

The spread of Type 2 Porcine Reproductive and Respiratory Syndrome Virus (PRRSV) in North America: A phylogeographic approach



Mang Shi^a, Philippe Lemey^b, Manreetpal Singh Brar^c, Marc A. Suchard^{d,e,f}, Michael P. Murtaugh^g, Susy Carman^h, Sylvie D'Allaireⁱ, Benjamin Delisleⁱ, Marie-Ève Lambertⁱ, Carl A. Gagnonⁱ, Li Ge^h, Yihan Qu^j, Dongwan Yoo^k, Edward C. Holmes^{a,l}, Frederick Chi-Ching Leung^{c,m,*}

^a Sydney Emerging Infections & Biosecurity Institute, School of Biological Sciences and Sydney Medical School, The University of Sydney, NSW 2006, Australia

^b Department of Microbiology and Immunology, Rega Institute, KU Leuven, B-3000 Leuven, Belgium

^c School of Biological Sciences, The University of Hong Kong, Hong Kong, China

^d Department of Biostatistics, Jonathan and Karin Fielding School of Public Health, University of California, Los Angeles, United States

^e Department of Biomathematics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, United States

^f Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, United States

^g Department of Veterinary and Biomedical Sciences, University of Minnesota, St. Paul, MN, United States

^h University of Guelph, Laboratory Services Division, Animal Health Laboratory, Box 3612, Guelph, ON, Canada N1H 6R8

ⁱ Faculté de Médecine Vétérinaire, Université de Montréal, C.P. 5000, St-Hyacinthe, Québec, Canada J2S 7C6

^j Faculty of Agriculture and Environment, University of Sydney, NSW 2006, Australia

^k College of Veterinary Medicine, University of Illinois at Urbana-Champaign, Urbana, United States

^l Fogarty International Center, National Institutes of Health, Bethesda, United States

^m Bioinformatics Center, Nanjing Agricultural University, Nanjing, China

ARTICLE INFO

Keywords:

Phylogeography

PRRSV

Evolution

Recombination

Molecular epidemiology

ABSTRACT

The emergence and spread of Type 2 Porcine Reproductive and Respiratory Syndrome virus (Type 2 PRRSV) in North America is heavily influenced by the multiple site production system used in the hog industry. However, it is unclear how anthropogenic factors such as this have shaped the current spatial distribution of PRRSV genotypes. We employed Bayesian phylogeographic analyses of 7040 ORF5 sequences to reveal the recent geographical spread of Type 2 PRRSV in North America. The directions and intensities in our inferred virus traffic network closely mirror the hog transportation. Most notably, we reveal multiple viral introductions from Canada into the United States causing a major shift in virus genetic composition in the Midwest USA that went unnoticed by the regular surveillance and field epidemiological studies. Overall, these findings provide important insights into the dynamics of Type 2 PRRSV evolution and spread that will facilitate programs for control and prevention.

© 2013 Elsevier Inc. All rights reserved.

Introduction

Porcine Reproductive and Respiratory Syndrome virus (PRRSV) is a positive-sense RNA virus that poses a continual threat to domestic swine populations. There are two genotypes of PRRSV, denoted Types 1 and 2. In North America, Type 2 PRRSV is the dominant genotype in the field (Shi et al., 2010a, 2010b). Over the past 20 years, the genetic diversity of Type 2 PRRSV has greatly expanded as a result of major changes in the modern swine industry (Murtaugh et al., 2010). A recent study classified Type 2 PRRSV into nine distinct

lineages based on a comprehensive collection of ORF5 sequences. With the exception of strains related to live attenuated vaccine strains the North America field viruses fall within two clades, denoted Lineages 1–2 and Lineages 6–9, respectively (Shi et al., 2010a). The diversity of Type 2 PRRSV is further expanded when field isolates from Ontario, Canada, are included in the phylogeny, as most of the non-vaccine PRRSVs from Ontario fall within Lineages 1–2 instead of Lineages 6–9 (Brar et al., 2011).

As a virus infecting domestic pigs, the spread of PRRSV over long distances is heavily influenced by human activities such as hog transportation and artificial insemination. This often causes the rapid and distance-independent dispersal of viruses (Goldberg et al., 2000; Shi et al., 2010a), and hence is an impediment to disease prevention and control. Our previous analysis of Type 2 PRRSV in the United States revealed high intensity movement

* Corresponding author at: School of Biological Sciences, The University of Hong Kong, Hong Kong, China. Tel.: +852 2299 0825; fax: +852 2857 4672.

E-mail address: fcleung@hkuc.hku.hk (F. Chi-Ching Leung).

surrounding the Midwest and highly asymmetrical movement in and out of North Carolina (Shi et al., 2010a). While these results imply that PRRSV dispersal is influenced by hog transportation within multi-site production systems, they are not intended to demonstrate any changes in regional genetic diversity that are associated with these dispersals, nor were they able to explain the uneven distribution of Lineage 1–2 within the United States compared to Lineage 6–9. In particular, previous studies lack PRRSVs samples from Canada and so cannot provide a complete picture of viral spread in North America.

In this study, we apply a Bayesian phylogeographic inference methodology to Type 2 PRRSV in North America to reveal the key patterns of viral spread (Lemey et al., 2009). Importantly, this analysis allows a formal examination of different models for viral dispersion within a comparative probabilistic framework. In addition, we constructed a framework that allows phylogeographic model parameters to be estimated simultaneously from multiple sets of phylogenies, in doing so overcoming the difficulties inherent in analyzing data sets of this scale. Hence, our new phylogeographic tool offers a unique opportunity to investigate the evolution and spread of Type 2 PRRSV in North America.

Results

Recombination detection and description

Six recombination events were identified using RDP (Table 2): three were inter-lineage recombination between Lineages 1 and 2; two involved vaccine strains as potential parental lineage(s); and the sixth was a recombination between two sublineages within Lineage 1. The number and location of breakpoints were determined in similarity plots, which revealed similar breakpoint locations for distinct recombination events. For example, recombinant groups R1 and R2 had a common breakpoint at nt 350 (Fig. 1A–C); whereas MB2573 and QC123–QC131 shared two breakpoint locations: one at nt 126, and the other at around nt 430 (Table 2). These positions therefore appear to be potential hot spots for recombination within ORF5.

Notably, three of the six recombination events (i.e. R1, R2 and R3) had abundant recombinant descendants with a wide range of isolation dates (Table 2; Fig. 1E and F), suggesting they had been circulating for an extensive time period. Geographically, these recombination variants were identified in different regions of Canada and the United States. Their parental groups, however, were all identified to be Ontario field strains (Table 2), which suggests Ontario as the most likely location of the initial recombination events.

Genotyping of PRRSV circulating in Canada

Among the 1130 Canadian PRRSV analyzed here, 69.6% were genotyped as Lineage 1–2 (including the three recombinant groups

R1, R2 and R3), with the remainder belonging to Lineage 5–9 (Table 1). However, further detailed classification suggested that 99.2% of the Lineage 5–9 Canadian sequences belonged to two of the vaccine-related sublineages (i.e. MLV and ATP-related). In contrast, only two sequences from Ontario and two from Manitoba were classified within the non-vaccine diversity of Lineages 8 and 9 (Table 1). Surprisingly, the majority of the Manitoba sequences were vaccine related (93.8%). The remainder fell at diverse locations on the phylogenetic trees including Lineages 1, 2, 8, and 9. Finally, no Canadian sequences were genotyped as Asian lineages.

A reference phylogenetic tree was inferred after inclusion of novel genetic variants (Fig. 2). The tree topology was largely congruent to those produced previously (Brar et al., 2011; Shi et al., 2010a), differing only in the placement of Lineage 2, although this was not supported by strong posterior probabilities (Fig. 2). Although the backbone remained largely unchanged, the genetic diversity of Lineages 1–2 was greatly expanded with the inclusion of the additional previously uncharacterized Canadian sequences. Within these two lineages, the USA clusters (Sublineages 1.1–1.9 and part of Lineage 2) identified from previous studies were all nested within the extent of genetic variation observed in Canada (Fig. 2), indicating an earlier establishment of these two lineages in Canada. Indeed, the ancestral locations for the internal nodes were all inferred to be Canadian ($P > 0.99$) in the ancestral state reconstruction analyses.

Within Lineages 1–2, the Canadian PRRSV population was further subdivided into large and geographically specific clades established in either Ontario or Quebec (Fig. 2). Specifically, Lineage 2 Canadian sequences were almost exclusively from Ontario (93.3%), whereas Lineage 1 had three large sub-clades: two from Ontario and one from Quebec. Nevertheless, despite the general population subdivision between the two provinces, more than one Ontario cluster was found within the Quebec clade, and vice versa (Fig. 2). These clusters, although limited in sample size, clearly represent viral traffic of PRRSV between the two provinces.

Recent (late 1990s–2010) PRRSV circulation network in North America

We used Bayes factors to compare four phylogeographic models which describe different scenarios for the spread of PRRSV in North America (Table 3). Among the four models tested, the stepping stone model which disallows virus flow between non-neighboring state/provinces yielded significantly lower marginal likelihood support compared to unconstrained models of viral spread, suggesting relatively frequent virus movements between non-neighboring states/provinces. Among the non-stepping stone models, that allowing asymmetric virus movement between neighboring regions/provinces (i.e. the irreversible model) showed significant improvement over the reversible model which does not, indicating that the direction of movements between geographic regions was highly asymmetric.

A more comprehensive view of PRRSV circulation was obtained through estimating the instances of virus flow between pairs of regions in the ancestral histories. Indeed, virus circulation was highly asymmetrical among regions (Fig. 3). In addition, Lineages 1–2 exhibited significantly different dynamics compared to Lineages 6–9. In Lineages 1–2, viral circulation was centered in Ontario and the Midwestern regions of United States (i.e. Lake States, Corn Belt, and Northern Plains), with apparently greater virus flow from Ontario to the Corn Belt and Lake States regions and among the Corn Belt, Lake States and Northern Plains regions (Fig. 3A). In contrast, PRRSV seldom traveled from the United States to Canada, which suggests that, on these data, virus movement between the two countries was largely unidirectional. Interestingly, cross-country viral gene flow occurred

Table 1
Genotyping of Canadian PRRSV samples.

Lineage	Major distribution	Ontario	Quebec	Manitoba
1–2	Canada, USA	499	273	15
3–4	Asia	0	0	0
5 ^a	Cosmopolitan	115 (115)	30 (30)	200 (200)
6–9 ^b	USA, Asia	31 (29)	5 (5)	58 (56)
Total		645	308	273

^a Number of samples related to vaccine strain IngelvacMLV are shown in parenthesis.

^b Number of samples related to vaccine strain IngelvacATP are shown in parenthesis.

Table 2
PRRSV recombination events documented in this study.

Recombinant strain/lineage	No. of sequences (year; place)	Breakpoint position (nt)	Parental region (A/B)	Parental lineage (sublineage)	Closest strain (year; place)
R1	70 (2002~2007; NT, CB)	350	1–350 (A) 351–603 (B)	Lineage 1 (SL 1.1–1.4) Lineage 2	G9838583 (1998; ON) G9938224 (1998; ON)
R2	64 (1998~2010; ON, QC, CB, LS)	350	1–350 (A) 351–603 (B)	Lineage 1 (SL 1.1–1.4) Lineage 2	G9838583 (1998; ON) R9800215 (1998; ON)
R3	39 (2003~2009; ON, QC)	390	1–390 (A) 391–603 (B)	Lineage 1 (SL 1.1–1.4) Lineage 2	G0511959 (2005; ON) G9938224 (1998; ON)
MB1118940	1 (2008; MB)	212, 510	1–212, 510-603 (A) 213–509 (B)	Lineage 8 (ATP related) Lineage 5 (MLV related)	MB1185008 (2009; MB) MB1017521 (2007; MB)
MB2573	1 (2002; MB)	126, 434	1–126, 434-603 (A) 127–433 (B)	Lineage 5 (MLV related) Lineage 1	MB1163474 (2009; MB) MB6912 (2005; MB)
QC123, QC131	2 (2007; QC)	126, 421	1–126, 421-603 (A) 127–420 (B)	Lineage 1 (SL 1.6) Lineage 1 (SL 1.7–1.9)	QC078 (2006; QC) QC068 (2006; QC)

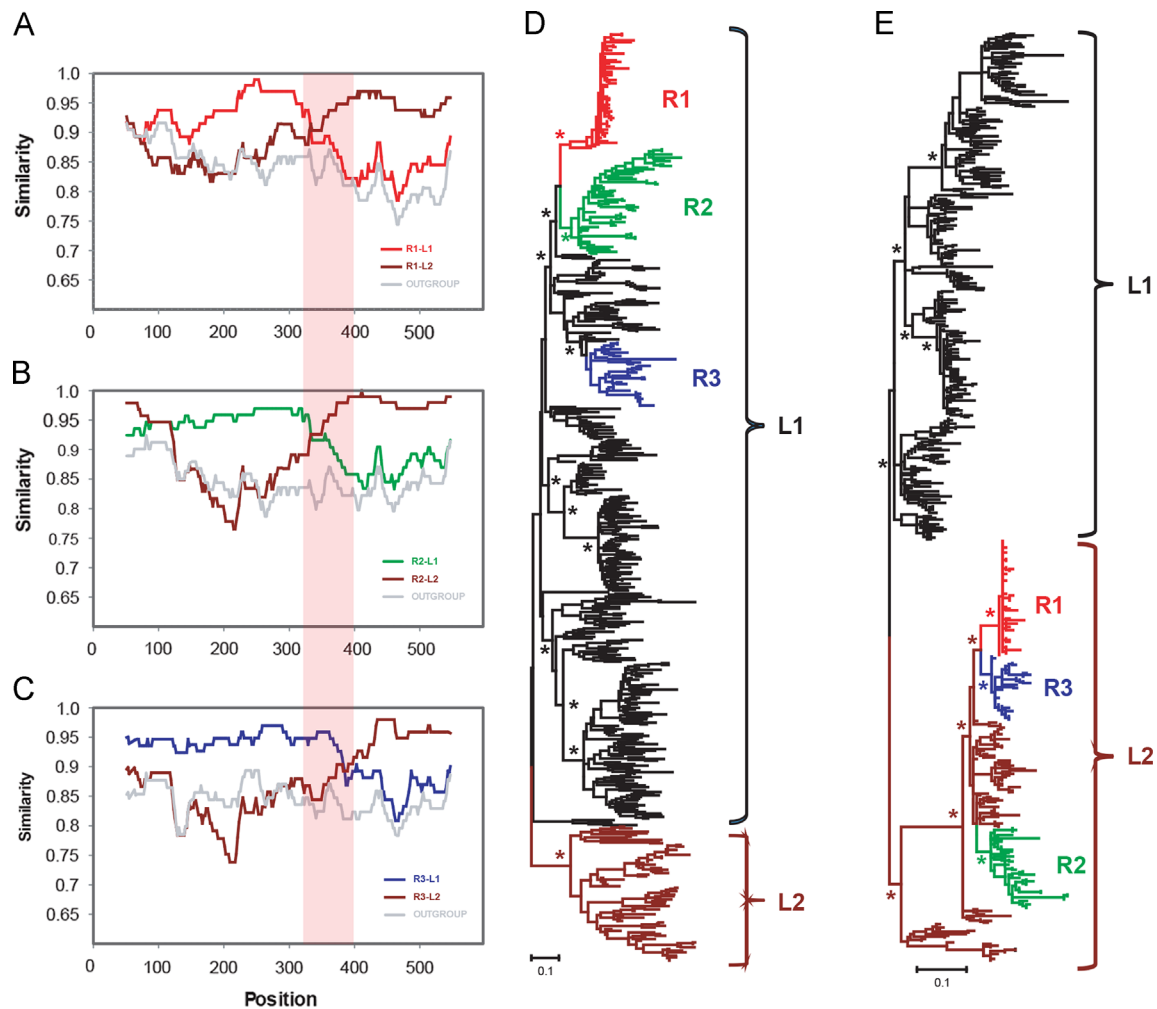


Fig. 1. Characterization of three recombinant events between PRRSV Lineages 1 and 2. The similarity plots generated by SIMPLOT are shown on the left panel of the figure: (A) Recombinant Group 1, (B) Recombinant Group 2, and (C) Recombinant Group 3. Within each plot, different colored lines represent the similarity comparisons of the recombinant to its two potential parental strains as well as an outgroup (gray). The red shaded area marks the potential location of recombination breakpoints. The phylogenetic trees of (D) parental region A (1–350 nt) and (E) parental region B (400–603 nt) of the three recombinants are shown on the right panel of the figure. Asterisks depict nodes with greater than 0.9 posterior probability support. For clarity, only Lineages 1 and 2 ORF5 sequences are included in the phylogenetic analyses. The three recombinant groups are labeled with different colors to distinguish from each other and from their parental groups. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

more frequently in the USA than among Canadian provinces, compatible with the more localized nature of swine management in Canada. Unlike Lineages 1–2, Lineages 6–9 virus circulation did not involve Canada. Within the United States intensive viral movement was

observed among all the major hog production regions (Fig. 3B). Of these regions, Appalachia had highly asymmetric virus movements, primarily unidirectional to the Corn Belt, compatible with its role as an exporting center of feeder pigs. Indeed, the virus traffic in Lineages 6–9

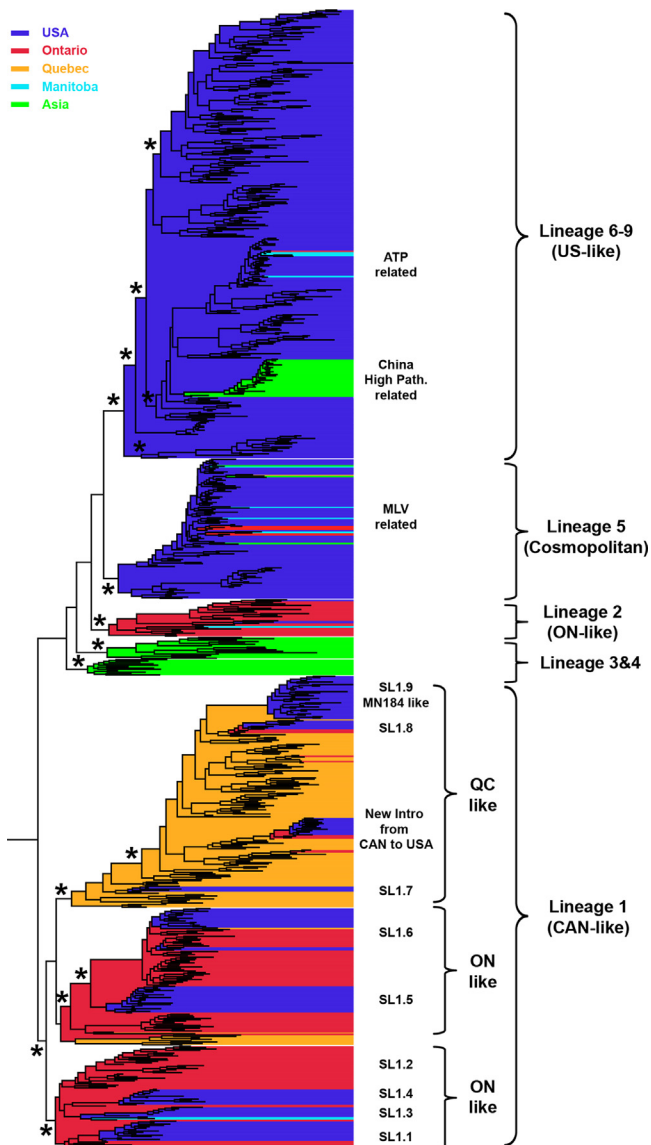


Fig. 2. Overall reference phylogenetic tree based on ORF5 sequences from different lineages of Type 2 PRRSV ($n=818$). The tree was inferred using MrBayes and was mid-point rooted for clarity only. Asterisks depict nodes with greater than 0.9 posterior probability support. The phylogeny is shaded with different colors that indicate the most likely ancestral locations of specific clades. Strain classification information and the annotations of the vaccine sublineages are shown on the left of the phylogeny.

Table 3

Bayes Factor (BF) comparisons of different phylogeographic models for the spread of PRRSV.

Model ^a	Ln marginal likelihood	Log BF ^b
Reversible	−6526.594	76.62
Reversible step	−6703.011	–
Irreversible	−6206.51	215.63
Irreversible step	−6379.891	140.33

^a Four different phylogeographic models were tested: reversible, reversible step, irreversible and irreversible step.

^b Log Bayes factors are used to compare reversible step (the most restricted model) to the other phylogeographic models. The best-fit (irreversible) model is highlighted in bold.

was almost a mirror image of the hog movement within the USA, as shown by the significant association between the virus flow and hog flow ($R=0.71$ and $P<0.0001$). A similar association was seen for

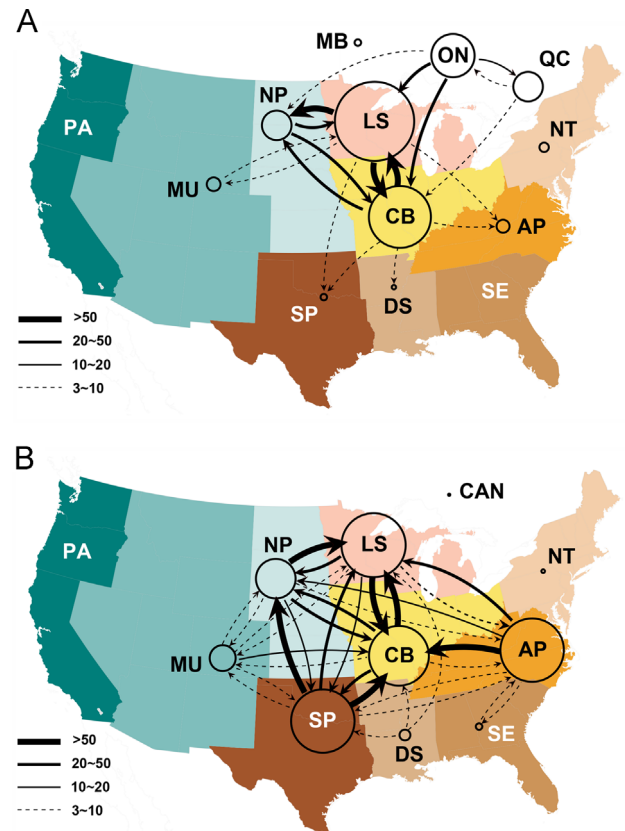


Fig. 3. Schematic showing North American PRRSV circulation among swine production regions, including CAN (Canada), Manitoba (MB), Ontario (ON), Quebec (QC), AP (Appalachia), CB (Corn Belt), LS (Lake States), MU (Mountain), NP (Northern Plain), NT (Northeast), and SP (Southern Plain). The size of the circle reflects the PRRSV sample size in each region, while the thickness of lines with arrows indicates the frequency of inter-regional virus flows. The picture of inter-regional PRRSV circulation was separated into two parts: (A) Lineages 1–2 PRRSV flows, where both Canada and United States were involved; and (B) Lineages 5–9 PRRSV flows, almost exclusively within the United States.

Lineages 1–2, although with a little weaker statistical support ($r=0.45$ and $P<0.0001$).

A shift in the predominant viral genotype in the Midwest United States

In the last 10 years, a significant shift in the genetic composition of non-vaccine related PRRSV occurred in the Midwest regions of the United States (Fig. 4A). Before 2000, Lineages 6–9 sequences were clearly dominant in Midwest regions, while Lineages 1–2 sequences were rare. However, after 2001, the sampling frequency of Lineages 1–2 sequences gradually increased. The trend continued uninterrupted, and as a result 77% of the most recent samples collected from Midwest regions belong to Lineages 1–2. The same trend was observed for each of the three regions (Fig. 4B–D), and the most pronounced shift was noted in the Lake States where the percentage of Lineages 6–9 dropped to 15% in the most recent sampling. Nevertheless, the shift in genotype composition from Lineages 6–9 to Lineages 1–2 was limited to the Midwest regions only. For other regions in the United States, Lineages 6–9 remained the dominant PRRSV variant over the past decade (data not shown).

PRRSV movements between Canada and the United States

We further tested the hypothesis of unidirectional PRRSV flow from Canada to United States within a likelihood framework.

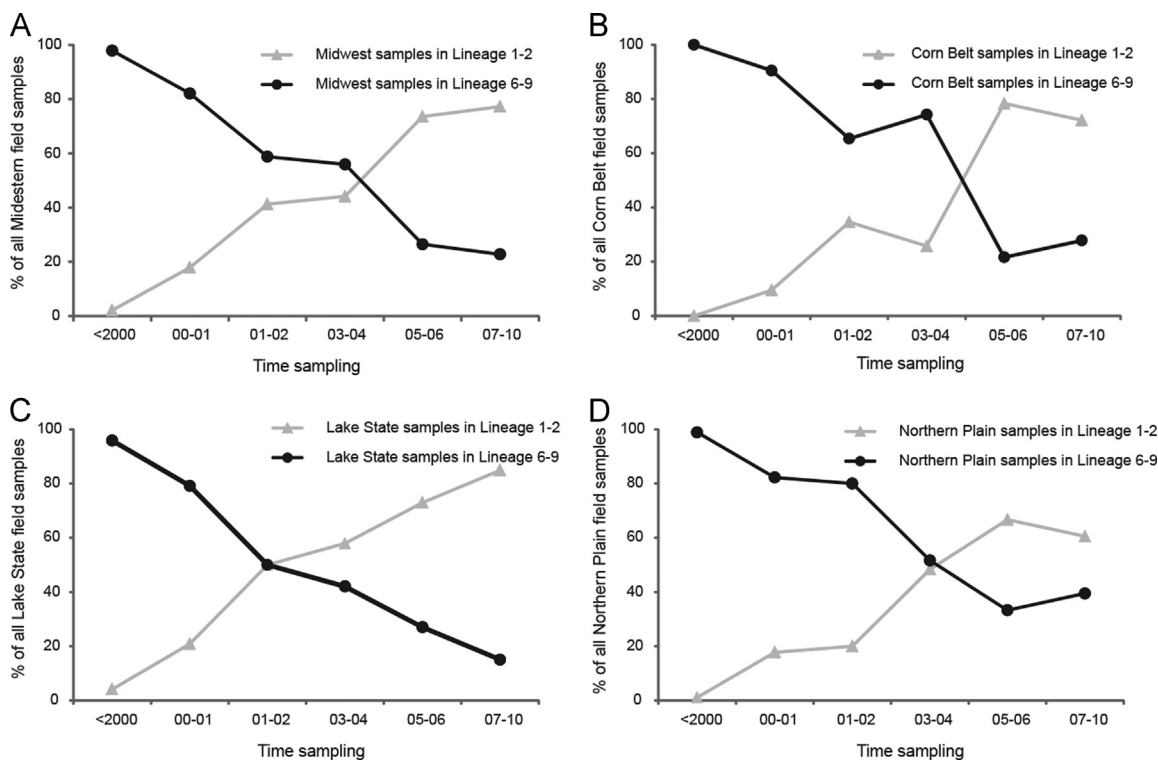


Fig. 4. Shift in the genetic composition of Type 2 PRRSV in (A) the Midwest (including the Corn Belt, Lake States and Northern Plains regions), (B) Corn Belt, (C) Lake States, and the (D) Northern Plains over time. Within each subplot, the proportion of Lineages 1–2 viruses in all field samples is represented by a gray triangle while the proportion of Lineages 6–9 viruses is represented by black dots.

Table 4
Comparisons of models describing PRRSV flow between Canada and USA.

Model	Parameters ^a	Estimates	Log likelihood ^b
One rate (symmetrical)	$r_{CU}=r_{UC}$	$r=0.5467$	-237.0108
Unidirectional from USA to Canada	$r_{CU}=0, r_{UC}$	$r_{UC}=1.4678$	-358.095
Unidirectional from Canada to USA	$r_{CU}, r_{UC}=0$	$r_{CU}=1.1157$	-213.2359**
Two rates (asymmetrical)	r_{CU}, r_{UC}	$r_{CU}=1.0897$ $r_{UC}=0.0982$	-212.2933**

^a Rate of PRRSV flow, r_{CU} represents the rate from Canada to USA, r_{UC} the opposite direction.

^b Models that fit significantly better compared to the One rate model in a likelihood ratio test ($P < 0.00001$) are indicated with.**

Among the four models tested, both the unidirectional model (from Canada to the United States) and the two rates model showed significantly better fit to the data compared to the equal rate model (Table 4). Hence, the pattern of virus dispersal is clearly asymmetric. Moreover, the likelihood of the two-rate model was not significantly higher than the unidirectional model (from Canada to the United States), which suggests that virus flow is mostly unidirectional. Hence, the virus movement from the United States to Canada is rare compared to that occurring in the opposite direction.

Our BEAST phylogenetic analysis suggested that of the numerous introductions from Canada to United States, many died out quickly (data not shown). However, a few introductions whose transmission chains were maintained over time accounted for most of the infections by Canada-like (Lineages 1–2) PRRSV. Introductions from Canada that led to successful transmission chains in the USA are summarized on a time-line in Fig. 5. These introductions first took place in the late 1990s and have occurred sporadically ever since. Of particular interest is a recent introduction that occurred around mid-2007, and although this cluster is set within the Quebec-like clade, its immediately ancestor can be

traced back to Ontario field isolates which bore the RFLP pattern 1-3-4 or 1-4-4 (Fig. 2). Upon introduction, this variant expanded quickly in both genetic diversity and geographic range. Moreover, a few isolates from the Appalachia region (North Carolina) were also found in this cluster, which suggests this variant also established a pattern of transmission and circulation in Appalachia.

Discussion

We employed Bayesian phylogeographic analyses of 7040 North American PRRSV samples to reveal the geographic dispersion of Type 2 PRRSV in North America. Our results reveal that the current distribution of PRRSV genetic diversity was formed in two stages: (i) the virus population was subdivided geographically into ‘Canada-like’ (Lineages 1–2) and ‘USA-like’ (Lineages 6–9) clades in earlier years, perhaps because of limited movement of viruses between the Canada and the United States at that time and (ii) having established this phylogeographic subdivision, Canada-like viruses were then introduced on several occasions into the United States beginning in the mid-1990s until the present day. The first

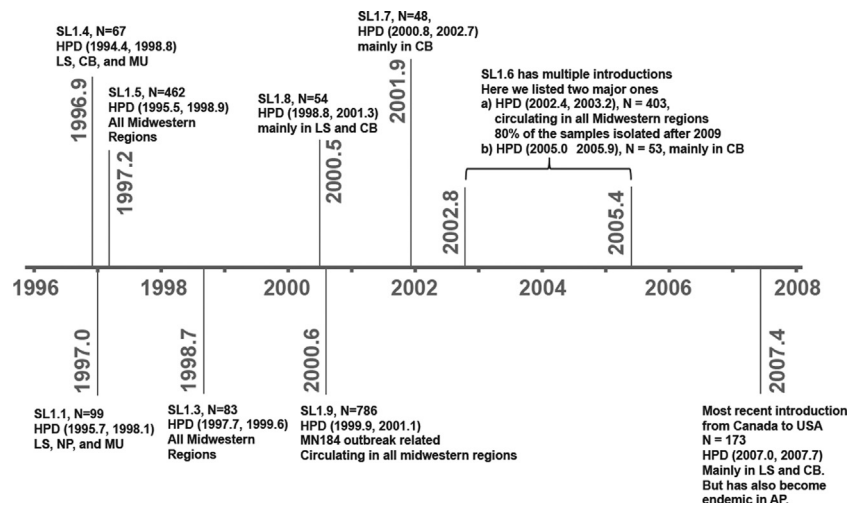


Fig. 5. A time-line of PRRSV introductions from Canada that have established successful transmission chains in the United States. The introduction dates are the inferred times to Most Recent Common Ancestry (tMRCAs, given as fractional years) of monophyletic clusters of USA viruses. Each introduction was labeled with sampling size, the 95% Highest Posterior Density (HPD) intervals of tMRCAs, and geographic distributions.

stage of early geographic separation is supported by observations that the ancestral nodes of Lineages 6–9 and Lineages 1–2 are inferred to be in United States and Canada, respectively (Fig. 2), and that despite continuous sampling since 1991, Lineages 1–2 viruses were not observed in the United States until 1998 (Han et al., 2006) with all earlier members of these two lineages obtained from Canada (Shi et al., 2010a). The unidirectional nature of viral gene flow from Canada to the United States is particularly striking with, for example, multiple USA clusters found in Lineages 1–2 (Canadian origin) while only a few Canadian field isolates clustered within Lineages 6–9 (United States origin).

A key question is to determine what could have caused the largely unidirectional PRRSV flow from Canada to the United States and why has this cross-border circulation occurred only recently. Strikingly, the patterns of virus flow between Canada and the United States bear a strong resemblance to the hog movements between the two countries. There are a substantial number of growing pigs imported from Canada to the United States, driven by economic incentives to reduce the cost of feeder pigs and capitalize on favorable foreign exchange rates (Shields and Mathews Jr., 2003). The destinations are low-cost feeder sites mainly in Corn Belt and Lake States. Interestingly, these regions are also the destination for viruses introduced from Ontario, as depicted from our reconstructed PRRSV circulation network. Moreover, the timing of increased hog shipment from Canada (statistics retrieved from: <http://www5.statcan.gc.ca/cimt-cicm/>) is compatible with the time when Canadian-type PRRSV started to appear in the United States. In contrast, the hog exportation out of the United States has never occurred in large numbers (Shields and Mathews Jr., 2003). Hence, our findings suggest that hog transportation can have a major effect on virus movement between the two countries.

Also of note in this context is fact that Manitoba, one of the most prominent hog exporters in Canada (see: <http://www5.statcan.gc.ca/cimt-cicm/>), does not play a major role in our reconstructed PRRSV circulation network, despite the large number of Manitoba ORF5 sequences analyzed here. It is most likely due to the fact that most of these Manitoba sequences are genotyped to vaccine-associated lineages (Lineage 5 and Sublineage 8.9). To reveal the circulation network from vaccine-associated lineages is impracticable because (i) it is extremely difficult to distinguish vaccine-derived from field viruses and (ii) the dynamics of vaccine-derived viruses are different from that of field ones. Therefore, although we excluded vaccine-associated

lineages from the reconstructed network, we cannot exclude that Manitoba was involved in the PRRSV spread between Canada and the United States through the medium of hog transportation.

In some areas there appears to be a link between virus spread and hog transportation that relates to the widely implemented inter-regional multiple-site hog production system. Like Ontario and Manitoba (Shields and Mathews Jr., 2003), Appalachia is another major region for exporting feeder pigs and whose virus outflow greatly exceeds inflow (Fig. 3). Conversely, the Corn Belt, which represents the major destination for hogs shipped from all other regions, has the highest frequency of virus in-flows, and also serves as a locality where diverse viral strains accumulate and are then transmitted to other regions. Overall, the correlation between virus flow and hog flow is high ($r=0.71$ and $P<0.0001$ for Lineages 6–9; $r=0.45$ and $P<0.0001$ for Lineages 1–2), indicating hog flow has substantial predictive power for the pattern of virus spread. The lower correlation coefficient for Lineages 1–2 is most likely due to the fact that hog transportation includes southern and eastern regions (especially Southern Plain and Appalachia) in the United States, but virus traffic does not extend to these regions. Indeed, the ‘Canada-like’ viruses are relatively scarce in all US regions with the exception of the Midwest.

A role for hog transportation on viral phylogeography has been observed for a number of other swine pathogens, notably porcine circovirus 2 (PCV2) (Firth et al., 2009) and swine influenza virus (Nelson et al., 2011). In the case of PCV2, phylogeographic studies identified Canada as the major exporter of the viruses while the United States as a major importer (Firth et al., 2009) – a direction that agrees well with the findings of our study although with limited and biased sampling. As for swine influenza virus, the impact of hog transportation is reflected in the observation that the virus flow from southern United States into the Midwest is much greater than in the opposite direction (Nelson et al., 2011). Again, a similar pattern is reflected in our estimated PRRSV circulation network (Fig. 3). Collectively, these studies on unrelated viruses validate our results from PRRSV and hence underscore how the human-mediated transportation of animals can have a major impact on disease distributions.

An important finding of our study is the major impact of Canada-like PRRSV on mid-western regions of the United States. Not only did these viruses become endemic, but over the past 10 years they have gradually replaced the original local virus populations (i.e. Lineages 6–9 or USA-like) as the dominant type. In the USA, Lineages 1–2 genotypes (e.g. RFLP 1-8-4 and 1-22-2 types),

appear to be more virulent than endemic Lineages 6–9 genotypes, growing to higher levels in pigs and are more readily shed in aerosols (Han et al., 2006). It is not known if the virulent characteristics were present at the time of introduction or were acquired afterwards. The genetic variant responsible for RFLP 1–8–4 outbreak (SL 1.9) was never found in Canada, suggesting its post-introduction emergence. In either event, the responsible variants are more virulent, grow to higher levels in pigs and are more readily shed in aerosols (Han et al., 2006), characteristics that are consistent with increased fitness (17). In addition, it is possible that these strains are less susceptible to immunological protection by current attenuated vaccines. However, further evidence is required to examine whether vaccine-based control of PRRS is better in regions without Lineages 1–2 strains than with the strains.

Currently, the PRRSVs of Canadian origin are mainly circulating in Canada and the Midwest regions of the United States. Although some have spread to the Northeast, Appalachia, and Southern Plain, most transmission chains in these regions appear to have resulted in dead-ends, and others are still small in scale as judged from the currently available sampling. It is possible that some PRRSV strains may eventually become endemic, displacing the local variant and becoming dominant in these regions, as has occurred already in the upper Midwest with Canada-like PRRSV. A new shift in genetic composition might take place when an infectious variant emerges in one region and is transmitted rapidly through the network and thereafter initiates invasion dynamics outwardly. Although prediction is inherently difficult, it is helpful to recognize the source and direction of spread for disease control measures. To achieve this, thorough and continuous surveillance of virus genetic sequences from different geographic locations is needed.

Finally, our study revealed the circulation of inter-lineage recombinant strains in the field. Although recombination events between field PRRSV strains are not uncommon, most are ‘singleton’ events (i.e. represented by single sequences) (Shi et al., 2010b). However, we show that a small number of ‘circulating recombinants’ became established and may be expanding in geographic range, indicating that they are at least of the same fitness as their parental types. Indeed, two of these circulating recombinants have spread from their place of origin (Ontario) to Midwest regions and Northeast in the United States (Table 2). Since the recombination breakpoints are located within ORF5 it is possible that they also exhibit antigenic differences. However, in the absence of full genomic data it is not possible to fully understand the role played by recombination in the maintenance and transmission of PRRSV.

Materials and methods

Sequence data collection

To analyze the phylogeography of North America Type 2 PRRSV in as much detail as possible, we expanded a previous data set ($n=8624$) of ORF5 sequences (Shi et al., 2010a) to include a large-scale ORF5 sequence collection from Canada as well as a recent update of the available sequence data from the United States. The Canadian sequences (Table 1) were mainly collected from the provinces of Manitoba ($n=273$), Quebec ($n=308$), and Ontario ($n=645$), and a proportion have been described previously (Brar et al., 2011; Delisle et al., 2012). These Canadian sequences were provided by the Animal Health Laboratory of the University of Guelph and the Faculty of Veterinary Medicine of the University of Montreal. The sequence update from the United States ($n=2374$) was downloaded from the PRRSV database (<http://prsvdb.org/>) in

February 2011. Most of these USA sequences were uploaded by the Minnesota Veterinary Diagnostic Laboratory, the South Dakota Animal Disease Research & Diagnostic Laboratory, and the Iowa State University Veterinary Diagnostic Laboratory during the years 2009–2011. All sequences were aligned using the MUSCLE software package v3.6 (Edgar, 2004) with default settings, followed by manual adjustment.

Genotyping

To incorporate 3600 new sequences into the established classification system (Shi et al., 2010a), a simple similarity based approach was used. We first calculated the genetic distances between each query and a set of reference sequences provided by Shi et al. (2010a) for which lineage and sublineage information was known; once a highly similar sequence was found among the reference sequences, we assigned the reference lineage/sublineage classification to that query. To avoid the case where one query was typed to more than one lineage/sublineage, we set the similarity thresholds to 95% and 97.5% for lineage and sublineage level typing, respectively. Each classified query sequence was subsequently added to the reference sequence set. This process was repeated until no new sequence could be added to the system. Sequences that could not be genotyped were regarded as novel/undescribed strains/clusters. From these sequences, representative but non-recombinant sequences were selected and combined with the previously described reference data set ($n=550$) (Shi et al., 2010a) to form a new reference data set ($n=818$, Supplementary Material 1). Based on this new reference data set, a reference tree was constructed using a Bayesian inference method implemented in MrBayes v3.2 (Ronquist et al., 2012) under a general time-reversible (GTR) model with discrete gamma distributed rate variation among sites (Γ_4) and a proportion of invariable sites (I). The lineage/sublineage assignment of novel sequences was performed after inspecting the topology of the reference tree. Bayesian (MrBayes) phylogenetic trees were also inferred for updated sublineage level alignments as well as for novel diversities.

Recombination analyses

The new sequences in this study were combined with the reference data set for recombination screening. We used three methods implemented in the Recombination Detection Program (RDP) v4.16 (Martin et al., 2010) to identify potential recombinants – MaxChi (Smith, 1992), Chimaera (Posada, 2002), and Geneconv (Padidam et al., 1999). Recombination events detected by all three methods with default parameters were considered as potential recombination events. To confirm the recombination signal and to estimate the approximate breakpoint locations, selected sequences of the recombinant and parental lineages were extracted for similarity plot analysis implemented in SIMPLOT v3.5.1 (Lole et al., 1999), in which sequence similarity is plotted against alignment position using a sliding window procedure. The window size was set to 100 nt and the step size to 2 nt, and the genetic distances were calculated using the Kimura 2-Parameter substitution model (Kimura, 1980). To investigate the phylogenetic origins of different regions of the potential recombinants, a maximum likelihood (ML) phylogeny (employing the GTR+ Γ_4 +I substitution model) (Guindon et al., 2010) was inferred for each non-recombinant region. If a recombinant contained multiple breakpoints, the regions for which the query sequence shares the same parental lineages were concatenated before phylogeny estimation.

Biogeographic regions

The North American continent was divided into 13 regions based on geography and commonality in farm production activities. They

regions are; Ontario (ON), Quebec (QC) and Manitoba (MB) from Canada, and Lake States (LS), Corn Belt (CB), Northern Plains (NP), Northeast (NT), Appalachia (AP), Southeast (SE), Delta States (DS), Southern Plains (SP), Mountain (MU), and Pacific (PA) from the United States (Shields and Mathews Jr., 2003). These regions were also determined based on USDA description of agricultural regions (Shields and Mathews Jr., 2003). Note, we used 'regions' instead of 'states' to avoid the problem of over-parameterization in the model-based phylogeographic analyses (Sanmartin et al., 2010; Yang, 2006). The Canadian provinces were treated as separate geographic regions since they have little overlap in hog production systems.

Data set for phylogeographic analyses

To characterize PRRSV flow in North America, we adapted the sublineage level ORF5 sequence alignments for phylogeographic analysis. ORF5 is used because it has the largest and most comprehensive sequence collection from North America. From our complete data set ($n=12,329$), we first excluded all vaccine-associated sublineages, including Sublineage 5.1 (MLV-associated) and Sublineage 8.9 (ATP-associated), as the more structured distribution of vaccine compared to field strains may complicate our phylogeographic analyses. Second, relatively diverse sublineages with many representatives were subdivided based on the tree topology into two or more smaller alignments to increase computational speed. Third, for all sublineages we excluded early (before 1997) sequences which did not represent the recent transmission dynamics in North America. Fourth, non-North American sublineages ($n=3$) and orphan sublineages ($n=2$) within a single geographic region were excluded from the analyses. Finally, recombinant sequences ($n=4$) were not included unless the entire cluster was originated from a single recombination event. In total, we obtained 42 well supported 'monophyletic clusters' from our sublineage level trees, most of which covered a period from early 2000s to 2010. Each cluster contained 100–250 sequences which in sum produced a total data set of 7040 sequences in the phylogeographic analyses.

BEAST phylogenetic analysis

We first estimated time-scaled posterior tree distributions for each of the 42 monophyletic clusters using the Markov chain Monte Carlo (MCMC)-based Bayesian analysis implemented in the BEAST software package v1.7.5 (Drummond et al., 2012). In all cases, we used the SRD (HKY112+CP112+Gamma112) substitution model (Shapiro et al., 2006), an uncorrelated lognormal relaxed molecular clock model, and Bayesian Skyride coalescent prior. The MCMC sampling was run until convergence could be safely assumed after an initial burn-in period. We used Tracer v1.5 to monitor the mixing of MCMC and to ensure all model parameters have effective sample sizes (ESS) larger than 200. The tree distributions were subsequently used as empirical tree samples for Bayesian phylogeographic analyses as described in (Lemey et al., 2012).

Model-based phylogeographic analyses for inter-regional viral movement

To analyze inter-regional PRRSV movement, we used a Bayesian phylogeographic inference approach implemented in BEAST. Within this framework, four models were established representing different biogeographic scenarios for the spread of PRRSV: (i) a reversible model (Rev) in which dispersal rates varied between different pairs of regions, but with a single rate determining the intensity of movement between each pair independent of the directionality (i.e. the rate from A to B is the same as that from B to

A); (ii) an irreversible model (Irrev), in which the virus dispersal rate was both region specific and directional; and (iii and iv) stepping stone and sequential-dispersal models, where dispersal rates between non-adjacent regions are constrained to be zero, indicative of no viral traffic between these regions. The dispersal rates between adjacent regions were modeled as symmetric and asymmetric resulting in the 'Rev step' and 'Irrev step' models, respectively.

Under these models, the phylogeography of all ($n=7040$) North America PRRSV field samples was inferred across the 42 sets of empirical tree samples and model fit was assessed using harmonic mean estimates of the marginal likelihood (Suchard et al., 2003). The posterior expected numbers of inter-regional PRRSV transitions between pairs of regions were subsequently summarized from node location estimates in the posterior distribution of phylogeographic histories estimated under the best-fit model by Bayesian model test. The dispersal processes for Lineages 1–2 and 6–9 were analyzed separately.

Correlation between PRRSV flow and hog flow

We used a Pearson product-moment correlation coefficient to measure the correlation between PRRSV flow and hog flow. PRRSV flow was represented by estimated expected numbers of inter-regional PRRSV transitions between pairs of regions, whereas the hog flow was inferred from USDA records of inter-regional hog shipments (available at <http://naldc.nal.usda.gov/download/38888/PDF>). Unfortunately, we do not have contemporaneous information on the number of hog transportations between Canada and the United States. Therefore all correlation analyses are restricted to the United States alone.

Model-based phylogeographic analyses for inter country virus movement

A likelihood-based approach (Pagel, 1994) was used to examine the extent of gene flow of Lineages 1 and 2 between Canada and the United States. To reduce the size of the Lineages 1 and 2 data set, we clustered the highly similar sequences (>98%) from the same country and randomly chose one to represent each cluster. This exercise resulted in an alignment of 723 sequences (434 from USA and 289 from Canada), for which we inferred an ML tree using PhyML. To test the hypothesis of virus flow between the two countries, we reduced the phylogeographic regions to two (i.e. Canada and the United States), and set up four models with different constraints on the transition rate: (1) One rate model: same dispersal rate in both directions; (2) Unidirectional USA: the dispersal rate from Canada to the United States was set to zero; (3) Unidirectional Canada: the dispersal rate from the United States to Canada was set to zero; and (4) Two rate model: separate rate parameters for two directions. The comparisons of models were carried out using likelihood ratio tests.

Acknowledgments

This work was supported by grants from the Research Grant Council of the Hong Kong Government (GRF770612) and HKU Strategic Research Theme of Infection and Immunology to FCL and from the European Community's Seventh Framework Programme (FP7/2007–2013) under Grant Agreement no. 278433 and ERC (Grant agreement no. 260864) to PL. MS is supported by an International Postgraduate Research Scholarship from the University of Sydney, while ECH is supported by an NHMRC Australia Fellowship. Special thanks are given to Kay Faaberg, National

Animal Disease Center, Ames, IA, and Trevor Wennblom, University of Minnesota, Minneapolis, MN, for database maintenance.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.virol.2013.08.028>.

References

- Brar, M.S., Shi, M., Ge, L., Carman, S., Murtaugh, M.P., Leung, F.C., 2011. Porcine reproductive and respiratory syndrome virus in Ontario, Canada 1999 to 2010: genetic diversity and restriction fragment length polymorphisms. *J. Gen. Virol.* 92 (6), 1391–1397.
- Delisle, B., Gagnon, C.A., Lambert, M.E., D'Allaire, S., 2012. Porcine reproductive and respiratory syndrome virus diversity of Eastern Canada swine herds in a large sequence dataset reveals two hypervariable regions under positive selection. *Infect. Genet. Evol.* 12 (5), 1111–1119.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29 (8), 1969–1973.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32 (5), 1792–1797.
- Firth, C., Charleston, M.A., Duffy, S., Shapiro, B., Holmes, E.C., 2009. Insights into the evolutionary history of an emerging livestock pathogen: Porcine circovirus 2. *J. Virol.* 83 (24), 12813–12821.
- Goldberg, T.L., Hahn, E.C., Weigel, R.M., Scherba, G., 2000. Genetic, geographical and temporal variation of porcine reproductive and respiratory syndrome virus in Illinois. *J. Gen. Virol.* 81 (1), 171–179.
- Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59 (3), 307–321.
- Han, J., Wang, Y., Faaborg, K.S., 2006. Complete genome analysis of RFLP 184 isolates of porcine reproductive and respiratory syndrome virus. *Virus Res.* 122 (1–2), 175–182.
- Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16 (2), 111–120.
- Lemey, P., Rambaut, A., Bedford, T., Faria, N., Bielejec, F., Baele, G., Russell, C.A., Smith, D.J., Pybus, O.G., Brockmann, D., Suchard, M.A., 2012. The seasonal flight of influenza: a unified framework for spatiotemporal hypothesis testing. (<http://arxiv.org/pdf/1210.5877.pdf>), 12 December, 2012.
- Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A., 2009. Bayesian phylogeography finds its roots. *PLoS Comput. Biol.* 5 (9), e1000520.
- Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus Type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J. Virol.* 73 (1), 152–160.
- Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuve, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26 (19), 2462–2463.
- Murtaugh, M.P., Stadejek, T., Abrahante, J.E., Lam, T.T., Leung, F.C., 2010. The ever-expanding diversity of porcine reproductive and respiratory syndrome virus. *Virus Res.* 154 (1–2), 18–30.
- Nelson, M.I., Lemey, P., Tan, Y., Vincent, A., Lam, T.T.Y., Detmer, S., Viboud, C., Suchard, M.A., Rambaut, A., Holmes, E.C., Gramer, M., 2011. Spatial dynamics of human-origin H1 influenza A virus in North American swine. *PLoS Pathol.* 7 (6), e1002077.
- Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. *Virology* 265 (2), 218–225.
- Pagel, M., 1994. Detecting correlated evolution on phylogenies – a general method for the comparative-analysis of discrete characters. *Proc. R. Soc. London B Biol. Sci.* 255 (1342), 37–45.
- Posada, D., 2002. Evaluation of methods for detecting recombination from DNA sequences: empirical data. *Mol. Biol. Evol.* 19 (5), 708–717.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61 (3), 539–542.
- Sanmartin, I., Anderson, C.L., Alarcon, M., Ronquist, F., Aldasoro, J.J., 2010. Bayesian island biogeography in a continental setting: the Rand Flora case. *Biol. Lett.* 6 (5), 703–707.
- Shapiro, B., Rambaut, A., Drummond, A.J., 2006. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Mol. Biol. Evol.* 23 (1), 7–9.
- Shi, M., Lam, T.T., Hon, C.C., Murtaugh, M.P., Davies, P.R., Hui, R.K., Li, J., Wong, L.T., Yip, C.W., Jiang, J.W., Leung, F.C., 2010a. Phylogeny-based evolutionary, demographical, and geographical dissection of North American Type 2 porcine reproductive and respiratory syndrome viruses. *J. Virol.* 84 (17), 8700–8711.
- Shi, M., Lam, T.T.Y., Hon, C.C., Hui, R.K.H., Faaborg, K.S., Wennblom, T., Murtaugh, M.P., Stadejek, T., Leung, F.C., 2010b. Molecular epidemiology of PRRSV: a phylogenetic perspective. *Virus Res.* 154 (1–2), 7–17.
- Shields, D.A., Mathews Jr., K.H., 2003. Interstate livestock movements. United States Department of Agriculture-Economic Research Service Outlook Report, LDPM10801, USDA, Washington, DC, (<http://www.ers.usda.gov/publications/ldp/jun03/ldpm10801/ldpm10801.pdf>), (accessed 22.08.09).
- Smith, J.M., 1992. Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34 (2), 126–129.
- Suchard, M.A., Kitchen, C.M., Sinsheimer, J.S., Weiss, R.E., 2003. Hierarchical phylogenetic models for analyzing multipartite sequence data. *Syst. Biol.* 52 (5), 649–664.
- Yang, Z., 2006. Computational Molecular Evolution. Oxford University Press, Oxford.