding the replicase gene, showed that PCV2e is a circovirus and is not the product of a recombination event that combined ORF2 with genetic material from another source (Table 2 and Fig. 5). In addition, whole genome similarity analysis of PCV2 showed that PCV2e was not created by recombination within PCV2 family members. Interestingly, nearly all PCV2 genomes except for PCV2e flowed together as a single group, indicating that recombination did not give rise to PCV2e and suggests that recombination is unlikely to contribute significantly to evolution of PCV2 (Fig. 5).

#### 4. Discussion

Phylogenetic analysis of a PCV2 ORF2 sequence database in a veterinary diagnostic laboratory that serves the midwestern US swine industry revealed the presence of a novel PCV2 genotype that was first observed in 2006, and an abundance of PCV2b isolates with limited genetic diversity even though it was widely distributed throughout the U.S. national herd in 2006, suggesting a recent spread of the genotype prior to the appearance of vaccines (Puvanendiran et al., 2011). Widespread distribution of a genetically conserved virus in the absence of clinical disease could have been accomplished by spread through infected semen or breeding stock. PCV2a genotypes also were common and widely distributed in U.S. swine but show substantially more genetic diversity, as has been described previously (Franzo et al., 2015; Xiao et al., 2015). Both PCV2a and PCV2b sequences have been obtained every year since 2006 at the MVDL, regardless of the sampling frequency, which peaked in 2006 and 2007, then increased again in 2012 and 2013 (Fig. 6).

Presence of a distant PCV2 genotype in the U.S. at the beginning of PCVAD outbreaks in 2006 and sporadic isolation to the present time was unexpected. The marked divergence from other PCV2 genotypes indicates an ancient origin whose source is yet to be revealed. Additional analysis of diagnostic laboratory databases containing extensive inventories of viral and microbial genetic information may reveal further insights. The value of screening existing diagnostic databases for genetic diversity in animal pathogens is evident in the case of PCV2, a pathogen that does not reproducibly cause disease in experimental studies, and is present in swine herds without obvious disease (Allan et al., 2007; Guo et al., 2012; Opriessnig et al., 2010, 2008; Puvanendiran et al., 2011; Shen et al., 2010). The PCV2e isolates identified here were recovered from diagnostic submission associated with PCVAD and would have not been recognized as novel without sequencing. Diagnostic sequencing of typical PCVAD cases independently at Iowa State University shows that PCV2e isolates are detected with routine diagnostic PCR that also detects other isolates (Harmon et al., 2015).

The extra length of PCV2e capsid is due to 12 nucleotides present at the 3' end of ORF2, relative to 234 aa capsids, giving rise to a terminal 7 amino acid sequence, PPPLSYM, that is unlike the consensus NPK and PK termini of 234 and 233 aa capsids, respectively. We note

that PCV2a and PCV2b ORF2 encodes 233 aa capsid exclusively, while PCV2c contains only the 234 aa form and PCV2d genotypes contain 233 or 234 aa capsids. A simple model to account for these changes is that PCV2e is the prototype, with a single 12 nt deletion giving rise to a progenitor of 2a–d; and that PCV2c and PCV2d acquired a second, 3 nt deletion. The unique carboxyl terminus of PCV2e capsid does not appear to affect the predicted 3-dimensional protein structure compared to 233 or 234 aa length capsids modeled using Raptor X (Kallberg et al., 2012).

## Acknowledgements

The work was funded in part by support from Boehringer Ingelheim Vetmedica (BIV) and by the USDA National Institute of Food and Agriculture, multistate project MIN-63-112. XW was supported by a MnDRIVE Global Food Ventures fellowship. BD is an undergraduate student at University of Minnesota, Morris. MPM receives consulting income from BIV. The relationship has been reviewed and managed by the University of Minnesota in accordance with its conflict of interest policies.

## References

- Allan, G.M., McNeilly, F., Kennedy, S., Daft, B., Clarke, E.G., Ellis, J.A., Haines, D.M., Meehan, B.M., Adair, B.M., 1998. Isolation of porcine circovirus-like viruses from pigs with a wasting disease in the USA and Europe. *J. Vet. Diagn. Invest.* 10 (1), 3–10.
- Allan, G.M., Caprioli, A., McNair, I., Lagan-Tregaskis, P., Ellis, J., Krakowka, S., McKillen, J., Ostanello, F., McNeilly, F., 2007. Porcine circovirus 2 replication in colostrum-deprived piglets following experimental infection and immune stimulation using a modified live vaccine against porcine respiratory and reproductive syndrome virus. *Zoonoses Public Health* 54 (5), 214–222.
- Chae, C., 2015. An emerging porcine circovirus type 2b mutant (mPCV2b) originally known as PCV2d. *Vet. J.* 203 (1), 6–9.
- Dupont, K., Nielsen, E.O., Baekbo, P., Larsen, L.E., 2008. Genomic analysis of PCV2 isolates from Danish archives and a current PMWS case-control study supports a shift in genotypes with time. *Vet. Microbiol.* 128 (1–2), 56–64.
- Ellis, J., Hassard, L., Clark, E., Harding, J., Allan, G., Willson, P., Strokappe, J., Martin, K., McNeilly, F., Meehan, B., Todd, D., Haines, D., 1998. Isolation of circovirus from lesions of pigs with postweaning multisystemic wasting syndrome. *Can. Vet. J.* 39 (1), 44–51.
- Etherington, G.J., Dicks, J., Roberts, I.N., 2005. Recombination analysis tool (RAT): a program for the high-throughput detection of recombination. *Bioinformatics* 21 (3), 278–281.
- Franzo, G., Cortey, M., Olvera, A., Novosel, D., De Castro, A.M., Biagini, P., Segales, J., Drigo, M., 2015. Revisiting the taxonomical classification of Porcine circovirus type 2 (PCV2): still a real challenge. *Virology* 12 (1), 131.
- Guo, L., Fu, Y., Wang, Y., Lu, Y., Wei, Y., Tang, Q., Fan, P., Liu, J., Zhang, L., Zhang, F., Huang, L., Liu, D., Li, S., Wu, H., Liu, C., 2012. A porcine circovirus type 2 (PCV2) mutant with 234 amino acids in capsid protein showed more virulence in vivo, compared with classical PCV2a/b strain. *PLoS One* 7 (7), e41463.
- Haley, C., Wagner, B., Puvanendiran, S., Abrahante, J., Murtaugh, M.P., 2011. Diagnostic performance measures of ELISA and quantitative PCR tests for porcine circovirus type 2 exposure using Bayesian latent class analysis. *Prev. Vet. Med.* 101 (1–2), 79–88.
- Harmon, K.M., Gauger, P.C., Zhang, J., Pineyro, P.E., Dunn, D.D., Chriswell, A.J., 2015. Whole-genome sequences of novel Porcine circovirus type 2 viruses detected in swine from Mexico and the United States. *Genome Announc.* 3 (6), e01315-15.
- Kallberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., Xu, J., 2012. Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* 7 (8), 1511–1522.
- Opriessnig, T., Ramamoorthy, S., Madson, D.M., Patterson, A.R., Pal, N., Carman, S., Meng, X.J., Halbur, P.G., 2008. Differences in virulence among porcine circovirus type 2 isolates are unrelated to cluster type 2a or 2b and prior infection provides heterologous protection. *J. Gen. Virol.* 89 (Pt. 10), 2482–2491.
- Opriessnig, T., Prickett, J.R., Madson, D.M., Shen, H.G., Juhan, N.M., Pogranichniy, R.M., Meng, X.J., Halbur, P.G., 2010. Porcine circovirus type 2 (PCV2)-infection and re-inoculation with homologous or heterologous strains: virological, serological, pathological and clinical effects in growing pigs. *Vet. Res.* 41 (3), 31.
- Patterson, A.R., Opriessnig, T., 2010. Epidemiology and horizontal transmission of porcine circovirus type 2 (PCV2). *Anim. Health Res. Rev.* 11 (2), 217–234.
- Puvanendiran, S., Stone, S., Yu, W., Johnson, C.R., Abrahante, J., Jimenez, L.G., Griggs, T., Haley, C., Wagner, B., Murtaugh, M.P., 2011. Absence of porcine circovirus type 1 (PCV1) and high prevalence of PCV 2 exposure and infection in swine finisher herds. *Virus Res.* 157 (1), 92–98.
- Segales, J., Allan, G.M., Domingo, M., 2005. Porcine circovirus diseases. *Anim. Health Res. Rev.* 6 (2), 119–142.
- Segales, J., 2012. Porcine circovirus type 2 (PCV2) infections: clinical signs, pathology and laboratory diagnosis. *Virus Res.* 164 (1–2), 10–19.
- Shen, H., Wang, C., Madson, D.M., Opriessnig, T., 2010. High prevalence of porcine circovirus viremia in newborn piglets in five clinically normal swine breeding herds in North America. *Prev. Vet. Med.* 97 (3–4), 228–236.
- Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10 (3), 512–526.
- Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U. S. A.* 101 (30), 11030–11035.
- Tamura, K., Stecher, G., Peterson, D., Filipitski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30 (12), 2725–2729.
- Wang, F., Guo, X., Ge, X., Wang, Z., Chen, Y., Cha, Z., Yang, H., 2009. Genetic variation analysis of Chinese strains of porcine circovirus type 2. *Virus Res.* 145 (1), 151–156.
- Xiao, C.T., Halbur, P.G., Opriessnig, T., 2015. Global molecular genetic analysis of Porcine circovirus type 2 (PCV2) sequences confirms the presence of four main PCV2 genotypes and reveals a rapid increase of PCV2d. *J. Gen. Virol.* 96 (7), 1830–1841.