Virology **289,** 174–179 (2001)
doi:10.1006/viro.2001.1102, available online at http://www.idealibrary.com on IDEAL®

RAPID COMMUNICATION

A Molecular Clock Dates the Common Ancestor of European-type Porcine Reproductive and Respiratory Syndrome Virus at More Than 10 Years before the Emergence of Disease

Roald Forsberg,* Martin B. Oleksiewicz,‡ Anne-Mette Krabbe Petersen,† Jotun Hein,*
Anette Bøtner,‡ and Torben Storgaard‡[1]

*Department of Ecology and Genetics and †Department of Statistics, University of Aarhus, Denmark; and
‡Danish Veterinary Institute for Virus Research, Lindholm, Denmark

Received May 1, 2001; returned to author for revision June 25, 2001; accepted July 19, 2001

The disease caused by porcine reproductive and respiratory syndrome virus (PRRSV) emerged independently and almost simultaneously in Europe (1990) and North America (1987). The original reservoir of the virus and the date it entered the pig populations is not known. In this study, we demonstrate an accurate molecular clock for the European PRRSV ORF 3 gene, place the root in the genealogy, estimate the rate of nucleotide substitution, and date the most recent common viral ancestor of the data set to 1979; more than 10 years before the onset of the European epidemic. Based on these findings, we conclude that PRRSV virus most likely entered the pig population some time before the epidemic emergence of the virus, and hence, that emergence of European-type PRRSV is not the result of a recent species transmission event. Together, our results show that ORF3 sequencing is a valuable epidemiologic tool for examining the emergence and spread of PRRSV in Europe. As such, the panel of well-characterized and highly divergent ORF3 sequences described in this study provides a reference point for future molecular epidemiologic studies.

Key Words: molecular clock; PRRSV; ORF 3; evolution; molecular epidemiology.

INTRODUCTION

Porcine reproductive and respiratory syndrome virus (PRRSV) is the causative agent of a newly emerged disease with still on-going epidemics in populations of domestic pigs (*1*). There is no prior knowledge in veterinary records of this disease, which emerged almost simultaneously in North America and Western Europe in the late 1980s and early 1990s, respectively. Despite the concurrence in emergence and disease symptoms, the virus strains from the two continents have proven to be strikingly different with only 55–70% nucleotide similarity in the different viral genes (*2–5*). This has led to the conclusion that the two lineages must have evolved separately from a very distant common ancestor before the emergence into the pig populations on the two continents (*5, 6*). The novelty of the disease, and the evolutionary distance between the two virus lineages, has prompted the working hypothesis that the emergence of the virus was a result of two recent and independent species transmission events on the two continents (*6*). However, the original reservoir of the virus prior to the onset of the current epidemic, and whether this reservoir was the same on both continents, is unknown. It is also unknown when a potential species shift may have occurred from the original reservoir to domesticated pigs.

The ORF 3 gene of European-type PRRSV encodes a virion-associated structural glycoprotein (*7*). A previous study has shown a highly accurate molecular clock for the accumulation of base substitutions in this gene within the Danish virus population (*8*). This was done by including time information into an approach based on linear regression on genetic distances. However, due to phylogenetic correlation, the statistical properties of distance-based methods are not well known, and caution must be taken when basing conclusions on these.

In the current study, we expanded the original data set of Oleksiewicz *et al.* (*8*) to incorporate the European PRRSV prototype, the Lelystad isolate (*3*), and seven English isolates previously sequenced (*9*). In addition, the complete ORF 3 from five recent Danish isolates and five Italian isolates were sequenced as part of this study and included in the analysis (Table 1). The two latter countries have been shown to harbor the majority of European-type PRRSV genetic diversity (*8, 10* and T. Storgaard, unpublished results).

Virus isolation, RNA extraction, and sequencing procedures were essentially as described in Oleksiewicz *et al.* (*8*), differing only in a change of the reverse primer for

[1] To whom correspondence and reprint requests should be addressed at Novo Nordisk A/S, Novo Nordisk Park, 2760 Måløv, Denmark. Fax: +45 44 43 45 58. E-mail: TStS@novonordisk.com.

0042-6822/01 $35.00

TABLE 1

Details of the Isolates Used

Isolate	Country	Isolation date	Accession number	Journal reference
111/92	Denmark	09/04/92	AF171671	Oleksiewicz *et al.*, 2000
48/92-1	Denmark	20/03/92	AF171690	Oleksiewicz *et al.*, 2000
14474B	Denmark	04/03/96	AF171675	Oleksiewicz *et al.*, 2000
12654	Denmark	19/12/95	AF171672	Oleksiewicz *et al.*, 2000
12985	Denmark	04/01/96	AF171673	Oleksiewicz *et al.*, 2000
5767-6	Denmark	10/02/95	AF171692	Oleksiewicz *et al.*, 2000
20567 A	Denmark	26/03/97	AF171678	Oleksiewicz *et al.*, 2000
21191	Denmark	09/05/97	AF171679	Oleksiewicz *et al.*, 2000
24554/97	Denmark	19/11/97	AF171680	Oleksiewicz *et al.*, 2000
25434/98	Denmark	15/01/98	AF171681	Oleksiewicz *et al.*, 2000
28639/98	Denmark	16/06/98	AF171684	Oleksiewicz *et al.*, 2000
32-10/92	Denmark	17/03/92	AF171687	Oleksiewicz *et al.*, 2000
26666	Denmark	18/03/98	AF171683	Oleksiewicz *et al.*, 2000
34/92	Denmark	17/03/92	AF171688	Oleksiewicz *et al.*, 2000
38/8	Denmark	18/03/92	AF171689	Oleksiewicz *et al.*, 2000
18009/4	Denmark	28/09/92	AF171677	Oleksiewicz *et al.*, 2000
146/92	Denmark	24/09/92	AF171676	Oleksiewicz *et al.*, 2000
228 A	Denmark	16/11/93	AF171691	Oleksiewicz *et al.*, 2000
5544 A	Denmark	01/02/95	AF171696	Oleksiewicz *et al.*, 2000
6501	Denmark	22/03/95	AF171693	Oleksiewicz *et al.*, 2000
6504	Denmark	22/03/95	AF171694	Oleksiewicz *et al.*, 2000
6617	Denmark	29/03/95	AF171695	Oleksiewicz *et al.*, 2000
13759 B	Denmark	26/01/96	AF171674	Oleksiewicz *et al.*, 2000
26371 B	Denmark	04/03/98	AF171682	Oleksiewicz *et al.*, 2000
31540	Denmark	18/11/98	AF171685	Oleksiewicz *et al.*, 2000
31690 A	Denmark	25/11/98	AF171686	Oleksiewicz *et al.*, 2000
32413	Denmark	08/01/99	AF303361	This study
32351	Denmark	05/01/99	AF303360	This study
32929	Denmark	01/02/99	AF303362	This study
33792	Denmark	11/03/99	AF303363	This study
34229	Denmark	26/03/99	AF303365	This study
3391/93	Italy	24/11/93	AF303364	This study
2481/97	Italy	24/04/97	AF303359	This study
2029/97	Italy	08/04/97	AF303358	This study
3943/96	Italy	21/06/96	AF303366	This study
974/98	Italy	10/02/98	AF303367	This study
BE1	United Kingdom	18/02/93	L77913	Drew *et al.*, 1997
H3	United Kingdom	21/06/91	L77915	Drew *et al.*, 1997
NO1	United Kingdom	25/03/92	L77923	Drew *et al.*, 1997
LE1	United Kingdom	11/05/92	L77921	Drew *et al.*, 1997
L2	United Kingdom	5/10/92	L77919	Drew *et al.*, 1997
HA1	United Kingdom	18/06/92	L77917	Drew *et al.*, 1997
NY4	United Kingdom	22/06/94	L77925	Drew *et al.*, 1997
Lelystad	Netherlands	1991	M96262	Meulenberg *et al.*, 1993

the N-terminal amplification to 5′ATGCGTCGAGAAA-CAGGGCT3′ to successfully amplify this more diverse sample set of PRRSV isolates. The sequencing of the five Italian isolates identified ORF 3 deletion mutants similar to those previously reported among Danish PRRSV isolates (*8*). In addition, two of the Italian isolates (2029/97 and 3943/96, Table 1) had premature stop codons in the carboxyl-terminal part of ORF 3. This observation, together with our previous published data (*8*), suggest that the carboxyl-terminal part of the European type PRRSV ORF 3 has a nonessential role in the viral life cycle.

The newly generated sequences corresponding to the complete ORF 3 (Table 1) were aligned to the previously published ORF 3 sequences (*3, 8, 9*) using Clustal X (*11*), and adjusted manually around the region previously described to hold deletions (*8*). All sites in the alignment containing gaps were removed, yielding a final alignment of 747 nucleotide sites for analysis [alignment available via the Internet (www.bioinf.au.dk/~roald/)]. Based on this alignment, the genealogy of the isolates was estimated using the PAUP* program (*12*). A neighbor-joining tree was constructed upon which maximum likelihood (ML) parameters of the HKY+ Γ substitution model were estimated (*13, 14*). This model incorporates the

previously shown transition bias in PRRSV (*15*) and rate heterogeneity among sites, which is a common feature of protein coding regions (*14*). Estimated parameters were then used in an extensive heuristic topology search under the ML criterion. Ten random stepwise addition trees were used as the starting points for the search. The searches converged on five topologies, which only had very minor branching order differences of essentially zero-length branch lengths in the Lelystad-like clade and the very compact clades of Danish isolates (Fig. 1).

For hypothesis testing and parameter estimation, we used a genealogy-based likelihood approach (*13*), which incorporates the correlation of shared evolutionary history between isolates. Recent advances in the construction of models of molecular evolution have allowed for the inclusion of time-information from noncontemporaneous samples into a maximum likelihood framework (*16*). This permits the testing of hypotheses about the relation between calendar time and the rate of evolution, and estimation of substitution rate parameters, all within the statistically well-founded likelihood theory. The models employed here (Table 2) are implemented in the TipDate program (*16*) and are as follows: (i) A model of an unconstrained tree $M_{UNCONST}$ (Table 2). Under this model no assumptions are made about the amount of calendar time occurring on each branch of the tree, or the constancy of substitution rates (*13*). The lengths of all branches are therefore allowed to vary as free parameters of the model. (ii) A model of a clock-constrained tree with time information M_{CLOCK} (*16*). This model incorporates the time information of the isolates (Table 2) and positions tips at a distance from the root, which is dependent on their date of origin and the simultaneously estimated substitution rate parameter (μ). The interdependency caused by these constraints means that all branch lengths are no longer free to vary. These can now be expressed by the height of the nodes in the tree, which means a reduction of the parameter space. To introduce the clock constraints into the model, a root must be placed in the relating tree. As no close outgroup exists, the root of the complete genealogy was estimated by trying all possible positions. The placing of the root that yielded the highest likelihood score was then chosen.

A likelihood ratio test between these two models then becomes a test of the molecular clock hypothesis. As shown in Table 2, the clock hypothesis did not yield a significantly worse fit to the data when compared to the more parameter-rich hypothesis of an unconstrained tree. This result agrees with the previous results from Oleksiewicz *et al.* (*8*) in demonstrating an accurate molecular clock for the evolution of the ORF 3 gene in European-type PRRSV. The rate of substitution was estimated as 5.8×10^{-3} substitutions per site per year (95% confidence interval $4.8–6.9 \times 10^{-3}$). This high rate of substitution and the large number of sequences allowed us to estimate with high fidelity that the most recent common ancestor (MRCA) of these isolates existed in 1979 (95% confidence interval covers the period 1976–1981). The presence of PRRSV in European domestic swine has not been demonstrated prior to the onset of the current epidemic. If the current pan-European epidemic was caused by only one interspecies transmission event followed by rapid intraspecies transmission, we would expect the sequences to be related by a star-like genealogy, with an estimated time of the MRCA around 1990. From the dated genealogy (Fig. 1), it can be seen that several lineages existed around the onset of the current epidemic in 1990. This means that several transmission chains must have persisted and connected the pre-epidemic reservoir of the virus to the viral isolates of this sample set. Two scenarios could explain this pattern: (i) The pre-epidemic reservoir of PRRSV was another species, e.g., mice (*6*) or mallard ducks (*17*), from which several interspecies transmissions to domesticated pigs occurred at the time of or following the onset of the epidemic in 1990; or (ii) PRRSV acquired pigs as hosts a long time prior to the current epidemic and circulated and diversified in an isolated pig population; from this population genetically diverse viruses were introduced into the larger European swine populations by several intraspecies transmission events, hereby giving rise to the current pan-European epidemic. At the molecular level, a strong dependency on the specific host-cell environment is observed in the gene products involved in all viral life stages (*18*). This observation, and the fact that interspecies transmission of viruses are rarely observed, would indicate that species shifts are very rare events in the evolutionary history of most viruses and that the second of the above scenarios is the most likely explanation for our observations.

From the genealogy (Fig. 1) it can be seen that both the viral isolates within the clade containing only Danish isolates and the viral isolates within the clade containing the Lelystad-like isolates diverged around the onset of the current epidemic. However, the Italian isolates all diverged even earlier and harbor much of the genetic diversity of the sample set. High genetic diversity among Italian PRRSV isolates has previously been observed in the ORF 5 and 7 genes (*10* and T. Storgaard, unpublished results). The high genetic diversity of Italian PRRSV isolates indicates that more of the genetic diversity of the

FIG. 1. Genealogy of European-type PRRSV isolates. Branch lengths are optimized under the clock model M_{CLOCK}. Geographical location of isolates is indicated: DK, Denmark; IT, Italy; NL, The Netherlands; UK, U.K. The bar shows time in years. The most recent common ancestor (MRCA) and the onset of the European PRRSV epidemic are indicated by arrows on the bar.

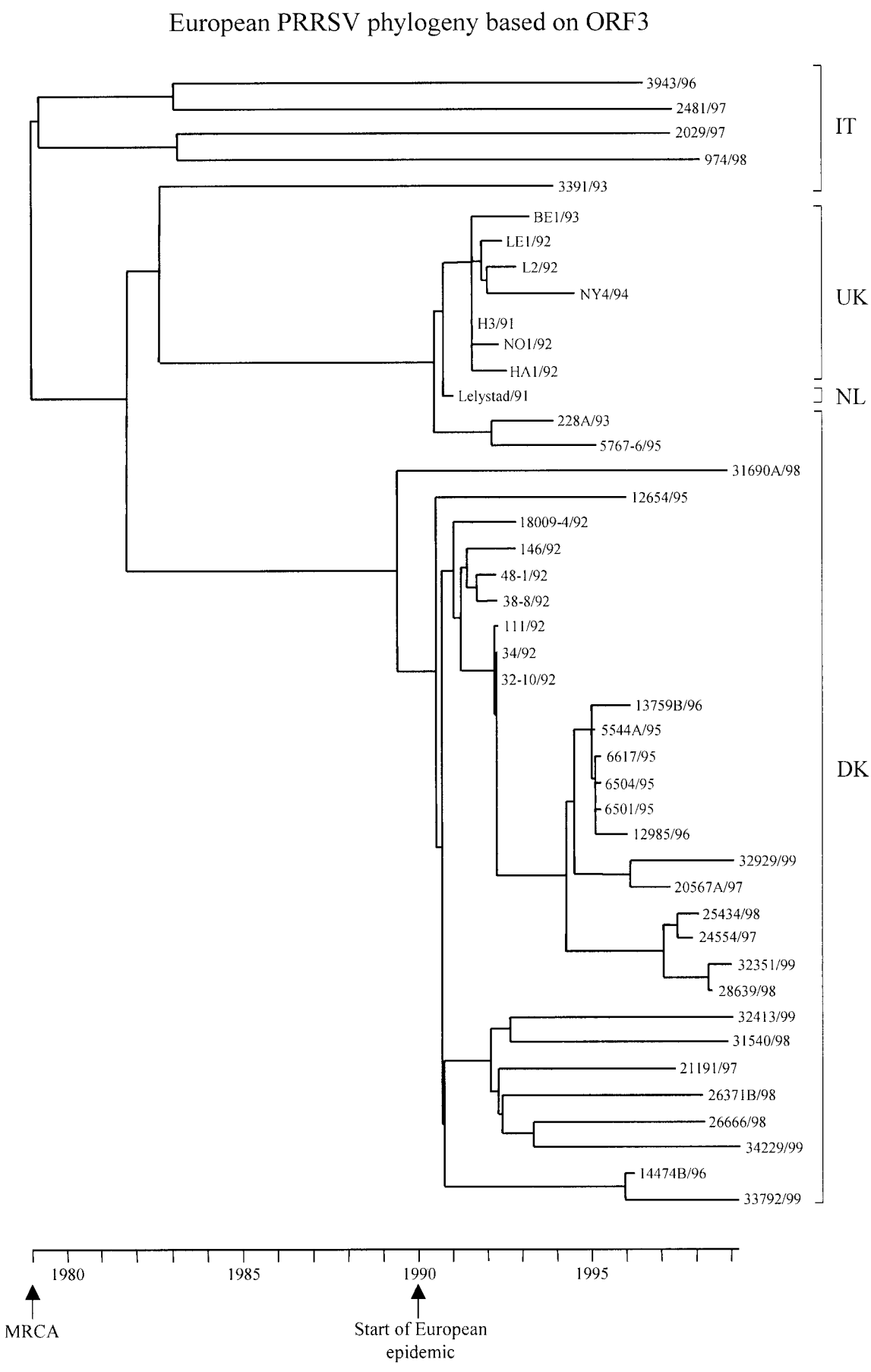
European PRRSV phylogeny based on ORF3
3943/96
2481/97
2029/97
974/98
IT
3391/93
BE1/93
LE1/92
L2/92
NY4/94
H3/91
NO1/92
HA1/92
UK
Lelystad/91
NL
228A/93
5767-6/95
31690A/98
12654/95
18009-4/92
146/92
48-1/92
38-8/92
111/92
34/92
32-10/92
13759B/96
5544A/95
6617/95
6504/95
6501/95
12985/96
32929/99
20567A/97
25434/98
24554/97
32351/99
28639/98
32413/99
31540/98
21191/97
26371B/98
26666/98
34229/99
14474B/96
33792/99
DK
1980
1985
1990
1995
MRCA
Start of European epidemic

TABLE 2

Models of Evolution

Model	Model parameters[a]	d.f.[b]	Log-likelihood[c]
M_{CLOCK}	T, α, κ, μ, $h_1 \ldots h_{n-1}$	$n = 44$	−4442.43
$M_{UNCONST}$	T, α, κ, $b_1 \ldots b_{2n-3}$	$2_{n-3} = 85$	−4423.94

[a] T, topology of the genealogy; α, rate parameter of the discrete gamma distribution; κ, kappa parameter of the HKY model describing transition bias; μ, substitution rate parameter; $h_1 \ldots h_{n-1}$, heights of internal nodes; $b_1 \ldots b_{2n-3}$ branch lengths.

[b] d.f., degrees of freedom; n, number of sequences.

[c] Two times the difference in the log-likelihoods was calculated and evaluated in a χ^2_{41} distribution ($P = 0.65$).

original reservoir has been transferred to the Italian swine population than to the remaining European swine populations investigated here. This would suggest that the swine population of Italy has been more exposed to introduction of viruses from the original pig reservoir due to, e.g., geographical proximity or trading pattern. Hence, we find it likely based on the current study that PRRSV circulated in an isolated pig population for some time before the onset of the current Western European epidemic. In this isolated swine population the virus diverged and was later transferred to the large Western European swine populations on several occasions. These transmission events seeded the current epidemic of the disease in Western Europe, and for unknown reasons, transmission events seem to have been most prevalent in the Italy.

The finding that the MRCA existed in an isolated population more than 10 years prior to the emergence of PRRSV-induced disease in domestic pigs is by no means surprising given the current understanding of emerging viral infections such as HIV (*19*).

As in all studies that date evolutionary events by a molecular clock, the estimation procedure invoked here is based on the assumption that the constant rate of substitution observed in the sampling interval (1991–1999) may be extrapolated back in time. If the rate of substitution differed between the sampling interval and the presampling interval (before 1991), this would affect the estimate of the date of the MRCA. A higher rate of substitution in this interval would bring the estimated time of the MRCA closer to the emergence of the epidemic in 1990 and potentially question our hypothesis, and a slower rate would bring the date of the MRCA further back in time and be conservative to our hypothesis. The rate we have estimated and extrapolated back in time to date the MRCA corresponds to the highest rates previously published for glycoproteins of other RNA viruses such as HIV and Influenza A (*20, 21*). Hence, a rate much higher than this in the presampling interval does not seem plausible. Furthermore, the estimated date of the MRCA may be moved even further back as more genetic diversity of the European-type PRRSV is discovered. Our estimate of 1979 as the age of the MRCA is therefore very conservative.

In summary, the current study confirmed the presence of a highly accurate molecular clock in the ORF3 gene previously reported by Oleksiewicz *et al.* (*8*) and extended the findings to also include Italian, English, and Dutch isolates. This allowed us to estimate that the most recent common ancestor of European-type PRRSV existed before 1981. We conclude that if European-type PRRSV is a result of an interspecies transmission to pigs from an unknown host, it is most likely, due to the presumed rarity of such events, that this interspecies transmission took place before 1981. Based on these findings, we suggest that investigations aimed at unraveling the origins of PRRSV should have two foci: (i) to identify candidate species which may have harbored PRRSV before it acquired pigs as hosts and factors which may have facilitated a potential species shift. Such studies should focus on the time period prior to the date (1981) of the MRCA reported here, and (ii) to investigate potential reservoirs where the virus may have existed in the period between a potential species shift and the onset of the current epidemic, and any environmental factors that may have facilitated the transmission out of these reservoirs. Such studies should focus on the time period from the MRCA (1981) until the start of the PRRSV epidemic in domestic pigs (late 1980s) and factors that may explain the geographical difference in genetic diversity observed. To have any explanatory power for a species shift and the transmission of PRRSV out of the putative pre-epidemic reservoir, candidate factors should have a global effect to explain the concurrent epidemics in North America and Europe.

ACKNOWLEDGMENTS

P. Normann is thanked for technical assistance and M. H. Schierup is thanked for helpful discussions. P. Cordioli and G. Sala are acknowledged for providing Italian PRRSV isolates. Part of this work was done while Roald Forsberg visited the laboratory of Dr. Allen Rodrigo at the School of Biological Sciences, University of Auckland. This work was partly supported by a grant from the Danish Agricultural and Veterinary Research Council (9901522) given to Torben Storgaard.

REFERENCES

1. Snijder, E. J., and Meulenberg, J. J. (1998). The molecular biology of arteriviruses. *J. Gen. Virol.* **79,** 961–979.
2. Meng, X. J., Paul, P. S., Halbur, P. G., and Lum, M. A. (1995). Phylogenetic analyses of the putative M (ORF 6) and N (ORF 7) genes of porcine reproductive and respiratory syndrome virus (PRRSV): Implication for the existence of two genotypes of PRRSV in the U.S.A. and Europe. *Arch. Virol.* **140,** 745–755.
3. Meulenberg, J. J. M., Hulst, M. M., De Meijer, E. J., Moonen, P. L. J. M., Den Besten, A., De Kluyver, E. P., Wensvoort, G., and Moormann, R. J. M. (1993). Lelystad virus, the causative agent of porcine epidemic abortion and respiratory syndrome (PEARS), is related to LDV and EAV. *Virology* **192,** 62–72.
4. Murtaugh, M. P., Elam, M. R., and Kakach, L. T. (1995). Comparison

of the structural protein coding sequences of the VR-2332 and Lelystad virus strains of the PRRS virus. *Arch. Virol.* **140,** 1451–1460.

5. Nelsen, C. J., Murtaugh, M. P., and Faaberg, K. S. (1999). Porcine reproductive and respiratory syndrome virus comparison: Divergent evolution on two continents. *J. Virol.* **73,** 270–280.
6. Plagemann, P. G. W. (1996). "Field's Virology" (B. N. Fields, D. M. Knipe, and P. M. Howley, Eds.), 3rd ed., pp. 1105–1120. Lippincott-Raven Publishers, Philadelphia, PA.
7. Nieuwstadt, A. V., Meulenberg, J. M., Essen-Zandbergen, A. V., Petersen-den, B. A., Bende, R. J., Moormann, R. M., Wensvoort, G., Van, N. A., and Van-Essen, Z. A. (1996). Proteins encoded by open reading frames 3 and 4 of the genome of Lelystad virus (*Arteriviridae*) are structural proteins of the virion. *J. Virol.* **70,** 4767–4772.
8. Oleksiewicz, M. B., Botner, A., Toft, P., Grubbe, T., Nielsen, J., Kamstrup, S., and Storgaard, T. (2000). Emergence of porcine reproductive and respiratory syndrome virus deletion mutants: Correlation with the porcine antibody response to a hypervariable site in the ORF 3 structural glycoprotein. *Virology* **267,** 135–140.
9. Drew, T. W., Lowings, J. P., and Yapp, F. (1997). Variation in open reading frames 3, 4 and 7 among porcine reproductive and respiratory syndrome virus isolates in the UK. *Vet. Microbiol.* **55,** 209–221.
10. Suarez, P., Zardoya, R., Martin, M. J., Prieto, C., Dopazo, J., Solana, A., and Castro, J. M. (1996). Phylogenetic relationships of European strains of porcine reproductive and respiratory syndrome virus (PRRSV) inferred from DNA sequences of putative ORF-5 and ORF-7 genes. *Virus Res.* **42,** 159–165.
11. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997). The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25,** 4876–4882.
12. Swofford, D. L. (2000). PAUP*. Phylogenetic analysis using parsimony (*and other methods), Version 4 ed. Sinauer Associates, Sunderland, MA.
13. Hasegawa, M., Kishino, H., and Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22,** 160–174.
14. Yang, Z. (1993). Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* **10,** 1396–1401.
15. Storgaard, T., Oleksiewicz, M., and Botner, A. (1999). Examination of the selective pressures on a live PRRS vaccine virus. *Arch. Virol.* **144,** 2389–2401.
16. Rambaut, A. (2000). Estimating the rate of molecular evolution: Incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* **16,** 395–399.
17. Zimmerman, J. J., Yoon, K. J., Pirtle, E. C., Wills, R. W., Sanderson, T. J., and McGinley, M. J. (1997). Studies of porcine reproductive and respiratory syndrome (PRRS) virus infection in avian species. *Vet. Microbiol.* **55,** 329–336.
18. Strauss, J. H., and Strauss, E. G. (1999). Viral RNA replication. With a little help from the host. *Science* **283,** 802–804.
19. Zhu, T., Korber, B. T., Nahmias, A. J., Hooper, E., Sharp, P. M., and Ho, D. D. (1998). An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature* **391,** 594–597.
20. Leitner, T., and Albert, J. (1999). The molecular clock of HIV-1 unveiled through analysis of a known transmission history. *Proc. Natl. Acad. Sci. USA* **96,** 10752–10757.
21. Sugita, S., Yoshioka, Y., Itamura, S., Kanegae, Y., Oguchi, K., Gojobori, T., Nerome, K., and Oya, A. (1991). Molecular evolution of hemagglutinin genes of H1N1 swine and human influenza A viruses. *J. Mol. Evol.* **32,** 16–23.