COMPARATIVE MOLECULAR EPIDEMIOLOGY AND MOLECULAR EVOLUTION OF SWINE VIRUSES IN THE MIDWESTERN UNITED STATES

Goldberg TL[1], Weigel RM[1], Hahn EC[1], Scherba G[1]

[1]University of Illinois, Department of Veterinary Pathobiology, 2001 South Lincoln Avenue, Urbana IL. 61820

Modes of transmission of swine viral diseases between farms, over distance and time, are poorly understood. Molecular epidemiology has the ability to offer insights into the epidemic spread of such diseases. This study employs various molecular epidemiological techniques, including restriction fragment length polymorphism analysis (RFLP), nucleic acid sequencing and phylogeographic inference, to the study of two economically-important swine viral diseases in the midwestern United States: porcine reproductive and respiratory syndrome virus (PRRSV) and pseudorabies virus (PrV).

Fifty-five PRRSV field isolates and 16 PrV isolates were analyzed. In each virus, an immunodominant membrane-associated glycoprotein was chosen for gene sequencing (the ORF5 protein in PRRSV and the gC gene in PrV). Additionally, the entire PrV genome was analyzed using RFLP. Traditional epidemiological data were also collected, including geographic locations of farms from which samples were collected, dates of sample collection, species of origin of the sample (PrV only), and data on the clinical aspects of the disease on the farms.

Overall levels of genetic variability differed markedly between the viruses, with PRRSV ORF5 displaying approximately seven times more variation at the nucleotide level than PrV gC. Neither virus showed evidence of an association between the geographic proximity of farms of origin and genetic similarity of isolates, as measured by statistical tests of matrix correlation and phylogenetic analysis. The spread of PRRSV and PrV via "distance-limited" processes of interfarm transmission (e.g. wind, wildlife vectors) therefore seems unlikely. The long-distance movement of animals or viruses by people appears to be a better explanation for the current distribution of genetic types across the landscape for both viruses. Genetic distance between isolates was significantly correlated with time separating the collection of those isolates for PrV only. In the case of PrV, the samples of which represent a several-month regional "outbreak," genetic techniques therefore appear able reliably to reconstruct the temporal spread of the virus.

In the case of PrV, the availability of both RFLP and sequence data allowed direct comparison of the two techniques. Genetic distances between isolates derived from both methods were correlated by an *r* of approximately 0.45. RFLP-based distance was most strongly associated with time separating the collection of isolates. However, sequence data also yielded a strong genetic/temporal correlation when the sequence data were edited to exclude all nucleic acid changes that would lead to amino acid changes (synonymous changes), thereby filtering out the effects of natural selection. Conversely, nucleic acid changes which did translate into amino acid changes (nonsynonymous changes) were most reliable for predicting the species of origin of the isolate. These results draw attention to the necessity of considering selective forces when using sequence data to address epidemiological questions. They also indicate that very different epidemiological questions can be answered with different subsets of the same sequence data.