ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

SHORT COMMUNICATION

doi: 10.1111/age.12522

Genomic investigation of piglet resilience following porcine epidemic diarrhea outbreaks

F. Bertolini*, J. C. S. Harding†, B. Mote‡, A. Ladinig§, G. S. Plastow¶ and M. F. Rothschild*

*Department of Animal Science, Iowa State University, Ames, IA, 50011-3150, USA. †Department of Large Animal Clinical Sciences, Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, SK, S7N 5B4, Canada. ‡Department of Animal Science, University of Nebraska, Lincoln, NE, 68583-0908, USA. §University Clinic for Swine, Department for Farm Animals and Veterinary Public Health, University of Veterinary Medicine Vienna, Wien, 1210, Austria. ¶Livestock Gentec, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, T6G 2R3, Canada.

Summary

Porcine epidemic diarrhea virus (PEDV) belongs to the Coronaviridae family and causes malabsorptive watery diarrhea, vomiting, dehydration and imbalanced blood electrolytes in pigs. Since the 1970s, PED outbreaks have become a source of problems in pig producing countries all over the world, causing large economic losses for pig producers. Although the infection in adults is not fatal, in naïve suckling piglets mortality is close to 100%. In this study, we investigated genome-wide differences between dead and recovered suckling piglets from commercial farms after PED outbreaks. Samples from 262 animals (156 dead and 106 recovered) belonging to several commercial lines were collected from five different farms in three different countries (USA, Canada and Germany) and genotyped with the porcine 80K SNP chip. Mean F_{st} value was calculated in 1-Mb non-overlapping windows between dead and recovered individuals, and the results were normalized to find differences within the comparison. Seven windows with high divergence between dead and recovered were detected—five on chromosome 2, one on chromosome 4 and one on chromosome 15—in total encompassing 152 genes. Several of these genes are either under- or overexpressed in many virus infections, including Coronaviridae (such as SARS-CoV). A total of 32 genes are included in one or more Gene Ontology terms that can be related to PED development, such as Golgi apparatus, as well as mechanisms generally linked to resilience or diarrhea development (cell proliferation, ion transport, ATPase activity). Taken together this information provides a first genomic picture of PEDV resilience in suckling piglets.

Keywords F_{st}, piglets, porcine epidemic diarrhea virus, resilience

Porcine epidemic diarrhea virus (PEDV) belongs to the Coronaviridae family; this family includes a wide variety of viruses that affect humans and other animals, causing respiratory and gastroenteric diseases (Weiss & Navas-Martin 2005). As with many coronaviruses, PEDV has a limited host range and tropism, infecting only the small intestine of pigs (Song & Park 2012; Jung *et al.* 2014). This infection causes acute intestinal disease during which the infected enterocytes rapidly develop necrosis, leading to a villous atrophy (Debouck *et al.* 1981; Ducatelle *et al.* 1982; Jung *et al.* 2014) and development of malabsorptive watery diarrhea, vomiting, dehydration and imbalanced blood electrolytes (Jung & Saif 2015). The major route of transmission for PEDV infection is fecal–oral (Turgeon *et al.* 1980; Utiger *et al.* 1995; Riley 2007) even if aerosolized PEDV remains infectious (Alonso *et al.* 2014). It is known that PEDV, as well as other swine coronaviruses, such as porcine TGEV (transmissible gastroenteritis virus), gains entry in the host cells through interaction with alanyl (membrane) aminopeptidase, encoded by the *ANPEP* gene (Oh *et al.* 2003; Li *et al.* 2007). This receptor is largely expressed by enterocyte cells, and its density has been correlated to the replication rate of the virus *in vitro* (Nam & Lee 2010). Since discovery of the virus during the 1970s in Europe (Pensaert *et al.* 1978), PED outbreaks have become an increasing source of problems in many swine breeding countries all over the world, causing severe problems as

Address for correspondence

F. Bertolini, Department of Animal Science, Iowa State University, Ames, IA 50011-3150, USA.
E-mail: fbert@iastate.edu

Accepted for publication 29 September 2016

© 2016 Stichting International Foundation for Animal Genetics,

well as large economic losses for pork producers (Song & Park 2012; Snelson 2014). It is widely understood that, although the infection in adults may not be fatal, PEDV infection in unprotected piglets under 3 weeks of age results in severe disease with mortality reaching 100% on many farms (Pensaert *et al.* 1978; Shibata *et al.* 2000). The degree to which the small percentage of naïve suckling piglets recover during PED outbreaks in the wider industry is unknown, as is the biological mechanisms involved, but two main hypothesis can be suggested: (i) survival can be related to variation in the intestinal receptor used by PEDV to gain entry to intestinal epithelial cells (the *ANPEP* gene) or (ii) as summarized by Schneider & Ayres (2008) and Ayres & Schneider (2012), survival and recovery can be related to particular host immune responses that enhance viral clearance, increase cell epithelial regeneration rate or reduce viral replication rate.

In this study, our aim was to investigate genome-wide differences between dead and recovered suckling piglets from commercial farms during PED outbreaks. The study was conducted using piglets younger than 3 weeks of age from farms during the acute phase of a PED outbreak (as soon after herd diagnosis as physically possible). As depopulation is one of the main approaches used to deal with outbreaks, it is difficult to obtain good information on survival rates, or indeed samples, to test these hypotheses. To address this, we collected samples from 262 animals belonging to several commercial terminal crossbred lines from five different farms in three different countries (USA, Canada and Germany). Among these animals, 156 animals died as a result of severe diarrhea and dehydration and 106 animals successfully recovered (Table S1). Specific piglets were collected to ensure that balanced numbers of dead and recovered piglets were collected from each farm, with the exception of one farm in the USA from which the numbers were not balanced. Moreover, in three of the five farms (one farm located in USA and the farms located in Germany) dead and recovering piglets belonged to the same litters (1–2 dead and 1–2 recovered per litter). Each animal was genotyped with the GeneSeek Genomic Profiler 80K SNP chip (Geneseek–Neogen), which contains 68 517 SNPs. For the analysis, only SNPs mapped to the autosomes and with a call rate >90 were considered. In total, 57 422 SNPs passed the filtering steps, and missing SNPs were imputed using BEAGLE 3.3.3 (Browning & Browning 2007). ADMIXTURE software (Alexander *et al.* 2009) was used to investigate population stratification. Following the software guidelines, the cross-validation (CV) procedure was applied to choose the best K for the model, where K is the number of (sub) populations that was assumed for the analysis. Within a range of 10 K values, 1–10, $K = 8$ showed the lowest CV error and therefore was considered a sensible model. The admixture plot considering $K = 8$ is reported in Fig. S1. As expected, the samples appeared to be stratified, with more similarity between the two US farms. Due to the expected low number of recovering piglets from each farm, detailed analysis by farm or within breed or population strata could not be considered. Therefore, we focused our attention on a comparison of all dead with all recovered, attempting to detect common windows of divergence. A multidimensional scaling plot was obtained using PLINK 1.7 (Purcell *et al.* 2007), comparing dead and recovered piglets (Fig. S2). The plot clearly shows a lack of clustering between the two groups. To find windows that were more divergent within the two groups, F_{st} was calculated for each SNP, adapting the formula reported by Karlsson *et al.* (2007):

$$F_{st_k} = N_k/D_k,$$

where k is the SNP marker k, with frequency $p_1^{[k]}$, $p_2^{[k]}$;

$$N_k = p_1^{[k]}(q_2^{[k]} - q_1^{[k]}) + p_2^{[k]}(q_1^{[k]} - q_2^{[k]}); \text{and}$$

$$D_k = p_1^{[k]}q_2^{[k]} + q_1^{[k]}p_2^{[k]} = N_k + p_1^{[k]}q_1^{[k]} + p_2^{[k]}q_2^{[k]}$$

Then, a mean F_{st} value (F_{stm}) was calculated in 1-Mb non-overlapping windows, and the result was Z transformed using a derived formula provided by Rubin *et al.* (2010):

$$ZF_{stm_i} = (F_{stm_i} - \mu F_{stm})/\sigma F_{stm},$$

where i is the 1-Mb window, μF_{stm} is the mean and σF_{stm} is the standard deviation.

The distribution of the frequency for the normalized values is reported in Fig. S3. Values lower than 6 represent 0.2% of the empirical distribution of all the normalized values and were the most divergent between the two groups and, therefore, were considered as significant. The normalized plot is reported in Fig. 1. A total of seven windows had values that were higher than the threshold: five located on chromosome 2 (positions 66–67, 67–68, 75–76, 77–78 and 79–80 Mb), one window located on chromosome 4 (position 53–54 Mb), and one on chromosome 15 (position

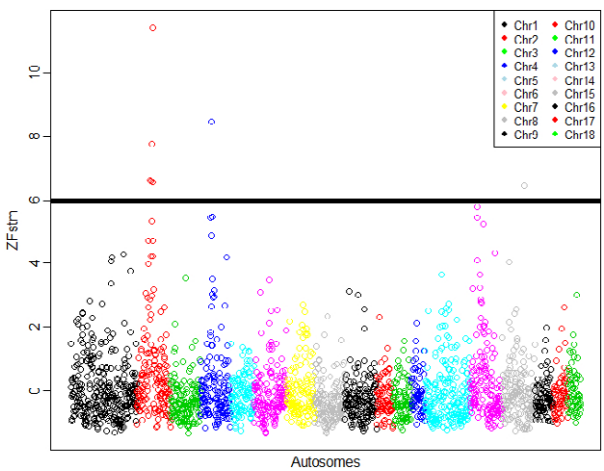

Figure 1 Normalized F_{st} plot. ZF_{st} values are reported on the Y axis; on the X axis, chromosomes are labelled in different colors (as indicated by the legend). The black line across the plot indicates the fixed threshold.

105–106 Mb). The number of SNPs included in the selected windows ranged from eight to 14 (Table S2). The *ANPEP* gene, known to be the receptor of PEDV and a possible candidate for PEDV resistance (Nam & Lee 2010), is located on chromosome 7 at position 60 240 145–60 262 914. The window that contains this gene had a ZF_{stm} score of 0.2, which is far below the threshold chosen for significance. Therefore this gene, which has previously been suggested to be linked to the replication rate of the virus (Nam & Lee 2010), did not appear to be related to PEDV recovery in this study. More specific analyses need to be conducted to establish if this gene does play a role in the recovery of suckling piglets after acute infection, because the number of markers and the linkage disequilibrium in the considered windows can affect the significance above all when only one gene in the window is involved.

After the completion of the genomic analysis, immunoglobulin G (IgG) in the serum of all animals was measured using either an indirect ELISA (Swinecheck® PED; Biovet Inc.) or an indirect immunofluorescence assay (Iowa Veterinary Diagnostic Laboratory), depending on the farm (data not shown). Half (52/100) of the German pigs, including both dead ($n = 11$) and recovered ($n = 41$), tested ELISA positive. Because all dead pigs were 8 days old or younger at sampling, positive ELISA tests were indicative of maternally derived antibodies. In spite of the fact that the recovered pigs were exposed to PEDV during their first week of life, serum PEDV IgG was not measured in German pigs until 14–21 days of age. At that age, the majority of the piglets likely had naturally seroconverted (Opriessnig *et al.* 2014); hence, it was not possible to distinguish whether a positive response was the result of maternal antibodies or an adaptive antibody-mediated immune response. Furthermore, it is known that the Swinecheck PEDV ELISA cross-reacts with porcine deltacoronavirus (personal communication, A. Ambagala, June 24, 2016). Re-analysis of the genomic data with and without the ELISA positive pigs still identified the most interesting regions on the same chromosomes (chromosomes 2: 55–56 Mb and 86–90 Mb, 4: 53–54 Mb and 15: 105–106 Mb; for more details see Fig. S4), and hence, the results obtained appear reliable. Moreover, for this specific case, the imbalance between dead and survived and across farms (Germany contributed with 39 dead piglets and nine recovered) could influence the results. As both dead and recovered pigs in these cohorts were ELISA positive, it still suggests they responded differentially to the infection during the acute phase of the disease and are useful in this genomic analysis.

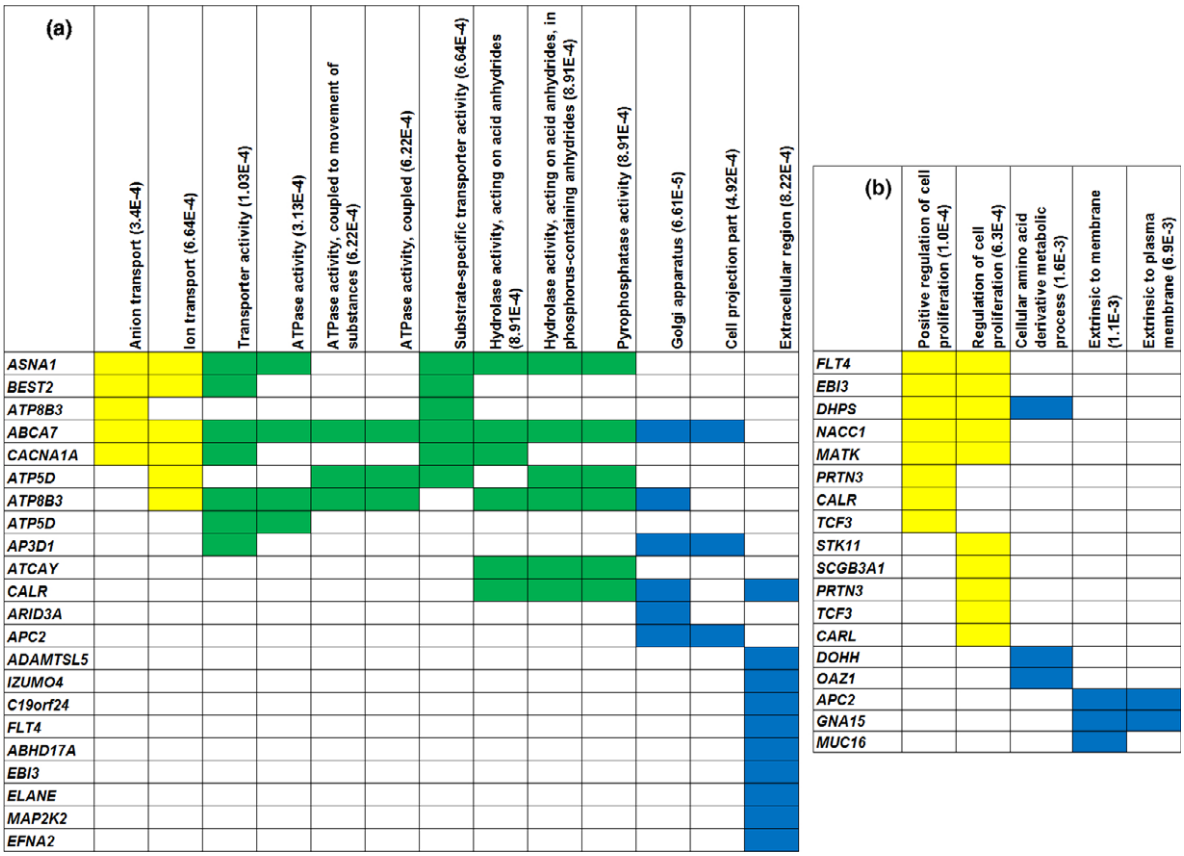

Figure 2 A Gene Ontology (GO) term table with the genes included in each GO term detected with GORILLA (a) and ENRICHR (b); *P*-values ≤0.001. In the plots, output for biological processes (yellow), functions (green) and components (blue) are reported.

The seven selected windows were then screened to find annotated genes located 200 kb before and after the windows, according to Sscrofa 10.2 and the annotation provided by Ensembl (http://www.ensembl.org/). To avoid ambiguous windows, we considered only protein-coding genes (not pseudogenes or miRNAs) that were officially named, uniquely orthologous in mammalian species or named in the Uniprot database. A total of 152 genes fit these requirements, with only one gene located in the window on chromosome 4 and the majority (40 genes) located on chromosome 2: 74.8–76.2 Mb. The list of the annotated genes for each window is reported in Table S2. The detected genes were analyzed to investigate their role during virus infection and proliferation.

All genes were screened to find evidence of gene perturbation and differential gene expression after virus–host protein interaction using the online tool ENRICHR (Chen *et al.* 2013). This analysis was performed because it is known that host targets of viral proteins reside in networks in proximity to products of disease susceptibility genes (Goh *et al.* 2007; Zhong *et al.* 2009; Gulbahce *et al.* 2012). Then, ENRICHR and the GORILLA tool (Eden *et al.* 2007, 2009), with *Homo sapiens* as a reference, were used to find Gene Ontology (GO) terms of functions, processes or components for the considered genes. Only GO terms with $P \leq 0.001$ were considered, which is the minimum threshold allowed by GORILLA. Several genes that have been found to be underexpressed or overexpressed in many viral infections, including coronaviruses (such as SARS-CoV), are shown in Table S3. A total of 32 genes were included in one or more GO terms (Fig. 2). The GO term outputs are related mainly to regulation of cell proliferation, ion transport, Golgi apparatus, ATPase activity, hydrolase activities and pyrophosphatase activity. Interestingly, several of these GO terms could be directly linked to PEDV proliferation and diarrhea development, as follows: (i) Cell proliferation has been linked to the ability of the organism to overcome virus or bacterial tissue destruction (Schneider & Ayres 2008); (ii) ion channels, particularly K+ and Ca++ channels, are necessary for intestinal homeostasis, and their functions are altered during diarrhea in several species including humans (Field 2003) and pigs (Moeser & Blikslager 2007); (iii) ATPase activity is required for the function of several ion channels (reviewed by Gouaux & Mackinnon 2005); and (iv) the assembly of the PED virus occurs by budding through intra-cytoplasmic membranes, such as the endoplasmic reticulum and Golgi apparatus (Ducatelle *et al.* 1982).

In conclusion, we did not find evidence of a direct association between the *ANPEP* gene and PEDV resilience in naïve piglets. Instead, the picture provided by the F_{st} analyses seems to support the associations between recovery and host responses that could influence cell epithelial regeneration rate, virus replication rate and the general consequences of virus infection.

Conflicts of interest

The authors declare they do not have any conflicts of interest. Data reported in this work will be available one year after the publication of the manuscript at the following link: http://www.animalgenome.org/repository/pub/ISU2016.0617/.

Acknowledgements

The authors appreciate the financial support of the US Pig Genome Coordination Program, particularly Chris Tuggle and Cathy Ernst, and the State of Iowa. Funding supporting the collection and processing of samples from PED outbreak farms was provided by the Genome Alberta's 2014 Program on Research and Innovation Leading to a Rapid Genomics Response to the Porcine Epidemic Diarrhea Virus (PEDV).

References

Alexander D.H., Novembre J. & Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655–64.

Alonso C., Goede D.P., Morrison R.B., Davies P.R., Rovira A., Marthaler D.G. & Torremorell M. (2014) Evidence of infectivity of airborne porcine epidemic diarrhea virus and detection of airborne viral RNA at long distances from infected herds. *Veterinary Research* **45**, 73.

Ayres J.S. & Schneider D.S. (2012) Tolerance of infections. *Annual Review of Immunology* **30**, 271–94.

Browning S.R. & Browning B.L. (2007) Rapid and accurate haplotype phasing and missing data inference for whole genome association studies using localized haplotype clustering. *American Journal of Human Genetics* **81**, 1084–97.

Chen E.Y., Tan C.M., Kou Y., Duan Q., Wang Z., Meirelles G.V., Clark N.R. & Ma'ayan A. (2013) ENRICHR: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128.

Debouck P., Pensaert M. & Coussement W. (1981) The pathogenesis of an enteric infection in pigs, experimentally induced by the coronavirus-like agent, Cv-777. *Veterinary Microbiology* **6**, 157.

Ducatelle R., Coussement W., Debouck P. & Hoorens J. (1982) Pathology of experimental CV777 coronavirus enteritis in piglets II: electron microscopic study. *Veterinary Pathology* **19**, 57–66.

Eden E., Lipson D., Yogev S. & Yakhini Z. (2007) Discovering motifs in ranked lists of DNA sequences. *PLoS Computational Biology* **3**, e39.

Eden E., Navon R., Steinfeld I., Lipson D. & Yakhini Z. (2009) GORILLA: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48.

Field M. (2003) Intestinal ion transport and the pathophysiology of diarrhea. *Journal of Clinical Investigation* **111**, 931–43.

Goh K.I., Cusick M.E., Valle D., Childs B., Vidal M. & Barabási A.L. (2007) The human disease network. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8685–90.

Gouaux E. & Mackinnon R. (2005) Principles of selective ion transport in channels and pumps. *Science* **310**, 1461–5.

Gulbahce N., Yan H., Dricot A. *et al.* (2012) Viral perturbations of host networks reflect disease etiology. *PLoS Computational Biology* **8**, e1002531.

Jung K. & Saif L.J. (2015) Porcine epidemic diarrhea virus infection: etiology, epidemiology, pathogenesis and immunoprophylaxis. *Veterinary Journal* **204**, 134–43.

Jung K., Wang Q., Scheuer K.A., Lu Z., Zhang Y. & Saif L.J. (2014) Pathology of US porcine epidemic diarrhea virus strain PC21A in gnotobiotic pigs. *Emerging Infectious Diseases* **20**, 662–5.

Karlsson E.K., Baranowska I., Wade C.M. *et al.* (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nature Genetics* **39**, 1321–8.

Li B.X., Ge J.W. & Li Y.J. (2007) Porcine aminopeptidase N is a functional receptor for the PEDV coronavirus. *Virology* **365**, 166–72.

Moeser A.J. & Blikslager A.T. (2007) Mechanisms of porcine diarrheal disease. *Journal of the American Veterinary Medical Association* **231**, 56–67.

Nam E. & Lee C. (2010) Contribution of the porcine aminopeptidase N (CD13) receptor density to porcine epidemic diarrhea virus infection. *Veterinary Microbiology* **144**, 41–50.

Oh J.S., Song D.S. & Park B.K. (2003) Identification of a putative cellular receptor 150 kDa polypeptide for porcine epidemic diarrhea virus in porcine enterocytes. *Journal of Veterinary Science* **4**, 269–75.

Opriessnig T., Xiao C.T., Gerber P.F., Zhang J. & Halbur P.G. (2014) Porcine epidemic diarrhea virus RNA present in commercial spray-dried porcine plasma is not infectious to naïve pigs. *PLoS One* **9**, e104766.

Pensaert M.B. & de Bouck P. (1978) A new coronavirus-like particle associated with diarrhea in swine. *Archives of Virology* **58**, 243–7.

Purcell S., Neale B., Todd-Brown K. *et al.* (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, **81**, 559–75.

Riley S. (2007) Large-scale spatial-transmission models of infectious disease. *Science* **316**, 1298–301.

Rubin C.J., Zody M.C., Eriksson J. *et al.* (2010) Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature* **464**, 587–91.

Schneider D.S. & Ayres JS. (2008) Two ways to survive infection: what resistance and tolerance can teach us about treating infectious diseases. *Nature Reviews Immunology* **8**, 889–95.

Shibata I., Tsuda T., Mori M., Ono M., Sueyoshi M. & Uruno K. (2000) Isolation of porcine epidemic diarrhea virus in porcine cell cultures and experimental infection of pigs of different ages. *Veterinary Microbiology* **72**, 173–82.

Snelson H. (2014) *PEDv Update*. Retrieved from www.aasv.org/news/story.php?id=6989.

Song D. & Park B. (2012) Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemiology, diagnosis, and vaccines. *Virus Genes* **44**, 167–75.

Turgeon D.C., Morin M., Jolette J., Higgins R., Marsolais G. & DiFranco E. (1980) Coronavirus-like particles associated with diarrhea in baby pigs in Quebec. *Canadian Veterinary Journal* **21**, 100–xxiii.

Utiger A., Tobler K., Bridgen A. & Ackermann M. (1995) Identification of the membrane protein of porcine epidemic diarrhea virus. *Virus Genes* **10**, 137–48.

Weiss S.R. & Navas-Martin S. (2005) Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiology and Molecular Biology Reviews* **69**, 635–64.

Zhong Q., Simonis N., Li Q.R. *et al.* (2009) Edgetic perturbation models of human inherited disorders. *Molecular Systems Biology* **5**, 321.

Supporting information

Additional supporting information may be found online in the supporting information tab for this article:

Figure S1 Admixture plot considering $K = 8$ for the analyzed data.

Figure S2 Multidimensional scaling plot of all dead and recovered piglets considering the first two dimensions.

Figure S3 Distribution of the frequency of the ZF_{stm} values in 1-Mb windows calculated comparing the dead with the recovered suckling piglets.

Figure S4 Normalized F_{st} plot without the samples that were positive based on the ELISA test.

Table S1 Number of dead and recovered piglets from the PEDV outbreak in each considered farm.

Table S2 Information about the windows with $ZF_{stm} > 6$.

Table S3 Virus changes, upregulated and downregulated from the Gene Expression Omnibus (GEO).